© Copyright 2015

Michael Curtis Grier Carlson

The lifestyles of viruses in the sunlit ocean revealed through isolation, genomics and infection dynamics

Michael Curtis Grier Carlson

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Gabrielle Rocap, Chair

E. Virginia Armbrust

John Baross

Program Authorized to Offer Degree:

School of Oceanography

University of Washington

Abstract

The lifestyles of viruses in the sunlit ocean revealed through isolation, genomics and infection dynamics

Michael Curtis Grier Carlson

Chair of the Supervisory Committee:

Dr. Gabrielle Rocap

Oceanography

Abstract

In the sunlit ocean, heterotrophic and photosynthetic microbes are prey for earth's most abundant predators. Viruses are regulators of microbial communities, terminating phytoplankton blooms, altering biogeochemical cycles through cell lysis, and driving evolution through horizontal gene transfer. However, viral infection in the marine environment is unquantified. My research seeks understand infection in the ocean through a model marine host-virus system and through coupled studies of virus diversity and gene expression in the field. The model host were diatoms, which contribute to global biogeochemical cycles, form the base of coastal food webs, and account for approximately one fifth of global photosynthesis. The first chapter details the isolation and characterization of a novel virus that infects the toxigenic diatom *Pseudo-nitzschia multiseries*. The virus had the largest genome of any single stranded DNA (ssDNA) virus isolated, with a gene complement atypical for ssDNA viruses. The virus had a broad host range, infecting hosts

from the distantly related centric and pennate diatom groups. The infection dynamics were host dependent, varying orders of magnitude in burst size from 220 to 9,300 viruses cell^{-1} and between 12-60 hours in latent period. The second chapter assesses the interactions of numerous *Pseudo-nitzschia* hosts and natural viral communities by isolating 41 new *Pseudo-nitzschia* and challenging them with 20 environmental virus samples. *Pseudo-nitzschia* cultures were infected in 8% of the crosses. Viral infectivity intensified during the summer when the most permissive hosts were isolated, and viral communities were most infective of co-occurring hosts. Genotyping of the *Pseudo-nitzschia* cultures revealed subgroups of genetically identical hosts that were differentially infected suggesting that diatom diversity is an impediment to viral termination of diatom blooms. In the third chapter, viral infection was measured by comparing viral metagenomes to host metatranscriptomes on a coastal to open ocean transect. A small fraction of the total viral population was actively infecting. Viral groups such as the archeoviruses, virophage, and ssDNA viruses had different distributions and were differentially enriched in particle-associated and free-living metatranscriptomes. Viral expression was not correlated with virus to host ratios suggesting that microbial populations were at different stages of infection. Finally, auxiliary metabolic genes in *Synechococcus* phage were differentially expressed between stations despite similar community composition suggesting that auxiliary genes help viruses succeed under certain conditions. By linking culture based and culture-independent explorations of viral ecology, this research highlights the diversity of viruses in the ocean and the lifestyles that have made them successful.

Acknowledgements

There are truly not enough ways to thank all the people that have contributed to the work in this thesis and to the person who wrote it. Of course, none of this work would be written today without funding from the University of Washington, the National Science Foundation, the National Oceanic and Atmospheric Administration, and the Environmental Protection Agency. Unfortunately, clearly delineating the impact of individual people on this science or on myself is next to impossible. The people that I worked with in the Center for Environmental Genomics were unparalleled collaborators and friends. My committee believed in my abilities and encouraged my progress when I doubted the path before me. The undergraduates that worked with me, who are all far more talented than myself, provided the adventure, energy, and inspiration necessary to complete this research. I would not have been able to endure a Ph.D. without incredible friends, from those that were beside me daily to those that thought I studied coral cancer. I cannot be thankful enough of the unconditional support of my family who enabled me to pursue my dreams even from my earliest years. Finally, my adviser, Gabrielle Rocap, has undoubtedly shaped my life and my mind with wisdom, patience, and kindness.

Table of Contents

List of Figures

List of Tables

Introduction

Viruses are earth's most and ancient and diverse predators (Nasir and Caetano-Anolles, 2015). They likely arose in the primordial ocean, multiple times, in tandem with the evolution of their cellular hosts (Koonin et al. 2006). They have kept signatures of their antiquity in their simple structures, a genome encapsidated in protein and sometimes lipids. Their genomes are comprised of single stranded or double stranded RNA or DNA (and sometimes a fusion of the two (Roux et al. 2013)) and span four orders of magnitude in size, from 1.7kb to 2.5 Mb (Finsterbusch and Mankertz 2009, Philippe et al 2013). Similarly, viruses themselves range in size from 17 to 600nm. No common gene unifies all viruses. However, viruses are all unified by their goal as parasites to replicate and produce more progeny viruses at the expense of their host. Viruses achieve this goal through many different types of lifestyles. Viral infection is a complex process where a virus binds to a host cell, migrates inside, and commandeers the host's metabolic machinery to replicate. Lysis concludes infection, releasing progeny viruses into the environment. Some viruses take detours from this path by integrating into their hosts' genomes (lysogeny), delaying their infection (latency), or infecting but budding off new viruses without killing the host (chronic infection). Together, these genomic, morphological, and functional characteristics shape the ecology of viruses and their hosts.

Life in the ocean presents substantial challenges for viruses. The ocean is patchy at many levels, driven by physical and chemical processes that profoundly influence marine microbial populations. At the global scale, temperature, light, and nutrients have created regimes that support different consortia of microbes (Longhurst et al. 1995, Follows and Dutkiewicz 2011). At the mesoscale, features can concentrate or disperse microbes by transporting water between biomes, such as from coastal regions to the open ocean in the case of eddies or from deep water

to surface water during upwelling (Monier et al., 2014). Zooming in even further, patchiness can be found even at the scale of a micrometer (Azam 1998, Stocker 2012). A single colony of the nitrogen fixing cyanobacterium *Trichodesmium* or a diatom seems like the city of Manhattan compared to viruses which would be ant sized (Finkel et al. 2009). In total, these processes create an ever-changing microbial landscape in a diffusive environment. For viruses in the ocean, finding a host may seem like a chance encounter, but viruses have been successful in this environment for billions of years.

The first marine virus was cultured in 1955 (Spencer, 1955), but only with the acknowledgement in 1989-1990 (Bergh *et al.*, 1989; Proctor and Furhman, 1990) of viruses being abundant, diverse, and ecologically relevant biotic entities in the ocean, did marine virology emerge as an integral discipline within oceanography. Viruses have since been identified as the most abundant predators on earth. In the ocean, where they number on average 10^7 per ml of seawater, viruses are thought to structure microbial communities as dominant contributors to cell death through lysis (Suttle 2007, Breitbart 2012). Also, viral lysis of hosts alters the global biogeochemical cycles mediated by microbes because of the release of cell contents into the water column. Finally, viruses are responsible for driving evolution through two mechanisms. First, evolution is influenced through predator-prey dynamics, in which hosts and viruses are constantly changing variable regions in their genomes to evade viral predation or attack new hosts (Stern and Sorek, 2011). Second, viruses transfer genetic material between organisms, allowing both viruses and hosts to explore new niches through new gene complements (Breitbart 2012). Therefore, viruses in the marine environment link microbial ecology, global biogeochemical cycles, and evolution.

Two contrasting methodological routes have sought to understand viral function and diversity: virus isolation and culture independent techniques, particularly metagenomics. The culturing of model host-virus systems such as *Vibrio* (Nakanishi *et al.,* 1966), *Prochlorococcus* (Sullivan *et al.,* 2003), *Synechococcus* (Waterbury and Valois, 1993), *Pelagibacter* (Zhao *et al.*, 2013), *Micromonas* (Cottrell and Suttle, 1991), *Phaeocystis* (Brussaard *et al.*, 2004), *Aureococcus* (Garry *et al.*, 1998) and *Emiliania huxleyi* (Schroeder *et al.,* 2002) have allowed detailed exploration of host-virus dynamics and demonstrated complex evolutionary patterns where hosts and viruses transfer genes to and from each other and fuel coevolution through the evolutionary arms race. These model host-virus systems have revealed strategies of how a virus infects and manipulates its host's physiology during infection (Lindell *et al.*, 2007; Vardi *et al.*, 2009). Recently, the genomic information from these systems has shed light on a sea of unknowns uncovered by metagenomics, a more recent and efficient method of capturing complex *in situ* dynamics (Angly *et al.*, 2006; Sharon *et al.*, 2009; Zhao *et al.*, 2013). Metagenomics has revealed equally interesting stories of viral diversity (Breitbart *et al.*, 2002), evolution (Alperovitch-Lavy *et al.*, 2011), and host dynamics (Rodriguez-Brito *et al.*, 2010) without the dependence on the culturability of both the host and the virus. However, most sequences in viral metagenomes are unable to be identified (Hurwitz and Sullivan, 2013). My research seeks to understand how viruses impact one of the most important groups of organisms in the ocean: diatoms.

Diatoms are the most abundant and genetically diverse group of marine eukaryotic phytoplankton (Armbrust, 2009). They contribute to approximately 20% of global photosynthesis (Field *et al.*, 1998) and are ubiquitously distributed in the marine environment. Because of this, diatoms are critical to global biogeochemical cycling (Nelson *et al.*, 1995). Their

silica cell walls or frustules cause them to be exported out of the euphotic zone upon death fueling the sequestration of carbon, silica, and other nutrients at depth, known as the biological pump (Smetachek, 2000). The success of diatoms in the oceans has been attributed to genomic plasticity allowing for diversification and the occupation of a wide range of niches (Bowler *et al.*, 2008). Diatom genomes are a mélange of genes that have come from a diversity of organisms including the ancestral heterotrophic eukaryote, photosynthetic endosymbiotic cyanobacteria, numerous non-phototrophic bacteria, and viruses (Armbrust et al, 2004, Bowler *et al.*, 2008). Horizontal gene transfer, therefore, has played an essential role in the evolution of the diatoms. Finally, diatoms form the base of productive food webs as major primary producers. One genus of diatoms, *Pseudo-nitzschia*, is known for its ability to produce a neurotoxin, domoic acid, which can be biomagnified through food webs and can cause ecosystem disruptive effects and public health concerns (Bates *et al.*, 1989; Scholin *et al.*, 2000). These reasons make diatoms one of the most important microbes in the ocean. Furthermore, understanding the processes that structure diatom communities, influence diatom death, removal, and sinking, and shape diatom genomes is critical. Viruses play an important role in phytoplankton mortality, but understanding of the extent to which virus impact diatom biology is in its infancy.

The first diatom viruses were isolated and characterized a decade ago (Nagasaki *et al.*, 2004). Since then the number has grown to 15, and they have been isolated on 4 genera of diatoms, *Rhizosolenia*, *Chaetoceros*, *Thalassionema*, all centric diatoms, and *Asterionellopsis*, a pennate diatom (Nagasaki *et al.*, 2004, Shirai *et al.*, 2008, Tomaru *et al.,* 2009, Tomaru *et al.*, 2012, Nagasaki *et al.*, 2005, Bettarel *et al.*, 2005, Eissler *et al.,* 2009, Tomaru *et al.*, 2008, Tomaru *et al.*, 2011, Tomaru *et al.*, 2012, Tomaru *et al.,* 2013a, Tomaru *et al.,* 2013b, Kimura *et al.*, 2013, Toyoda *et al.*, 2012, Kimura and Tomoru, 2015). The diatom viruses have small

capsids typically less that 30 nm in diameter and burst sizes in the thousands of viruses per cell (Tomaru and Kimura 2015). They have narrow host ranges, often only infecting the original host (Tomaru and Kimura 2015). All diatom viruses have fallen into two groups based on their nucleic acid content of either single stranded RNA or single stranded DNA.

The genomes of the diatom viruses have been one of their most interesting characteristics. For the RNA viruses, all are positive sense ssRNA viruses that are closely related to other RNA viruses infecting protists (Tomaru *et al.*, 2013). In general, positive sense ssRNA viruses have small genomes, ranging from 4-26kb in length that frequently have poly(A) tails (Lang *et al.*, 2009). The hallmark characteristics of this class of viruses are their helicase gene, capsid protein, and RNA-dependent RNA polymerase (Rdrp), the main phylogenetic marker for this group of viruses (Lang *et al.*, 2009). RNA viruses that replicate using Rdrp are known for having near perilous mutation rates (Duffy *et al.*, 2008). During infection, for example, high rates of mutations result in quasispecies of progeny viruses, which helps the viruses survive in diverse environments (Duffy *et al.*, 2008). A handful of mutations found in the capsid gene of an ssRNA virus that infects the marine protist *Heterocapsa* were critical for determining the virus' host range and infectivity (Nakayama *et al.*, 2013). This highlights the importance of capturing sequence diversity in the ssRNA viruses as it can significantly affect virus-host infection dynamics.

The diatom DNA viruses with sequenced genomes form a new group of ssDNA viruses, the *Bacillariodnavirus* group (Toyoda *et al.*, 2012). The overarching characteristics of eukaryotic ssDNA viruses are genomes typically smaller than 6kb encoding less than 6 genes (Rosario *et al.*, 2012) including capsid genes, the replication initiator protein (Rep), and rolling circle replication protein (RCR) which have been commonly used to infer phylogenetic relatedness in

the eukaryotic ssDNA viral groups (Gibbs *et al.*, 2006; Koonin and Ilyina, 1992). Based on phylogenies of the Rep gene, the diatom viruses are distantly related to Circoviruses (Nagasaki *et al.*, 2005), which infect vertebrates, and Geminiviruses, which infect higher plants (Labonte and Suttle, 2013). While it may seem like ssDNA viruses have limited evolutionary space to explore due to small genome sizes and gene complement, ssDNA viruses are an incredibly diverse group at the sequence level. Their evolution is driven through two mechanisms. First, their rate of mutation can be on par with that of ssRNA viruses (Duffy *et al.*, 2008). This is reflected in ssDNA virus taxonomy, for example, with Anelloviruses that have within-genus genome nucleotide identity differences up to 80% (Rosario *et al.*, 2012). Second, ssDNA viruses are known for high frequencies of recombination leading to chimeric genomes (Lefeuvre *et al.*, 2009). This gene swapping is also an efficient mechanism facilitating the transfer of genes to host genomes. One example of a gene gained by recombination is a Rep protein from a nanovirus, which is a group of ssDNA viruses distinct from the diatom viruses, found in the nuclear genome of *Phaeodactylum tricornutum* (Liu *et al.*, 2011). The existence of the Nanovirus-like Rep gene in the genome of *Phaeodactylum* suggests that there may be other groups of ssDNA viruses besides the Bacillariodnaviruses that can infect diatoms, but have yet to be isolated.

Culture-dependent techniques have yielded a first and important glimpse at diatom virus diversity (Kimura and Tomaru, 2015). On the other hand, a handful of diatom virus-like genomes or genes have been assembled from metagenomes from viral communities (Culley *et al.*, 2006; Culley *et al.*, 2014, McDaniel *et al.*, 2014). Labonte et al. (2013) used methods to specifically target ssDNA viruses in the environment and were able to assemble over 590 new ssDNA viral genomes. However, none of these genomes were closely related to the known

diatom virus genomes. This highlights the need for more reference genomes of diatom viruses for phylogenetic comparisons. The scarcity of sequenced diatom viruses impedes insight into diatom-virus diversity and dynamics in the field.

Little is known about diatom-infecting viruses' influence on host community composition. Only one study has surveyed the abundance of known-diatom infecting viruses (Tomaru *et al.*, 2011a). However, viruses have been shown to affect host physiology and community abundance of several other harmful algal species. At the end of one mesocosm study in which the coccolithophorid *E. huxleyi* bloomed, virus-like particle abundance increased dramatically and viruses known to infect *E. huxleyi* also increased in abundance (Schroeder *et al.*, 2003). Glycosphingolipids, used as biomarkers of viral infection, also coincided with decreased *E. huxelyi* abundance in a transect through a coccolithophore dominated bloom (Vardi *et al.*, 2009). Finally, tracking a bloom of *Aureococcus* showed that the number of VLPs increased over the bloom and water from the bloom was able to lyse cultured *Aureococcus* (Gastrich *et al.*, 2004).

This thesis seeks to understand the diversity and function of viruses in the marine environment, focusing on viruses that infect diatoms. The approach in these three chapters bridges the gap that lies between culture based and culture-independent methods of environmental virology. Chapter one details the isolation and characterization of a novel diatom virus with a unique ssDNA genome and a broad host range. It lays the groundwork for a model diatom-virus system and expands the boundaries of viral genomic diversity. Chapter two looks to the field to understand the patterns of interactions between diatoms and their viruses. It explores temporal and spatial scales of infection as well as host permissivity to infection using diatoms of the genus *Pseudo-nitzschia*. Finally, chapter three assesses active viral infection in the ocean by

integrating viral metagenomes with cellular metatranscriptomes. Together they depict the varied lifestyles of viruses and how viruses control microbial communities in the ocean.

References

Alperovitch-Lavy, A., I. Sharon, F. Rohwer, E.-M. Aro, F. Glaser, R. Milo, N. Nelson, and O. Béjà. 2011. Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. Environ. Microbiol. **13**: 24–32.

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. a Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. a Suttle, and F. Rohwer. 2006. The marine viromes of four oceanic regions. PLoS Biol. **4**: e368.

Armbrust, E. V., J. a Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. a Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. a Rynearson, M. a Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar. 2004. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science **306**: 79–86.

Armbrust, E. V. 2009. The life of diatoms in the world's oceans. Nature **459**: 185–92.

Azam, F. 1998. Microbial control of cceanic carbon flux: the plot thickens. Science. **280**: 694–696.

Bates, S. S., C. J. Bird, S. W. de Freitas, R. Foxall, M. Gilgan, L. A. Hanic, G. R. Johnson, A. W. McCulloch, P. Odense, R. Pocklington, M. A. Quilliam, P. G. Sim, J. C. Smith, D. V. Subba Rao, E. C. D. Todd, J. A. Walter, and J. L. C. Wright. 1989. Pennate diatom Nitzschia pungense as the primary source of domoic acid, a toxin in the shellfish from Eastern Prince Edward Island, Canada. Can. J. Fish. Aquat. Sci. **46**: 1203–1215.

Bettarel, Y., J. Kan, K. Wang, K. E. Williamson, S. Cooney, S. Ribblett, et al. (2005). Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom Chaetoceros cf. gracilis. Aquatic Microbial Ecology 40(2): 103- 114.

Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. a Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M.-P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J.-L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. a Rynearson, J. Schmutz, H.

Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, and I. V Grigoriev. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456**: 239–44.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U. S. A. **99**: 14250–5.

Breitbart, M. 2012. Marine Viruses: Truth or Dare. Ann. Rev. Mar. Sci. **4**: 425–448.

Brussaard, C. P. D., S. M. Short, C. M. Frederickson, and C. a Suttle. 2004. Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton Phaeocystis globosa (Prymnesiophyceae). Appl. Environ. Microbiol. **70**: 3700–5.

Cottrell, M. T. and C. A. Suttle (1991). Wide-spread occurence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankter, Micromonas pusilla. Marine Ecology-Progress Series 78(1): 1-9.

Culley, A. I., A. S. Lang, and C. a Suttle. 2006. Metagenomic analysis of coastal RNA virus communities. Science **312**: 1795–8.

Culley, A. I., J. A. Mueller, and M. Belcaid. 2014. The Characterization of RNA Viruses in Tropical Seawater Using Targeted PCR and Metagenomics. , doi:10.1128/mBio.01210-14.

Duffy, S., L. a Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. **9**: 267–76.

Eissler, Y., K. Wang, F. Chen, K. E. Wommack and D. W. Coats (2009). Ultrastructural characterization of the lytic cycle of an intracellular virus infecting the diatom Chaetoceros cf. wighamii (Bacillariophyceae) from Chesapeake Bay, USA Journal of Phycology 45(4): 787-797.

Field, C. B., M. J. Behrenfeld, and J. T. Randerson. 1998. Primary Production of the Biosphere : Integrating Terrestrial and Oceanic Components. Science (80). **281**: 237–241.

Finkel, Z. V., J. Beardall, K. J. Flynn, A. Quigg, T. A. V. Rees, and J. A. Raven. 2010. Phytoplankton in a changing world: cell size and elemental stoichiometry. J. Plankton Res. **32**: 119–137.

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

Follows, M. J., and S. Dutkiewicz. 2011. Modeling diverse communities of marine microbes. Ann. Rev. Mar. Sci. **3**: 427–451.

Garry, R. T., and E. M. Cosper. 1998. Characterization of a lytic virus infectious to the bloom-forming Aureococcus anophagefferens (Pelagophyceae). J. Phycol. **621**: 616–621.

Gastrich, M. D., J. A. Leigh-Bell, C. J. Gobler, O. R. Anderson, S. W. Wilhelm and M. Bryan (2004). Viruses as potential regulators of regional brown tide blooms caused by the alga, Aureococcus anophagefferens. Estuaries 27(1): 112-119.

Gibbs, M. J., V. V Smeianov, J. L. Steele, P. Upcroft, and B. a Efimov. 2006. Two families of rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes. Mol. Biol. Evol. **23**: 1097–100.

Hurwitz, B. L., and M. B. Sullivan. 2013. The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology F. Thompson [ed.]. PLoS One **8**: e57355.

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

Kimura, K., and Y. Tomaru. 2015. Discovery of Two Novel Viruses Expands the Diversity of Single-Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. Appl. Environ. Microbiol. **81**: 1120–1131.

Labonté, J. M., and C. a Suttle. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J. 1–9.

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

Lefeuvre, P., J.-M. Lett, a Varsani, and D. P. Martin. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. **83**: 2697–707.

Lindell, D., J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen, W. R. Hess, G. M. Church, and S. W. Chisholm. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature **449**: 83–6.

Liu, H., Y. Fu, B. Li, X. Yu, J. Xie, J. Cheng, S. a Ghabrial, G. Li, X. Yi, and D. Jiang. 2011. Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. BMC Evol. Biol. **11**: 276.

Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill. 1995. An estimate of global primary production in the ocean from satellite radiometer data. J. Plankton Res. **17**: 1245–1271.

Koonin, E. V, and T. V Ilyina. 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol. **73**: 2763–6.

Koonin, E. V, T. G. Senkevich, and V. V Dolja. 2006. The ancient Virus World and evolution of cells. Biol. Direct **1**: 29.

McDaniel, L. D., K. Rosario, M. Breitbart, and J. H. Paul. 2014. Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. Environ. Microbiol. **16**: 570–85.

Monier, A., J. Comte, M. Babin, A. Forest, A. Matsuoka, and C. Lovejoy. 2014. Oceanographic structure drives the assembly processes of microbial eukaryotic communities. ISME J. 1–13.

Nagasaki, K., Y. Tomaru, N. Katanozaka, Y. Shirai, K. Nishida, S. Itakura, and M. Yamaguchi. 2004. Isolation and Characterization of a Novel Single-Stranded RNA Virus Infecting the Bloom-Forming Diatom Rhizosolenia setigera. Appl. Environ. Microbiol. **70**: 704–711.

Nagasaki, K., Y. Tomaru, Y. Takao, K. Nishida, Y. Shirai, H. Suzuki, et al. (2005). Previously unknown virus infects marine diatom. Applied and Environmental Microbiology 71(7): 3528-3535.

Nakayama, N., A. Fujimoto, H. Kawami, Y. Tomaru, N. Hata, and K. Nagasaki. 2013. High Interaction Variability of the Bivalve-Killing Dinoflagellate Heterocapsa circularisquama Strains and Their Single-Stranded RNA Virus HcRNAV Isolates. Microbes Environ. **28**: 112–119.

Nasir, A., and G. Caetano-anollés. 2015. A phylogenomic data-driven exploration of viral origins and evolution. , doi:10.1126/sciadv.1500527.

Nakanishi, H., Y. Iida, K. Maeshima., T. Teramoto, Y. Hosaka and M. Ozaki (1966). Isolation and properties of bacteriophages of Vibrio parahaemolyticus. Biken J 9: 149-157.

Nelson, D. M., P. Trrguer, and M. A. Brzezinski. 1995. Production and dissolution of biogenic silica in the ocean : Revised global estimates , comparison with regional data and relationship to biogenic sedimentation. Global Biogeochem. Cycles **9**: 359–372.

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

Philippe, N., M. Legendre, G. Doutre, Y. Couté, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J.-M. Claverie, and C. Abergel. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science **341**: 281–6.

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

Rodriguez-Brito, B., L. Li, L. Wegley, M. Furlan, F. Angly, M. Breitbart, J. Buchanan, C. Desnues, E. Dinsdale, R. Edwards, B. Felts, M. Haynes, H. Liu, D. Lipson, J. Mahaffy, A. B. Martin-Cuadrado, A. Mira, J. Nulton, L. Pasić, S. Rayhawk, J. Rodriguez-Mueller, F. Rodriguez-Valera, P. Salamon, S. Srinagesh, T. F. Thingstad, T. Tran, R. V. Thurber, D. Willner, M. Youle, and F. Rohwer. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J. **4**: 739–51.

Rosario, K., S. Duffy, and M. Breitbart. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. **157**: 1851–71.

Roux, S., F. Enault, G. Bronner, D. Vaulot, P. Forterre, and M. Krupovic. 2013. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. Nat. Commun. **4**: 2700.

Scholin, C. A., F. Gulland, G. J. Doucette, S. Benson, M. Busman, F. P. Chavez, J. Cordaro, R. Delong, A. De Vogelaere, J. Harvey, M. Haulena, K. L. I, T. Lipscomb, S. Loscutoff, L. J. Lowenstine, R. M. Iii, P. E. Miller, W. A. Mclellan, P. D. R. Moeller, C. L. Powell, T. Rowles, P. Silvagni, M. S. I, T. Spraker, V. Trainer, and F. M. Van Dolah. 2000. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. **403**: 80–84.

Schroeder, D. C., J. Oke, G. Malin, and W. H. Wilson. 2002. Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects Emiliana huxleyi. Arch. Virol. **147**: 1685–98.

Sharon, I., A. Alperovitch, F. Rohwer, M. Haynes, F. Glaser, N. Atamna-Ismaeel, R. Y. Pinter, F. Partensky, E. V Koonin, Y. I. Wolf, N. Nelson, and O. Béjà. 2009. Photosystem I gene cassettes are present in marine virus genomes. Nature **461**: 258–62.

Smetacek, V. 1999. Diatoms and the ocean carbon cycle. Protist **150**: 25–32.

Spencer, R. 1955. A marine bacteriophage. Nature **175**: 690–691.

Stern, A., and R. Sorek. 2011. The phage-host arms race: shaping the evolution of microbes. Bioessays **33**: 43–51.

Stocker, R. 2012. Marine Microbes See a Sea of Gradients. Science (80). **338**: 628–633.

Sullivan, M. B., J. B. Waterbury, and S. W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature **424**: 1047–51.

Suttle, C. a. 2007. Marine viruses--major players in the global ecosystem. Nat. Rev. Microbiol. **5**: 801–12.

Tomaru, Y., N. Fujii, S. Oda, K. Toyoda and K. Nagasaki. 2011a. Dynamics of diatom viruses on the western coast of Japan. Aquatic Microbial Ecology 63(3): 223-230.

Tomaru, Y., Y. Shirai, H. Suzuki, T. Nagumo and K. Nagasaki. 2008. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom Chaetoceros dehilis. Aquatic Microbial Ecology 50(2): 103-112.

Tomaru, Y., Y. Takao, H. Suzuki, T. Nagumo, K. Koike and K. Nagasaki. 2011b. Isolation and Characterization of a Single-Stranded DNA Virus Infecting Chaetoceros lorenzianus Grunow. Applied and Environmental Microbiology 77(15): 5285-5293.

Tomaru, Y., Y. Takao, H. Suzuki, T. Nagumo and K. Nagasaki. 2009. Isolation and Characterization of a Single-Stranded RNA Virus Infecting the Bloom-Forming Diatom Chaetoceros socialis. Applied and Environmental Microbiology 75(8): 2375-2381.

Tomaru, Y., K. Toyoda, K. Kimura, N. Hata, M. Yoshida and K. Nagasaki. 2012. First evidence for the existence of pennate diatom viruses. Isme Journal 6(7): 1445-1448.

Tomaru, Y., K. Toyoda, K. Kimura, Y. Takao, K. Sakurada, N. Nakayama, and K. Nagasaki. 2013. Isolation and characterization of a single-stranded RNA virus that infects the marine planktonic diatom Chaetoceros sp. (SS08-C03). Phycol. Res. **61**: 27–36.

Tomaru, Y., K. Toyoda, H. Suzuki, T. Nagumo, K. Kimura, and Y. Takao. 2013. New single-stranded DNA virus with a unique genomic structure that infects marine diatom Chaetoceros setoensis. Sci. Rep. **3**: 3337.

Tomaru, Y., N. Fujii, S. Oda, K. Toyoda, and K. Nagasaki. 2011. Dynamics of diatom viruses on the western coast of Japan. Aquat. Microb. Ecol. **63**: 223–230.

Toyoda, K., K. Kimura, N. Hata, N. Nakayama, K. Nagasaki, and Y. Tomaru. 2012. Isolation and characterization of a single-stranded DNA virus infecting the marine planktonic diatom Chaetoceros sp. (strain TG07-C28). Plankt. Benthos Res. **7**: 20–28.

Vardi, A., B. a S. Van Mooy, H. F. Fredricks, K. J. Popendorf, J. E. Ossolinski, L. Haramaty, and K. D. Bidle. 2009. Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. Science **326**: 861–5.

Waterbury, J. B., and F. W. Valois. 1993. Resistance to co-occurring phages enables marine synechococcus communities to coexist with cyanophages abundant in seawater. Appl. Environ. Microbiol. **59**: 3393–9.

Zhao, Y., B. Temperton, J. C. Thrash, M. S. Schwalbach, K. L. Vergin, Z. C. Landry, M. Ellisman, T. Deerinck, M. B. Sullivan, and S. J. Giovannoni. 2013. Abundant SAR11 viruses in the ocean. Nature **494**: 357–60.

Chapter 1

A giant single stranded DNA virus with an origin in double stranded DNA phage

1.1 Abstract

Single stranded DNA viruses are some of the most successful yet simple pathogens on the planet, found in every environment and infecting organisms from all three domains of life. One of the ocean's most diverse and abundant primary producers, diatoms, are one of the few known hosts for ssDNA viruses in the ocean. A ssDNA virus was isolated and characterized using the diatoms *Pseudo-nitzschia multiseries* and *Thalassiosira pseudonana*. The *P. multiseries* DNA virus (PmDNAV) has two linear genome segments, 9,682 and 28,403 nt in length, the largest genome of any ssDNA virus. It also has an unusual gene complement for ssDNA viruses including a DNA polymerase B indicating novel replication dynamics for an ssDNA virus. The virus had no homology to previously sequenced ssDNA viruses and instead was more closely related to double stranded DNA viruses and prophage that are integrated into bacterial genomes. The virus indicates that a new type of ssDNA viruses originated from dsDNA viruses. These characteristics suggest that the PmDNAV has a unique lifestyle for ssDNA viruses and represents a new lineage of ssDNA viruses.

Viruses, earth's most ancient predators, likely arose in the primordial ocean multiple times, in tandem with the evolution of their cellular hosts (Koonin et al. 2006, Nasir and Caetano-Anolles, 2015). Single stranded DNA (ssDNA) viruses infect all three domains of life and are prevalent in nearly every environment sampled . They typically have genomes smaller than 10kb, which can be segmented (Rosario et al. 2009) and replicate via a distinctive rolling

circle replication (RCR) mechanism using proteins that are unlike the DNA polymerases found in cellular life and other double stranded DNA (dsDNA) viruses (Gibbs et al., 2006; Koonin and Ilyina, 1992). High mutation rates of their replication proteins (Duffy et al. 2008) and frequent recombination (Lefeuvre et al., 2009) drive the evolution of ssDNA viruses. These viruses have polyphyletic origins (Krupovic, 2013). Bacteria-infecting ssDNA viruses are thought to predate dsDNA viruses (Nasir and Caetano-Anolles, 2015), while eukaryote-infecting ssDNA viruses originated from the recombination of various bacterial plasmids with ssRNA viruses multiple times (Koonin et al. 2015). The eukaryotic marine phytoplankton group diatoms are one of the few known hosts for ssDNA viruses in the ocean (Nagasaki and Bratbak 2010), which is the largest reservoir of viruses on the planet (Suttle 2005). Here we detail the characteristics of a new diatom-infecting virus.

During a toxic bloom of the pennate diatom *Pseudo-nitzschia* at Sunset Beach, Oregon, USA in August 2009, the viral fraction (<0.2 um) of water was infectious to cultures of *Pseudo-nitzschia multiseries* CLnn-16. The virus was purified with a plaque assay and three dilution-to-extinction experiments. When treated with heat, UV, or filtered through a 0.02 μm membrane, the infectious potential of the lysates was eliminated. Viral particles with an icosahedral head that measured 50 nm (s.d. ± 1.8 nm, n=10) in diameter and lacked a tail or an outer membrane were observed in electron micrographs of *P. multiseries* CLnn-16 culture lysates using negative staining (figure 1a). In addition to infecting *P. multiseries,* the virus also infected the centric diatom *Thalassiosira pseudonana* CCMP1335.

Viral nucleic acids were extracted from lysates of both *P. multiseries* CLnn-16 and *T. pseudonana* CCMP1335. At least three fold more ssDNA than dsDNA was detected from viral nucleic acid extracts, and no RNA was detected (figure 1b). Digestion of viral nucleic acids with

DNase I (cleaves dsDNA and ssDNA) yielded signals below detection limits for both dsDNA (<0.5 ng ml^{-1}) and ssDNA (<1 ng ml^{-1}) (figure 1b). Digestion with S1 nuclease (cleaves ssDNA and ssRNA) also resulted in undetectable signal (figure 1b). Furthermore, when viral nucleic acids were sheared and ligated into plasmids, no *E. coli* colonies had viral inserts (n=48). Together, these data confirmed that the virus is a single stranded DNA virus, and it was named accordingly as *Pseudo-nitzschia multiseries* DNA virus (PmDNAV).

The viral genome was assembled from two sources: ssDNA extracted and converted to dsDNA from lysates of *P. multiseries* GGB1 and of *T. pseudonana* CCMP1335. The *T. pseudonana* DNA was sequenced with PacBio technology and the *P. multiseries* DNA was sequenced with Sanger sequencing of paired-end clone libraries. Assembly of the Pac Bio reads resulted in 3 contigs, each with an average coverage of 3000x. The Sanger reads were 99.8% identical to Pac Bio sequence and, when mapped to the Pac Bio assembly, linked two contigs to generate two linear genome segments of 28,403 and 9,682 nucleotides (nt) in length (figure 2a). Both segments had 54% GC content. Terminal inverted repeats at both ends of the larger genome segment indicated the segment was closed. The second smaller segment had one terminal repeat. No mate pairs extended off the end of either segment. In total, 24% of the DNA is double stranded because of the terminal repeats, accounting for the signal detected using a dsDNA fluorescing dye (figure 1b). The PmDNAV has the largest genome, a combined 38,085 nt, of any known ssDNA virus. The larger genome segment alone is ~4,000 bases longer than the next longest ssDNA virus, the *Aeropyrum* coil-shaped virus (Mochizuki et al., 2012), and ~4x larger than the typical <10kb genomes of the majority of ssDNA viruses (Rosario et al. 2009).

To determine whether the two segments represented two separate viral genomes, several attempts were made to separate them. First, both genome segments were detected using segment-

specific quantitative polymerase chain reaction at every endpoint in 15 independent dilution to extinction experiments using axenic *T. pseudonana* CCMP1335. The probability that both genomes were detected at every end point if they were independent was $p<0.05$ (supplemental table 1). Second, both segments purified between 1.35 and 1.4 g ml^{-1} in cesium chloride density dependent gradients and thus were unable to be separated based on their densities (supplemental figure 1). Third, both segments increased approximately 5x in copy number simultaneously with the appearance of virus-like particles at 32 and 60 hours post infection in one step growth curves performed on axenic *P. multiseries* PC9 and *T. pseudonana* CCMP 1335, respectively (supplemental figure 2). Thus, our interpretation of these data is that the two genome segments are contained in one virus.

The genome had 59 putative open reading frames (ORFs), of which 35 were similar to previously sequenced organisms and 12 had predicted functions (figure 2b, supplemental table 2). No genes were found that are common to previously sequenced ssDNA viruses such as the hallmark RCR endonucleases or capsid genes (Koonin et al. 2015). Genes involved in replication, specifically, a replication initiator gene, a transcriptional regulator gene, a helicase, and DNA polymerase B (polB), were located on the larger genome segment (figure 2b). Also, an HNH-endonuclease, which cleaves DNA at specific sequence patterns, suggested that the virus has the potential for horizontal gene transfer and genome rearrangement. Genes encoding viral structural proteins and maturation were located on the smaller genome segment (figure 2b). While the majority of ssDNA viruses have capsid genes with a double jelly roll fold (Abrescia et al., 2012), the PmDNAV capsid genes are predicted belong to a different structural lineage as they have HK97 fold domains (Pietila et al. 2011). A peptidase gene next to the capsid gene may be involved in the cleavage and maturation of the capsid genes to form the virion, and a gene

with an acetyltransferase domain may further modify viral proteins. Together, the replication and structural genes encoded on the separate genome segments complement each other by performing the functions necessary for viral infection, replication, and assembly.

The presence of these genes and particularly a polB is rare for ssDNA viruses. Only one other ssDNA virus, the *Bombyx mori* densovirus, encodes a protein-primed polB, which it may have acquired from the recombination between a parvovirus (ssDNA) and a dsDNA virus derived transposon (Krupovic and Koonin, 2014). It is likely the polymerase on the PmDNAV is DNA primed, and uses the inverted terminal repeats similar to the linear ssDNA parvoviruses to begin replication (Cotmorel, 1996). One of the major differences between ssDNA and dsDNA virus polymerases is that DNA polymerase B is, on average, less error prone than the RCR endonucleases of ssDNA viruses (Banach-Orlwsha et al. 2005). The rolling circle strategy is hypothesized to impose a genome size threshold on ssDNA viruses because of the rate of accumulation of deleterious mutations (Duffy et al. 2008), thus the genome size of the PmDNAV may not be constrained by errors during replication with a polB. In total, based on the content of both replication and structural genes, the PmDNAV appears functionally distinct from all other ssDNA viruses.

The functional differences between the PmDNAV and other ssDNA viruses likely lead to different ecological strategies. The other diatom-infecting ssDNA viruses have narrow host ranges, often infecting only the original host used for isolation, burst sizes up to 22,000 viruses cell^{-1}, and encode potentially highly error-prone RCR endonucleases (Kimura and Tomaru, 2015). Mutations may to help alter the host range of ssDNA viruses (Duffy et al., 2008), and thus, with high burst sizes, may be a strategy to acquire new hosts and propagate in the water. The PmDNAV, on the other hand, produced 270 and 200 viruses cell^{-1} upon lysis of axenic *P.*

multiseries PC9 and *T. pseudonana* CCMP 1335, respectively (supplemental figure 2), and replicates with the less error prone polB. It also has a broad host range, able to infect both pennate (*Pseudo-nitzschia*) and centric diatoms such as *Thalassiosira, Ditylum*, Skeletonema, *Chaetoceros*, *Odontella*, and *Corethron* (supplemental table 3). Ultimately, this strategy is different than the other diatom infecting viruses; driven by different mechanisms that allow viruses to find new hosts, which might be rooted in their different polymerases and thus mutation rates.

The PmDNAV is phylogenetically unlike all other previously sequenced ssDNA viruses. The viral genome had no homology to previously sequenced ssDNA viruses based on blast with an e-value of $1e^{-2}$ (supplemental table 2). Genes that did have homology to previously sequenced organisms were either similar to bacteria which are likely putative prophage (28 genes) or dsDNA viruses (7 genes). Five of the 7 genes related to dsDNA viruses clustered together on the larger genome segment and were all involved in replication. Phylogenies indicated that two of these, polB and a helicase, were most closely related to Rhizobium phage (figure 2c). Rhizobium phage, which infect bacteria that are endosymbiotic with plants, are capable of horizontal gene transfer (Finan et al. 1984) suggesting, in addition to the endonuclease, that the PmDNAV may also share this capability. Genes outside this genomic island such as one of the capsid genes on the smaller genome segment and the replication initiator gene were more closely related to genes in bacteria, probably endogenous viruses, not to Rhizobium phage as seen with polB and the helicase genes (figure 2c). The differing evolutionary lineages and distinctive clustering of the genes on the PmDNAV indicate that the virus may have resulted from a recombination of two dsDNA viruses: an endogenous virus and a virus similar to Rhizobium phage. However, since the PmDNAV has an ssDNA genome, it is evidence that, at one point, a double stranded DNA

virus converted into a single stranded DNA virus.

The genomic characteristics and provenance of the PmDNAV do not fit within the current framework of the evolution of eukaryotic ssDNA viruses from the integration of ssRNA viruses with plasmids or within the hypothesis that bacteria infecting ssDNA viruses are an ancient group of viruses predating dsDNA viruses (Koonin et al. 2015, Nasir and Caetano-Anolles, 2015). Thus the PmDNAV likely represents a new lineage of ssDNA viruses and illuminates a novel pathway that has lead to a group of ssDNA viruses that appears to have arisen from dsDNA viral elements. Ultimately, the PmDNAV may not just represent a new group of viruses, but suggests that the conversion from a single stranded to double stranded genome may be a more widespread phenomenon. Other viral ssDNA viral groups arising from dsDNA viruses may be populating the environment that have yet to be discovered.

1.2 Acknowledgements

Funding was provided by the Pacific Northwest Center for Human Health and Ocean Sciences (National Institute of Environmental Health: P50 ES012762, the National Science Foundation: OCE-0434087), a grant from the National Science Foundation: OCE-1356779, and the University of Washington Royalty Research Fund A65810 to GR. MC was funded by the Pacific Northwest Consortium Predoctoral Traineeship in Oceans and Human Health and an EPA STAR fellowship. A Mary Gates Research Scholarship and an NSF REU fellowship supported KRF. We would also like to thank Benjamin Van Biber for critical early experiments suggestive of the existence of a *Pseudo-nitzschia* virus, Terence Leach for performing some of the infectivity condition experiments. Finally, we would like to thank the Armbrust lab, Bates Lab, and Leila Tirichine for providing cultures for this study.

1.3 References

Abrescia, N. G. A., D. H. Bamford, J. M. Grimes, and D. I. Stuart. 2012. Structure Unifies the Viral Universe. Annu. Rev. Biochem. **81**: 795–822.

Banach-Orlowska, M., I. J. Fijalkowska, R. M. Schaaper, and P. Jonczyk. 2005. DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in Escherichia coli. Mol. Microbiol. **58**: 61–70.

Duffy, S., L. a Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. **9**: 267–76.

Cotmorel, S. F. 1996. Parvovirus DNA Replication. Cold Spring Harbor Laboratory Press, 799-813.

Finan, T. M., E. Hartweig, K. LeMieux, K. Bergman, G. C. Walker, and E. R. Signer. 1984. General transduction in *Rhizobium meliloti*. J. Bacteriol. **159**: 120–4.

Gibbs, M. J., V. V Smeianov, J. L. Steele, P. Upcroft, and B. a Efimov. 2006. Two families of rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes. Mol. Biol. Evol. **23**: 1097–100.

Kimura, K., and Y. Tomaru. 2015. Discovery of Two Novel Viruses Expands the Diversity of Single-Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. Appl. Environ. Microbiol. **81**: 1120–1131.

Koonin, E. V., V. V. Dolja, and M. Krupovic. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology **479-480**: 2–25.

Koonin, E. V, T. G. Senkevich, and V. V Dolja. 2006. The ancient Virus World and evolution of cells. Biol. Direct **1**: 29.

Koonin, E. V, and T. V Ilyina. 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. J. Gen. Virol. **73 (Pt 10**: 2763–6.

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

Krupovic, M., and E. V. Koonin. 2014. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. **13**: 105–115.

Lefeuvre, P., J.-M. Lett, a Varsani, and D. P. Martin. 2009. Widely conserved recombination patterns among single-stranded DNA viruses. J. Virol. **83**: 2697–707.

Mochizuki, T., M. Krupovic, G. Pehau-Arnaudet, Y. Sako, P. Forterre, and D. Prangishvili. 2012. Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. Proc. Natl. Acad. Sci. **109**: 13386–13391.

Nagasaki, K., and G. Bratbak. 2010. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. Virology 92–101.

Nasir, A., and G. Caetano-anollés. 2015. A phylogenomic data-driven exploration of viral origins and evolution. , doi:10.1126/sciadv.1500527

Pietilä, M. K., P. Laurinmäki, D. A. Russell, C. Ko, D. Jacobs-sera, and R. W. Hendrix. 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. , doi:10.1073/pnas.1303047110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1303047110

Rosario, K., S. Duffy, and M. Breitbart. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol. **157**: 1851–71.

Suttle, C. a. 2005. Viruses in the sea. Nature **437**: 356–61.

1.4 Figure Legends

Figure 1.1 The morphology and nucleic acid content of the *Pseudo-nitzschia multiseries* DNA virus. a) 2 virus particles of the *Pseudo-nitzschia multiseries* DNA virus from a CLnn-16 lysate negatively stained and visualized with transmission electron microscopy. Scale bars are 100nm. b) Quantification of viral nucleic acids using Quant-iT high sensitivity DNA assays (blue), ssDNA assays (green), and RNA assays (orange) undigested and after digestion with DNAase 1 (cleaves dsDNA and ssDNA) and S1 nuclease (cleaves ssDNA and ssRNA). Colored lines denote the source of the nucleic acids, either extraction from lysates of *T. pseudonana* CCMP 1335 (red) or *P. multiseries* CLnn-16 (cyan). Error bars represent the standard deviation of triplicate treatments. Boxes below the dashed line represent values that were below the limit of detection for the assay.

Figure 1.2 The genomic characteristics of the *Pseudo-nitzschia multiseries* DNA virus. a) A schematic of the architecture of the two single stranded linear genome segments showing the inverted terminal repeats. b) The genome of the PmDNAV. Dashed lines indicate inverted terminal repeats. Open reading frames (ORFs) are indicated by arrows. ORFs are colored by their similarity to previously sequences organisms: white denotes ORFs with no significant similarity to previously sequenced organisms, red denotes ORFs similar to double stranded viruses, and blue denotes ORFs similar to bacteria, likely prophage. ORFs with putative functions are labeled while ORFs without functions are not. Maximum likelihood phylogenies of genes representing the two different sources of genetic material in the PmDNAV genome c) DNA polymerase B (on the large genome segment representing dsDNA viruses) and d) a capsid gene (on the small genome segment representing bacteria/prophage). The PmDNAV is labeled in purple, bacterial sequences in blue, dsDNA viruses in red, and ssDNA viruses in green. Statistical significance of the branching pattern is passed on 100 bootstrap replications of the original alignment. Scale bars indicate genetic distance.

1.5 Supplemental figure legends

Supplemental figure 1. Genome segment quantification in density fractions from cesium chloride gradients. Bars represent the abundance of genome segment 1 (a) or genome segment 2 (b) in each fraction from either 1.2, 1.3, 1.4, 1.5, and 1.7 g ml^{-1} or 1.25, 1.35, 1.45, 1.55, 1.7 g ml^{-1} gradients.

Supplemental figure 2. One step growth curves of axenic *P. multiseries* PC9 (a,c), and axenic *T. pseudonana* CCMP1335(b,d) infected with the *Pseudo-nitzschia multiseries* DNA virus. Solid

lines represent control cultures while dashed line represent virus infected cultures. Closed squares indicate host abundance. Open circles indicate viruses per cell. Abundance in figures a-c is based on epifluorescent microscope counts. Genome copies (c,d) were determined with qPCR.

Supplemental figure 3. Maximum likelihood phylogenetic trees of the (a) replication initiator protein and (b) helicase. The PmDNAV is labeled in purple, bacterial sequences in blue, dsDNA viruses in red, and ssDNA viruses in green.

1.6 Supplemental tables

Supplemental table 1.1. Genome detection and infectious units in dilution-to-extinction end points

Experiment	Infectious viruses ml^{-1} at end point*	Genome 1	Genome 2
1	1.51	✔	✔
2	0.73	✔	✔
3	0.73	✔	✔
4	0.73	✔	✔
5	4.23	✔	✔
6	0.73	✔	✔
7	1.51	✔	✔
8	0.73	✔	✔
9	2.31	✔	✔
10	2.36	✔	✔
11	0.73	✔	✔
12	1.51	✔	✔
13	2.36	✔	✔
14	1.51	✔	✔
15	0.73	✔	✔

✔ genome detected with qPCR in endpoint lysate
* determined by most probable number assays

Genome segment 1	e value	% identity	Taxa	Function	Pfam domain	Size (amino acids)
ORF 1	n/a					65
ORF 2	n/a					72
ORF 3	n/a					72
ORF 4	n/a					65
ORF 5	7.00E-34	37	Xylella phage Sano	hypothetical protein Sano_09		220
ORF 6	n/a					78
ORF 7	n/a					161
ORF 8	n/a					32
ORF 9	n/a					59
ORF 10	2.00E-24	55	Sphingomonas sanxanigenens	HNH endonuclease	HNH	90
ORF 11	n/a					145
ORF 12	4.00E-20	42	Hyphomonas sp. CY54-11-8	hypothetical protein		181
ORF 13	n/a					36
ORF 14	6.00E-63	38	Rhizobium sp. Root1203	virulence protein E	VirE	732
ORF 15	8.00E-08	40	Leptospirillum ferrooxidans	chromosomal replication initiator protein DnaA	Bac_DnaA_C	105
ORF 16	2.00E-07	41	Acinetobacter calcoaceticus	TetR family transcriptional regulator	TetR_N	76
ORF 17	4.00E-24	44	Edwardsiella phage PEi21	hypothetical protein		125
ORF 18	2.00E-166	51	Rhizobium phage vB_RglS_P106B	putative helicase	ResIII, Helicase_C	535
ORF 19	8.00E-04	33	Rhizobium phage vB_RglS_P106B	hypothetical protein P106B_54		196
ORF 20	n/a					59
ORF 21	3.00E-49	35	Rhizobium phage vB_RglS_P106B	hypothetical protein	DUF2800	366
ORF 22	1.00E-45	39	Agrobacterium phage 7-7-1	hypothetical protein	DUF2815	263
ORF 23	n/a					66
ORF 24	0	46	Edwardsiella phage PEi21	putative DNA polymerase		641
ORF 25	5.00E-104	44	Cupriavidus sp. UYPR2.512	hypothetical protein		393
ORF 26	n/a					133
ORF 27	n/a					80
ORF 28	1.00E-53	64	Maricaulis maris	peptidase M15	Peptidase_M15_3	151
ORF 29	5.00E-12	46	alpha proteobacterium Mf 1.05b.01	hypothetical protein		76
ORF 30	7.00E-47	29	Afipia sp. P52-10	hypothetical protein		1030
ORF 31	1.00E-06	31	Rhodobacter sphaeroides	hypothetical protein		130
ORF 32	n/a					223
ORF 33	n/a					106

				hypothetical protein		
ORF 35	2.00E-06	52	Methylobacterium aquaticum	Maq22A_c02400		237
ORF 36	3.00E-07	27	Agrobacterium sp. LC34	hypothetical protein		80
ORF 37	n/a					61
ORF 38	n/a					166
ORF 39	n/a					237
ORF 40	2.00E-06	37	Halomonas meridiana prophage PSSB64-02	hypothetical protein		258
ORF 41	n/a					106
Genome segment 2						
ORF 1	4.00E-10	36	Hyphomicrobium nitrativorans NL23	capsid protein	Phage_capsid	107
ORF 2	2.00E-54	48	Providencia sneebia	primosome assembly protein PriA		232
ORF 3	3.00E-30	46	Hoeflea phototrophica	peptidase U35	Peptidase_S78	142
ORF 4	9.00E-123	47	Janthinobacterium lividum	phage capsid protein		428
ORF 5	n/a					55
ORF 6	3.00E-21	52	Microgenomates bacterium	hypothetical protein		97
ORF 7	1.00E-30	35	Nitratireductor basaltis	hypothetical protein		213
ORF 8	2.00E-07	42	Enterobacteria phage RB49	hypothetical protein RB49p096		117
ORF 9	7.00E-41	52	Bradyrhizobium sp. YR681	acetyltransferase	Hexapep	164
ORF 10	4.00E-70	53	Bradyrhizobium sp. WSM1417	hypothetical protein		211
ORF 11	5.00E-27	38	Pseudomonas syringae	hypothetical protein		205
ORF 12	n/a					142
ORF 13	2.00E-33	44	Devosia soli	HK97 family phage protein	HK97-gp10_like	155
ORF 14	3.00E-07	34	Nesiotobacter exalbescens	hypothetical protein	DUF3168	87
ORF 15	1.00E-19	37	Rhodovulum sp. PH10	hypothetical protein		150
ORF 16	3.00E-28	44	Ochrobactrum sp. CDB2	hypothetical protein		144
ORF 17	n/a					50
ORF 18	2.00E-46	36	Ochrobactrum sp. EGD-AQ16	hypothetical protein		330

Group	Genus	Species	Strain	Isolation date	Isolation location	Virus: Host	Infection phenotype	Days to lysis
Raphid Pennates	*Pseudo-nitzschia*	*P. multiseries*	CLnn-16[◊]	Feb 2007	Lab mating/Eastern Canada	8	+	0.5
			CLnn-17[◊]	Feb 2007	Lab mating/Eastern Canada	4.5	+	6
			CLnn-21[◊]	Feb 2007	Lab mating/Eastern Canada	n.a.	+	3
			PC2[◊]	Jul 2010	Penn Cove, Washington, USA	7.1	+	8
			PC4[◊]	Jul 2010	Penn Cove, Washington, USA	17	-	
			PC9[◊]	Jul 2010	Penn Cove, Washington, USA	32	+	1.5
			GGA2	Jun 2010	Golden Gardens, Washington, USA	6.0	-	
			GGB1[◊]	Jun 2010	Golden Gardens, Washington, USA	5.1	+	6
		P. pungens c.f. pungens	GGC3[†]	Jun 2010	Golden Gardens, Washington, USA	41	-	
			PnC2[†]	Jun 2009	Friday Harbor, Washington, USA	15	-	
			SunA2[†]	Aug 2009	Sunset Beach, Oregon, USA	19	+	3
		P. pungens c.f. cingulata	GGA1[¤]	Jun 2010	Golden Gardens, Washington, USA	33	+	9
			GGA3	Jun 2010	Golden Gardens, Washington, USA	25	-	
			GGC2[¤]	Jun 2010	Golden Gardens, Washington, USA	65	+	11
			PnC4	Jun 2009	Friday Harbor, Washington, USA	27	-	
		P. fraudulenta	OC1[Δ]	Jun 2011	Newport Beach, California, USA	14	-	
			OC2[Δ]	Jun 2011	Newport Beach, California, USA	5.6	+	6
			OC3[Δ]	Jun 2011	Newport Beach, California, USA	1.6	-	
			OC4[Δ]	Jun 2011	Newport Beach, California, USA	2.3	-	
		P. delicatissima	K2	Jun 2006	Sequim Bay, Washington, USA	n.a.	-	
	Cylindrotheca	*C. closterium*	RCC 1713	May 2008	English Channel	7.3	-	
	Navicula	*N. transitans*		n.a.	n.a.	n.a.	-	
	Amphora	*sp.*	CCMP 2910	Jun 2006	Crinan Harbor, Scotland, UK	25	-	
	Phaeodactylum	*P. tricornutum*	Pt1	2003	Blackpool, England, UK	n.a.	-	
			Pt2	1910s	English Channel, UK	n.a.	-	
			Pt3	1930s	English Channel, UK	n.a.	-	
			Pt4	1951	Tvarminne, FIN	n.a.	-	
			Pt5	1972	West Dennis, Massachusetts, USA	n.a.	-	
			Pt6	1956	Woods Hole, Massachusetts, USA	n.a.	-	
			Pt7	1952	Long Island, Massachusetts, USA	n.a.	-	
			Pt8	Mar 1987	Vancouver, British Columbia, CA	n.a.	-	
			8Tc	n.a.	n.a.	n.a.	-	
			Pt9	Feb 1981	Guam, USA	n.a.	-	
			Pt10	2000	Dalian, Yellow Sea	n.a.	-	
			18.6	n.a.	n.a.	n.a.	-	

Polar Centrics	*Odontella*	*sp.*		July 2010	Newport, Oregon, USA	49	+	7
	Thalassiosira	*T. pseudonana*	CCMP 1007	1964	Chincoteague, Virginia, USA	8.4	+	6
			CCMP 1011	n.a.	St. Croix, US Virgin Islands	5.1	+	2
			CCMP 1012	1965	Swan River Estuary, Perth, AUS	7.1	-	
			CCMP 1013	1973	Conwy, Gwynedd, Wales, UK	6.6	+	4
			CCMP 1014	Nov 1971	North Pacific Gyre (open ocean)	5.1	+	4
			CCMP 1015	1985	San Juan Island, Washington, USA	13.2	+	4
			CCMP 1335	Sep 1958	Moriches Bay, NY, USA	5.3	+	4
			RCTP-1		Adriatic Sea	4.7	+	4
		T. rotula	St01-0501	n.a.	n.a.	9.2	+	5
		T. weissflogii	CCMP 1048	1976	East Falmouth, Massachusetts USA	16	+	4
	Ditylum	*D. brightwellii*	LB	n.a.	Long Beach, Washington, USA	2.9	+	1
	Skeletonema	*S. c.f. costatum*	RCC1716	May 2008	English Channel	n.a.	+	7
	Chaetoceros	*C. socialis*	CCMP205	Dec 1982	Narragansett Bay, Rhode Island, USA	2.7	+	4
Radial Centrics	*Corethron*	*C. hystrix*	CCMP308	May 1984	Fladenground, North Sea	25	+	n.a.
Heterokonts	*Bolidomonas*	*B. pacifica*	CCMP 1866	Dec 1994	South Pacific (open ocean)	3.9	-	
	Aureoumbra	*A. lagunensis*	CCMP 1507	Jun 1992	Laguna Madre, Texas, USA	1.9	-	
	Nannochloropsis	*N. oculata*	CCMP 525	1968	n.a.	3.4	-	
	Heterosigma	*H. akashiwo*	CCMP 452	1952	Long Island Sound, USA	18	-	
Red Algae	*Porphyridium*	*P. cruentum*	CCMP1328	1957	Woods Hole, Massachusetts, USA	0.49	-	

+ Lysed
- Resistant
*Virus:host ratios were determined by microscope based counts of the virus inoculum and the phytoplankton culture. Infectious units per cell vary from counted virus:cell ratios. For example, 1 infectious unit per cell for CLnn-16 and PC9 are 8 and 32 respectively.
n.a. not available
† ◊ Δ ¤ ITS1 region 100% identical

1.7 Supplemental methods

Algal culture conditions

All cultures were maintained in f/2 medium + silica (Guillard and Ryther, 1962; Guillard, 1975) at 13 °C at an irradiance of 26 μmol photons m^{-2} s^{-1} with cool white fluorescent illumination on a 16:8 light-dark cycle. Growth was monitored daily via chlorophyll-*a* fluorescence with a Turner 10-AU Fluorometer in glass test tubes or on a Spectra MAX m2 plate reader (Molecular Devices) in 48 well microtiter plates (Corning Inc. Costar).

Virus isolation

Surface seawater, 50L, was collected during a *Pseudo-nitzschia* bloom from Sunset Beach, Oregon (46°05'53N, 123°56'43W) on August 1, 2009. The water was pre-filtered through a 0.8 μm (Millipore) filter and concentrated to ~1L using a 30 kDa pore size tangential flow filtration system (Millipore). The concentrated sample was filtered through a 0.2 μm filter (Millipore) and the filtrate stored at 4 °C.

Cultures of *P. multiseries* CLnn-16, 25ml, were inoculated with 7 ml of concentrated environmental sample and monitored via chlorophyll-*a* fluorescence. Control cultures were inoculated with 7 ml of medium. Four days post-inoculation, 12 ml of the culture volume was filtered onto a 0.2 μm (Millipore) filter. The filter was then stamped onto 1% (wt/vol) low melting point agarose (Invitrogen) in f/2 medium plates to create lawns of *Pseudo-nitzschia*. The plates were incubated as described above temperature and light conditions and monitored for plaque formation. Plugs of agarose containing plaques were inoculated into exponentially growing *P. multiseries* CLnn-16. These cultures were monitored via chlorophyll-a fluorescence relative to a control, which was an agarose plug that contained healthily growing culture. To

ensure the virus was clonal, it was passaged through three dilution to extinctions (Suttle, 1993). Exponentially growing *P. multiseries* CLnn-16 was inoculated with viral lysate in replicates of 5 over 8 10-fold dilution series in 48 well plates. Wells in which culture death was observed and had the most dilute virus concentrations were 0.2 μm filtered and used to inoculate the next round of dilution to extinction experiments. The virus was defined as clonal after three rounds of dilution to extinction. A one liter volume of exponentially growing *P. multiseries* CLnn-16 was inoculated with the virus, lysed, 0.2 μm filtered and stored at 4 °C as a working stock for all later experiments.

Virus morphology

Virus samples were concentrated for electron microscopy with polyethylene glycol (PEG) precipitation. One liter of late exponential phase *Pseudo-nitzschia* CLnn-16 was infected with 30 ml viral working stock and cells were harvested after 2 days. The lysate was centrifuged at 4,000 x g for 15 minutes to pellet cell debris. The supernatant was then passed sequentially through 0.8, 0.45, and 0.2 μm filters. Polyethylene glycol 8000 (Sigma-Aldrich) was added to a final concentration of 10% wt/vol, incubated overnight at 4 °C, spun at ~25,000 x g for 1 hour and the pellet resuspended in 3 ml TE.

Copper electron microscope grids, carrying a Formvar film stabilized with carbon and glow discharged, were used to negative stain virus particles. Viral lysates were floated atop the grid and allowed to settle for 1 minute after which the grid was wicked dry. Grids were stained for 1 minute with a 5 μl drop of 2% uranyl acetate (pH 4) and wicked dry. Grids were viewed using a LEO 912 transmission electron microscope with a Proscan 2048 X 2048 pixel CCD camera. Virus size was measured using Olympus Soft Imaging System (SIS) software.

Nucleic acid characterization

PEG precipitated virus samples were digested with 50U DNase (Ambion Ribopure-Bacteria) for 30 minutes at 37 °C to remove contaminating host and bacterial DNA, and then extracted using a phenol:chloroform extraction (Thurber *et al.*, 2009). Nucleic acids were tested for bacterial contamination with polymerase chain reaction (PCR) using universal bacterial 16S rRNA primers (Lane, 1991) For detection of *Pseudo-nitzschia* genomic DNA, eukaryotic PCR primers 18SF and 5.8SR were used as in Hubbard et al., 2008 to amplify the full-length ITS1 sequence. Only viral DNA that was free from bacterial and eukaryotic contamination was used for further analysis.

Viral nucleic acids were quantified using Quant-iT Assays (Invitrogen) to detect dsDNA, ssDNA, and RNA. First, viral nucleic acids were quantified each of the three Quant-iT Assays. Next, the Nucleic acids were digested with 0.67 U DNase (Ambion) and 3.33 U S1 nuclease (Promega) in 50 µl reactions at 37 °C for 1 hour. Controls consisted of reactions containing buffer and nuclease and buffer only served as negative controls. Digestions were performed in triplicate, and the nucleic acid concentration quantified with both the dsDNA assay and the ssDNA assay.

Viral nucleic acids were mechanically sheared using a Hydroshear (Digilab) to 1500 bp and ligated into pUC 2.4 plamids using the TOPO TA cloning kit. Chemically competent *E. coli* were transformed with the plasmids and plated on 50 µg kanamycin Luria brother agar plates. Plates were incubated at 37 °C for 24 hours. Colonies were counted and screened for inserts using PCR with M13 primers and running the resulting PCR products on an electrophoresis gel.

Genome sequencing and analysis

Cultures of *P. multiseries* GGB1 and *Thalassiosira pseudonana* CCMP1335 were inoculated with the virus and lysed. Lysates were purified by sequential filtering and PEG precipitatation and DNA was extracted and tested for contamination as described above. Purified viral DNA was converted to double-stranded DNA (dsDNA) with rolling circle amplification (Genomiphi). Viral dsDNA from the *P. multiseries* culture was mechanically sheared into 1500bp fragments and clone libraries were constructed using the TOPO TA cloning kit (Invitrogen) using the manufacturers instructions. Insert DNA in plasmids from 384 colonies was Sanger-sequenced from both ends using the T7 and T3 primer sites. Sequence reads were trimmed for vector sequence and quality and reads that overlapped with their mates were assembled into contigs. After quality trimming, 475 sequences had an N50 of 713bp.

Viral DNA from the *T. pseudonana* culture was sequenced using PacBio resulting in 92,368 reads with an average length of 5,679bp. PacBio reads were assembled using the Heirarchical Genome Assembly Process v2.0 with quiver polishing using the default options. Contigs were assembled using Sequencher v4.9. To verify the assembly, Sanger sequencing reads were mapped to the PacBio contigs using Sequencher. Open reading frames (ORFs) were called using GeneMark.hmm v3.0.2 (Besemer and Borodovsky, 1997) and annotated based on the top 5 BLAST consensus (e-value = 1e-3) with at least 50% sequence identity with hit sequences, and bit score >50 against nr (November 30, 2015). ORFs not meeting these criteria were labeled hypothetical.

Phylogenies of the polymerase B (polB), capsid, replication initiator, and helicase genes were constructed using RAxML v 8.0.23 (Stamatakis, 2012). Briefly, reference protein sequences were downloaded from GenBank and aligned with Muscle v 3.8.31 (Edgar, 2004). A

maximum likelihood tree was inferred from 20 starting trees using the amino acid substitution model models VTF for polB, RTREV for the capsid gene, WAGF for the replication intiator gene, and BLOSUM62F for the helicase gene.

Quantification of viral genomes

Genome segment-specific qPCR primers were designed using Primer3 (Rozen and Skaletsky 2000). Genome fragment 1 was targeted by the primers G1H-F: 5'-TAGCGATGGACAGCAAACAG-3' and G1H-R: 5'-AGGAGCGATTTGACGTGTTC-3'. Genome fragment 2 was targeted by the primers G2e-F: 5'-ACCGACAAAACAGGAGCATC-3' and G2E-R: 5'- TCCGCAATCTCTTGACTGTG-3'. Primers were tested *in silico* for cross amplification of untargeted genome regions. Standards for qPCR were generated by PCR amplification of viral DNA consisting of 95 °C for 5 minutes followed by 30 cycles of 95 °C for 30 s, 60 °C for 45 s, and 72 for 60 s. Reactions consisted of 1X reaction buffer, 1.5mM $MgCl_2$, .8 mM dNTPs, 1μM forward and review primer, 10ng viral template, and 1 U *Taq* polymerase in 20 μl final volume. Amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter) and ligated into pCR2.1 vector and cloned into *Escheria coli* using the TOPO TA Cloning Kit. Positive colonies were grown overnight at 37 °C in Luria broth medium and the plasmids harvested using Plasmid Miniprep Kit (Qiagen). Isolated plasmids were linearized with *XbaI* (New England Biolabs) at 37 °C for 2 hours. Linearization was confirmed with gel electrophoresis, and the linearized plasmid was cleaned with AMPure beads. DNA concentration was quantified with a Qubit HS DNA assay and used to calculate the number of copies μl^{-1} used in the standards.

qPCR consisted of 20 µl reactions with 5 µl of sample template or linearized plasmid standard, 300 nmol forward and reverse primer, 0.5 µl ROX reference dye (Invitrogen), and 10 µl of iQ Supermix (BioRad). Reactions were carried out in a ABI One Step Plus Real Time Thermocycler with an initial incubation at 95 °C for 3 minutes followed by 40 cycles of 95 °C for 30 s, 58 °C for 45 s, and 72 for 60 s. A step-wise melt curve followed the cycling conditions to verify a single melt peak. A dilution series of standards and no-template controls were amplified at the same time as unknown samples and copy numbers in unknown samples were calculated relative to the standard curve by ABI One Step Plus software.

Verification of viral genome segment relationship

Axenic *T. pseudonana* CCMP 1335 cultures were inoculated in 15 independent experiments with a 1:2 dilution series of viral lysate, with each dilution having 8 replicates. Experiments were monitored in 48 well microplates (Corning) for 3 weeks. Most probable number (MPN) tables determined the number of viruses in lysed wells from the most dilute virus inoculum. DNA from these wells was extracted using a phenol-chloroform extraction and both genome segments were quantified using qPCR.

Viral lysate, 4ml, was adjusted with cesium chloride to 1.12 g ml^{-1} and loaded onto two different columns, one with density gradients with 1 ml layers of 1.2, 1.3, 1.4, 1.5, and 1.7 g ml^{-1} and the second with 1 ml layers of 1.25, 1.35, 1.45, 1.55, and 1.7 g ml^{-1} in thin walled ultracentifuge tubes (Beckman). The columns were centrifuged at ~80,000 xg (22,000rpm) for 2 hours at 4 °C in a Beckman SW 41 Ti rotor. Each layer (1mL) was carefully extracted using a 20 gauge needle and syringe. Viral DNA was extracted as above. The abundance of each genome segment was quantified using qPCR.

Host range

Algal cultures in supplemental table 2 were grown to mid-exponential phase. Seven serial ten-fold dilutions of the virus stock (100 µl) were inoculated into 900 µl algal cultures in quadruplicate in 48-well plates. Four control wells used sterile filtered and autoclaved natural seawater as inoculum. 1 ml algal culture was preserved with 40 µl acidified Lugol's solution at time 0 for cell counting on a Sedgwick Rafter slide or haemocytometer to determine virus to host ratios (Wildlife Supply Company). Infected cultures were monitored via fluorescence until control cultures declined, approximately two weeks.

Virus to host ratio determination

The infectious units per volume for one-step growth curves were determined by a fine-scale dilution series. A range of viral lysates, from 50% to 0.1% of total volume, was inoculated in quadruplicate into exponentially growing axenic *P. multiseries* PC9 or axenic *T. pseudonana* CCMP 1335. Control wells had filtered f/2 medium added to 10% of the total volume. Chlorophyll-a fluorescence measurements were taken every four hours until the cultures had lysed as defined by a decrease of half the maximum fluorescence. At t=0, 1ml of culture was preserved for cell counts in 40 µl acidified Lugol's solution. The appropriate viral to host ratio was predicted to lie between the dilution that exhibited a uniform rate of decline and the dilution in which the rate of decline was variable over the infection, indicating the entire culture was not infected. At least 300 cells were quantified by counting on a Sedgwick-Rafter or haemocytometer (Wildlife Supply Company). One milliliter of the virus stock solution was gently filtered onto a 0.02 µm filter (Whatman), stained for 15 minutes with a 2X concentration

of SYBR gold, and visualized in the FITC channel using a Nikon Eclipse 80i epifluorescent microscope and Photometrics Cool Snap HQ^2 camera (Thurber *et al.*, 2009). For virus counts, 10 random fields per filter were captured at 1000X magnification and counted.

One step growth curve

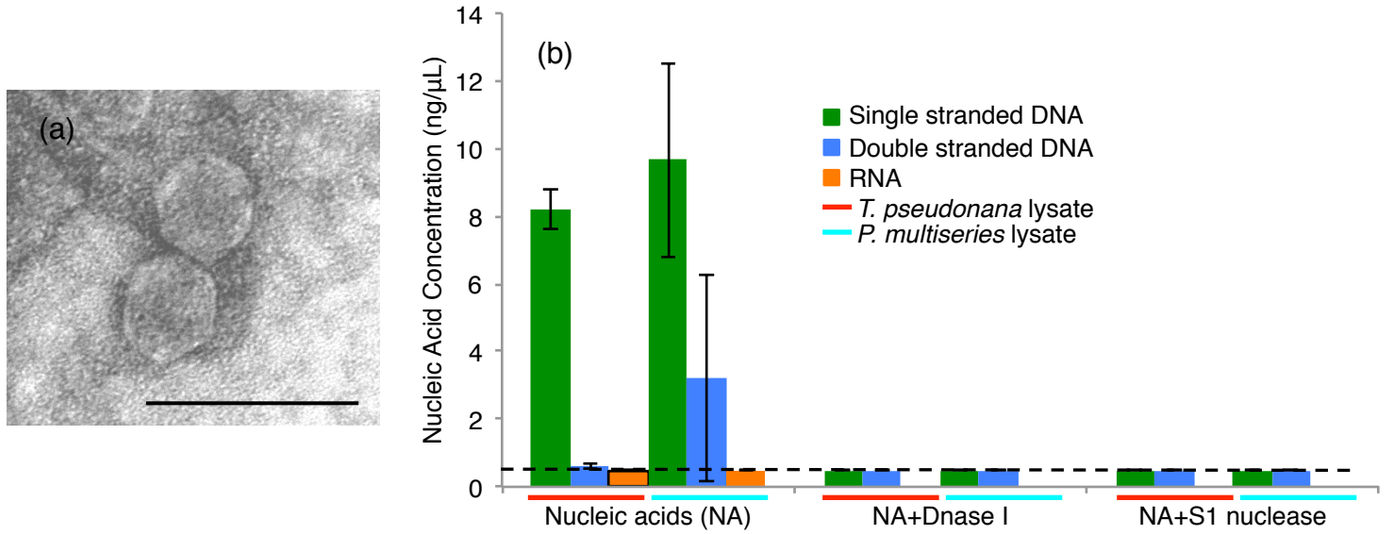
Triplicate exponentially growing 30 ml cultures of axenic *P. multiseries* PC9 or axenic *T. pseudonana* CCMP 1335 were inoculated with either the working stock viral lysate at the appropriate concentration for that strain or control lysate that had been heated at 90 °C for 15 minutes and UV irradiated for 15 minutes at the same volume. In all experiments, the cultures were regularly sampled for cell counts, and virus counts. Five hundred µl of culture was fixed with paraformaldehyde to a final concentration of 2%, and viruses were counted using epifluorescent microscopy as detailed above. For diatom counts, 10 random fields per filter were counted at 100x magnification, with unlysed diatoms detected by the fluorescence of an intact nucleus. Both genome segments were quantified with qPCR as described above.

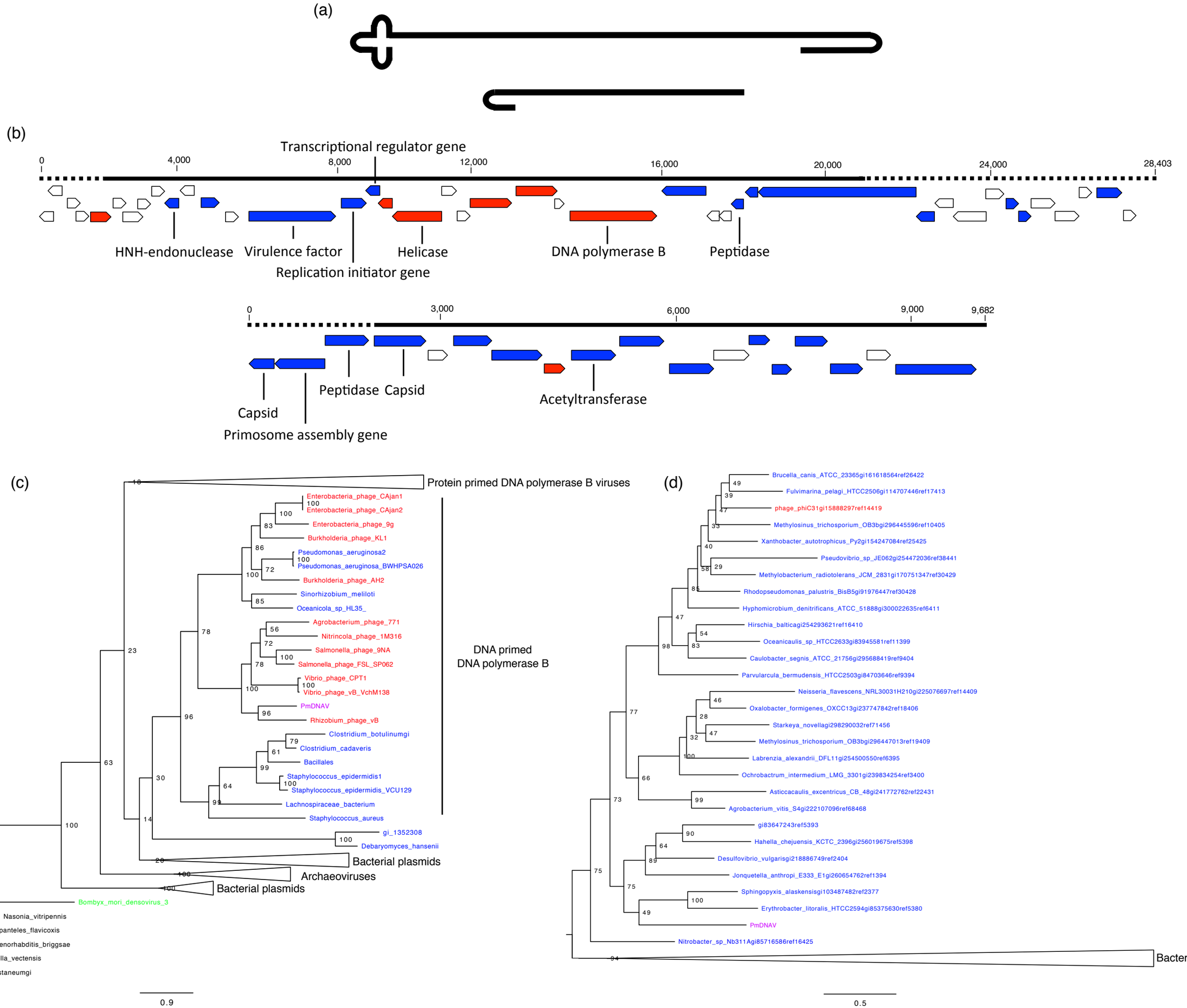
In the *P. multiseries* PC 9 experiment, samples from one replicate of the infected cultures were fixed with 1% glutaraldehyde and stored at 4 °C before DNA extraction. In the *T. pseudonana* experiment, samples from each infected and control replicate were first passed through a 1.2 µm filter. DNA from the free virus containing filtrate was then extracted.

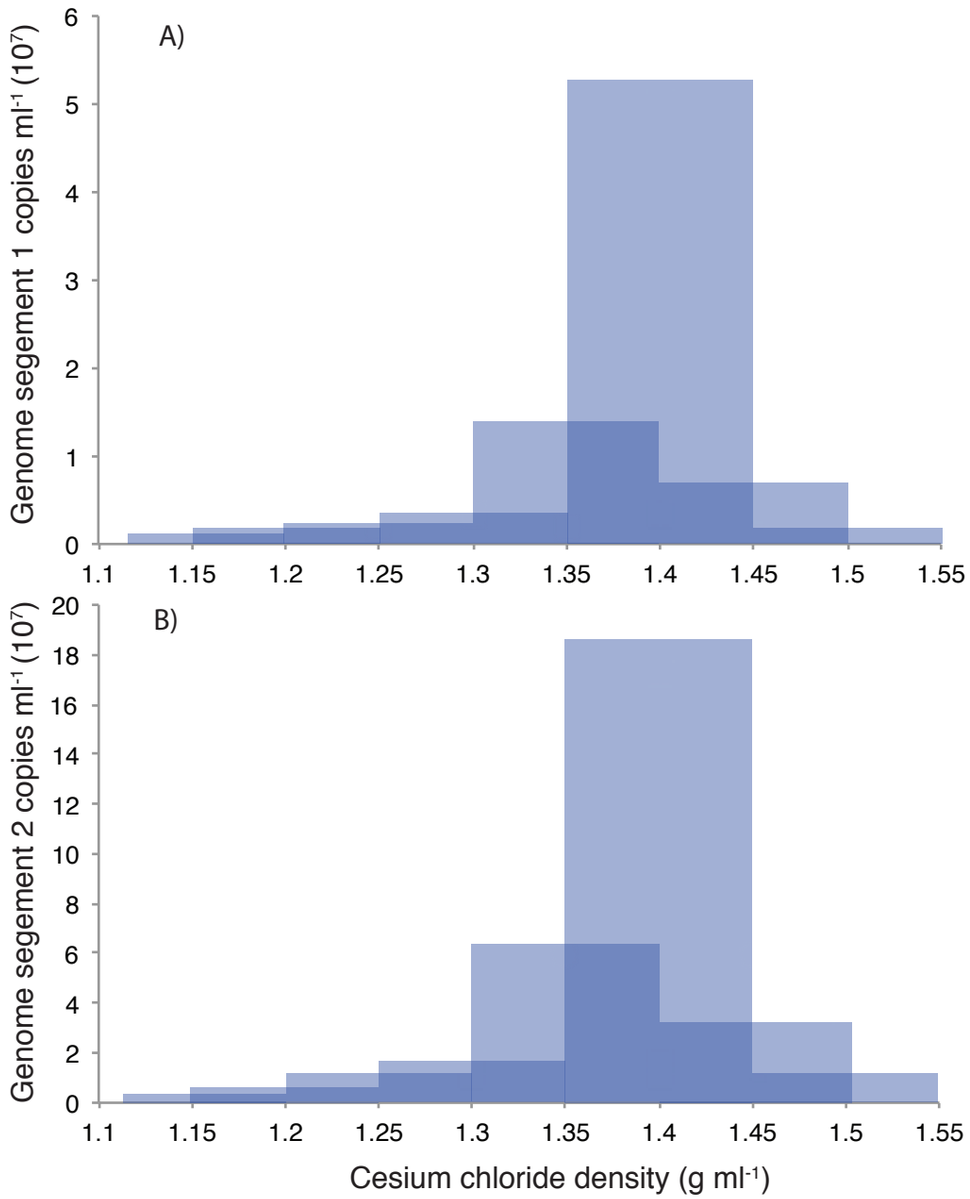
Pseudo-nitzschia *culture isolation and genotyping*

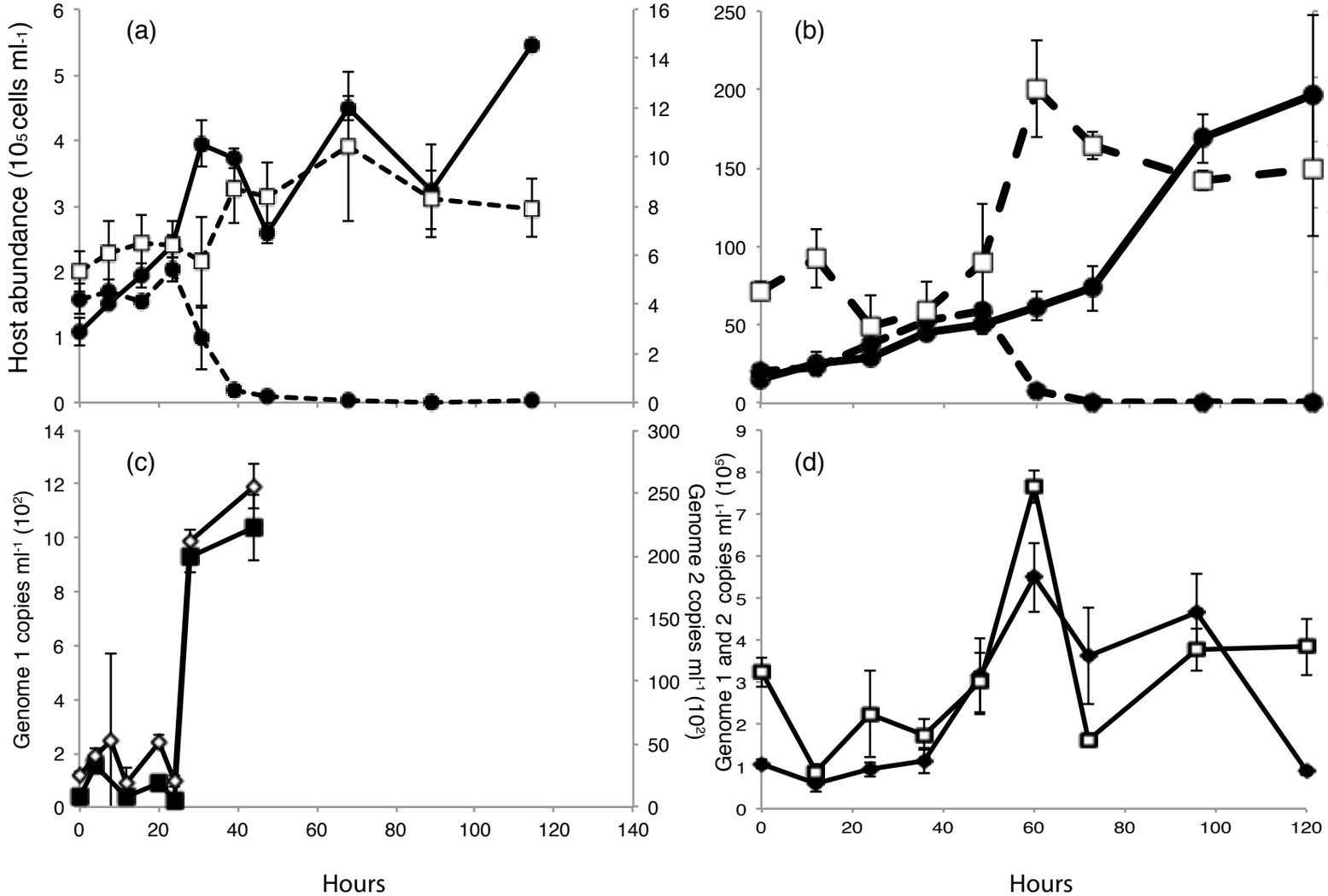
Pseudo-nitzschia cultures were isolated from 20 µm mesh net tow samples from Penn Cove, WA (48°13'24N 122°41'14W), Golden Gardens, WA (47°41'27N 122°24'16W), East Sound, WA (48°40'49N 122°53'57W), Sequim, WA (48°404'12N 123°02'33W), Sunset Beach,

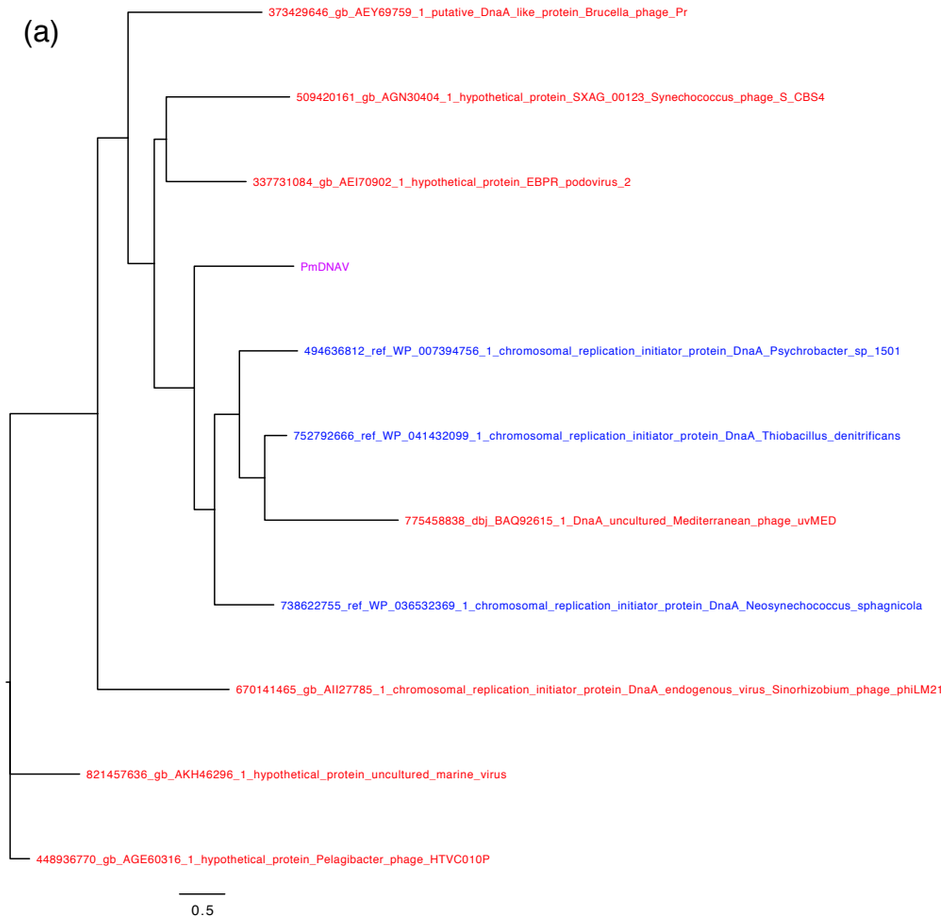
OR (46°05’53N, 123°56’43W) and Newport Beach, CA (33°35'54N 117°54'03W) by micropipetting single cells. Isolates were deemed uni-algal after 3 transfers where only one cell type was observed based on light microscopy. For ITS1 sequencing, cultures were collected on a 2 μm filter and DNA was extracted using a DNeasy Plant Mini Kit (Qiagen). The ITS1 region was amplified using primers 18SF and 5.8SR (Hubbard et al., 2008). PCR products were purified and sequenced at University of Washington Department of Biochemistry DNA Sequencing Facility. Sequences were identified taxonomically based on greater than 97% similarity to sequences of SEM verified cultures in GenBank. MUSCLE (Edgar, 2004) was used for alignments and pairwise percent identity calculations. Sequences are deposited in Genbank under accession numbers KF015190-KF015204.

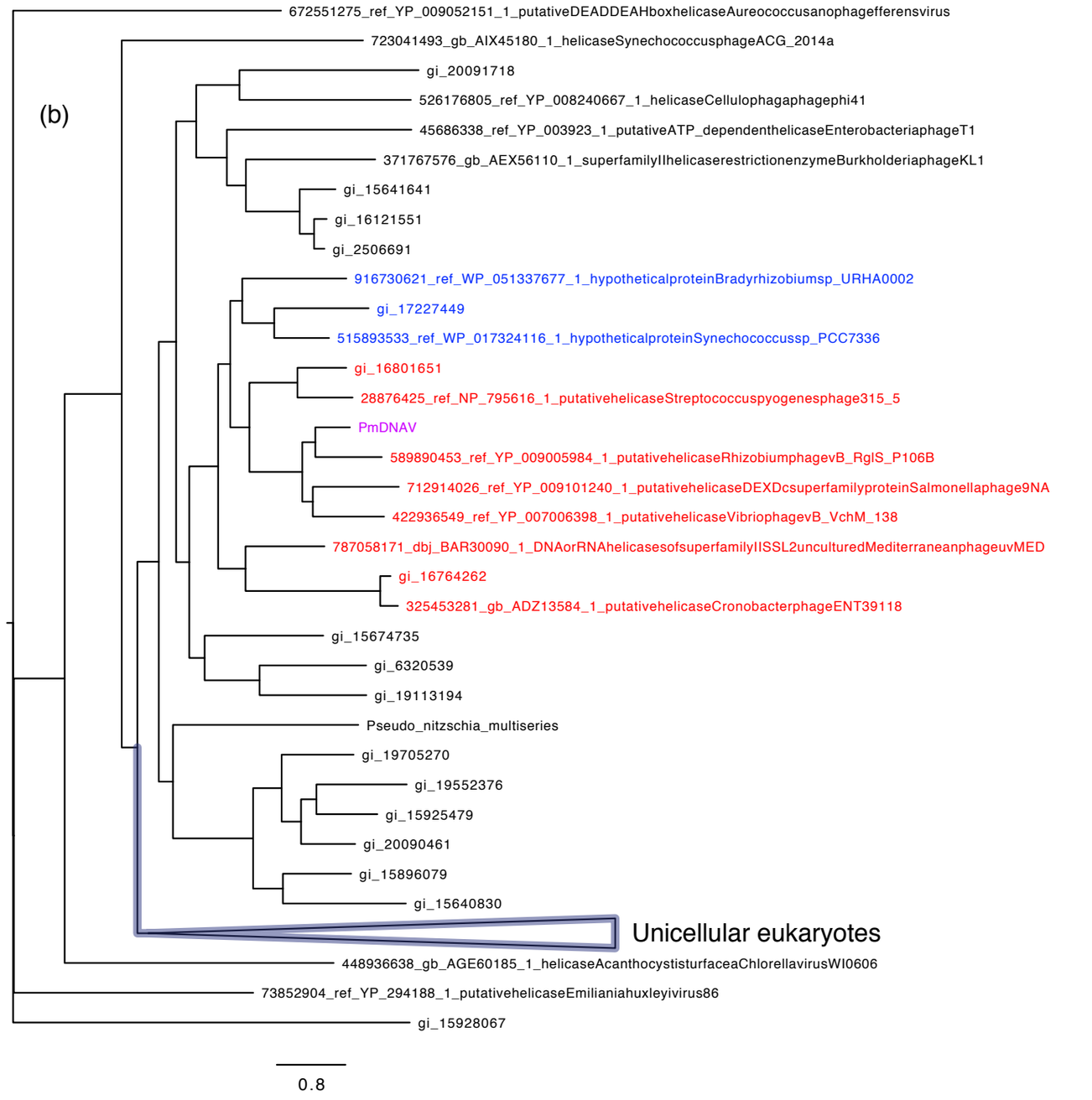

(a)
(b)
Nucleic Acid Concentration (ng/µL)
14
12
10
8
6
4
2
0
Single stranded DNA
Double stranded DNA
RNA
T. pseudonana lysate
P. multiseries lysate
Nucleic acids (NA)
NA+Dnase I
NA+S1 nuclease


(a)
(b)
Transcriptional regulator gene
0
4,000
8,000
12,000
16,000
20,000
24,000
28,403
HNH-endonuclease
Virulence factor
Replication initiator gene
Helicase
DNA polymerase B
Peptidase
0
3,000
6,000
9,000
9,682
Peptidase
Capsid
Capsid
Primosome assembly gene
Acetyltransferase
(c)
Protein primed DNA polymerase B viruses
DNA primed
DNA polymerase B
Enterobacteria_phage_CAjan1
Enterobacteria_phage_CAjan2
Enterobacteria_phage_9g
Burkholderia_phage_KL1
Pseudomonas_aeruginosa2
Pseudomonas_aeruginosa_BWHPSA026
Burkholderia_phage_AH2
Sinorhizobium_meliloti
Oceanicola_sp_HL35_
Agrobacterium_phage_771
Nitrincola_phage_1M316
Salmonella_phage_9NA
Salmonella_phage_FSL_SP062
Vibrio_phage_CPT1
Vibrio_phage_vB_VchM138
PmDNAV
Rhizobium_phage_vB
Clostridium_botulinumgi
Clostridium_cadaveris
Bacillales
Staphylococcus_epidermidis1
Staphylococcus_epidermidis_VCU129
Lachnospiraceae_bacterium
Staphylococcus_aureus
gi_1352308
Debaryomyces_hansenii
Bacterial plasmids
Archaeoviruses
Bacterial plasmids
Bombyx_mori_densovirus_3
Nasonia_vitripennis
Glyptapanteles_flavicoxis
Caenorhabditis_briggsae
Nematostella_vectensis
Tribolium_castaneumgi
0.9
(d)
Brucella_canis_ATCC_23365gi161618564ref26422
Fulvimarina_pelagi_HTCC2506gi114707446ref17413
phage_phiC31gi15888297ref14419
Methylosinus_trichosporium_OB3bgi296445596ref10405
Xanthobacter_autotrophicus_Py2gi154247084ref25425
Pseudovibrio_sp_JE062gi254472036ref38441
Methylobacterium_radiotolerans_JCM_2831gi170751347ref30429
Rhodopseudomonas_palustris_BisB5gi91976447ref30428
Hyphomicrobium_denitrificans_ATCC_51888gi300022635ref6411
Hirschia_baltica gi254293621ref16410
Oceanicaulis_sp_HTCC2633gi83945581ref11399
Caulobacter_segnis_ATCC_21756gi295688419ref9404
Parvularcula_bermudensis_HTCC2503gi84703646ref9394
Neisseria_flavescens_NRL30031H210gi225076697ref14409
Oxalobacter_formigenes_OXCC13gi237747842ref18406
Starkeya_novellagi298290032ref71456
Methylosinus_trichosporium_OB3bgi296447013ref19409
Labrenzia_alexandrii_DFL11gi254500550ref6395
Ochrobactrum_intermedium_LMG_3301gi239834254ref3400
Asticcacaulis_excentricus_CB_48gi241772762ref22431
Agrobacterium_vitis_S4gi222107096ref68468
gi83647243ref5393
Hahella_chejuensis_KCTC_2396gi256019675ref5398
Desulfovibrio_vulgarisgi218886749ref2404
Jonquetella_anthropi_E333_E1gi260654762ref1394
Sphingopyxis_alaskensisgi103487482ref2377
Erythrobacter_litoralis_HTCC2594gi85375630ref5380
PmDNAV
Nitrobacter_sp_Nb311Agi85716586ref16425
Bacteria
0.5


A)
Genome segement 1 copies ml-1 (107)
6
5
4
3
2
1
0
1.1
1.15
1.2
1.25
1.3
1.35
1.4
1.45
1.5
1.55
B)
Genome segement 2 copies ml-1 (107)
20
18
16
14
12
10
8
6
4
2
0
Cesium chloride density (g ml-1)


(a)
(b)
(c)
(d)
Host abundance (10$_5$ cells ml^{-1})
Viruses cell^{-1} (10^2)
Genome 1 copies ml^{-1} (10^2)
Genome 2 copies ml^{-1} (10^2)
Genome 1 and 2 copies ml^{-1} (10^5)
Hours
Hours


(a)
373429646_gb_AEY69759_1_putative_DnaA_like_protein_Brucella_phage_Pr
509420161_gb_AGN30404_1_hypothetical_protein_SXAG_00123_Synechococcus_phage_S_CBS4
337731084_gb_AEI70902_1_hypothetical_protein_EBPR_podovirus_2
PmDNAV
494636812_ref_WP_007394756_1_chromosomal_replication_initiator_protein_DnaA_Psychrobacter_sp_1501
752792666_ref_WP_041432099_1_chromosomal_replication_initiator_protein_DnaA_Thiobacillus_denitrificans
775458838_dbj_BAQ92615_1_DnaA_uncultured_Mediterranean_phage_uvMED
738622755_ref_WP_036532369_1_chromosomal_replication_initiator_protein_DnaA_Neosynechococcus_sphagnicola
670141465_gb_AII27785_1_chromosomal_replication_initiator_protein_DnaA_endogenous_virus_Sinorhizobium_phage_phiLM21
821457636_gb_AKH46296_1_hypothetical_protein_uncultured_marine_virus
448936770_gb_AGE60316_1_hypothetical_protein_Pelagibacter_phage_HTVC010P
0.5


(b)
672551275_ref_YP_009052151_1_putativeDEADDEAHboxhelicaseAureococcusanophagefferensvirus
723041493_gb_AIX45180_1_helicaseSynechococcusphageACG_2014a
gi_20091718
526176805_ref_YP_008240667_1_helicaseCellulophagaphagephi41
45686338_ref_YP_003923_1_putativeATP_dependenthelicaseEnterobacteriaphageT1
371767576_gb_AEX56110_1_superfamilyIIhelicaserestrictionenzymeBurkholderiaphageKL1
gi_15641641
gi_16121551
gi_2506691
916730621_ref_WP_051337677_1_hypotheticalproteinBradyrhizobiumsp_URHA0002
gi_17227449
515893533_ref_WP_017324116_1_hypotheticalproteinSynechococcussp_PCC7336
gi_16801651
28876425_ref_NP_795616_1_putativehelicaseStreptococcuspyogenesphage315_5
PmDNAV
589890453_ref_YP_009005984_1_putativehelicaseRhizobiumphagevB_RglS_P106B
712914026_ref_YP_009101240_1_putativehelicaseDEXDcsuperfamilyproteinSalmonellaphage9NA
422936549_ref_YP_007006398_1_putativehelicaseVibriophagevB_VchM_138
787058171_dbj_BAR30090_1_DNAorRNAhelicasesofsuperfamilyIISSL2unculturedMediterraneanphageuvMED
gi_16764262
325453281_gb_ADZ13584_1_putativehelicaseCronobacterphageENT39118
gi_15674735
gi_6320539
gi_19113194
Pseudo_nitzschia_multiseries
gi_19705270
gi_19552376
gi_15925479
gi_20090461
gi_15896079
gi_15640830
Unicellular eukaryotes
448936638_gb_AGE60185_1_helicaseAcanthocystisturfaceaChlorellavirusWI0606
73852904_ref_YP_294188_1_putativehelicaseEmilianiahuxleyivirus86
gi_15928067
0.8

Chapter 2

Pseudo-nitzschia challenged with co-occurring viral communities display diverse infection phenotypes

2.1 Abstract

Viruses are catalysts of biogeochemical cycling, architects of microbial community structure, and terminators of phytoplankton blooms. Diatoms are eukaryotic phytoplankton that contribute to carbon export and productive food webs in the ocean. The impact of viruses on diatom abundance and community composition is unknown. Diatom-virus dynamics were explored by sampling every month at 2 coastal and estuarine locations in Washington state, USA resulting in 41 new isolates of the pennate diatom *Pseudo-nitzschia* and 20 environmental virus samples. We conducted a total of 820 pair-wise crosses of the *Pseudo-nitzschia* isolates and viral communities. Viral communities infected *Pseudo-nitzschia* isolates in 8% of the crosses and were most susceptible to infection when the host and viral communities were isolated from the same sample. Isolates ranged in their permissivity to infection with some isolates not infected by any viral samples and others infected by up to 10 viral communities. Isolates that were infected by the most viral communities yielded the highest viral titers (as high as 16000 infectious units ml^{-1}). Titers were host dependent, as titers for one viral sample on 8 different hosts spanned 4 orders of magnitude. A comparison of *Pseudo-nitzschia* Internal Transcribed Spacer 1 sequences revealed multiple subgroups of genetically identical hosts that were differentially infected. Monospecific blooms of *Pseudo-nitzschia* were composed of multiple viral infection phenotypes. This phenotypic diversity may be a defense against viral termination of diatom blooms. The interactions between *Pseudo-nitzschia* and the viral communities highlight the

diversity of diatoms and emphasize the complexity and variability of diatom-virus dynamics in the ocean.

2.2 Introduction

In the ocean, viral infection links microbial community structure, biogeochemical cycling, and microbial evolution (Breitbart, 2012). Viruses regulate marine phytoplankton communities by impacting host abundance and diversity through cell lysis (Weitz and Wilhelm, 2012). Viruses and their hosts are thought to cycle dynamically, with encounter rates favoring infection of dominant microbial taxa, which are removed due to lysis and then supplanted by new microbial populations that fill the vacant ecological niche (Thingstad, 2000). These 'Kill-the-Winner' dynamics have important, but often cryptic scales of interaction are thought to occur at varying temporal, spatial, and taxonomic levels (Thingstad 2000; Thingstad et al. 2015). Understanding the scales of host-virus interactions is critical for accurately quantifying viral contributions to microbial mortality.

Host permissivity to viral infection and viral host range are important mechanisms that underlie kill-the-winner dynamics and directly affect the success of viruses in the ocean. Cultured marine host-virus systems suggest that viruses range from generalists to specialists, while hosts range in their susceptibility to viral infection from highly permissive to resistant; the hierarchical ordering of these properties in hosts and viruses is known as nestedness (Flores et al. 2011). Additionally, hosts and viruses can coevolve to form modules defined by viruses capable of infecting only hosts within a given module. These patterns have been observed in wild host-virus communities. Phage isolated from a transect across the Atlantic were most infective of co-occurring host bacteria and formed modules by geographic separation (Flores et al. 2012). On the

other hand, Waterbury and Valois (1993) demonstrated that *Synechococcus* phage titers over two years at the same location were not inversely correlated with *Synechococcus* abundance and thus were unimportant in controlling co-occurring cyanobacteria populations. These divergent results may be due to the small sample sizes of isolation based studies and the timing of host population cycling: isolated hosts may be in the process of being removed by their co-occurring viruses, or they may represent the supplanting microbial population that is resistant to the dominant viruses in the water. Thus, co-occurring resistance and susceptibility fluctuate in Kill-the-Winner dynamics such that both scenarios are plausible.

The dramatic boom and bust life cycles of eukaryotic phytoplankton pose both challenges and opportunities for viruses. Eukaryotic phytoplankton blooms reach high cell densities, are often composed of few species, which may be excellent conditions for viral infection (Brussaard 2004; Armbrust 2009). Viral termination of blooms has been observed in eukaryotic phytoplankton–virus systems such as *Emiliania, Phaeocystis, Heterosigma, Aureococcus,* and *Micromonas* (Bratbak et al. 1993; Cottrell and Suttle 1995; Tarutani et al. 2000; Gobler et al. 2004; Baudoux et al. 2006; Vardi et al. 2009, Lehahn et al., 2014). Under non-bloom conditions, viruses of eukaryotic phytoplankton must survive times of host scarcity since the propagation of viruses relies on contact rates between hosts and viruses. Viruses may rely on alternative strategies of propagation such lysogeny and latent infections (McDaniel et al. 2002, Thyrhaug et al. 2003), or sequestration in sediments (Tomaru et al. 2011a) during these times. Ultimately, the reduction of viral abundance during times of host scarcity may be a mechanism that eventually allows phytoplankton to increase in abundance without immediate infection by viruses.

Diatoms are a group of diverse and ubiquitously distributed eukaryotic phytoplankton that exhibit a boom-and-bust life cycle. They dominate primarily in temperate coastal and polar

oceans where they can form massive blooms, which fuel carbon export and productive food webs (Nelson et al. 1995). *Pseudo-nitzschia* is a cosmopolitan genus within the diatoms consisting of 37 described species (Lelong et al. 2012; Trainer et al. 2012). *Pseudo-nitzschia* is particularly known for the ability to produce the neurotoxin domoic acid, which can be biomagnified through food webs and can disrupt ecosystems and create public health concerns (Bates et al. 1989; Scholin et al. 2000). Toxin production varies by species (Trainer et al. 2012), underscoring the importance of *Pseudo-nitzschia* community structure for understanding toxic bloom formation.

The first diatom viruses were isolated and characterized only a decade ago (Nagasaki et al. 2004) and since then the number of diatom viruses has grown to 15, isolated on 4 genera, the centric diatoms *Rhizosolenia*, *Chaetoceros*, and the pennate diatoms, *Asterionellopsis* and *Thalassionema* (Bettarel et al. 2005, Eissler et al. 2009, Kimura and Tomaru, 2015). All diatom viruses have fallen into two groups based on their nucleic acid content, either single stranded RNA or single stranded DNA. This is in contrast to the majority of model eukaryotic phytoplankton – virus systems that involve large double stranded DNA viruses, primarily from the Phycodna- and Megaviridae families (Nagasaki and Bratbak 2010; Moniruzzaman et al. 2014). Furthermore, the host ranges of diatom viruses are narrow. Only a few diatom viruses, such as CdebDNAV and RsRNAV, have been shown to infect multiple hosts, all within the same species (Nagasaki et al. 2004; Tomaru et al. 2008; Kimura and Tomaru 2015).

Diatom viruses are genomically and functionally different than other viruses that infect other photosynthetic marine eukaryotes. Diatoms exhibit similar boom-and-bust dynamics other photosynthetic eukaryotes. Therefore are the dynamics between diatoms and their viruses similar in their capacity to control diatom populations and terminate blooms? Diatom viral infectivity

based on titers of virus concentration performed on one strain of *Chaetoceros gracilis* fluctuated seasonally, reaching a maximum during the early spring when treated with environmental viral communities from Chesapeake Bay (Bettarel et al. 2005). Similarly, viral infection of one strain of *Chaetoceros tenuissimus* consistently peaked in the late summer and fall from water and sediment samples taken from coastal Japan (Tomaru et al. 2011a). Additionally, Tomaru et al. (2011a) designed qPCR primers that were specific to *C. tenuissimus* and *C. salsugineum*, both of which have been used to isolate viruses (Shirai et al. 2008, Nagasaki et al. 2005). Host and environmental virus abundance were monitored over the span of several years with qPCR and titers, however, no correlation between fluctuations in the *Chaetoceros* species abundance and virus abundance were found (Tomaru et al. 2011a). While only three diatom cultures total have been tested, these results highlight an incongruity between diatom and virus dynamics, which may be occurring on unknown scales.

In this study, we assessed the temporal and spatial scales of diatom–virus interactions and quantified patterns of infection and host permissivity. *Pseudo-nitzschia* were isolated and challenged with viral communities taken every month from two locations in the Pacific Northwest. The resulting patterns in infectious crosses between *Pseudo-nitzschia* and the virus communities were used to understand how viral infection changed between locations and in time, how infection patterns correlated with host genotype, and how these interactions might shape *Pseudo-nitzschia* communities in the field.

2.3 Methods

Environmental virus sample collection and purification

Samples were collected at Penn Cove, Washington (48.2397, -122.6795) and Grays Harbor, Washington (46.7625, -124.0898) (figure 1) monthly from April 2013 to April 2014 except from November to February, when sampling occurred at Penn Cove in November and January and at Grays Harbor in December and February. Approximately 15L of surface water was filtered through triplicate 3.0 μm and 0.2 μm 147mm polyethersulfone filters (Millepore) with a peristaltic pump within 2 hours of sampling. The filters were frozen at -80°C for molecular analysis. The viruses in the filtrate were precipitated by adding iron chloride (1g L^{-1}) and incubating for 12 hours at 13°C in the dark (John et al. 2011). The iron-precipitated viruses were collected by filtering on to 2.0μm and 0.2μm 147 polycarbonate filters (Millepore). The viruses were resuspended from both filters by incubating them with 0.2M ascorbate-0.25M EDTA-Mg2-0.25M Tris-HCL for at least 24 hours with periodic shaking (John et al. 2011). The resuspended viruses were purified to remove Fe-EDTA complexes by adding 10% poly ethylene glycol (PEG) and 2% NaCl final concentrations and incubating at 4°C for 24 hours. The samples were then spun at 12000 xg for 20 minutes at 4°C to pellet the viruses. The supernatant was removed and the pellet washed three times with modified SM buffer (0.4M NaCl, 0.02M MgSO4, 0.05M Tris, pH 7.5). Washing and pelleting were repeated twice. The PEG was removed according to Colombet and Sime-ngando (2012) by adding 2M KCl and incubating the samples on ice and removing the supernatant. The purified virus samples were stored at 4°C.

Pseudo-nitzschia community analysis

Subsections of the 3.0 um filters collected during sampling events were extracted using a Qiagen DNeasy minikit. The Internal Transcribed Spacer 1 (ITS1) was amplified for Automated Ribosomal Intergenic Spacer Analysis (ARISA) using triplicate PCR amplifications of the

purified DNA of environmental DNA or culture DNA using primers PnALLF and 6FAM-PnALLR for 34 cycles with conditions described in Hubbard et al. (2008). Reactions were purified with a Qiaquick PCR Purification Kit, and sequenced at the University of Washington Fred Hutchinson Research Center. Two blanks, one water and one negative control from the ARISA PCR, were used to establish a noise baseline for each run. ARISA reactions were confirmed to be in linear amplification phase at cycle 34 for semiquantitive analysis using iQ SYBR green supermix (BioRad) with the same DNA and primer concentrations and cycling conditions as the reactions above.

Electropherograms of the ARISA runs were analyzed using PeakScanner (Life Technologies). Peaks were called if they were above a 50:1 signal to noise ratio, between 120 and 300 bp in length, and represented at least 3% of the total fluorescence. Peak area was summed over two base pair bins and divided by total fluorescence, which gave a relative abundance. Peaks were identified using predicted ITS lengths from the *Pseudo-nitzschia* isolates as well as from reported ITS lengths in Hubbard et al. 2008 and Hubbard et al. 2014. The resulting community profiles were analyzed using the statistics package Primer-6 (Clarke and Warwick 2001). Similarity between community profiles was calculated using both Bray-Curtis and Jaccard matrices. Matrices were clustered and tested for significance using a SIMPROF test.

Isolation and identification of Pseudo-nitzschia

A 10 um net was hand-towed through the water for approximately 5 minutes during each sampling event. Single chains of *Pseudo-nitzschia* were picked with a pipette and purified through three washes with f/20 medium. All cultures in this study were maintained in f/2 medium + silica at 13 °C at an irradiance of 26.7 μmol photons m^{-2} s^{-1} with cool white

fluorescent illumination on a 16:8 light-dark cycle. Isolated strains were verified by light microscopy to be unialgal but not axenic.

Cultures were grown to mid-exponential phase and centrifuged at 5000 x g for five minutes to pellet cells. DNA was extracted with a DNeasy plant minikit (Qiagen) according to the manufacturer's protocol. The ITS1 amplification was based on the methods in Hubbard et al. (2008). PCR primers Euk-18SF and Euk-5.8SR were used to amplify the full-length ITS1 sequence of the *Pseudo-nitzschia* strains. PCR amplicons were purified with Qiaquick PCR Purification Kit and sequenced using Euk-18SF and Euk-5.8SR primers with Sanger sequencing at Genewiz (Seattle, WA) and University of Washington High Throughput Sequencing Center (Seattle, WA). Sequences were identified taxonomically based on greater than 97% sequence identity to sequences of scanning electron micrograph (SEM) verified cultures in GenBank. MUSCLE (Edgar 2004) was used for alignments and pairwise percent identities calculations. Sequences have been deposited in GenBank under accession numbers KR053126-KR053164.

Infections of Pseudo-nitzschia *isolates with environmental samples*

Pseudo-nitzschia culture growth was monitored via chlorophyll-a fluorescence on a Turner AU-10 fluorometer. All infection crosses were performed in 48 well plates (Corning) in replicates of 5 with 1 ml exponentially growing culture, between 10-30% of their maximum relative fluorescence units, and 20 µl of purified environmental virus sample. For each cross, a control culture was amended with 20 µl f/2 media. Additionally, control *Pseudo-nitzschia* cultures were inoculated with 20 µl of virus samples that were UV irradiated (100 cm from a Philips TUV 36 T5 SP UV bulb) for 15 minutes. Crosses were maintained under the same temperature and light conditions as above. Culture growth in well plates was measured via

chlorophyll-a fluorescence on a Spectramax M2 Plate Reader (Molecular Devices). Treated wells that declined by more than half of their maximum fluorescence while control cultures were still healthy were scored as infected.

10-fold dilutions, ranging from 10^{0}-10^{-7}, of the environmental virus samples were inoculated into 1 ml cultures of exponentially growing *Pseudo-nitzschia* in replicates of 5. The growth and death of the *Pseudo-nitzschia* in titer experiments was monitored as above via chlorophyll-a fluorescence. Concentrations of infectious units per unit volume were determined based on most probable number (MPN) tables. Infectious units ml^{-1} seawater were calculated assuming 100% retention of viral infectivity during filtering, flocculation, and storage. To get infectious units ml^{-1} seawater, MPN values were multiplied by the 10^{6} to account for concentration of viruses from whole water to the final viral community sample.

Statistical analysis of infection networks

Statistical structure of the infection network generated from the crosses was tested using the BiMat package developed by Flores et al. (2014) in MatLab. Tests of modularity, using the Adaptive Brim algorithm, and nestedness, using NODF (nestedness measure based on overlap and decreasing fills) were compared to 10000 equiprobable randomized networks for statistical significance. Correlation between modules and ITS1 genotype, location, time, and infection permissivity were tested by comparing the Shannon diversity index of modules based on the predetermined categories to modules with randomly assigned categories (Flores et al. 2014).

2.4 Results

Environmental conditions and Pseudo-nitzschia *community structure*

Samples for viral communities and *Pseudo-nitzschia* isolates were taken every month from April 2013 to April 2014 (except for December and February in Penn Cove and November and January in Grays Harbor) at two sites (figure 1). Penn Cove is a shallow (20m depth) inlet in the Puget Sound estuary and Grays Harbor is located on the Pacific coast of Washington State. In total, 20 environmental virus communities were sampled and 41 *Pseudo-nitzschia* strains were isolated. *Pseudo-nitzschia* were isolated successfully during summer months when water temperatures were warm (13-17 °C), nutrient concentrations were low (<6μM NO_3^-) (figure 2a and b). Nitrate concentrations were positively correlated with phosphate and silicate concentrations at Grays Harbor and Penn Cove, respectively ($p<0.001$).

Eight species of *Pseudo-nitzschia* were isolated and identified based on ≥97% sequence identity of the ITS1 region to SEM verified *Pseudo-nitzschia* strains: *P. pungens, P. multiseries, P australis, P. delicatissima, P. americana, P. hasleana,* and two unknown species. *P. pungens* was isolated most frequently in 8 samples and comprised 28/41 strains isolated. Out of 22 isolates from Penn Cove, 21 isolates were *P. pungens* and *P. multiseries* (figure 2a), while out of 20 isolates from Grays Harbor, 17 isolates were *P. pungens*, *P. australis,* and *P. delicatissima* (figure 2b). Two strains, *P. sp.* GH10 and *P. sp.* PC33, were unable to be assigned species identification based on their ITS1 sequence. *P. sp.* GH10 ITS1 was 88% similar to *P. seriata* and had a 138 base pair ITS1 fragment when amplified using *Pseudo-nitzschia* specific primers. The full length ITS1 region of *P. sp.* PC33 was unable to be amplified. Both isolates were confirmed via light microscopy to be *Pseudo-nitzschia*.

Pseudo-nitzschia community composition at the two sites was determined by ARISA targeting the *Pseudo-nitzschia* ITS1 region (Hubbard et al. 2008). *Pseudo-nitzschia* were detectable in 15 of 20 samples (figure 2c and d). *Pseudo-nitzschia* communities from both Grays

Harbor and Penn Cove throughout the year were composed of 3 or more species except for a monospecific bloom of *P. pungens* at Penn Cove in June (figure 2c). *P. pungens* and *P. subpacifica* were the most common species detected in the entire dataset, present in 4 of 7 samples from Penn Cove and 4 of 8 samples from Grays Harbor. *Pseudo-nitzschia* communities did not cluster significantly by location or time.

Infection of Pseudo-nitzschia *host strains*

The 41 *Pseudo-nitzschia* isolates were crossed with the 20 environmental virus samples in replicates of 5 to create an infection network of 820 infectious crosses. In total, 68 *Pseudo-nitzschia* - virus community combinations showed signs of infection. *Pseudo-nitzschia* isolates inoculates with UV irradiated viral communities showed no signed of infection compared to medium-only controls. Overall 8% of crosses between *Pseudo-nitzschia* and virus communities were infective (figure 3). Hosts isolated from Penn Cove were infected by virus communities from Grays Harbor, and vice versa. Hosts and viruses that came from different times and locations were not significantly less infective than the total average. Crosses of *Pseudo-nitzschia* and virus communities from the same time and location resulted in infection at double the overall infection rate (Chi-square p=0.009) (figure 3).

Pseudo-nitzschia isolates were not infected uniformly in the initial survey of the *Pseudo-nitzschia* hosts crossed with the viral communities (figure 4). Thirteen host strains showed no detectable signs of infection from any environmental virus community. The 28 strains were infected more than once, and ranged from being infected by one viral community to up to 10 viral communities. Five *Pseudo-nitzschia* hosts, *P. pungens* GH29, *P. pungens* PC45, *P. australis* GH31, *P. pungens* GH23, and *P. pungens* GH20 were infected by 5 or more viral

communities, accounting for 48% of the total infectious crosses observed. Frequently, not all replicates died within one host-virus community cross (figure 4). Based on most probable number calculations in this survey, bounds of infectious units could be put on infectious crosses where between 1 – 4 replicates died (e.g. 1 replicate death = 2-11 infectious units ml^{-1}, 4 replicates death = 13-34 infectious units ml^{-1}). The samples that resulted in infection in all 5 replicates had titers that were at least 23 infectious units ml^{-1} but the upper bound of infectious units was unknown. Hosts with more infections had infectious crosses where more replicates died (linear regression $R^2 = 0.536$, $p<0.01$). The most infected host, for example, *P. pungens* GH20 had 8 out of 10 infectious crosses where all five replicates died (figure 4). Titers were performed to quantify the infectious units in these highly infectious crosses.

Host specific viral titers

More detailed changes in viral abundance over time were quantified by performing titers on *Pseudo-nitzschia* strains that represented a range of infection phenotypes. Host strains *P. pungens* PC45 and *P. pungens* GH20 were infected by 5 and 11 of the 20 viral communities and represented hosts with high numbers of infections. *P. sp* 1 GH10, *P. pungens* GH30, and *P. pungens* GH34 were infected by 2 or 3 of the 20 viral communities and were defined as representative of moderately infected hosts. *P. pungens* PC62, *P. pungens* PC40, and *P. australis* GH28 represented hosts with low numbers of infections. Viral titers were determined for each of these 9 hosts with every viral community. Measures of viral abundance varied by time and by host (figure 5a and b). Host strain *P. pungens* PC45 had high titers with three titers above 300 infectious units ml^{-1} that occurred predominantly during summer months. The highest viral infectivity of over 10^4 infectious units ml^{-1} seawater was observed on this strain crossed with the

July Penn Cove viral community (figure 5a). Interestingly, PC45 was isolated from the same water sample. However, strain *P. pungens* PC40, also isolated at the same time from the same water sample as PC45 had four orders of magnitude lower viral abundance when crossed with the same July Penn Cove virus community (figure 5a). This viral community did not infect the other 6 host strains on which titers were performed. In contrast with the observed high viral titers on PC45 in the summer at both locations, strain *P. pungens* GH20 had the highest titers in the fall and winter months at both locations. Host strains that were infected by the widest variety of viral samples (figure 4), such as PC45 and GH20 also had the highest maximum observed titers, 16000 and 540 infectious units ml^{-1}, respectively, compared with less infected strains like PC40 and GH28, which had lower maximum titers, 2 and 7.8 infectious units ml^{-1}respectively.

Patterns of viral infection by host genotype

The *Pseudo-nitzschia* hosts were grouped by ITS1 based species identification, ITS1 percent sequence identity, and according to the number of infectious crosses with the viral communities (figure 6). Five groups of isolates had 100% nucleotide identity at the ITS1 region (figure 6). Sixteen *P. pungens* strains with 100% identical ITS1 sequences consisted of 12 infection phenotypes. A second group of 8 P. pungens strains with a different ITS1 sequence consisted of 6 infection phenotypes. The phenotypes ranged from infected by multiple viral communities to not infected at all. This same pattern of diverse infection phenotypes within groups of isolates that 100% identical ITS1 sequences was observed in *P. multiseries* (3 infection phenotypes in 4 strains), *P. australis* (2 infection phenotypes in 3 strains), and *P. delicatissima* (2 infection phenotypes in 2 strains) (figure 6). In fact, of the 28 strains that were infected by at

least 1 viral community, only two strains *P. pungens* GH 14 and *P. americana* GH39 displayed the same infection phenotype, and they belonged to different species.

Pseudo-nitzschia–*virus infection network*

The interactions between the *Pseudo-nitzschia* and the viral communities were tested to see if there were statistically significant patterns of nestedness and/or modularity by comparing patterns in the host-virus network to 10^5 randomized equiprobable null models. First, the network was anti-nested (nestedness value (NODF) = 0.1023, z-score = 3.4193, percentile = 99.83, figure 7a). Second, the interactions between hosts and viruses occurred in modules (modularity value (Adaptive Brim Qb) = 0.5133, Z-score = -2.3491, percentile = 0.95, figure 7b). Reorganization to optimize modularity suggested 6 statistically significant modules, 3 of which had at least 12 infectious interactions between viruses and hosts while the other three had less than 5 interactions. Nested patterns can exist within modules, but only module 3 was significantly nested (Qb=0.53061, Z-score= 2.1241, percentile= 96.64). Interactions within modules did not group by location, time of sampling, ITS genotype, or host permissivity.

2.5 Discussion

Host specific viral interactions

Pseudo-nitzschia strains ranged in their susceptibility to the viral communities sampled in this study with some hosts showing no signs of infection from any of the viral communities tested while others were infected by multiple different communities. Hosts that were susceptible to more viral communities had higher titers than less infected strains (figure 4, 5). As seen with the July Penn Cove viral sample, for example, different hosts gave widely different titers ranging

from over 16,000 infectious units ml^{-1} to below the limit of detection (figure 5a). *Prochlorococcus* and its phage exhibit similar trends of differential susceptibility and titers by host, which are the result of different host specificities of infecting viruses (Dekel-Bird et al. 2014). Furthermore, different hosts enable the isolation of different viral assemblages (Dekel-Bird et al. 2014), underscoring the need for isolating viruses on a range of hosts in order to capture a better picture of viral diversity.

There are caveats to using titers to quantify infection of algae. The use of only permissive hosts for host titers and virus isolation may skew data of viral abundance and composition. Also, viruses may change their infection dynamics based on the number of viruses inoculated into a culture and infecting a host (Gutierrez et al. 2010), potentially leading to different infection dynamics. Finally, bacteria, which may be algacidal, may pass through a 0.2 μm filter (Lee et al. 2010) and lyse algal cultures. Ultimately, we were unable to differentiate between viral lysis and other pathogens.

The patterns of host specific interactions seen in the viral infectivity or titers did not follow host genotype determined by ITS1 sequence, as strains with identical ITS sequences displayed widely varying infection phenotypes (figure 6) and titers (figure 5). Similarly, isolated diatom viruses have been observed to infect some strains but not others within one species (Nagasaki et al. 2004; Tomaru et al. 2008; Kimura and Tomaru 2015). The variation by host seen here and in isolated diatom viruses suggest that trying to survey the impact of viruses on diatom communities using traditional phylogenetic markers will most likely be inconclusive. There is no known genetic marker diagnostic of infection phenotype in diatoms.

Time and location of host and viral community origin was the best predictor of infectivity with co-occurring hosts and viruses being the most infective. A similar study using 10

Synechoccocus isolates and 10 phage isolated from the same water as the hosts, found that hosts were more resistant to co-occurring viruses (Waterbury and Valois 1993). However, Flores et al. (2012) analyzed the interactions between 286 bacteria and 215 viruses isolated from a transect across the Atlantic and found that phages were more likely to kill hosts from the same location. In the framework of 'Kill-the-Winner' dynamics, both co-occurrence and resistance fluctuate dynamically as well (Thingstad 2000). The rate that new hosts fill vacant niches caused by viral infection may determine how frequently resistance or co-occurring infectivity is observed. Niches that are filled quickly by resistant taxa would result in a higher chance that sampling a host that is resistant to a virus from the same water. Consequently, fine temporal scale sampling is needed to resolve changing microbial populations. Additionally, host identification using markers that distinguish between taxa based on viral infection phenotype are needed to fully reveal the dynamic cycling of hosts and viruses in the environment.

Interestingly, viral communities from Grays Harbor could infect hosts from Penn Cove and vice versa (figure 3). Furthermore, the rates were no different from viruses infecting hosts at the same location, but at different times. Additionally, every virus community sample could infect at least one host. One explanation for these results may be a connectivity between viral and host populations in the Puget Sound and on the Washington coast. Based on hydrographic models of Puget Sound, surface water from Whidbey Basin, where Penn Cove is located, could reach the Washington coast in 15-30 days, while deep water from the coast would reach Whidbey on the order of 2 months at least (Babson et al. 2006). Thus, the transport of water between Puget Sound and the Washington coast occurs on temporal scales roughly similar to the frequency we sampled each individual location.

Anti-nestedness and viral host range

A major question about the ecology of diatom infecting viruses is, given the dramatic bloom and bust life style of diatoms, how are viruses propagated and successful? *Pseudo-nitzschia* communities sampled over the year became so rare that they were undetectable with ARISA 20% of the time, particularly during winter months (figure 2c and d). *Pseudo-nitzschia* ARISA can detect *Pseudo-nitzschia* at concentrations as low as 1 cell L^{-1} (Hubbard et al. 2014). Yet viruses from those months were infective of *Pseudo-nitzschia* isolates. There are two explanations for this disconnect between host abundance and viral infectivity. First, sediment samples have consistently been shown to have high viral infectivity from diatom viruses (Tomaru et al. 2011a). The sediments could be seed bank for diatom virus communities. Sediment resuspension or entrainment with upwelling, which occurs during turbulent mixing events particularly in the winter and spring in the Pacific Northwest (Hickey and Banas 2008), could be a mechanism for re-inoculating surface waters with viruses. This would allow diatom viruses to 'overwinter' during times of host scarcity. Second, diatom viruses may have broader host ranges (beyond a single species) than have been detected in culture studies to date, allowing for propagation on a wider range hosts to maintain their abundance in the water column.

The observed infections of *Pseudo-nitzschia* were the result of the integrated infectivity of the entire viral community. Typically, virus-host networks are composed of virus isolates crossed with isolated hosts (Weitz et al. 2012), resulting in a network where both the viral isolates' host ranges and host susceptibility is known and can then be tested for evolutionary and ecological patterns between hosts and their viruses. Here, the composition and abundance of viruses in the environmental communities was unknown. A high titer on a specific strain could be the result of many viruses at low abundance or one virus at high abundance. Similarly, the

infection pattern seen in any one viral community could be due to one virus with a broad host range or many viruses with narrow host ranges. The infection patterns by each viral community were highly variable from month to month. Statistical analyses indicate that the infection network is anti-nested (figure 7a). In contrast to nested networks, anti-nested networks are interaction poor, with hosts increasing incrementally in their number of interactions but viruses not following the incremental increase typical of nested patterns. Therefore, in order to have increasing host interactions, the interactions must come from a significantly more diverse set of viruses. Diversity in the viral communities could also be a function of the error prone RNA-dependent RNA polymerases and rolling circle replication mechanisms used by ssRNA and ssDNA viruses, respectively, that result in high mutation rates which alter viral host ranges (Duffy et al. 2008; Nakayama et al. 2013). The connection between host range, diatom virus diversity, and community composition is a critical, but unknown factor in diatom-virus ecology.

Diatom virus modulation of Pseudo-nitzschia *communities*

Diatoms are one of the most important groups phytoplankton in the ocean because of their ability to form massive blooms, the fate of which is critical to biogeochemical cycling (Smetachek, 1999). Over the year sampled in this study, *Pseudo-nitzschia* communities were diverse and transient. They were similar in composition and diversity to communities found previously throughout the Pacific Northwest in other years using the same ARISA technique (Hubbard et al. 2008, 2014). Diatom communities have been shown to be composed of populations of cells that are genetically distinct but nearly identical at the 18S, 5.8S, and ITS1 regions (Rynearson and Armbrust 2005). Thus the genetic diversity of these *Pseudo-nitzschia* communities is probably underestimated. On three occasions, June Penn Cove, September Penn

Cove, and August Grays Harbor, a single species (*P. pungens, P. cuspidata,* and *P. pungens* respectively) composed more than 60% of the *Pseudo-nitzschia* community. However, the *P. pungens* dominated community in August at Grays Harbor, for example, was composed of at least 4 different host phenotypes that were indistinguishable based on sequencing or fingerprinting (such as ARISA) of taxonomic markers. Based on the interactions between *Pseudo-nitzschia* isolates and the viral communities, we suggest these periods of single species dominated communities were actually much more phenotypically diverse.

The mechanisms that could lead to these diverse phenotypes remain unknown in diatoms. Bacteria may mediate resistance to infection in diatoms. For example, certain species of bacteria added to axenic cultures *Chaetoceros tenuissimus* prevented total lysis of the culture by the CtenRNAV (Kimura and Tomaru 2014). Differential viral resistance in *Prochlorococcus* was a result of genetic diversity found in the hypervariable regions of the hosts' genomes (Avrani et al. 2011). Similar genetic based resistance may also exist in the expansive diatom genomes (Armbrust et al. 2004; Bowler et al. 2008). Alternatively, predation by viruses may be stimulating phenotypic diversity in diatom communities through Red Queen dynamics where hosts and viruses are constantly evolving in response to each others changing predation or defenses (Van Valen 1973). With some strains being highly susceptible to infection, this pressure may be causing the phenotypic diversification seen in other, closely related strains.

In conclusion, the patterns of virus-*Pseudo-nitzschia* interactions may suggest a paradigm that differs from the role of viruses as terminators of phytoplankton blooms. Instead, diatom blooms are complex communities and diatom diversity, whether stimulated by viral predation or as a result of other resistance factors, may impede viral termination. The host specific interactions can lead to large variability in infectivity and virus titers. Furthermore, these patterns

do not follow host phylogeny as we can measure it with the ITS1 region. Methods that estimate community composition or abundance using these markers do not capture viral infection phenotype, as strains sharing 100% ITS1 identity isolated from the same water yielded different titers and susceptibilities to viral infection. Also, the infection patterns suggested connectivity between the communities of Puget Sound and the Washington coast, with viruses being most infective of their co-occurring hosts. Finally, not only did the diatom hosts display many different infection phenotypes, but the viral communities were also diverse and ephemeral. The taxonomic and temporal scales of diatom-virus interactions uncovered here illustrate the importance of permissivity and host range and emphasize the need to determine the cellular mechanisms of these attributes. Ultimately this will lead to a quantitative understanding of the impacts of viral infection on abundance and structure of wild diatom populations.

2.6 Acknowledgements

Funding was provided by a grant from the National Science Foundation: OCE-1356779 and the University of Washington Royalty Research Fund A65810 to GR. MCGC was funded by an EPA STAR fellowship. A Mary Gates Undergraduate Research Scholarship supported TSL and NDM and a Levinson fellowship additionally supported NDM. We would also like to thank Clara Fuchsman and Billy Brazelton for help collecting and processing samples and Kyle Frischkorn for thoughtful comments on the manuscript.

2.7 References

Armbrust, E. V. 2009. The life of diatoms in the world's oceans. Nature **459**: 185–92.

Armbrust, E. V., J. a Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. a Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M.

S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. a Rynearson, M. a Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science **306**: 79–86.

Avrani, S., O. Wurtzel, I. Sharon, R. Sorek, and D. Lindell. 2011. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. Nature **474**: 604–8.

Babson, a. L., M. Kawase, and P. MacCready. 2006. Seasonal and interannual variability in the circulation of Puget Sound, Washington: A box model study. Atmosphere-Ocean **44**: 29–45.

Bates, S. S., C. J. Bird, S. W. de Freitas, R. Foxall, M. Gilgan, L. A. Hanic, G. R. Johnson, A. W. McCulloch, P. Odense, R. Pocklington, M. A. Quilliam, P. G. Sim, J. C. Smith, D. V. Subba Rao, E. C. D. Todd, J. A. Walter, and J. L. C. Wright. 1989. Pennate diatom *Nitzschia pungense* as the primary source of domoic acid, a toxin in the shellfish from Eastern Prince Edward Island, Canada. Can. J. Fish. Aquat. Sci. **46**: 1203–1215.

Baudoux, A., A. Noordeloos, M. Veldhuis, and C. Brussaard. 2006. Virally induced mortality of *Phaeocystis globosa* during two spring blooms in temperate coastal waters. Aquat. Microb. Ecol. **44**: 207–217.

Bettarel, Y., J. Kan, K. Wang, K. Williamson, S. Cooney, S. Ribblett, F. Chen, K. Wommack, and D. Coats. 2005. Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros cf. gracilis*. Aquat. Microb. Ecol. **40**: 103–114.

Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. a Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin-Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M.-P. Oudot-Le Secq, C. Napoli, M. Obornik, M. S. Parker, J.-L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. a Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, and I. V Grigoriev. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature **456**: 239–44.

Bratbak, G., M. Heldal, and J. K. Egge. 1993. Termination of algal blooms: viral mortality of the marine coccolithophorid *Emiliania huxleyi*. Mar. Ecol. Prog. Ser. **93**: 39–48.

Breitbart, M. 2012. Marine Viruses: Truth or Dare. Ann. Rev. Mar. Sci. **4**: 425–448.

Brussaard, C. P. D. 2004. Viral control of phytoplankton populations-A review. The Journal of Eukaryotic Microbiology. **51**: 125–138.

Clarke, KR, Warwick RM. 2001. Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. PRIMER-E, Plymouth, 172pp.

Colombet, J., and T. Sime-ngando. 2012. Use of PEG , Polyethylene glycol , to characterize the diversity of environmental viruses. Curr. Microsc. Contrib. To Adv. Sci. Technol. 316–322.

Cottrell, M. T., and C. A. Suttle. 1995. Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. **40**: 730–739.

Dekel-Bird, N. P., G. Sabehi, B. Mosevitzky, and D. Lindell. 2014. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. Environ. Microbiol. doi:10.1111/1462-2920.12569.

Duffy, S., L. a Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. **9**: 267–76.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**: 1792–1797.

Eissler, Y., K. Wang, F. Chen, K. Eric Wommack, and D. Wayne Coats. 2009. Ultrastructural Characterization of the Lytic Cycle of an Intranuclear Virus Infecting the Diatom *Chaetoceros Cf. Wighamii*(Bacillariophyceae) From Chesapeake Bay, Usa1. J. Phycol. **45**: 787–797.

Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host-phage interactions. Proc. Natl. Acad. Sci. U. S. A. **108**: E288–97.

Flores, C. O., T. Poisot, S. Valverde, and J. S. Weitz. 2014. BiMat: a MATLAB package to facilitate the analysis and visualization of bipartite networks. arXiv **1406.6732v**.

Flores, C. O., S. Valverde, and J. S. Weitz. 2012. Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. ISME J. 1–13.

Gobler, C. J., S. Deonarine, J. Leigh-Bell, M. D. Gastrich, O. R. Anderson, and S. W. Wilhelm. 2004. Ecology of phytoplankton communities dominated by *Aureococcus anophagefferens*: the role of viruses, nutrients, and microzooplankton grazing. Harmful Algae **3**: 471–483.

Gutiérrez, S., M. Yvon, G. Thébaud, B. Monsion, Y. Michalakis, and S. Blanc. 2010. Dynamics of the multiplicity of cellular infection in a plant virus. PLoS Pathog. **6**, doi:10.1371/journal.ppat.1001113

Hickey, B., and N. Banas. 2008. Why is the Northern End of the California Current System So Productive? Oceanography **21**: 90–107.

Hubbard, K. a., G. Rocap, and E. V. Armbrust. 2008. Inter- and Intraspecific Community Structure Within the Diatom Genus *Pseudo-Nitzschia* (Bacillariophyceae). J. Phycol. **44**: 637–649.

Hubbard, K., C. Olson, and E. Armbrust. 2014. Molecular characterization of *Pseudo-nitzschia* community structure and species ecology in a hydrographically complex estuarine system (Puget Sound, Washington, USA). Mar. Ecol. Prog. Ser. **507**: 39–55.

John, S. G., C. B. Mendez, L. Deng, B. Poulos, A. K. M. Kauffman, S. Kern, J. Brum, M. F. Polz, E. a Boyle, and M. B. Sullivan. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ. Microbiol. Rep. **3**: 195–202.

Kimura, K., and Y. Tomaru. 2014. Coculture with marine bacteria confers resistance to complete viral lysis of diatom cultures. Aquat. Microb. Ecol. **73**: 69–80.

Kimura, K., and Y. Tomaru. 2015. Discovery of Two Novel Viruses Expands the Diversity of Single-Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. Appl. Environ. Microbiol. **81**: 1120–1131.

Lee, A., J. McVey, P. Faustino, S. Lute, N. Sweeney, V. Pawar, M. Khan, K. Brorson, and D. Hussong. 2010. Use of Hydrogenophaga pseudoflava penetration to quantitatively assess the impact of filtration parameters for 0.2-micrometer-pore-size filters. Appl Env. Microbiol **76**: 695–700.

Lehahn, Y., I. Koren, D. Schatz, M. Frada, U. Sheyn, E. Boss, S. Efrati, Y. Rudich, M. Trainic, S. Sharoni, C. Laber, G. R. DiTullio, M. J. L. Coolen, A. M. Martins, B. A. S. Van Mooy, K. D. Bidle, and A. Vardi. 2014. Decoupling Physical from Biological Processes to Assess the Impact of Viruses on a Mesoscale Algal Bloom. Curr. Biol. **24**: 2041–2046.

Lelong, A., H. Hegaret, P. Soudant, and S. S. Bates. 2012. *Pseudo-nitzschia* (Bacillariophyceae) species , domoic acid and amnesic shellfish poisoning : revisiting previous paradigms. **51**: 168–216.

Lundholm, N., N. Daugbjerg, and Ø. Moestrup. 2002. Phylogeny of the Bacillariaceae with emphasis on the genus *Pseudo - nitzschia* (Bacillariophyceae) based on partial LSU rDNA. Eur. J. Phycol. **37**: 115–134.

McDaniel, L., L. a Houchin, S. J. Williamson, and J. H. Paul. 2002. Lysogeny in marine *Synechococcus*. Nature **415**: 496.

Moniruzzaman, M., G. R. LeCleir, C. M. Brown, C. J. Gobler, K. D. Bidle, W. H. Wilson, and S. W. Wilhelm. 2014. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. Virology **466-467**: 60–70.

Nagasaki, K., Y. Tomaru, N. Katanozaka, Y. Shirai, K. Nishida, S. Itakura, and M. Yamaguchi. 2004. Isolation and Characterization of a Novel Single-Stranded RNA Virus Infecting the Bloom-Forming Diatom *Rhizosolenia setigera*. Appl. Environ. Microbiol. **70**: 704–711.

Nagasaki, K., and G. Bratbak. 2010. Isolation of viruses infecting photosynthetic and nonphotosynthetic protists. Virology 92–101.

Nagasaki, K., Y. Tomaru, Y. Takao, K. Nishida, Y. Shirai, H. Suzuki, and T. Nagumo. 2005. Previously Unknown Virus Infects Marine Diatom. Appl. Environ. Microbiol. **71**: 3528–3535.

Nakayama, N., A. Fujimoto, H. Kawami, Y. Tomaru, N. Hata, and K. Nagasaki. 2013. High Interaction Variability of the Bivalve-Killing Dinoflagellate *Heterocapsa circularisquama* Strains and Their Single-Stranded RNA Virus HcRNAV Isolates. Microbes Environ. **28**: 112–119.

Nelson, D. M., P. Trrguer, and M. A. Brzezinski. 1995. Production and dissolution of biogenic silica in the ocean : Revised global estimates , comparison with regional data and relationship to biogenic sedimentation. Global Biogeochem. Cycles **9**: 359–372.

Rynearson, T. a, and E. V. Armbrust. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom *Ditylum brightwellii*. Mol. Ecol. **14**: 1631–40.

Smetacek, V. 1999. Diatoms and the ocean carbon cycle. Protist **150**: 25–32.

Scholin, C. A., F. Gulland, G. J. Doucette, S. Benson, M. Busman, F. P. Chavez, J. Cordaro, R. Delong, A. De Vogelaere, J. Harvey, M. Haulena, K. L. I, T. Lipscomb, S. Loscutoff, L. J. Lowenstine, R. M. Iii, P. E. Miller, W. A. Mclellan, P. D. R. Moeller, C. L. Powell, T. Rowles, P. Silvagni, M. S. I, T. Spraker, V. Trainer, and F. M. Van Dolah. 2000. Mortality of sea lions along the central California coast linked to a toxic diatom bloom. **403**: 80–84.

Shirai, Y., Y. Tomaru, Y. Takao, H. Suzuki, T. Nagumo, and K. Nagasaki. 2008. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. Appl. Environ. Microbiol. **74**: 4022–7.

Tarutani, K., K. Nagasaki, and M. Yamaguchi. 2000. Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton: *Heterosigma akashiwo*. Appl. Environ. Microbiol. **66**: 4916–4920.

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

Thingstad, T. F., B. Pree, J. Giske, and S. Vage. 2015. What difference does it make if viruses are strain- , rather than species- specific? Front. Microbiol. doi:10.3389/fmicb.2015.00320

Thyrhaug, R., A. Larsen, T. F. Thingstad, and G. Bratbak. 2003. Stable coexistence in marine algal host-virus systems. Mar. Ecol. Prog. Ser. **254**: 27–35.

Tomaru, Y., Y. Shirai, H. Suzuki, T. Nagasaki, and T. Nagumo. 2008. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. Aquat. Microb. Ecol. **50**: 103–112.

Tomaru, Y., N. Fujii, S. Oda, K. Toyoda, and K. Nagasaki. 2011a. Dynamics of diatom viruses on the western coast of Japan. Aquat. Microb. Ecol. **63**: 223–230.

Trainer, V. L., S. S. Bates, N. Lundholm, A. E. Thessen, W. P. Cochlan, N. G. Adams, and C. G. Trick. 2012. *Pseudo-nitzschia* physiological ecology, phylogeny, toxicity, monitoring and impacts on ecosystem health. Harmful Algae **14**: 271–300.

Van Valen, L. 1973. A new evolutionary law. Evol. Theory **1**: 1–30.

Vardi, A., B. a S. Van Mooy, H. F. Fredricks, K. J. Popendorf, J. E. Ossolinski, L. Haramaty, and K. D. Bidle. 2009. Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. Science **326**: 861–5.

Waterbury, J. B., and F. W. Valois. 1993. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl. Environ. Microbiol. **59**: 3393–9.

Weitz, J. S., T. Poisot, J. R. Meyer, C. O. Flores, S. Valverde, M. B. Sullivan, and M. E. Hochberg. 2012. Phage–bacteria infection networks. Trends Microbiol. 1–10.

Weitz, J. S., and S. W. Wilhelm. 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. **8**: 2–9.

2.8 Figure legends

Figure 2.1 Locations of sampling. Penn Cove, located in the Puget Sound estuary, and Grays Harbor located on the coast of Washington state, USA. Inset map of North America shows the region of sampling.

Figure 2.2 *Pseudo-nitzschia* isolates obtained and community composition at (a, c) Penn Cove and (b, d) Grays Harbor from April 2013-April 2014. Time of sampling is shown in Julian day

and monthly increments on the x-axis. Solid lines are water temperature and dashed lines are nitrate concentration. *Pseudo-nitzschia* species are colored by phylogenetic clade (Lundholm et al. 2002, Guannel, unpublished data) with members of clade 1 represented by warm colors and clade 2 represented by cool colors. Unidentifiable ARISA fragments are represented in grayscale. Black bars represent ARISA fragments for isolated *Pseudo-nitzschia* with no species identification. * indicates months with no detectable *Pseudo-nitzschia*.

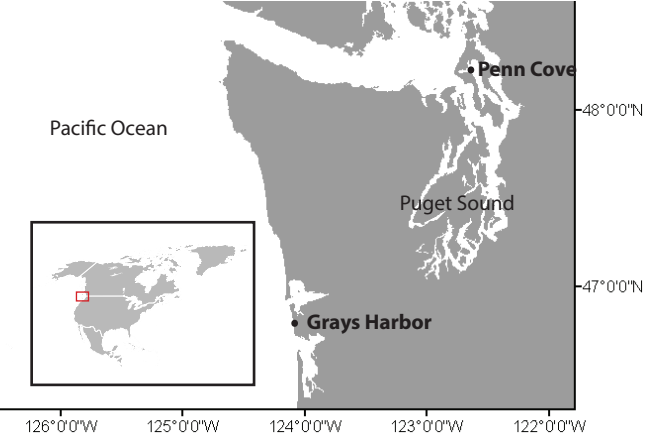
Figure 2.3 Infection frequency based on location and time of host isolation and virus community sample. * denotes a significant of p-value = 0.009 based on Chi-square test.

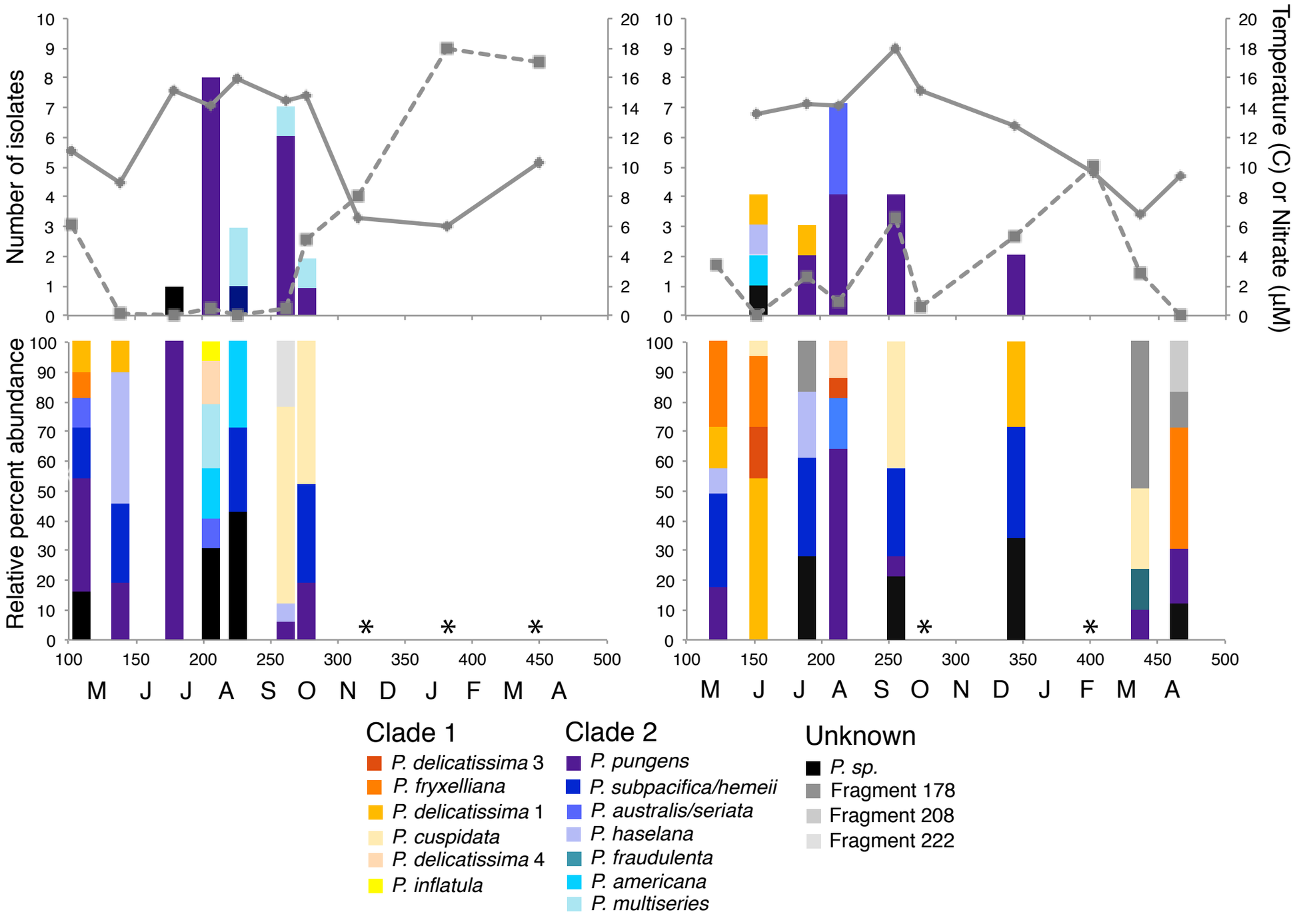
Figure 2.4 Total number of viral community samples that resulted in an infection by host strain. The infectious units ml^{-1} for each infectious cross was calculated using most probable number tables based on the number of replicates that died. Crosses with higher titer values are represented by darker colors.

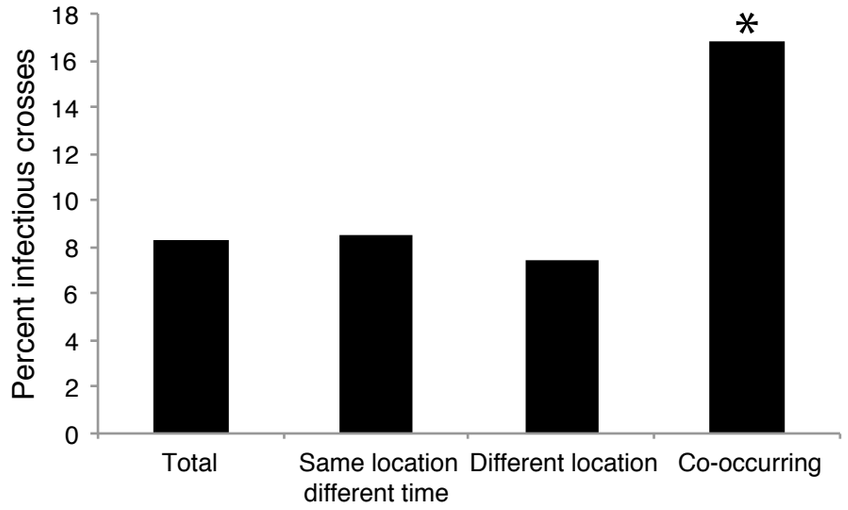
Figure 2.5 Titers of infectious units over time in Julian days with monthly increments at a) Penn Cove and b) Grays Harbor. Each of the 9 strains was crossed with each of the Penn Cove or Grays Harbor viral communities. Cool colors are hosts isolated from Penn Cove, warm colors are hosts isolated from Grays Harbor. Error bars are 95% confidence intervals from a 5 well MPN tables. Values below the limit of detection of 1.8 infectious units ml^{-1} are not shown.

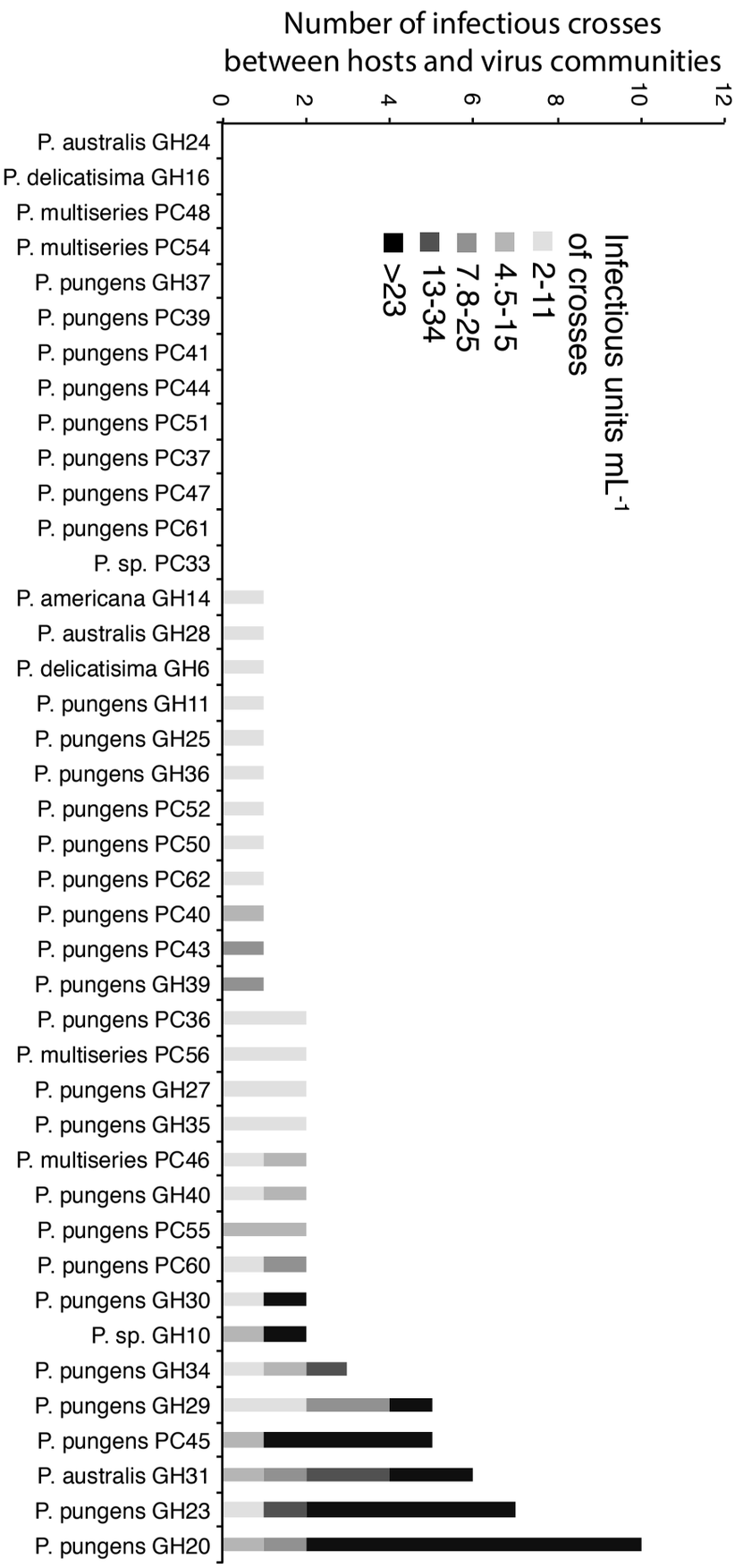
Figure 2.6 *Pseudo-nitzschia* – virus infection network. Filled black boxes represent infective crosses. Black outlines delineate groups of hosts that display identical ITS1 sequences.

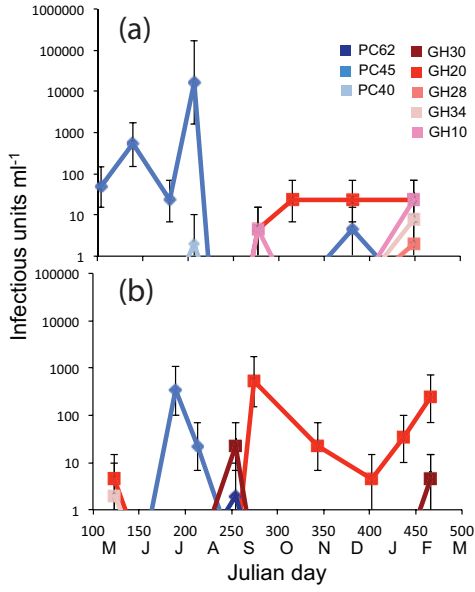
Figure 2.7 *Pseudo-nitzschia* – virus infection network sorted to maximize a) nestedness and b) modularity. Columns are hosts and rows are viral communities. Curved lines represent the nested isocline. Boxes represent statistically significant modules.

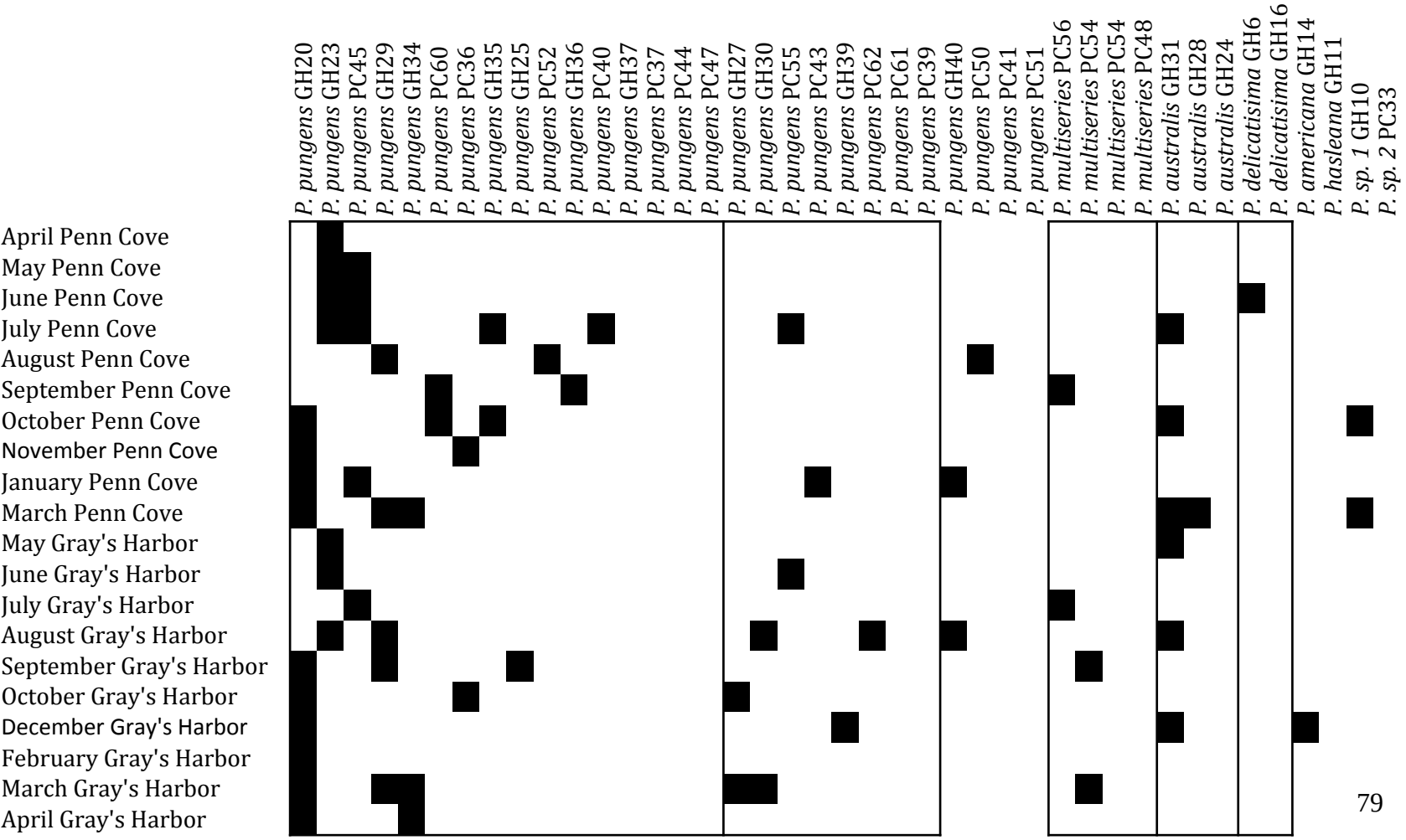

Penn Cove
Pacific Ocean
48°0'0"N
Puget Sound
47°0'0"N
Grays Harbor
126°0'0"W
125°0'0"W
124°0'0"W
123°0'0"W
122°0'0"W

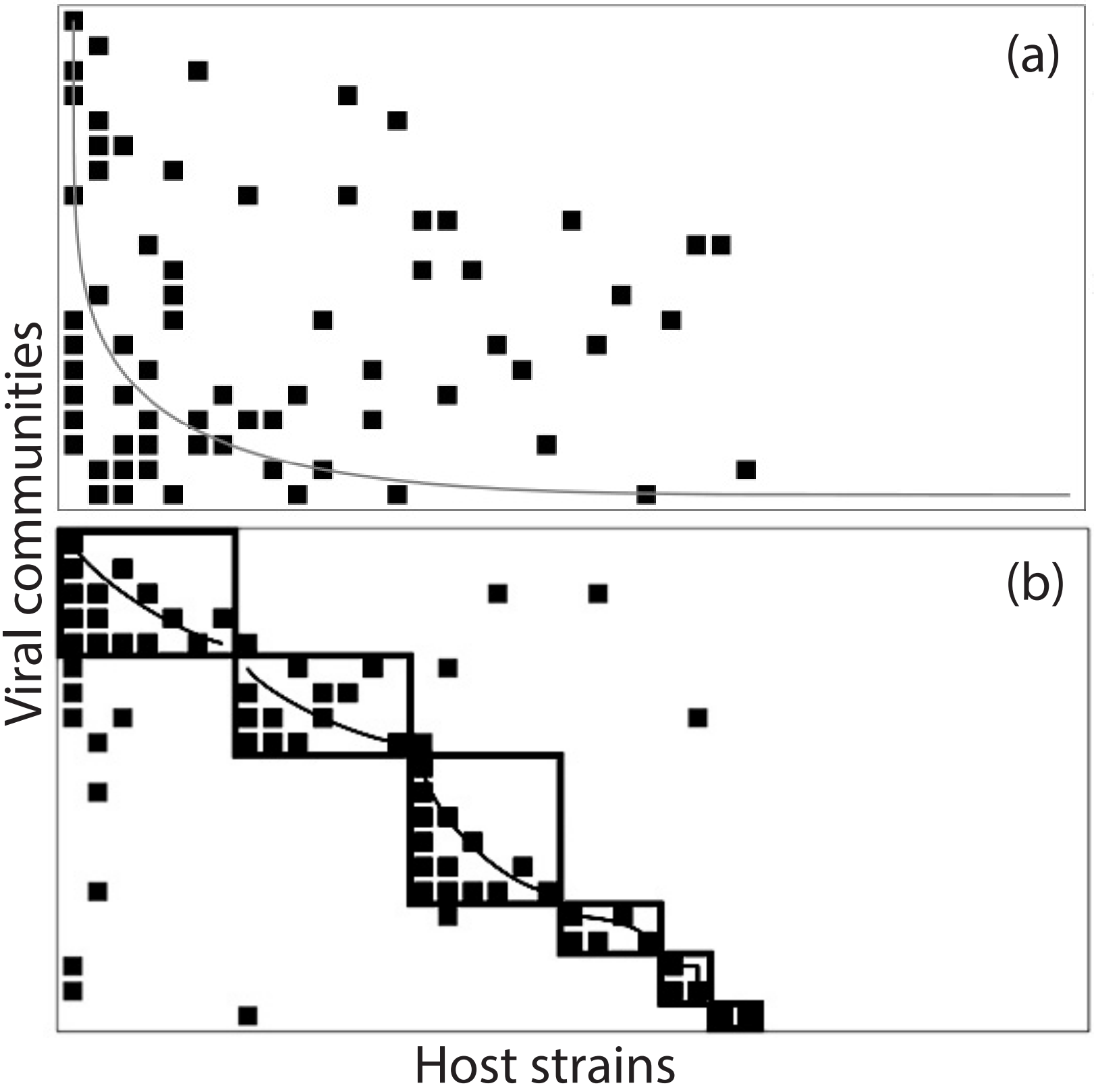

Number of isolates
Temperature (C) or Nitrate (μM)
Relative percent abundance
M J J A S O N D J F M A
Clade 1
P. delicatissima 3
P. fryxelliana
P. delicatissima 1
P. cuspidata
P. delicatissima 4
P. inflatula
Clade 2
P. pungens
P. subpacifica/hemeii
P. australis/seriata
P. haselana
P. fraudulenta
P. americana
P. multiseries
Unknown
P. sp.
Fragment 178
Fragment 208
Fragment 222


Percent infectious crosses
18
16
14
12
10
8
6
4
2
0
*
Total
Same location different time
Different location
Co-occurring


Number of infectious crosses
between hosts and virus communities
0
2
4
6
8
10
12
Infectious units mL-1
of crosses
2-11
4.5-15
7.8-25
13-34
>23
P. australis GH24
P. delicatisima GH16
P. multiseries PC48
P. multiseries PC54
P. pungens GH37
P. pungens PC39
P. pungens PC41
P. pungens PC44
P. pungens PC51
P. pungens PC37
P. pungens PC47
P. pungens PC61
P. sp. PC33
P. americana GH14
P. australis GH28
P. delicatisima GH6
P. pungens GH11
P. pungens GH25
P. pungens GH36
P. pungens PC52
P. pungens PC50
P. pungens PC62
P. pungens PC40
P. pungens PC43
P. pungens GH39
P. pungens PC36
P. multiseries PC56
P. pungens GH27
P. pungens GH35
P. multiseries PC46
P. pungens GH40
P. pungens PC55
P. pungens PC60
P. pungens GH30
P. sp. GH10
P. pungens GH34
P. pungens GH29
P. pungens PC45
P. australis GH31
P. pungens GH23
P. pungens GH20


(a)
(b)
PC62
PC45
PC40
GH30
GH20
GH28
GH34
GH10
Infectious units ml-1
1000000
100000
10000
1000
100
10
1
100
150
200
250
300
350
400
450
500
M
J
J
A
S
O
N
D
J
F
M
Julian day


P. pungens GH20
P. pungens GH23
P. pungens PC45
P. pungens GH29
P. pungens GH34
P. pungens PC60
P. pungens PC36
P. pungens GH35
P. pungens GH25
P. pungens PC52
P. pungens GH36
P. pungens PC40
P. pungens GH37
P. pungens PC37
P. pungens PC44
P. pungens PC47
P. pungens GH27
P. pungens GH30
P. pungens PC55
P. pungens PC43
P. pungens GH39
P. pungens PC62
P. pungens PC61
P. pungens PC39
P. pungens GH40
P. pungens PC50
P. pungens PC41
P. pungens PC51
P. multiseries PC56
P. multiseries PC54
P. multiseries PC54
P. multiseries PC48
P. australis GH31
P. australis GH28
P. australis GH24
P. delicatisima GH6
P. delicatisima GH16
P. americana GH14
P. hasleana GH11
P. sp. 1 GH10
P. sp. 2 PC33
April Penn Cove
May Penn Cove
June Penn Cove
July Penn Cove
August Penn Cove
September Penn Cove
October Penn Cove
November Penn Cove
January Penn Cove
March Penn Cove
May Gray's Harbor
June Gray's Harbor
July Gray's Harbor
August Gray's Harbor
September Gray's Harbor
October Gray's Harbor
December Gray's Harbor
February Gray's Harbor
March Gray's Harbor
April Gray's Harbor


(a)
(b)
Viral communities
Host strains

Chapter 3

Caught in the act: Infection by wild viruses observed through viral transcription

3.1 Abstract

Viruses are the most abundant and diverse predators in the ocean, and have been hypothesized to structure microbial communities and influence global biogeochemical cycles. Yet viral infection is rarely directly assessed in the environment. We sought to understand viral infection in the ocean by observing the transcription of viruses. Viral metagenomes were paired with size-fractionated cellular metatranscriptomes taken in the North Pacific ocean along Line P. A small fraction of the total virus community was observed to be actively transcribed; 16% at the most coastal station and only 0.2% in the most oceanic station. Viral abundance was compared to host abundance for the well-described virus groups of cyanophage, SAR11 phage, virophage and picoeukaryote viruses. In some cases, high virus:host ratios were observed without high transcription by the viruses, suggestive of a recent lysis event. On the other hand, several virus groups had above average expression, but a low virus:host ratio indicating that they were in the process of overtaking a dominant microbial taxon. Additionally, the infection dynamics of underexplored viral groups, such as archeoviruses and ssDNA viruses, indicated that new uncultured groups of viruses were transcriptionally active and shed light on who those viruses may have been infecting. *Synechococcus* phage were used as a model to understand how the genomic potential of viruses influences their success in the ocean. *Synechococcus* phage similar to three isolates were most abundant, and only 2 strains had significant expression. Suites of auxiliary genes were identified, but only at one station were host-derived photosynthesis genes

significantly expressed. This coincided with a cyanophage community with unusually high transcription intimating that auxiliary genes may provide a competitive advantage to viruses in the ocean. Directly observing viral infection revealed hidden dynamics of wild viral communities and underscored the necessity to understand these processes in order to quantify the contribution of viruses to microbial ecology and biogeochemistry.

3.2 Introduction

Viral infection and lysis in the marine environment have been proposed as major catalysts of biogeochemical cycling and processes which structure microbial communities. Because of the release of cell contents into the water column, viral lysis is estimated to move 150 Gt C y^{-1} from the particulate phase to the dissolved phase, which is an order of magnitude more than any other process contributes to carbon fluxes between environmental reservoirs (Suttle, 2005). However, the underlying estimates of viral infection that contribute to these global scale projections are poorly quantified. Estimates of viral contributions to mortality in microbial populations range from 2% to 100% (Cottrell and Suttle, 1995; Jacquet et al., 2002). The methods used to determine in situ infection rates lack adequate sensitivity to detect viral production or have significant biases (Brussaard, 2004). Epifluorescent microscopy, which can be used to count changes in viral abundance does not detect RNA, single stranded, or small viruses and may be missing half the viruses in the ocean (Steward et al. 2012). Since viruses lack a common genetic marker, most molecular tools can only target portions of the viral community.

The advent of metagenomics has allowed culture-independent analysis of entire *in situ* viral assemblages (Breitbart et al., 2002). Early viral metagenomes quickly revealed that viral genetic diversity was immense (Angly et al., 2006; Breitbart et al., 2002; Culley et al., 2006).

Furthermore, investigations targeting viral groups such as microphage, eukaryotic ssDNA viruses, positive sense RNA viruses, and dsDNA viruses have confirmed that viral diversity is high (Hopkins et al. 2014, Labonte et al. 2013, Culley et al 2014, Schmidt et al. 2014). However, viruses appear to be widely dispersed which contributes to lower global diversity (Angly et al. 2006, Bretibart et al. 2004, Brum et al. 2015). The 'every virus is everywhere' hypothesis (Breitbart and Rohwer, 2005) suggests that a diverse and cosmopolitan population of viruses serves as a reservoir for infection (Goldsmith et al. 2015). This raises the question of how this reservoir is maintained.

While metagenomics has enabled a more comprehensive view of viral communities, ~90% of viral sequences from metagenomes are novel (Angly et al. 2006, Hurwitz et al. 2013). Additionally, information about viral interactions with their hosts is nearly completely lost when only sequencing the viral fraction. Furthermore, certain viral types, such as single stranded viruses and RNA viruses, are still misrepresented or omitted in viral metagenomes because they require extra steps to convert to dsDNA needed for sequencing (Duhaime et al. 2012). When viromes do target these groups, dsDNA is often synthesized from random amplification, which preferentially targets small circular single stranded templates, and is not quantitative (Yilmaz et al. 2010).

The inability to measure infection in the environment still substantially hinders the field of viral ecology. Metaproteomic and metatranscriptomic approaches have serendipitously shed light on viral dynamics in the field. On a south Atlantic transect, viral proteins were the fourth most abundant peptide identified in bacterial proteomic samples (Morris et al., 2010). Despite the shift in microbial communities along this open ocean to coast transect, the fraction of viral protein spectra was split evenly between the two regimes (Morris et al., 2010). Two

interpretations of this data are viruses are being passively trapped on filters at equivalent concentrations in both regimes or that viral infection is equally prevalent in both environments despite changes in the host communities. Metatranscriptomic approaches paired with metagenomes have provided direct insight into the dynamics of viruses during the course of phytoplankton bloom. In mesocosm experiments, paired metagenomes (targeting the bacterial size fraction) and bacterial metatranscriptomes were taken at the bloom's peak and after the bloom (Gilbert et al., 2008). Post bloom, viral transcripts were 24.5 times higher than bloom peak, dominating the transcriptional signal in that sample (Gilbert et al., 2008). Even with poor identification of viral sequences, this indicates that metatranscriptomics can successfully capture active viral infection in the field.

We sought to understand viral infection in the ocean by observing the transcription of viruses. The Global-scale Microbial Interactions across Chemical Survey (GeoMICS) cruise in May 2012 combined organic and trace metal geochemistry with genomic techniques to survey how microbial processes impact ocean chemistry. The cruise sampled four stations on a transect from the coast of Washington state to the open North Pacific ocean along Line P. These stations traversed chemical gradients and changes in microbial communities. This provided an excellent opportunity to survey the corresponding changes in the viral community and identify and quantify actively infecting viruses. Four viral metagenomes were taken in coordination with cellular metatranscriptomes, providing a holistic view of viral infection.

3.3 Methods

Sample collection

Samples were collected along the Line P transect in the northeastern Pacific on the R/V

Thomas G. Thompson from May 16-21, 2012 as part of the Global-scale Microbial Interactions across Chemical Survey cruise. At four stations, P1 (48.58 N, 125.50 W), P4 (48.65 N, 126.67 W), P6 (48.74 N, 127.67 W) and P8 (48.82 N, 128.66 W), a conductivity temperature depth with 10L niskin bottles was used to collect water from 5m depth. Plankton were collected on a 2.0 μm pore-size filter (142 mm diameter polycarbonate) and a 0.2 μm pore-size filter (142 mm diameter Supor) following prefiltration through a 53 μm pore-size filter. Filters were flash frozen in liquid N_2 and stored at -80°C. 20L of the <0.2 μm filtrate was amended with iron chloride (1g L^{-1}) and incubated for at least 1 hour at room temperature to flocculate viruses (John et al. 2011). Viral flocculates were filtered onto 0.6 μm polycarbonate filters and stored at 4°C.

Virome construction, sequencing, and assembly

Viruses were resuspended from filters by incubating them with 20ml of 0.2M ascorbate-0.25M EDTA-Mg2-0.25M Tris-HCL for 24 hours with periodic shaking (John et al. 2011). Viral suspensions were incubated for 2 hours at room temperature with DNase 1 (100 U ml^{-1}, Ambion) to remove contaminating extracellular DNA. Viral DNA was extracted using a MoBio PowerSoil Total RNA isolation kit with the DNA elution column accessory, according to the manufacturers instructions. To convert single stranded DNA (ssDNA) viruses to double stranded DNA (dsDNA), complementary strands were synthesized using a modified random priming-mediated sequence-independent single-primer amplification (RP-SISPA) method designed to generate quantitative whole genome shotgun libraries (Djikeng et al. 2008, Culley et al. 2010). Triplicate 20 μl reactions for each sample and a no template control containing 10 μl template (or water), 0.2mM dNTPs, 0.5 mM DTT, 1 μM FR26RV-N primer (5'-GCCGGAGCTCTGCAGATATCNNNNNN -3'), 1mM $MgCl_2$, and 1X Klenow buffer were

heated to 94 °C for 3 minutes and snap cooled on ice. 2.5 U of Klenow Fragment, 3'-5' exo- (New England Biolands) was added to each reaction then incubated at 37 °C for 60 minutes followed by 75 °C for 10 minutes. Reactions were pooled, cleaned with an ethanol precipitation, and resuspended in T low-E (10mM Tris-HCl, 0.1 mM EDTA pH 8.0).

To construct Illumina shotgun metagenome libraries, viral DNA was sheared to 800-1000 bp using a Covaris nebulizer and cleaned with AmPure XP beads. Viral metagenomes (viromes) were constructed using the Rubicon ThruPLEX-FD Kit, which uses qPCR to linearly amplify low concentration DNA templates. Libraries were sequenced on an Illumina HiSeq 2500 with 150bp paired end reads at the Michigan State University Sequencing Center.

Virome reads were trimmed of adapter sequences and low quality regions and filtered to remove low quality reads with the Trimmomatic pipeline (Bolger et al., 2014). Sequences were *de novo* assembled into contigs by first using diginorm (Brown et al., 2012) followed by the velvet assembler (Zerbino and Birney, 2008) with k-mer length set at 25, 29, 33, 51 for P1, P4, P6, and P8 respectively and a minimum contig length of 200bp. The resulting contigs were cleaned of cellular contamination by first blasting against the RDP database. A custom database of taxa that were identified from the RDP search as well as known gene transfer agent producers was created to search and remove any contigs that were annotated with an e-value of 1e-20 or lower to these contaminants. These putative cellular contigs were then blasted against viral ref seq, and hits of < 1e-100 were blasted against nr and added back to the viromes if the top 5 blast hits were viral, which occurred with <1% of the contigs.

Contig annotation and binning

Contigs were annotated by the best blast hit against viral ref seq (searched on September 16th, 2015) at an e-value cutoff of 1e-5. Host:virus ratios were calculated by dividing normalized viral abundances by normalized 16S tag sequencing abundance (Limnardo and Worden, unpublished data).

Metatranscriptome construction and sequencing

RNA was prepared for sequencing as previously described (Durham et al 2015). Briefly, cells were lysed in 10 mL of Ambion lysis buffer (AM8540G) + 0.5 mL each of 0.5 and 0.1 mm zirconia beads (BioSpec Products, Inc.), and an internal standard 1,006 nt in length (Satinsky et al 2013) was added to the lysis buffer in known copy numbers. Total RNA was treated for DNA contamination using two successive treatments with Turbo-DNase (Ambion; Life Technologies) and cleaned/concentrated with an RNeasy MinElute Cleanup Kit (Qiagen). rRNAs were removed using a subtractive hybridization method (Stewart et al 2010) with biotinylated rRNA probes specific to the environmental samples (i.e. bacterial and archaeal 16S and 23S and eukaryotic 18S and 28S). Probe-bound RNAs were removed with strepdavidin-coated magnetic beads (New England Biolabs). Depleted RNA samples were linearly amplified using the MessageAmp II-Bacteria kit (Ambion; Life Technologies).

Amplified mRNAs were converted into cDNAs using the Superscript III First-Strand Synthesis System (Invitrogen; Life Technologies) followed by the NEBNext mRNA Second Strand Synthesis Module (New England Biolabs). cDNAs were then purified using either the PureLink PCR Purification Kit (Life Technologies) or the QIAquick PCR purification kit (Qiagen) followed by ethanol precipitation. cDNAs were sheared ultrasonically to ~250 base pair fragments and libraries were constructed for paired-end (2 x 150) sequencing using either the

Illumina MiSeq or HiSeq 2500 sequencing platform (Illumina Inc., San Diego, CA). Reads were joined using PANDAseq (Masella et al 2012), and paired reads were trimmed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). rRNA, tRNA, and internal standard sequences were removed using a BLASTn search against a database containing representative rRNA and tRNA sequences along with the internal standard sequence (bit score cutoff >50).

Metatranscriptome recruitment

BBMap version 35.x was used to map metatranscriptome reads to viral contigs. Contigs were indexed with a padding of 300 and reads with a minimum 98% identity were mapped to contigs. Each contig, therefore, had an RNA coverage and DNA coverage (from assembly), which were used to create a ratio of RNA:DNA for each contig.

Viral phylogenies, read placement, and diversity statistics

Phylogenies of the terminase large subunit (terL), DNA polymerase B (polB), ssDNA replication protein (vRep), major capsid proteins (MCP), and photosystem D1 (psbA) were constructed using RAxML v 8.0 (Stamatakis, 2012). Briefly, reference sequences were downloaded from Genbank and used to search proteins on virome contigs. Together, full or nearly full length virome proteins and reference sequences were aligned with Muscle v 3.8.31 (Edgar, 2004). A maximum likelihood tree was inferred from 20 starting trees using the amino acid substitution model model VTF for terL and polB, PMBF for vRep, VTF for MCP, and MTMAMF for psbA.

Viromes and metatranscriptomes were searched using blast at an e-value cutoff of 1e-10 with representative seed sequences from each of the above genes. Reads were trimmed according

to their blast alignment and aligned to the reference alignment using Parsimony-based Phylogeny-Aware Read Alignment program (PaPaRa) (Berger and Stamatakis, 2011). Reads were then mapped to the reference tree with the Evolutionary Placement Algorithm (EPA) (Berger et al. 2011).

Nodes and leaves from the tree were assumed to be the operational taxonomic units (OTUs). OTUs that did not recruit at least 0.1% of the total placed reads in from either the viromes or metatranscriptomes were removed. Since the terL phylogeny was designed to estimate bacteriophage diversity, bacteriophage OTUs were cut from the polB tree to prevent overrepresentation of bacteriophage taxa. Similarly, non-virophage nodes were removed from the MCP trees. Read abundance was normalized to the average gene lengths of each tree: terL (1000 aa), polB (600), vRep (400), MCP (275). Read abundances were then normalized to library size. OTUs that were ambiguous or unrelated to known groups were designated as unclassified viruses.

3.4 Results

Microbial community composition

Four stations were sampled along Line P in the northeastern Pacific on a transect from coastal waters to the open ocean between May 16-21, 2012 (figure 1a). Satellite derived chlorophyll concentrations for the month of May showed that the coastal station P1 had the highest chlorophyll concentrations, which decreased with distance from the coast at stations P4, P6, and P8. At the time of sampling, diatom abundance was highest at P1 where a bloom of *Pseudo-nitzschia pungens* dominated the community. Diatom abundance based on cell counts was lowest at P6 (Sara Bender, unpublished data, personal communication). Photosynthetic

picoeukaryotes were also most abundant at P1, and decreased at more oceanic stations. However, picophytoplankton became the dominant eukaryotic phytoplankton at stations P4, P6, and P8 (Francois Ribalet, unpublished flow cytometry data, personal communication). Cyanobacteria abundance followed an opposite trend from eukaryotic phytoplankton, being most abundant at the more oceanic station P6 and lowest in abundance at coastal P1 (Francois Ribalet, unpublished flow cytometry data, personal communication). *Synechococcus* clades I and IV composed >95% of the cyanobacterial communities at every station, and clade I was approximately 3 times more abundant than clade IV (Limnardo and Worden, unpublished 16S tag sequencing data).

Viral community composition and diversity

At each of the four stations, cellular metatranscriptomes from particle associated and free-living plankton communities were paired with semi-quantitative viral metagenomes (viromes) taken from the same water sample (figure 1b). Phylogenies of viral hallmark genes were used to assess viral diversity and abundance: the terminase large subunit (terL) represented bacteriophage, DNA polymerase B (polB) represented Nucleocytoplasmic Large DNA Viruses (NCLDV), archeoviruses, and two single stranded DNA (ssDNA) viruses, the major capsid protein (MCP) represented virophage, and the replication associated protein (vRep) represented ssDNA viruses. Sequences from contigs assembled from each station's virome were added to the trees to further capture viral diversity (supplemental figure 1). The leaves and nodes of the four phylogenies were designated as viral operational taxonomic units (OTUs) and 281 out of 972 possible OTUs recruited reads from the viral metagenomes or cellular metatranscriptomes. In the

viromes, 175 of the 281 OTUs were detected at all four stations. In the cellular metatranscriptomes, on the other hand, only 4 viral OTUs were detected at all four stations.

Viral abundance in the viromes did not correlate with high levels of transcription detected in the cellular metatranscriptomes. For example, SAR116 phage were the most abundant viruses at an average of 18.4% (±8.4%) of the total viral community (figure 2), yet represented only 1.2% of the total viral RNA transcripts compared to cyanophage (57.8% of transcripts), pelagiphage (6.6%), and unclassified phage (42.1%). Single stranded DNA (ssDNA) viruses were as abundant as bacteriophage composing 31.1% (±5.5%) of the community, but was the least expressed group of viruses representing only 2.8% of viral transcripts. They could be divided into 10 different groups based on their phylogeny. Each group was detected at every station in the viromes, but they were detected 41% of the time in the transcriptomes. The most detected was uncultured marine viruses at 7 or 8 transcriptomes while the least detected group was diatom viruses, which were never expressed. Viruses with polB including the NCLDV, archaeoviruses, and polB encoding ssDNA viruses were only 3.9% of the viral community, yet were comparable in their expression to ssDNA viruses and bacteriophage. Of that group, the photosynthetic picoeukaryote infecting phycodnaviruses were both the most abundant and the most expressed. Finally, virophage were both highly abundant and highly expressed, accounting for 28.1% (±9.0%) of the viral community and 77.6% (±14.6%) of the viral expression.

Connecting viral infection to host dynamics

Host abundance, viral abundance, and viral expression were examined for viral groups for which hosts have been identified in the marine environment. These host-viral groups are the *Synechoccocus* and *Synechoccocus* phage (Millard et al., 2009), *Pelagibacter* and pelagiphage

(Zhao et al., 2013), photosynthetic picoeukaryotes and picoeukaryote viruses (Nagasaki 2008), and Nucleocytoplasmic Large DNA viruses (NCLDVs) and virophage (Zhou et al., 2013). Additionally, diatoms were compared to eukaryotic single stranded DNA viruses, since diatoms are the only known hosts in the marine environment (Kimura and Tomaru, 2015), even though many of the eukaryotic ssDNA vRep sequences were not located within the known diatom virus clade (supplemental figure 1c). Virus to host ratios for cyanophage and picoeukaryote viruses were highest at P6, while pelagiphage, ssDNA viruses, and virophage had the greatest virus to host ratio at stations P1, P4, and P8 respectively (figure 3). Virus host ratios varied an order of magnitude or greater within each group at different stations (figure 3). The expression of each viral group varied by station. Cyanophage at P4 and pelagiphage at P1 had high virus:host ratios, but below average expression. However, most viral communities had low host:virus ratios but had higher than average expression. There were few instances where high virus:host ratios corresponded to high transcription, specifically P6 ssDNA viruses and P4 picoeukaryote viruses. For both of these instances, host abundance was the lowest observed across the entire transect. In fact, a negative logarithmic correlation existed between virus:host ratios and host abundance for all groups and stations (linear regression R^2=0.66, p=0.0001).

Strain specific expression differences

Viromes were assembled into contigs; over half of the reads from P4 and P6 viromes assembled, while 20% and 27% of P1 and P8 reads assembled respectively (supplemental table 1). A small minority of contigs from each virome recruited reads from the cellular free-living and particle attached metatranscriptomes (supplemental figure 2). The expressed fraction of the viral community decreased from the coastal station, at P1 with 16% of viral contigs expressed, to the

oceanic station, at P8 with 0.2% of contigs expressed. Viral contigs from each station recruited on average 0.48% (±0.16%) of reads from the corresponding free-living metatranscriptome versus 0.77% (±0.25%) of reads from the corresponding particle-attached metatranscriptome.

The viral groups examined above at P6 appeared to have consistently higher expression compared to other stations (figure 3). Examining expression of individual contigs revealed that P6 had a small group of contigs that were expressed orders of magnitude more than any other contig at the other stations (figure 4). *Synechococcus* phage, picoeukaryote viruses, and poxviruses were members of that highly expressed group (supplemental table 2). Most of these highly expressed contigs contained genes of hypothetical or unknown function. A putative structural protein and auxiliary metabolic gene (2OG-Fe(II) oxygenase) were identified on two of the cyanophage contigs and a putative ATP-binding cassette transporter containing contig from a poxvirus were highly expressed in both fractions (supplemental table 2).

Synechococcus phages were used as a model to explore strain specific differences in expression. A phylogeny of large terminase (*terL*) genes was constructed specifically to finely resolve taxonomic differences in *Synechococcus* phages using sequences from 34 complete *Synechococcus* phage genomes and other bacteriophage (supplemental figure 1a). Virome reads were mapped to the reference tree to identify and quantify *Synechococcus* phage at each station. Reads mapped to 29 of 34 isolates, and three isolates, S-RIP1, P60, and S-CRM01, recruited the 57% of reads (figure 5a). Furthermore, S-RIP1 and P60 represented 71% of the total *Synechococcus* phage transcripts. While both isolates recruited similar numbers of terL transcripts from stations P1, P4, and P6, at station P8 the P60 isolate recruited significantly more transcript reads.

Synechococcus phages are known to have different suites of host-derived auxiliary metabolic genes that are expressed during infection and that may benefit the viruses during replication (Labrie et al. 2013, Lindell et al. 2007, Bragg et al. 2009). Cyanophage auxiliary metabolic genes were detected in viromes at every station (figure 5b, supplemental table 2). Carbon and nucleotide metabolism, photosynthesis, and phosphate uptake related genes were the most abundant auxiliary metabolism genes at each station and occurred in similar ratios, except at station P1 (figure 5b). *PsbA* sequences, which encode the photosystem II D1 protein, in the viromes were phylogenetically related to previously sequenced cyanophage indicating that the photosynthesis genes were virally encoded (supplemental figure 3). However, the expression of these genes was dramatically different (figure 5c). Despite, P4 and P6 having nearly identical auxiliary metabolic gene abundances (figure 5b), photosynthesis genes were significantly more expressed at station P6 (figure 5c). In contrast, at station P4, an integrase, rather than a host-derived metabolism gene, was the most expressed auxiliary gene. At P6, 3 contigs out of 75 that were two standard deviations above the mean expression were photosynthesis related. These included a 2OG-Fe(II) oxygenase and photosystem II D1 protein (*psbA*), which were the 15th and 34th most expressed contigs in the free-living community, respectively, and a photosystem II D1 protein that was the 15th most expressed contig in the particle attached community. Interestingly, also in the top most expressed genes at P6 was a putative *ftsH2* gene from a picoeukaryote virus, which may help with the repair of photosystem D2 proteins (Komenda et al. 2010, Zaltsman et al. 2005).

3.5 Discussion

In 2007, Suttle suggested that viruses employ different life history strategies based on their replication dynamics and the population dynamics of their hosts to survive in the ocean. Viruses that have high burst sizes and replicate rapidly to take advantage of ephemeral host populations would be *r*-selected, while viruses that produce small numbers of progeny viruses upon lysis and co-exist with more stable host populations would be *K*-selected. This framework was based on culture-based observations of viral dynamics, but has not been evaluated in the environment. The changes in community composition of actively infecting viruses provided a window into the lifestyles of the viral taxa at these four stations. A small fraction of the viral community is actively infecting. This suggests that a few viral taxa out of the larger viral reservoir were proliferating. Several poorly understood viral taxa were caught in the act of infection.

Eukaryotic ssDNA viruses were abundant yet were the least transcriptionally active viral community at every station. Currently, diatoms are the only known hosts of eukaryotic ssDNA viruses in the marine environment (Kimura and Tomaru 2015). Curiously, despite the bloom of the diatom *Pseudo-nitzschia pungens* at P1, only 2 of 57 viral replication associated protein sequences from ssDNA viruses grouped with previously sequenced diatom viruses. This suggests that there may be considerably more diversity than the previously isolated diatom viruses. In fact, many of those sequences grouped with parvoviruses, which are linear ssDNA viruses, or an uncultured clade of ssDNA viruses, both of which have no known hosts, suggesting there may be other groups of ssDNA viruses that infect diatoms in addition to the currently isolated group or other hosts for ssDNA viruses in the marine environment besides diatoms. Diatom-infecting ssDNA viruses have been documented to have burst sizes in the range of thousands to tens-of-thousands per cell (Kimura and Tomaru, 2015). It appears that they may be examples of *r*-

selected organisms with high burst sizes that replicate rapidly to take advantage of ephemeral host populations.

In contrast, the other group of eukaryote infecting viruses which infect photosynthetic picoeukaryotes and heterotrophic unicellular protists, the NCLDV were proposed as *K*-selected viruses (Suttle, 2007) because they have lower burst sizes and their host populations are more stable. They were less abundant than the eukaryotic ssDNA viruses, but were more expressed than ssDNA viruses, suggesting that they employ a different strategy than the ssDNA viruses.

Virophage infect other viruses belonging to the NCLDV (LaScola et al. 2008). The environmental sequences from the viromes were not closely related to any of the isolated virophage, but instead related to virophage sequences from animal gut and freshwater metagenomes (Yau et al., 2011, Yutin et al., 2015). Using the hallmark genes to recruit virome and transcriptome reads revealed that they were among the must abundant viruses observed and accounted for the majority of viral transcription. Thus virophage are likely significant top down controls on NCLDV populations in the environment.

Archeovirus dynamics are poorly understood in the ocean. To date, archeoviruses have only been identified in marine sediments (Paul et al. 2014) and not in the water column. Two polymerase B sequences from these pelagic viromes were most closely related to extremophilic euryarcheal viruses. Additionally, Marine Group II euryarchaea were the most abundant archaeal taxa in the cellular community (data not shown). Thus, these viruses may be infecting Marine Group II archaea. The archeoviruses were detected in every virome, and transcriptionally active at all the stations except P8. The recent genome sequence and community analysis of a coastal metagenome revealed that marine group II euryarchaea are important components of microbial

communities, are motile, and may be associated with diatoms or large phytoplankton (Iverson et al. 2012), which were scarce at P8.

Viruses are architects of microbial community structure

Viral infection has been proposed as an important process, which structures microbial community composition and abundance. However, bottom up controls on microbial communities are excellent predictors of community structure independent of estimations of virally induced mortality (Follows et al., 2007). Host abundance, viral abundance, and viral transcription combined to reveal microbial communities engaged in different viral dynamics. The ratio of viruses to hosts varied up to 2 orders of magnitude between stations within a virus-host groups. Hypothetically, a high virus:host ratio is evidence of lysis, which reduces host numbers and increases viral abundance. High virus:host ratios were correlated with low host abundance. However, a high virus:host ratio did not necessarily coincide with high transcriptional activity of the viruses. The highest virus:host ratios for pelagiphage and cyanophage at P1 and P4, respectively, were linked to lower than average viral transcription. Few virus-host groups had both high expression and high virus:host ratios. Viruses with both high expression and high virus:host ratios may have been caught in the act of infecting a host population. In fact, many more viral populations had higher than average transcription, but low virus:host ratios. This is suggestive of impending lysis. These viral populations with different virus:host ratios and transcription are reminiscent of kill-the-winner dynamics, where relatively few viral taxa are actively infecting the most prolific hosts, which are then supplanted by new microbes (Thingstad et al. 2000). The lens of transcription highlights those viral taxa that are the current ‘winners’, those viruses that were once winners, and the bank of viruses that may succeed the current

winners. This provides a mechanistic link between kill-the-winner dynamics (Thingstad et al. 2000) and the viral seed bank hypothesis (Breitbart and Rohwer, 2005).

Auxiliary genes and the success of viruses

Kill the winner dynamics are thought to occur at the strain, rather than population level (Thingstad et al. 2015). While population level changes in host abundance were correlated with virus:host ratios, the taxonomic binning based on blast annotation did not provide enough resolution to determine if single strains of viruses were driving the overall transcriptional changes or if the entire viral population was just more infective. The transcription of individual contigs showed that P6 had a handful of contigs that were the highest expressed in the entire study. Picoeukaryote viruses, poxviruses, and cyanophage were the main members of that group. Intriguingly, these same groups at P6 had low or average virus:host ratios suggesting that these highly expressed viruses may have been in the process of removing a 'winning' host population. The importance of differences between strains and the effect those differences has on the success of viruses in the ocean is poorly understood.

Culture dependent work with cyanophage has shown important differences that occur at the strain level. Even though phage usually infect a narrow range of hosts, there are single hosts that are able to be infected by multiple phages (Dekel-bird et al 2014) and certain phages are able to infect multiple hosts (Sullivan et al. 2004, Flores et al. 2011). Therefore, cyanophages are likely competing with each other for hosts in the environment. Interestingly, in addition to lineage-defining viral genes, cyanophage genomes frequently encode host-derived and other auxiliary genes (Sullivan et al. 2005). These genes are involved in the hosts' metabolism, specifically in photosynthesis, light stress, phosphorous acquisition, and carbon metabolism

(Labrie et al. 2013). Viruses express these host-related genes during infection (Lindell et al. 2005, Lindell et al. 2007). Modeling studies suggest that host genes may allow the viruses to ultimately produce more progeny viruses under certain conditions (Bragg et al. 2008, Hellweger 2009). Not all cyanophage encode host metabolism genes or have the same suite of genes (Labrie et al. 2013, Dekel-bird et al. 2014). Therefore, we explored whether there were expression differences by strain and whether host metabolism genes might have played a role in the success of *Synechoccocus* phage at the four stations.

Using a phylogenetically enabled approach showed that *Synechoccocus* phage communities were composed of viruses similar to 29 different isolates, with viruses similar to 3 isolates dominating the community at every station except P1. Two of those three most abundant strains were highly expressed. One of those isolates, P60, does not have photosynthesis genes (Chen and Lu 2002) while the other, RIP-1, encodes the photosynthesis gene *psbA*. Cyanophage at P1 had a different ratio of auxiliary genes compared to the other stations, possibly explained by the differences in the phage community. P4 and P6 were remarkably similar in auxiliary gene abundance, virus abundance, and host community composition (Limnardo and Worden, unpublished data). Strikingly, however, photosynthesis genes had significantly higher expression at P6 and were, in fact, some of the highest expressed genes in the whole viral community. Integration genes were the most highly expressed at P4, on the other hand. These two communities lie on different quadrants of the virus:host ratio vs expression plot. The P4 community has a high virus:host ratio, but lower than average expression indicating that viruses in the community may have recently lysed a dominant *Synechococcus* strain. When host abundance is low, viruses may integrate in order allow them to continue to persist in the environment (McDaniel et al. 2002). On the other hand, P6 communities seem poised to

dethrone the current winning host taxa. Their expression of photosynthesis genes may have facilitated their success over non-photosynthesis containing phage. Additionally, the expression of these host genes is a window onto the stressors of the host (Lindell et al. 2007). A putative *ftsH2*, which is involved in photosynthesis and photosystem II repair *Synechocystis* and *Arabidopsis* (Komenda et al. 2010, Zaltsman et al. 2005) was also one of the highest expressed genes at P6, but was encoded by a picoeukaryote virus suggesting that other groups of viruses may be utilizing similar pathways to manipulate host metabolisms in similar environments. Therefore, these auxiliary host genes are likely important tools that viruses leverage for situational success in the marine environment.

Ultimately, any virus that is infecting a host and presumably making progeny viruses is successful as they have survived numerous obstacles such as diffusion, finding a host, and conquering host immunity. Pairing metatranscriptomes with metagenomes has provided a new insight onto essential questions about which viruses are successful and why they might be succeeding. This technique uncovered a link between two major conceptual models of viral ecology, viral seed bank and kill-the-winner dynamics. Directly quantifying viral infection is a critical and significant step towards understanding how viruses contribute to microbial community composition and biogeochemical cycling.

3.6 Acknowledgements

We would like to thank the captain and crew of the RV R/V *Thomas G. Thompson* as well as co-chief scientist of the GeoMICS cruise, Jim Moffett. Additionally, we would like to thank all the members GeoMICS working group and in particular, Sara Bender and Megan

Schatz for diatom counts, AJ Limnardo, Alex Worden for 16S tag sequencing data, Francois Ribalet for underway plankton data, and Jaci Saunders for obtaining the samples. Also, we are appreciative of the assistance of Clara Fuchsman and Cedar McKay in virome assembly. Funding was provided by a grant from the National Science Foundation: OCE-1356779 to GR and a University of Washington, School of Oceanography Innovation Grant to MC. MC was funded by an EPA STAR fellowship.

3.7 References

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. a Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. a Suttle, and F. Rohwer. 2006. The marine viromes of four oceanic regions. PLoS Biol. **4**: e368.

Berger, S. a, and A. Stamatakis. 2011. Aligning short reads to reference alignments and trees. Bioinformatics **27**: 2068–75.

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

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

Bragg, J. G., and S. W. Chisholm. 2008. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. PLoS One **3**: e3550.

Breitbart, M., and F. Rohwer. 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. **13**: 278–84.

Breitbart, M., J. H. Miyake, and F. Rohwer. 2004. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. **236**: 249–256.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U. S. A. **99**: 14250–5.

Brown, C. T., A. Howe, Q. Zhang, A. B. Pyrkosz, and T. H. Brom. 2012. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. 1–18.

Brum, J. R., J. C. Ignacio-espinoza, S. Roux, G. Doulcier, S. G. Acinas, A. Alberti, and S. Chaffron. ocean viral communities.

Brussaard, C. P. D. 2004. The Journal of Eukaryotic Microbiology. **51**: 125–138.

Chappell, P. D., L. P. Whitney, J. R. Wallace, A. I. Darer, S. Jean-Charles, and B. D. Jenkins. 2015. Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica from the northeast Pacific Ocean. ISME J. **9**: 592–602.

Chen, F., and J. Lu. 2002. Genomic Sequence and Evolution of Marine Cyanophage P60 : a New Insight on Lytic and Lysogenic Phages Genomic Sequence and Evolution of Marine Cyanophage P60 : a New Insight on Lytic and Lysogenic Phages †. Appl. Environ. Microbiol. **68**: 2589–2594.

Cottrell, M. T., and C. A. Suttle. 1995. Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate Micromonas pusilla. **40**: 730–739.

Culley, A. I., A. S. Lang, and C. a Suttle. 2006. Metagenomic analysis of coastal RNA virus communities. Science **312**: 1795–8.

Culley, A. I., C. A. Suttle, and G. F. Steward. 2010. Characterization of the diversity of marine RNA viruses. 193–201.

Culley, A. I., J. A. Mueller, and M. Belcaid. 2014. The Characterization of RNA Viruses in Tropical Seawater Using Targeted PCR and Metagenomics. , doi:10.1128/mBio.01210-14.Editor

Deher, T., J. G. Zeevaart, L. Wang, V. V Thakur, C. S. Leung, J. Tirado-, C. M. Bailey, R. a Domaoal, K. S. Anderson, and L. William. 2009. NIH Public Access. **130**: 9492–9499.

Dekel-Bird, N. P., G. Sabehi, B. Mosevitzky, and D. Lindell. 2014. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. Environ. Microbiol. **17**: 1286–1299.

Djikeng, A., R. Halpin, R. Kuzmickas, J. Depasse, J. Feldblyum, N. Sengamalay, C. Afonso, X. Zhang, N. G. Anderson, E. Ghedin, and D. J. Spiro. 2008. Viral genome sequencing by random priming methods. BMC Genomics **9**: 5.

Duhaime, M. B., L. Deng, B. T. Poulos, and M. B. Sullivan. 2012. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Environ. Microbiol. no–no.

Durham, B. P., S. Sharma, H. Luo, C. B. Smith, S. a. Amin, S. J. Bender, S. P. Dearth, B. a. S. Van Mooy, S. R. Campagna, E. B. Kujawinski, E. V. Armbrust, and M. A. Moran. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. Proc. Natl. Acad. Sci. **112**: 453–457.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**: 1792–1797.

Flores, C. O., J. R. Meyer, S. Valverde, L. Farr, and J. S. Weitz. 2011. Statistical structure of host-phage interactions. Proc. Natl. Acad. Sci. U. S. A. **108**: E288–97.

Follows, M. J., and S. Dutkiewicz. 2011. Modeling diverse communities of marine microbes. Ann. Rev. Mar. Sci. **3**: 427–451.

Gilbert, J. a, D. Field, Y. Huang, R. Edwards, W. Li, P. Gilna, and I. Joint. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One **3**: e3042.

Gobler, C. J., D. a. Hutchins, N. S. Fisher, E. M. Cosper, and S. Saøudo-Wilhelmy. 1997. Release and bioavailability of C, N, P Se, and Fe following viral lysis of a marine chrysophyte. Limnol. Oceanogr. **42**: 1492–1504.

Goldsmith, D. B., R. J. Parsons, D. Beyene, P. Salamon, and M. Breitbart. 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.

Hewson, I., S. Govil, D. Capone, E. Carpenter, and J. Fuhrman. 2004. Evidence of Trichodesmium viral lysis and potential significance for biogeochemical cycling in the oligotrophic ocean. Aquat. Microb. Ecol. **36**: 1–8.

Hopkins, M., S. Kailasan, A. Cohen, S. Roux, K. P. Tucker, A. Shevenell, M. Agbandje-McKenna, and M. Breitbart. 2014. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. ISME J. 1–11.

Huang, Y., B. Niu, Y. Gao, L. Fu, and W. Li. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics **26**: 680–682.

Hurwitz, B. L., and M. B. Sullivan. 2013. The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology F. Thompson [ed.]. PLoS One **8**: e57355.

Iverson, V., R. M. Morris, C. D. Frazar, C. T. Berthiaume, R. L. Morales, and E. V. Armbrust. 2012. Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. Science **335**: 587–90.

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

Keesing, F., L. K. Belden, P. Daszak, A. Dobson, C. D. Harvell, R. D. Holt, P. Hudson, A. Jolles, K. E. Jones, C. E. Mitchell, S. S. Myers, T. Bogich, and R. S. Ostfeld. 2010. Impacts

of biodiversity on the emergence and transmission of infectious diseases. Nature **468**: 647–652.

Kimura, K., and Y. Tomaru. 2015. Discovery of Two Novel Viruses Expands the Diversity of Single-Stranded DNA and Single-Stranded RNA Viruses Infecting a Cosmopolitan Marine Diatom. Appl. Environ. Microbiol. **81**: 1120–1131.

Komenda, J., J. Knoppová, V. Krynická, P. J. Nixon, and M. Tichý. 2010. Role of FtsH2 in the repair of Photosystem II in mutants of the cyanobacterium Synechocystis PCC 6803 with impaired assembly or stability of the CaMn4 cluster. Biochim. Biophys. Acta - Bioenerg. **1797**: 566–575.

La Scola, B., C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin, and D. Raoult. 2008. The virophage as a unique parasite of the giant mimivirus. Nature **455**: 100–4.

Labonté, J. M., and C. a Suttle. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J. 1–9.

Labrie, S. J., K. Frois-Moniz, M. S. Osburne, L. Kelly, S. E. Roggensack, M. B. Sullivan, G. Gearin, Q. Zeng, M. Fitzgerald, M. R. Henn, and S. W. Chisholm. 2013. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. Environ. Microbiol. **15**: 1356–76.

Lindell, D., J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen, W. R. Hess, G. M. Church, and S. W. Chisholm. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature **449**: 83–6.

Lindell, D., J. D. Jaffe, Z. I. Johnson, G. M. Church, and S. W. Chisholm. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. Nature **438**: 86–9.

Malits, a., U. Christaki, I. Obernosterer, and M. G. Weinbauer. 2014. Enhanced viral production and virus-mediated mortality of bacterioplankton in a natural iron-fertilized bloom event above the Kerguelen Plateau. Biogeosciences Discuss. **11**: 10827–10862.

Mizuno, C. M., R. Ghai, and F. Rodriguez-Valera. 2014. Evidence for metaviromic islands in marine phages. Front. Microbiol. **5**: 27.

Morris, R. M., B. L. Nunn, C. Frazar, D. R. Goodlett, Y. S. Ting, and G. Rocap. 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J. **4**: 673–85.

Philippe, N., M. Legendre, G. Doutre, Y. Couté, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J.-M. Claverie, and C. Abergel. 2013. Pandoraviruses: amoeba

viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science **341**: 281–6.

Schmidt, H. F., E. G. Sakowski, S. J. Williamson, S. W. Polson, and K. E. Wommack. 2014. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. ISME J. **8**: 103–14.

Stamatakis, a. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.

Steward, G. F., A. I. Culley, J. A. Mueller, and E. M. Wood-charlson. 2012. Are we missing half of the viruses in the ocean ? ISME **7**: 1–8.

Sullivan, M. B., J. B. Waterbury, and S. W. Chisholm. 2003. Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. Nature **424**: 1047–51.

Sullivan, M. B., M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm. 2005. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. PLoS Biol. **3**: e144.

Suttle, C. a. 2005. Viruses in the sea. Nature **437**: 356–61.

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

Thingstad, T. F., B. Pree, J. Giske, and S. Vage. 2015. What difference does it make if viruses are strain- , rather than species- specific? Front. Microbiol. , doi:10.3389/fmicb.2015.00320

Yilmaz, S., M. Allgaier, and P. Hugenholtz. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat. Methods **7**: 943–4.

Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**: 821–9.

Zhao, Y., B. Temperton, J. C. Thrash, M. S. Schwalbach, K. L. Vergin, Z. C. Landry, M. Ellisman, T. Deerinck, M. B. Sullivan, and S. J. Giovannoni. 2013. Abundant SAR11 viruses in the ocean. Nature **494**: 357–60.

Zhu, W., a. Lomsadze, and M. Borodovsky. 2010. Ab initio gene identification in metagenomic sequences. Nucleic Acids Res. **38**: e132–e132.

3.8 Figure legends

Figure 3.1 Sampling of metatranscriptomes and viromes along the Line P transect. a) Satellite derived chlorophyll concentrations for May 2012 and the locations of the 4 stations where metatranscriptomes and viromes were taken. b) Schematic of water sampling at each station showing fractionation of the same water used to create free-living and particle attached metatranscriptomes and viromes.

Figure 3.2 The abundance and expression of viral groups from each station. Reads from viromes and metatranscriptomes were identified and quantified by mapping to four phylogenetic trees of DNA polymerase B, major capsid protein, replication associated protein, and the terminase. Viral abundance is expressed as logarithm of read abundance after normalization for gene size and library size. Warm colors represent higher viral abundance or transcription while cool colors represent lower abundance or transcription. Black indicates no viruses were detected.

Figure 3.3 Comparison of virus:host ratio to viral expression. The x-axis is the virus host ratio plotted on a log scale. The y-axis is the log ratio of RNA coverage to DNA coverage centered at the average ratio for all expressed contigs. The shaded region represents the standard deviation around the mean expression. Viral groups with known hosts in the marine environment (colors) are plotted by station (shapes): red is diatoms and ssDNA viruses, yellow is photosynthetic picoeukaryotes and their viruses, dark blue is *Synechococcus* and their phage, cyan is SAR 11 and pelagiphage, green is NCLDVs and virophage. Circles are station 1, diamonds are station P4,

squares are station P6, and triangles are station P8. No expression was detected for P1 virophage represented by an unfilled circle at the lowest possible expression.

Figure 3.4 Per contig expression for expressed contigs at each station: a) P1 b) P4 c) P6 d) P8. Expression above 0 means contigs recruited more RNA reads than DNA reads. Circles are contigs expressed in the free-living fraction, triangles in the particle attached. Lines represent two standard deviations above the average expression for that station. Teal is for free-living expressed contigs, orange for particle associated.

Figure 3.5 Strain specific *Synechococcus* phage abundance and expression. a) Abundance (top) and expression (bottom) of *Synechococcus* phage strains based on phylogenetic placement of virome and metatranscriptome reads. b) Abundance of cyanophage auxiliary genes in viromes grouped by function at each station. c) Expression of cyanophage auxiliary genes with the standard deviation for each group. Error bars are standard deviation

3.9 Supplemental figure legends

Supplemental figure 3.1 Phylogenies of a) large terminase (terL), b) polymerase B (polB), c) replication associated proteins (vRep), and d) major capsid proteins (MCP). Red taxa are sequences from environmental viral contigs from this study.

Supplemental figure 3.2 Expression of the viral community. a) Percent of the total contigs that were expressed in red.

Supplemental figure 3.3 Phylogeny of cyanobacterial and cyanophage photosystem D1 proteins (psbA). psbA sequences from environmental virus contigs from this study are highlighted in red.

Supplemental table 3.1 Virome library, assembly and protein coding statistics

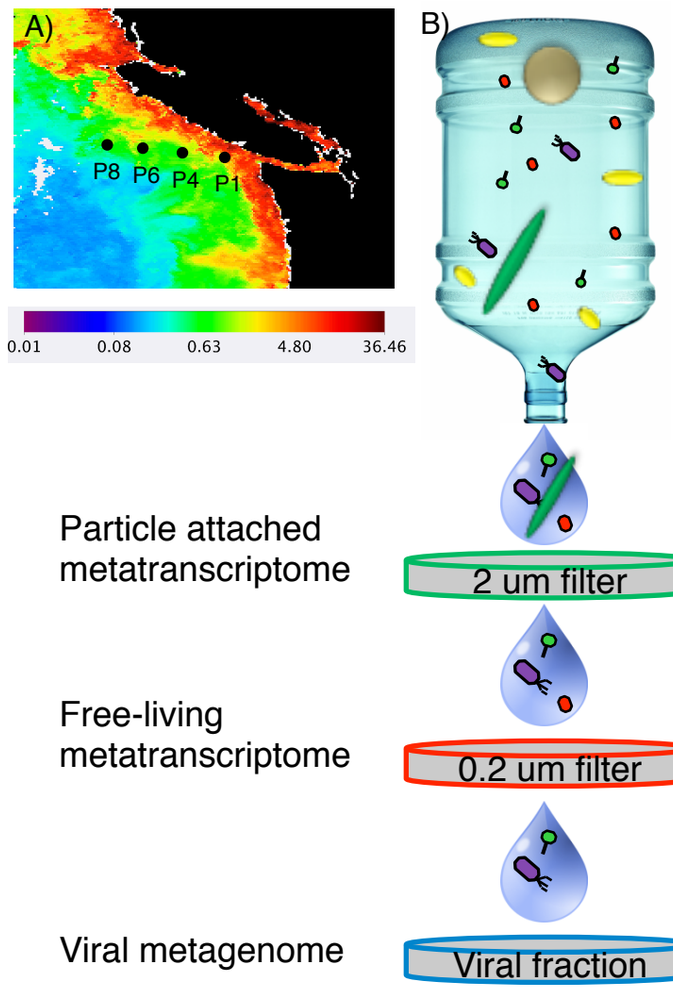
Station	# Reads	% reads in assembly	# contigs	Total contig length
P1	10,836,692	20	11,800	5,053,178
P4	9,008,045	57	101,748	38,718,766
P6	12,536,819	54	126,263	47,381,822
P8	14,271,211	27	73,065	26,160,382

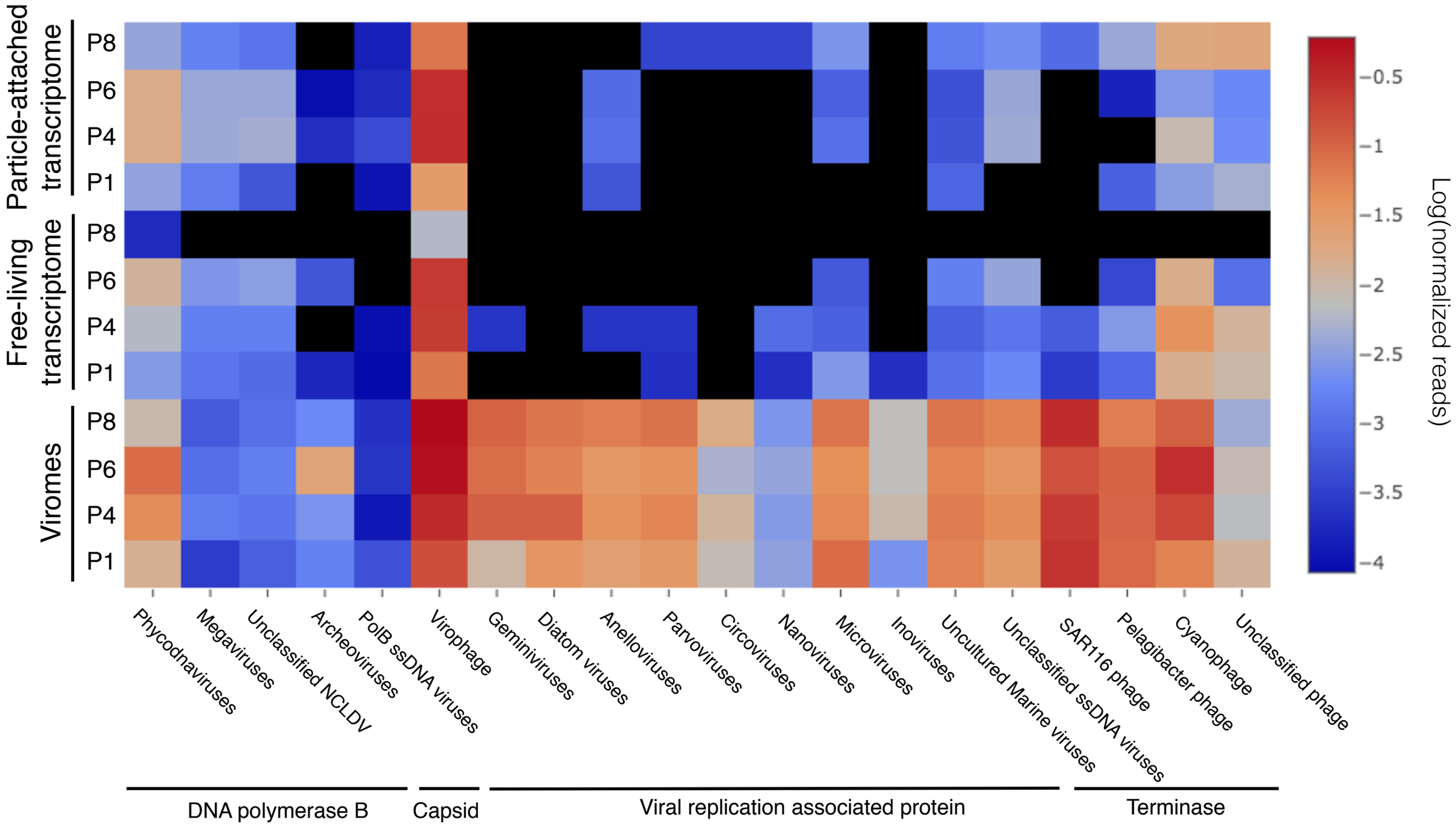
Supplemental table 3.2 Taxonomic and functional identification of contigs from the most highly expressed cluster at station P6

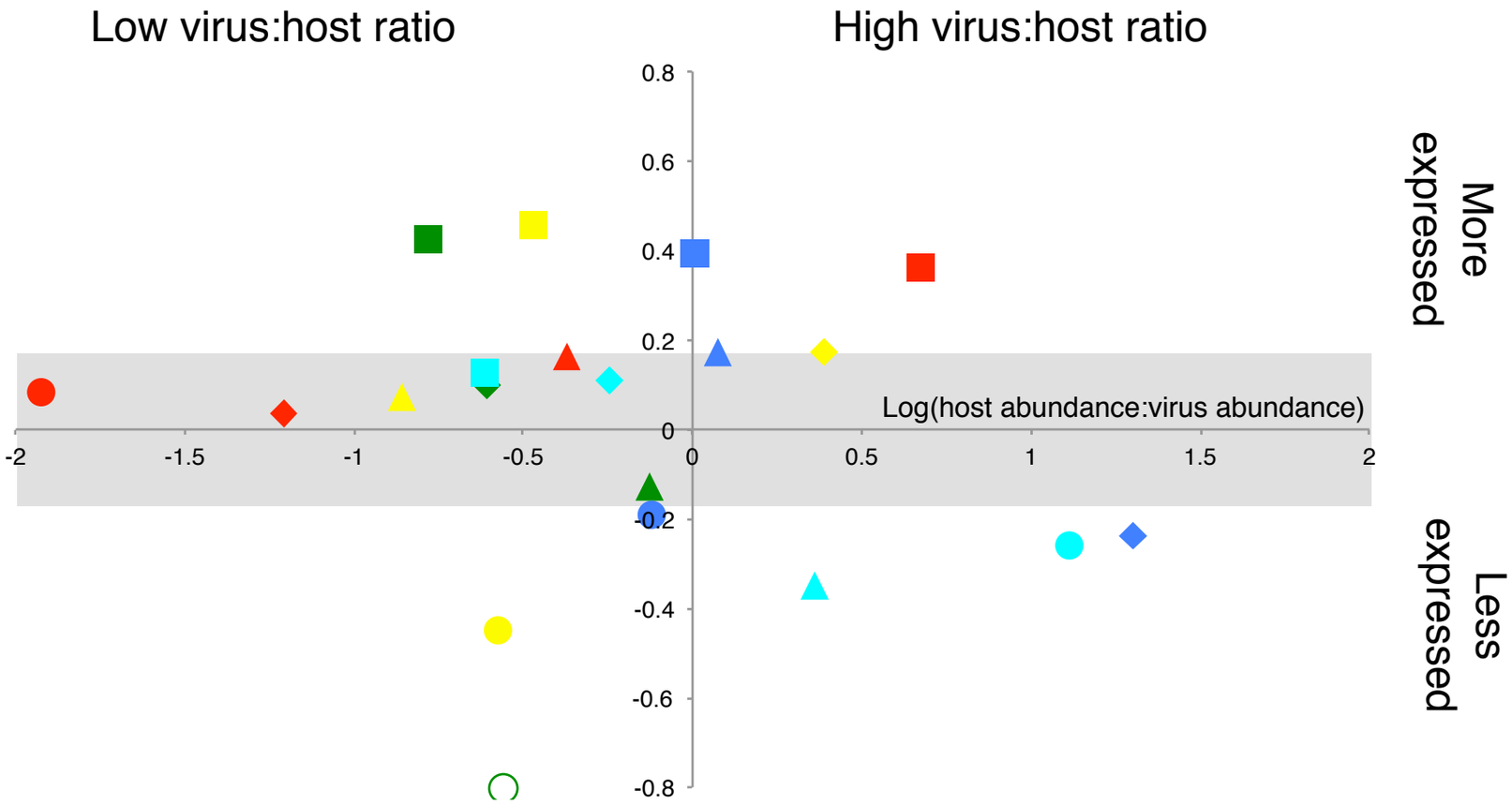
Contig Identifier	e-value	Functional annotation	Taxonomic annotation	Log(RNA:DNA)
Expressed in free-living fraction				
NODE_232530	No hit			6.54
NODE_196263	1.49E-35	hypothetical protein OTV1_016	Ostreococcus tauri virus 1	5.48
NODE_24168	1.70E-21	hypothetical protein Rm378p031	Rhodothermus phage RM378	5.43
NODE_281301	4.46E-06	hypothetical protein MpV1_166c	Micromonas sp. RCC1109 virus MpV1	5.28
NODE_281342	1.03E-06	putative ATP-binding cassette transporter	Amsacta moorei entomopoxvirus 'L'	5.23
NODE_285888	0.000505468	virion structural protein	Cyanophage S-TIM5	4.96
NODE_127075	No hit			4.96
NODE_271073	3.10E-28	hypothetical protein BpV1_139c	Bathycoccus sp. RCC1105 virus BpV1	4.95
NODE_301146	1.15E-18	hypothetical protein Syn7803C33_64	Synechococcus phage ACG-2014a	4.94
NODE_295832	1.61E-11	hypothetical protein BpV1_033	Bathycoccus sp. RCC1105 virus BpV1	4.91
NODE_151239	2.64E-53	hypothetical protein SXCG_00065	Synechococcus phage S-CAM8	4.9
NODE_272151	No hit			4.84
NODE_103848	No hit			4.84
NODE_221938	No hit			4.83
NODE_316992	6.42E-09	2OG-Fe(II) oxygenase	Synechococcus phage S-SM2	4.83
Expressed in particle attached fraction				
NODE_232530	No hit			6.07
NODE_281342	1.03E-06	putative ATP-binding cassette transporter	Amsacta moorei entomopoxvirus 'L'	5.28
NODE_172691	No hit			4.96
NODE_271073	3.10E-28	hypothetical protein BpV1_139c	Bathycoccus sp. RCC1105 virus BpV1	4.95

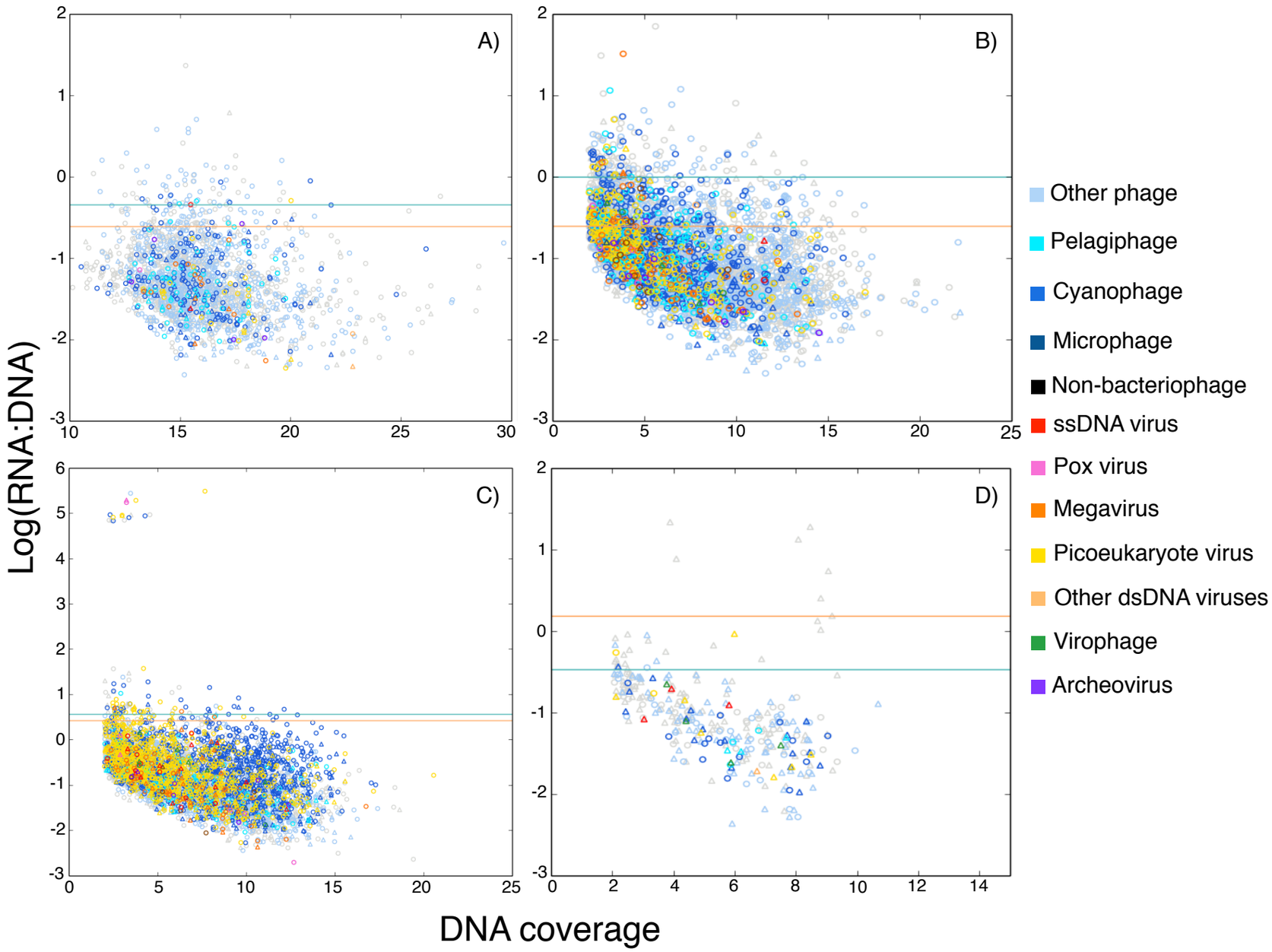
Supplemental table 3.3 *Synechococcus* phage auxiliary genes in each functional category

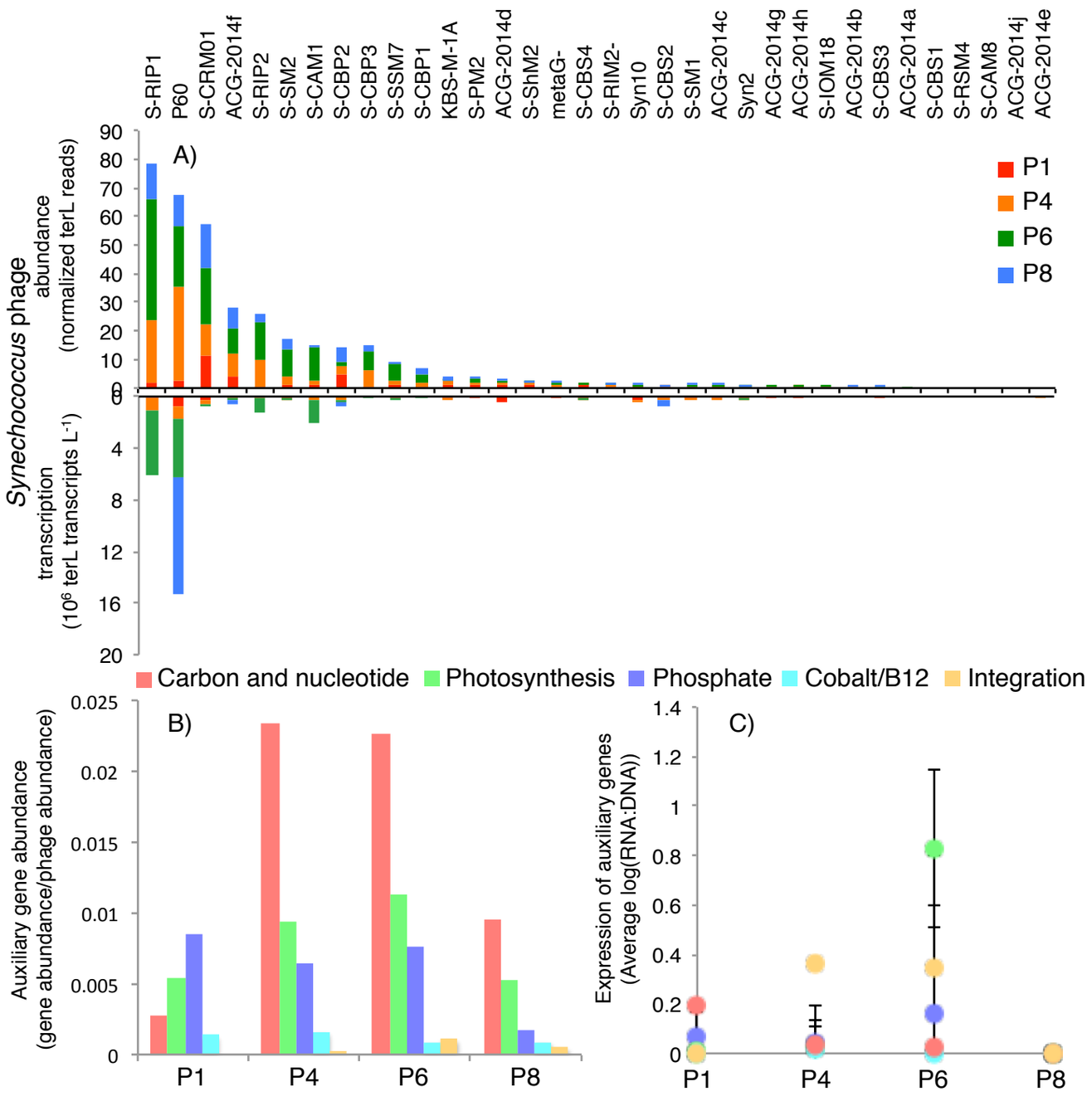
Functional Category	**Gene**	**Function**
Carbon and nucleotide metabolism	*talC*	transaldolase
	mazG	nucleotide pyrophosphatase
	purT	phosphoribosylglycinamide formyltransferase
	purM	phosphoribosylaminoimidazole synthetase
	nrdA/B/C	ribonucleotide reductase subunits A/B/C
Photosynthesis	*petF*	ferredoxin
	FECH	ferrochelatase
	hli	high light inducible protein
	psbA	photosystem II D1 protein
	psbD	photosystem II D2 protein
	petE	plastocyanin
	regA	photosynthetic apparatus regulatory protein
	2OG	2OG-Fe(II) oxygenase
Phosphorus metabolism	*phoH*	P-starvation inducible protein
	pstS	high affinity phosphate transporter
Cobalt/B12 metabolism	*cobS*	cobalt chelatase subunit
	cobT	porphyrin biosynthetic protein
Integration	*int*	integrase

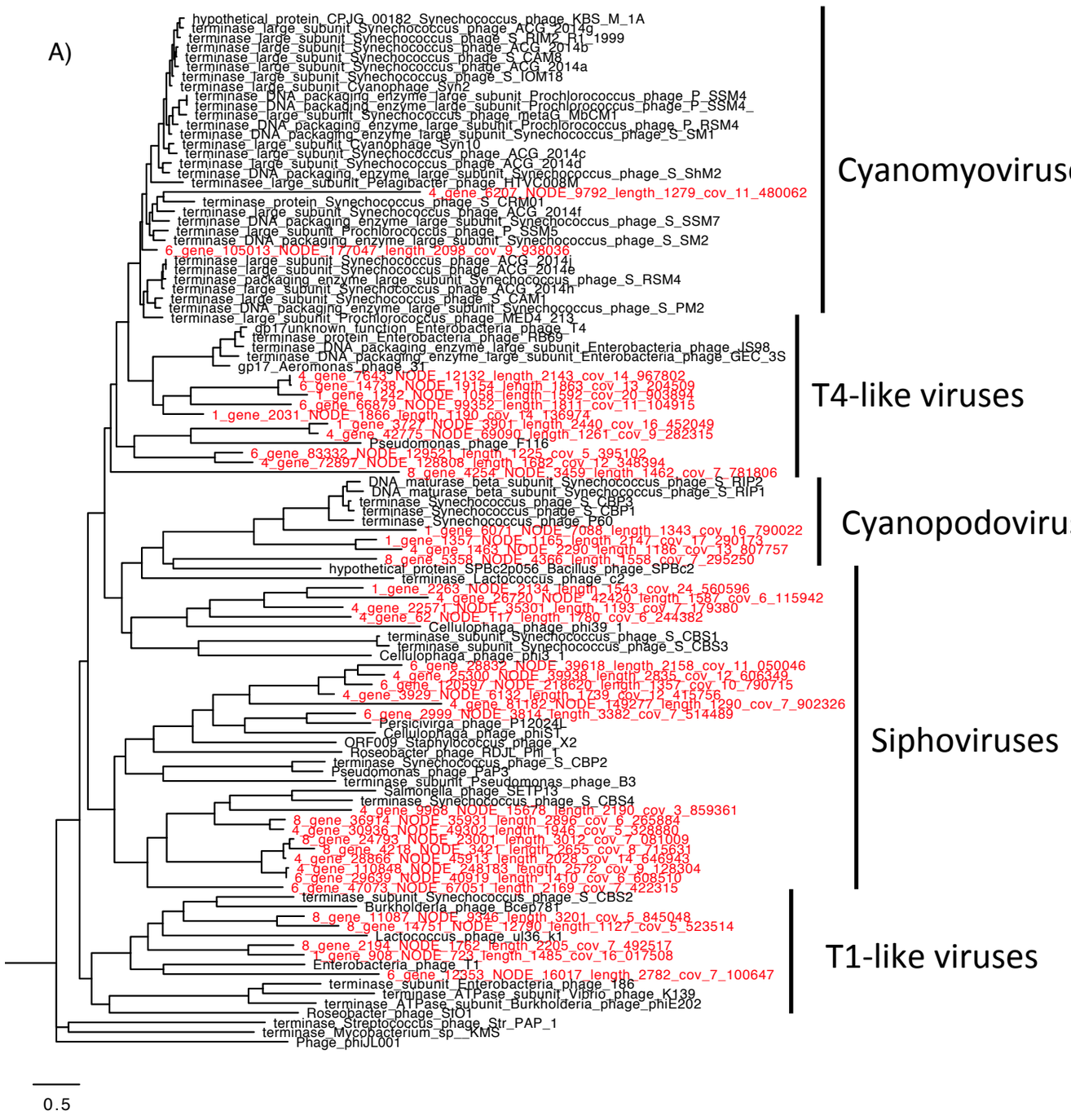

A)
P8
P6
P4
P1
0.01
0.08
0.63
4.80
36.46
B)
Particle attached
metatranscriptome
2 um filter
Free-living
metatranscriptome
0.2 um filter
Viral metagenome
Viral fraction


Particle-attached transcriptome
Free-living transcriptome
Viromes
P8
P6
P4
P1
P8
P6
P4
P1
P8
P6
P4
P1
−0.5
−1
−1.5
−2
−2.5
−3
−3.5
−4
Log(normalized reads)
Phycodnaviruses
Megaviruses
Unclassified NCLDV
Archeoviruses
PolB ssDNA viruses
Virophage
Geminiviruses
Diatom viruses
Anelloviruses
Parvoviruses
Circoviruses
Nanoviruses
Microviruses
Inoviruses
Uncultured Marine viruses
Unclassified ssDNA viruses
SAR116 phage
Pelagibacter phage
Cyanophage
Unclassified phage
DNA polymerase B
Capsid
Viral replication associated protein
Terminase

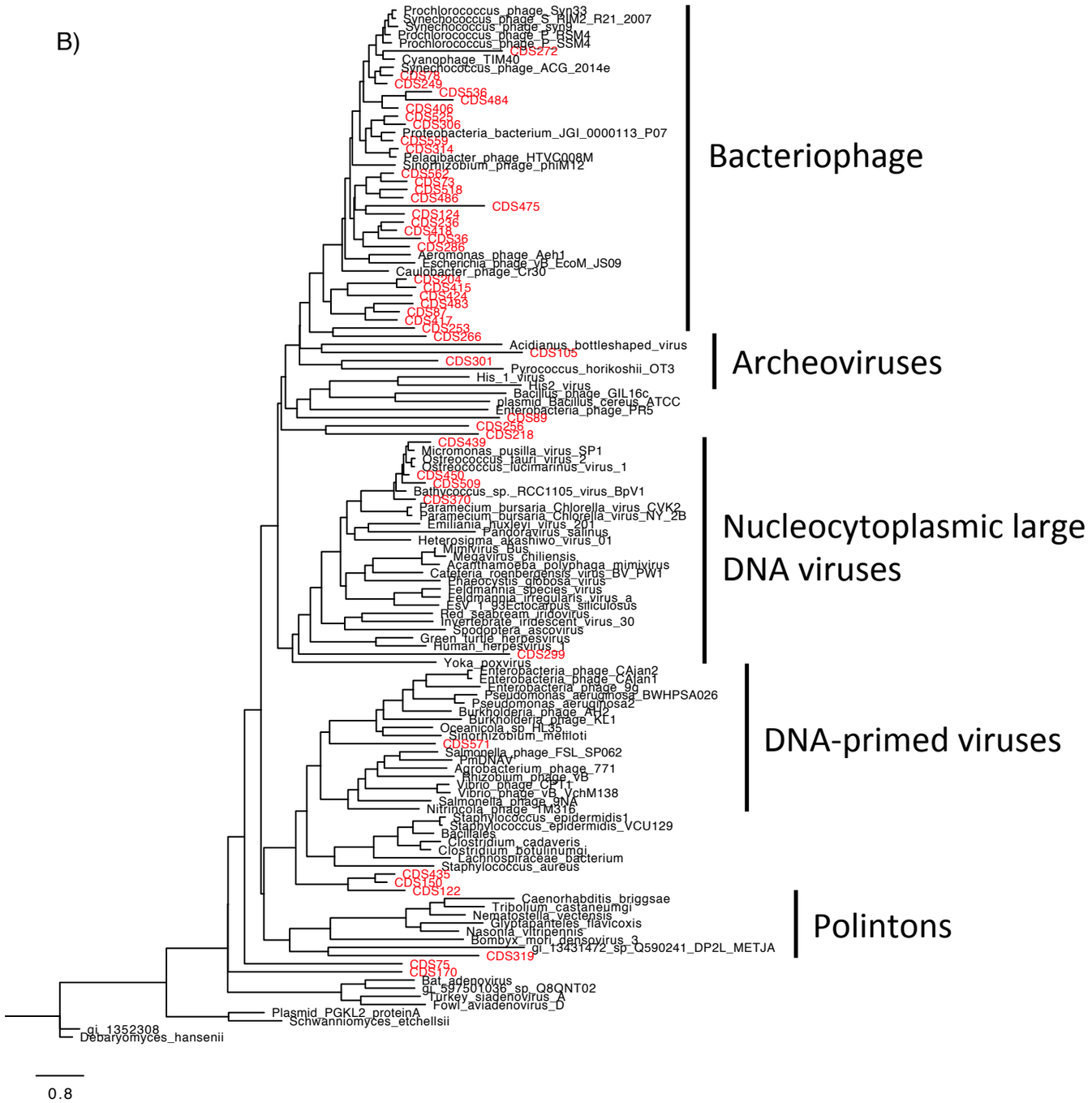

Low virus:host ratio
High virus:host ratio
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
Log(host abundance:virus abundance)
More expressed
Less expressed

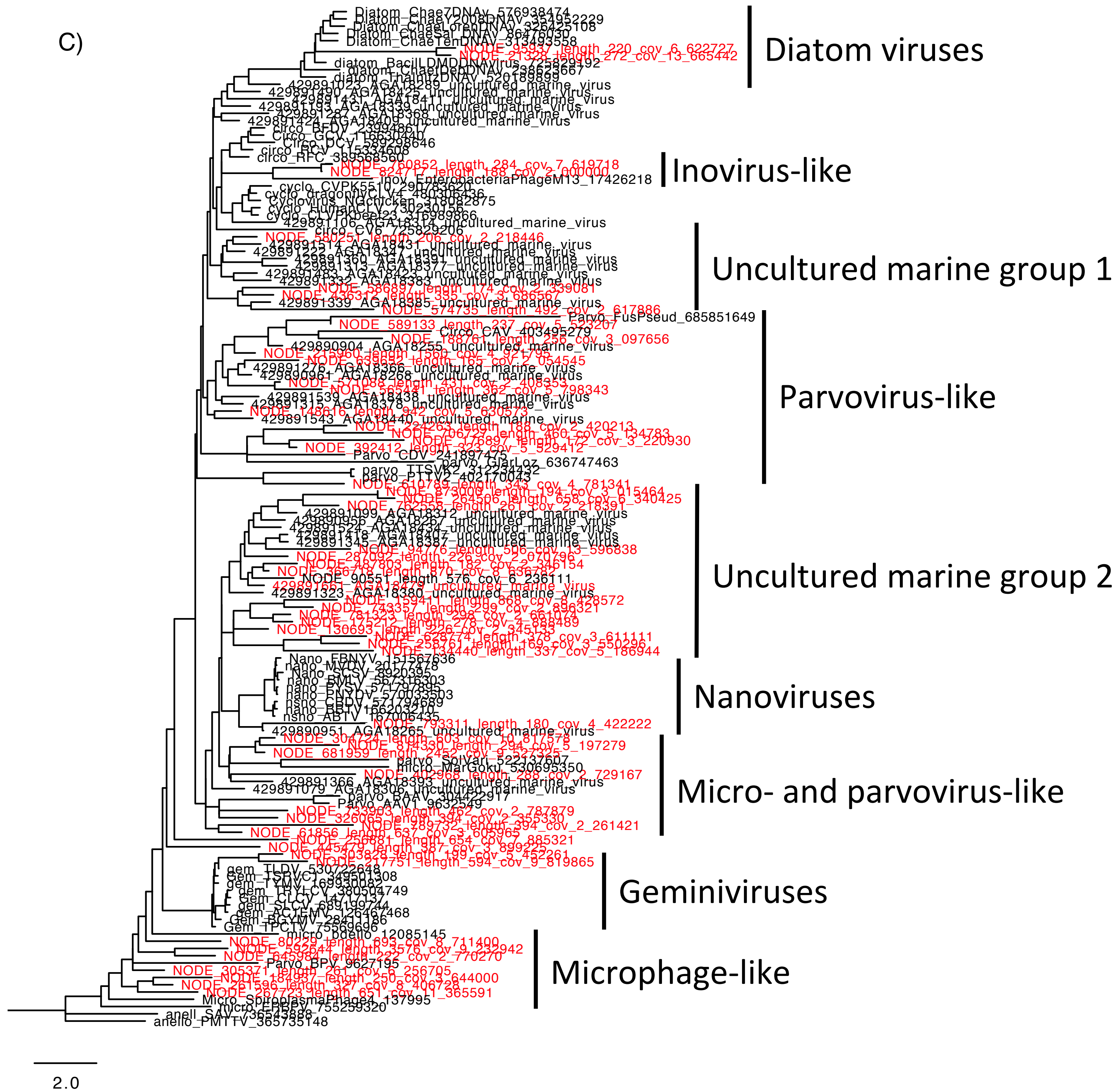

A)
B)
C)
D)
Log(RNA:DNA)
DNA coverage
Other phage
Pelagiphage
Cyanophage
Microphage
Non-bacteriophage
ssDNA virus
Pox virus
Megavirus
Picoeukaryote virus
Other dsDNA viruses
Virophage
Archeovirus


S-RIP1
P60
S-CRM01
ACG-2014f
S-RIP2
S-SM2
S-CAM1
S-CBP2
S-CBP3
S-SSM7
S-CBP1
KBS-M-1A
S-PM2
ACG-2014d
S-ShM2
metaG-
S-CBS4
S-RIM2-
Syn10
S-CBS2
S-SM1
ACG-2014c
Syn2
ACG-2014g
ACG-2014h
S-IOM18
ACG-2014b
S-CBS3
ACG-2014a
S-CBS1
S-RSM4
S-CAM8
ACG-2014j
ACG-2014e
A)
Synechococcus phage
abundance
(normalized terL reads)
transcription
(10^6 terL transcripts L^-1)
P1
P4
P6
P8
Carbon and nucleotide
Photosynthesis
Phosphate
Cobalt/B12
Integration
B)
Auxiliary gene abundance
(gene abundance/phage abundance)
C)
Expression of auxiliary genes
(Average log(RNA:DNA))

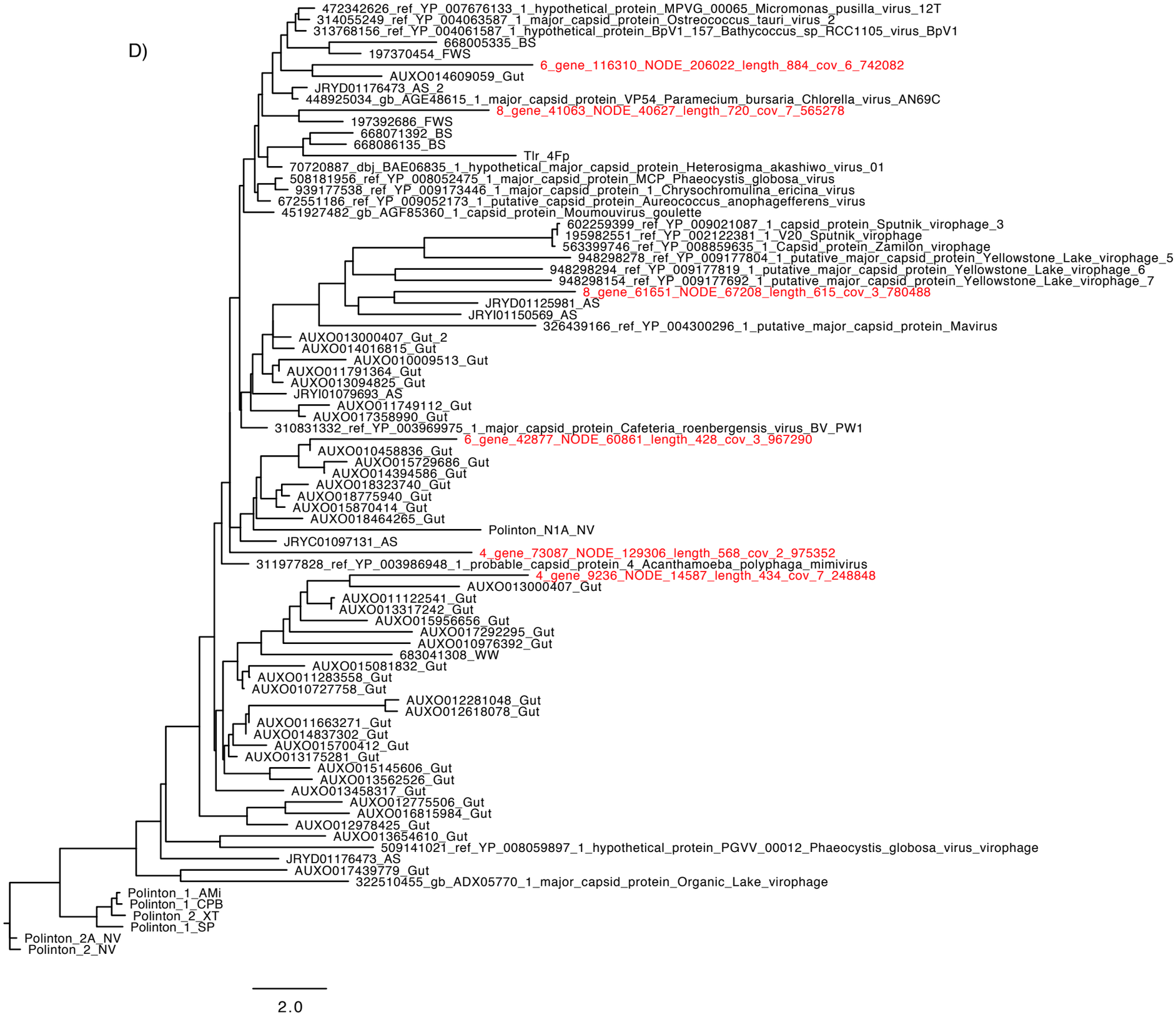

A)
hypothetical_protein_CPJG_00182_Synechococcus_phage_KBS_M_1A
terminase_large_subunit_Synechococcus_phage_ACG_2014g
terminase_large_subunit_Synechococcus_phage_S_RIM2_R1_1999
terminase_large_subunit_Synechococcus_phage_ACG_2014b
terminase_large_subunit_Synechococcus_phage_S_CAM8
terminase_large_subunit_Synechococcus_phage_ACG_2014a
terminase_large_subunit_Synechococcus_phage_S_IOM18
terminase_large_subunit_Cyanophage_Syn2
terminase_DNA_packaging_enzyme_large_subunit_Prochlorococcus_phage_P_SSM4
terminase_DNA_packaging_enzyme_large_subunit_Prochlorococcus_phage_P_SSM4_
terminase_large_subunit_Synechococcus_phage_metaG_MbCM1
terminase_DNA_packaging_enzyme_large_subunit_Prochlorococcus_phage_P_RSM4
terminase_DNA_packaging_enzyme_large_subunit_Synechococcus_phage_S_SM1
terminase_large_subunit_Cyanophage_Syn10
terminase_large_subunit_Synechococcus_phage_ACG_2014c
terminase_large_subunit_Synechococcus_phage_ACG_2014d
terminase_DNA_packaging_enzyme_large_subunit_Synechococcus_phage_S_ShM2
terminasee_large_subunit_Pelagibacter_phage_HTVC008M
4_gene_6207_NODE_9792_length_1279_cov_11_480062
terminase_protein_Synechococcus_phage_S_CRM01
terminase_large_subunit_Synechococcus_phage_ACG_2014f
terminase_DNA_packaging_enzyme_large_subunit_Synechococcus_phage_S_SSM7
terminase_large_subunit_Prochlorococcus_phage_P_SSM5
terminase_DNA_packaging_enzyme_large_subunit_Synechococcus_phage_S_SM2
6_gene_105013_NODE_177047_length_2098_cov_9_938036
terminase_large_subunit_Synechococcus_phage_ACG_2014i
terminase_large_subunit_Synechococcus_phage_ACG_2014e
terminase_packaging_enzyme_large_subunit_Synechococcus_phage_S_RSM4
terminase_large_subunit_Synechococcus_phage_ACG_2014h
terminase_large_subunit_Synechococcus_phage_S_CAM1
terminase_DNA_packaging_enzyme_large_subunit_Synechococcus_phage_S_PM2
terminase_large_subunit_Prochlorococcus_phage_MED4_213
gp17unknown_function_Enterobacteria_phage_T4
terminase_protein_Enterobacteria_phage_RB69
terminase_DNA_packaging_enzyme_large_subunit_Enterobacteria_phage_JS98
terminase_DNA_packaging_enzyme_large_subunit_Enterobacteria_phage_GEC_3S
gp17_Aeromonas_phage_31
4_gene_7643_NODE_12132_length_2143_cov_14_967802
6_gene_14738_NODE_19154_length_1863_cov_13_204509
1_gene_1242_NODE_1058_length_1597_cov_20_903894
6_gene_66879_NODE_99352_length_1811_cov_11_104915
1_gene_2031_NODE_1866_length_1190_cov_14_136974
1_gene_3727_NODE_3901_length_2440_cov_16_452049
4_gene_42775_NODE_69090_length_1261_cov_9_282315
Pseudomonas_phage_F116
6_gene_83332_NODE_129521_length_1225_cov_5_395102
4_gene_72897_NODE_128808_length_1682_cov_12_348394
8_gene_4254_NODE_3459_length_1462_cov_7_781806
DNA_maturase_beta_subunit_Synechococcus_phage_S_RIP2
DNA_maturase_beta_subunit_Synechococcus_phage_S_RIP1
terminase_Synechococcus_phage_S_CBP3
terminase_Synechococcus_phage_S_CBP1
terminase_Synechococcus_phage_P60
1_gene_8071_NODE_17088_length_1343_cov_16_790022
1_gene_1357_NODE_1165_length_2147_cov_17_290173
4_gene_1463_NODE_2290_length_1186_cov_13_807757
8_gene_5358_NODE_4366_length_1558_cov_7_295250
hypothetical_protein_SPBc2p056_Bacillus_phage_SPBc2
terminase_Lactococcus_phage_c2
1_gene_2263_NODE_2134_length_1543_cov_24_560596
4_gene_26720_NODE_42420_length_1587_cov_6_115942
4_gene_22571_NODE_35301_length_1193_cov_7_179380
4_gene_62_NODE_117_length_1780_cov_6_244382
Cellulophaga_phage_phi39_1
terminase_subunit_Synechococcus_phage_S_CBS1
terminase_subunit_Synechococcus_phage_S_CBS3
Cellulophaga_phage_phi3_1
6_gene_28832_NODE_39618_length_2158_cov_11_050046
4_gene_25300_NODE_39938_length_2835_cov_12_606349
6_gene_120597_NODE_218620_length_1357_cov_10_790715
4_gene_3929_NODE_6132_length_1739_cov_12_415756
4_gene_81182_NODE_149277_length_1290_cov_7_902326
6_gene_2999_NODE_3814_length_3382_cov_7_514489
Persicivirga_phage_P12024L
Cellulophaga_phage_phiST
ORF009_Staphylococcus_phage_X2
Roseobacter_phage_RDJL_Phi_1
terminase_Synechococcus_phage_S_CBP2
Pseudomonas_phage_PaP3
terminase_subunit_Pseudomonas_phage_B3
Salmonella_phage_SETP13
terminase_Synechococcus_phage_S_CBS4
4_gene_9968_NODE_15678_length_2190_cov_3_859361
8_gene_36914_NODE_35931_length_2896_cov_6_265884
4_gene_30936_NODE_49302_length_1946_cov_5_328880
8_gene_24793_NODE_23001_length_3012_cov_7_081009
8_gene_4218_NODE_3421_length_2655_cov_8_715631
4_gene_28866_NODE_45913_length_2028_cov_14_646943
4_gene_110848_NODE_248183_length_2572_cov_9_128304
6_gene_29639_NODE_40919_length_1410_cov_6_608510
6_gene_47073_NODE_67051_length_2169_cov_7_422315
terminase_subunit_Synechococcus_phage_S_CBS2
Burkholderia_phage_Bcep781
8_gene_11087_NODE_9346_length_3201_cov_5_845048
8_gene_14751_NODE_12790_length_1127_cov_5_523514
Lactococcus_phage_ul36_k1
8_gene_2194_NODE_1762_length_2205_cov_7_492517
1_gene_908_NODE_723_length_1485_cov_16_017508
Enterobacteria_phage_T1
6_gene_12353_NODE_16017_length_2782_cov_7_100647
terminase_subunit_Enterobacteria_phage_T86
terminase_ATPase_subunit_Vibrio_phage_K139
terminase_ATPase_subunit_Burkholderia_phage_phiE202
Roseobacter_phage_SIO1
terminase_Streptococcus_phage_Str_PAP_1
terminase_Mycobacterium_sp__KMS
Phage_phiJL001
0.5
Cyanomyoviruses
T4-like viruses
Cyanopodoviruses
Siphoviruses
T1-like viruses

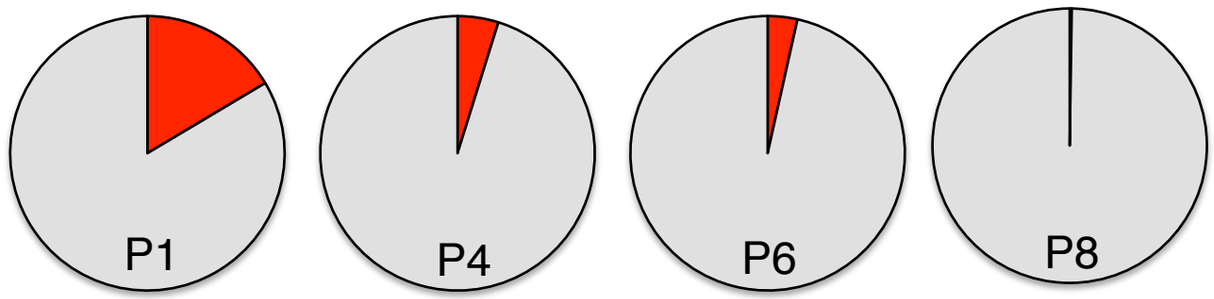
B)

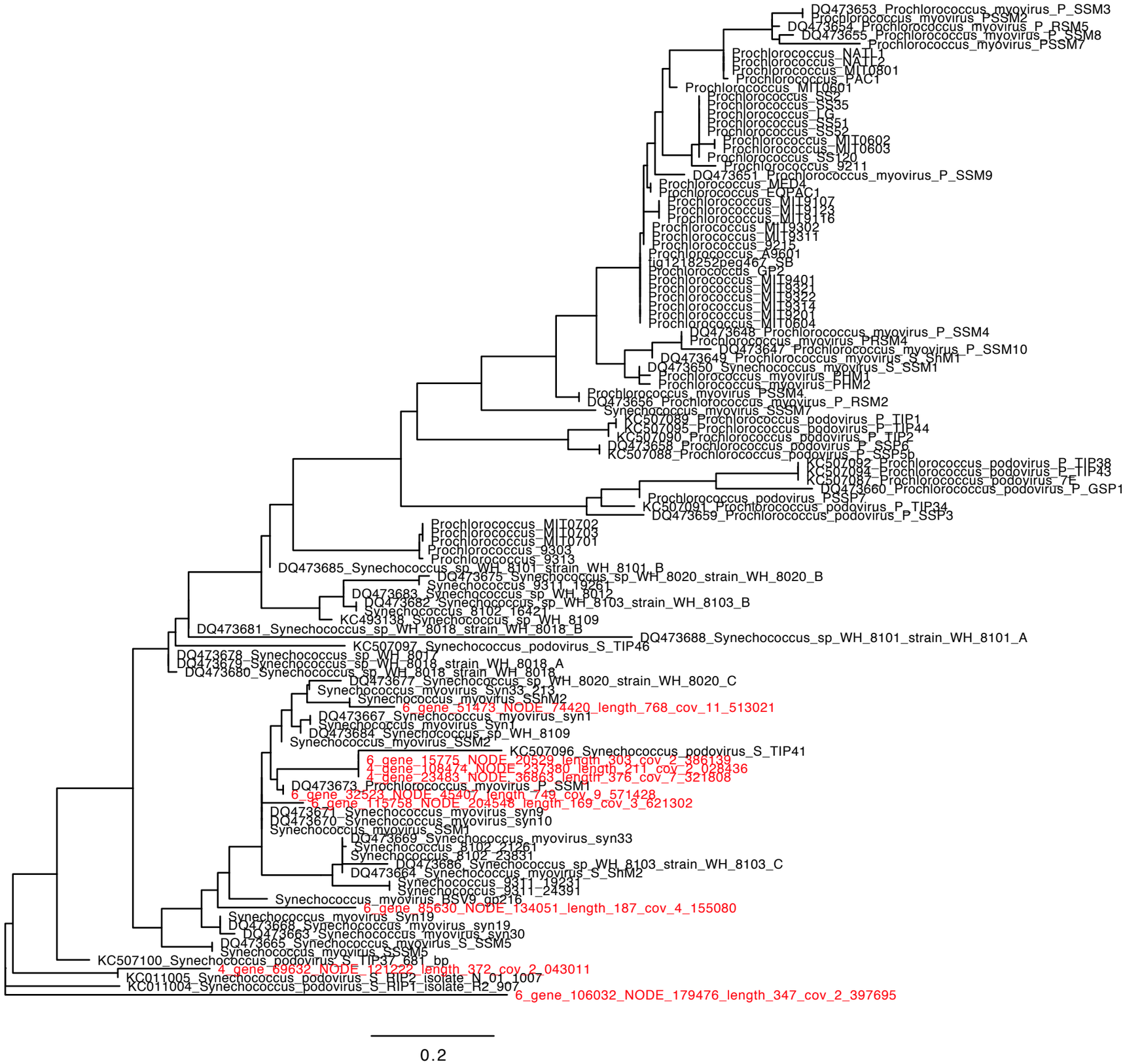

Prochlorococcus_phage_Syn33
Synechococcus_phage_S_PIM2_R21_2007
Synechococcus_phage_syn9
Prochlorococcus_phage_P_RSM4
Prochlorococcus_phage_P_SSM4
CDS272
Cyanophage_TIM40
Synechococcus_phage_ACG_2014e
CDS78
CDS249
CDS536
CDS484
CDS406
CDS525
CDS306
Proteobacteria_bacterium_JGI_0000113_P07
CDS559
CDS314
Pelagibacter_phage_HTVC008M
Sinorhizobium_phage_phiM12
CDS582
CDS573
CDS518
CDS486
CDS475
CDS124
CDS236
CDS418
CDS536
CDS286
Aeromonas_phage_Aeh1
Escherichia_phage_vB_EcoM_JS09
Caulobacter_phage_Cr30
CDS204
CDS415
CDS424
CDS483
CDS597
CDS417
CDS253
CDS266
Acidianus_bottleshaped_virus
CDS105
CDS301
Pyrococcus_horikoshii_OT3
His_1_virus
His2_virus
Bacillus_phage_GIL16c
plasmid_Bacillus_cereus_ATCC
Enterobacteria_phage_PR5
CDS589
CDS256
CDS218
CDS439
Micromonas_pusilla_virus_SP1
Ostreococcus_tauri_virus_2
Ostreococcus_lucimarinus_virus_1
CDS450
CDS509
Bathycoccus_sp._RCC1105_virus_BpV1
CDS370
Paramecium_bursaria_Chlorella_virus_CVK2
Paramecium_bursaria_Chlorella_virus_NY_2B
Emiliania_huxleyi_virus_201
Pandoravirus_salinus
Heterosigma_akashiwo_virus_01
Mimivirus_Bus
Megavirus_chiliensis
Acanthamoeba_polyphaga_mimivirus
Cafeteria_roenbergensis_virus_BV_PW1
Phaeocystis_globosa_virus
Feldmannia_species_virus
Feldmannia_irregularis_virus_a
EsV_1_93Ectocarpus_siliculosus
Red_seabream_iridovirus
Invertebrate_iridescent_virus_30
Spodoptera_ascovirus
Green_turtle_herpesvirus
Human_herpesvirus_1
CDS299
Yoka_poxvirus
Enterobacteria_phage_CAjan2
Enterobacteria_phage_CAjan1
Enterobacteria_phage_9g
Pseudomonas_aeruginosa_BWHPSA026
Pseudomonas_aeruginosa2
Burkholderia_phage_AH2
Burkholderia_phage_KL1
Oceanicola_sp._HL35
Sinorhizobium_meliloti
CDS571
Salmonella_phage_FSL_SP062
PmuNAV
Agrobacterium_phage_771
Rhizobium_phage_vB
Vibrio_phage_CP11
Vibrio_phage_vB_VchM138
Salmonella_phage_9NA
Nitrincola_phage_1M316
Staphylococcus_epidermidis1
Staphylococcus_epidermidis_VCU129
Bacillales
Clostridium_cadaveris
Clostridium_botulinumgi
Lachnospiraceae_bacterium
Staphylococcus_aureus
CDS435
CDS150
CDS122
Caenorhabditis_briggsae
Tribolium_castaneumgi
Nematostella_vectensis
Gryptapanteles_flavicoxis
Nasonia_vitripennis
Bombyx_mori_densovirus_3
gi_13431472_sp_Q590241_DP2L_METJA
CDS319
CDS75
CDS170
Bat_adenovirus
gi_59750103 6_sp_Q8QNT02
Turkey_siadenovirus_A
Fowl_aviadenovirus_D
Plasmid_PGKL2_proteinA
Schwanniomyces_etchellsii
gi_1352308
Debaryomyces_hansenii
0.8
Bacteriophage
Archeoviruses
Nucleocytoplasmic large DNA viruses
DNA-primed viruses
Polintons


C)
Diatom viruses
Inovirus-like
Uncultured marine group 1
Parvovirus-like
Uncultured marine group 2
Nanoviruses
Micro- and parvovirus-like
Geminiviruses
Microphage-like
2.0

D)


472342626_ref_YP_007676133_1_hypothetical_protein_MPVG_00065_Micromonas_pusilla_virus_12T
314055249_ref_YP_004063587_1_major_capsid_protein_Ostreococcus_tauri_virus_2
313768156_ref_YP_004061587_1_hypothetical_protein_BpV1_157_Bathycoccus_sp_RCC1105_virus_BpV1
668005335_BS
197370454_FWS
6_gene_116310_NODE_206022_length_884_cov_6_742082
AUXO014609059_Gut
JRYD01176473_AS_2
448925034_gb_AGE48615_1_major_capsid_protein_VP54_Paramecium_bursaria_Chlorella_virus_AN69C
8_gene_41063_NODE_40627_length_720_cov_7_565278
197392686_FWS
668071392_BS
668086135_BS
Tlr_4Fp
70720887_dbj_BAE06835_1_hypothetical_major_capsid_protein_Heterosigma_akashiwo_virus_01
508181956_ref_YP_008052475_1_major_capsid_protein_MCP_Phaeocystis_globosa_virus
939177538_ref_YP_009173446_1_major_capsid_protein_1_Chrysochromulina_ericina_virus
672551186_ref_YP_009052173_1_putative_capsid_protein_Aureococcus_anophagefferens_virus
451927482_gb_AGF85360_1_capsid_protein_Moumouvirus_goulette
602259399_ref_YP_009021087_1_capsid_protein_Sputnik_virophage_3
195982551_ref_YP_002122381_1_V20_Sputnik_virophage
563399746_ref_YP_008859635_1_Capsid_protein_Zamilon_virophage
948298278_ref_YP_009177804_1_putative_major_capsid_protein_Yellowstone_Lake_virophage_5
948298294_ref_YP_009177819_1_putative_major_capsid_protein_Yellowstone_Lake_virophage_6
948298154_ref_YP_009177692_1_putative_major_capsid_protein_Yellowstone_Lake_virophage_7
8_gene_61651_NODE_67208_length_615_cov_3_780488
JRYD01125981_AS
JRYI01150569_AS
326439166_ref_YP_004300296_1_putative_major_capsid_protein_Mavirus
AUXO013000407_Gut_2
AUXO014016815_Gut
AUXO010009513_Gut
AUXO011791364_Gut
AUXO013094825_Gut
JRYI01079693_AS
AUXO011749112_Gut
AUXO017358990_Gut
310831332_ref_YP_003969975_1_major_capsid_protein_Cafeteria_roenbergensis_virus_BV_PW1
6_gene_42877_NODE_60861_length_428_cov_3_967290
AUXO010458836_Gut
AUXO015729686_Gut
AUXO014394586_Gut
AUXO018323740_Gut
AUXO018775940_Gut
AUXO015870414_Gut
AUXO018464265_Gut
Polinton_N1A_NV
JRYC01097131_AS
4_gene_73087_NODE_129306_length_568_cov_2_975352
311977828_ref_YP_003986948_1_probable_capsid_protein_4_Acanthamoeba_polyphaga_mimivirus
4_gene_9236_NODE_14587_length_434_cov_7_248848
AUXO013000407_Gut
AUXO011122541_Gut
AUXO013317242_Gut
AUXO015956656_Gut
AUXO017292295_Gut
AUXO010976392_Gut
683041308_WW
AUXO015081832_Gut
AUXO011283558_Gut
AUXO010727758_Gut
AUXO012281048_Gut
AUXO012618078_Gut
AUXO011663271_Gut
AUXO014837302_Gut
AUXO015700412_Gut
AUXO013175281_Gut
AUXO015145606_Gut
AUXO013562526_Gut
AUXO013458317_Gut
AUXO012775506_Gut
AUXO016815984_Gut
AUXO012978425_Gut
AUXO013654610_Gut
509141021_ref_YP_008059897_1_hypothetical_protein_PGVV_00012_Phaeocystis_globosa_virus_virophage
JRYD01176473_AS
AUXO017439779_Gut
322510455_gb_ADX05770_1_major_capsid_protein_Organic_Lake_virophage
Polinton_1_AMi
Polinton_1_CPB
Polinton_2_XT
Polinton_1_SP
Polinton_2A_NV
Polinton_2_NV
2.0


P1
P4
P6
P8


DQ473653_Prochlorococcus_myovirus_P_SSM3
Prochlorococcus_myovirus_PSSM2
DQ473654_Prochlorococcus_myovirus_P_RSM5
DQ473655_Prochlorococcus_myovirus_P_SSM8
Prochlorococcus_myovirus_PSSM7
Prochlorococcus_NATL1
Prochlorococcus_NATL2
Prochlorococcus_MIT0801
Prochlorococcus_PAC1
Prochlorococcus_MIT0601
Prochlorococcus_SS2
Prochlorococcus_SS35
Prochlorococcus_LG
Prochlorococcus_SS51
Prochlorococcus_SS52
Prochlorococcus_MIT0602
Prochlorococcus_MIT0603
Prochlorococcus_SS120
Prochlorococcus_9211
DQ473651_Prochlorococcus_myovirus_P_SSM9
Prochlorococcus_MED4
Prochlorococcus_EQPAC1
Prochlorococcus_MIT9107
Prochlorococcus_MIT9123
Prochlorococcus_MIT9116
Prochlorococcus_MIT9302
Prochlorococcus_MIT9311
Prochlorococcus_9215
Prochlorococcus_A9601
tig1218252peg467_SB
Prochlorococcus_GP2
Prochlorococcus_MIT9401
Prochlorococcus_MIT9321
Prochlorococcus_MIT9312
Prochlorococcus_MIT9314
Prochlorococcus_MIT9201
Prochlorococcus_MIT0604
DQ473648_Prochlorococcus_myovirus_P_SSM4
Prochlorococcus_myovirus_PRSM4
DQ473647_Prochlorococcus_myovirus_P_SSM10
DQ473649_Prochlorococcus_myovirus_S_ShM1
DQ473650_Synechococcus_myovirus_S_SSM1
Prochlorococcus_myovirus_PHM1
Prochlorococcus_myovirus_PHM2
Prochlorococcus_myovirus_PSSM4
DQ473656_Prochlorococcus_myovirus_P_RSM2
Synechococcus_myovirus_SSSM7
KC507089_Prochlorococcus_podovirus_P_TIP1
KC507095_Prochlorococcus_podovirus_P_TIP44
KC507090_Prochlorococcus_podovirus_P_TIP2
DQ473658_Prochlorococcus_podovirus_P_SSP6
KC507088_Prochlorococcus_podovirus_P_SSP5b
KC507092_Prochlorococcus_podovirus_P_TIP38
KC507094_Prochlorococcus_podovirus_P_TIP43
KC507087_Prochlorococcus_podovirus_7E
DQ473660_Prochlorococcus_podovirus_P_GSP1
Prochlorococcus_podovirus_PSSP7
KC507091_Prochlorococcus_podovirus_P_TIP34
DQ473659_Prochlorococcus_podovirus_P_SSP3
Prochlorococcus_MIT0702
Prochlorococcus_MIT0703
Prochlorococcus_MIT0701
Prochlorococcus_9303
Prochlorococcus_9313
DQ473685_Synechococcus_sp_WH_8101_strain_WH_8101_B
DQ473675_Synechococcus_sp_WH_8020_strain_WH_8020_B
Synechococcus_9311_19261
DQ473683_Synechococcus_sp_WH_8012
DQ473682_Synechococcus_sp_WH_8103_strain_WH_8103_B
Synechococcus_8102_16421
KC493138_Synechococcus_sp_WH_8109
DQ473681_Synechococcus_sp_WH_8018_strain_WH_8018_B
DQ473688_Synechococcus_sp_WH_8101_strain_WH_8101_A
KC507097_Synechococcus_podovirus_S_TIP46
DQ473678_Synechococcus_sp_WH_8017
DQ473679_Synechococcus_sp_WH_8018_strain_WH_8018_A
DQ473680_Synechococcus_sp_WH_8018_strain_WH_8018
DQ473677_Synechococcus_sp_WH_8020_strain_WH_8020_C
Synechococcus_myovirus_Syn33_213
Synechococcus_myovirus_SSM2
6_gene_51473_NODE_74420_length_768_cov_11_513021
DQ473667_Synechococcus_myovirus_syn1
Synechococcus_myovirus_Syn1
DQ473684_Synechococcus_sp_WH_8109
Synechococcus_myovirus_SSM2
KC507096_Synechococcus_podovirus_S_TIP41
6_gene_15775_NODE_20529_length_303_cov_2_386139
4_gene_108474_NODE_237380_length_211_cov_2_028436
4_gene_23483_NODE_36863_length_376_cov_7_321808
DQ473673_Prochlorococcus_myovirus_P_SSM1
6_gene_32523_NODE_45407_length_749_cov_9_571428
6_gene_115758_NODE_204548_length_169_cov_3_621302
DQ473671_Synechococcus_myovirus_syn9
DQ473670_Synechococcus_myovirus_syn10
Synechococcus_myovirus_SSM1
DQ473669_Synechococcus_myovirus_syn33
Synechococcus_8102_21261
Synechococcus_8102_23831
DQ473686_Synechococcus_sp_WH_8103_strain_WH_8103_C
DQ473664_Synechococcus_myovirus_S_ShM2
Synechococcus_9311_19231
Synechococcus_9311_24391
Synechococcus_myovirus_BSV9_gp216
6_gene_85630_NODE_134051_length_187_cov_4_155080
Synechococcus_myovirus_Syn19
DQ473668_Synechococcus_myovirus_syn19
DQ473663_Synechococcus_myovirus_syn30
DQ473665_Synechococcus_myovirus_S_SSM5
Synechococcus_myovirus_SSSM5
KC507100_Synechococcus_podovirus_S_TIP37_681_bp
4_gene_69632_NODE_121222_length_372_cov_2_043011
KC011005_Synechococcus_podovirus_S_RIP2_isolate_N_01_1007
KC011004_Synechococcus_podovirus_S_RIP1_isolate_R2_907
6_gene_106032_NODE_179476_length_347_cov_2_397695
0.2

Conclusion

Viral ecology has blossomed in the genomic era, which has an enabled unprecendented depth and breadth of insight into viral dynamics. By far, the overwhelming headline has been how incredibly diverse viruses are (Brum et al. 2015). In every environment (Angly et al. 2006, Aguirre de Carcer et al. 2014, Yoshida et al. 2013, Cassman et al. 2012) and for every type of virus (Hopkins et al. 2014, Labonte et al. 2013, Culley et al 2014, Schmidt et al. 2014), previously unknown, uncultured viral groups are being discovered. Certainly, this theme of novel diversity is highlighted in this thesis. In chapter one, the *Pseudo-nitzschia multiseries* DNA virus (PmDNAV) expanded the current boundaries of single stranded DNA viruses (ssDNA) having the largest ssDNA genome and a novel polymerase not typically seen in ssDNA viruses. Chapter 2 reveals even diatom hosts have surprising microdiversity in their susceptibility to infection to the PmDNAV and whole viral communities. Finally, the viral communities on the GeoMICS transect in chapter 3 also followed the theme of diversity with ~70% of sequences not similar to previously sequenced viruses. Summarily, this diversity can be overwhelming and lends itself to the conclusion that there will never be enough viruses isolated or viral communities sequenced to capture the diversity of viruses in the ocean.

Despite this diversity, viruses are united functionally through infection. Focusing on this process, which requires understanding infection from both the host's and the virus's perspective begins to simplify this diversity and reveal emergent patterns in viral ecology. However, understanding infection at the microbial is not trival because microbes do not have easily identifiable symptoms. Working with model systems empowers our ability to understand environmental dynamics. This thesis attempted to create this bridge, focusing on diatoms as model system and exposed three overarching themes.

First, viruses are not just specialists for one host (Brussaard et al. 2004), but instead viruses lie on a spectrum of specialists to generalists and that spectrum influences their ecology. The isolation of the broad host range PmDNAV shattered the paradigm that diatom viruses, which had mostly only been isolated on one genus of diatom, *Chaetoceros*, are only specialists. Furthermore, diatoms of the genus *Pseudo-nitzschia* clearly ranged in their permissivity to viral infection. This permissivity may even be a significant factor influencing infection dynamics. This nested pattern of host-virus interactions affects the bloom and crash cycles of diatom populations. There is mounting evidence that diatom blooms, even monospecific ones, are composed of a diversity of hosts (Rynearson et al., 2005, Alexander et al. 2015). Varying susceptibility to viruses within a diatom bloom may be a defense against viral termination at a time when viruses would thrive. Viruses, including the PmDNAV, try to take advantage of these situations. They replicate quickly and have huge burst sizes. During times of host scarcity, the host dependent differential infection dynamics observed in the PmDNAV suggests that viruses may cobble together an existence with different infection dynamics until the next bloom arrives. Ultimately, perhaps there is trade-off between viruses with different types of host ranges, with specialists possibly dominating times of high productivity, but generalists being steadier in their abundance throughout the seasons.

Second, key viral genes may be as important for viral success as their host range. Host ranges and host permissivity are determined by specialized, hypervariable regions in both host and virus genomes (Weitz et al. 2012, Avrani et al. 2011). However, other genes, like the auxiliary metabolic genes in cyanophage or varying types of polymerases may be as essential for viral dynamics. In cyanophage, these genes are likely responsible for altering the infection process, changing latent periods and burst sizes (Lindell et al. 2005, Lindell et al., 2007). The

difference in mutation rates between rolling circle replication proteins and DNA polymerases such as that which was found in the PmDNAV, may influence genome size and generally affects the evolutionary trajectory of a virus (Duffy et al. 2008). Indeed, these non-host derived auxiliary genes that affect infection dynamics such polymerase processivity or error rates (Tabor et al., 2987, Banach-Orlowska et al. 2005) may be as important as the unique virally encoded host genes in defining a virus' niche.

Finally, viral population diversity and host community structure are linked through 'kill-the-winner' dynamics. The patterns of occurrence and resistance observed in chapter two hinted at the existence of this diversity sustaining and population structuring mechanism. Only when directly observing infection in chapter three could these dynamics be more clearly seen. Despite the strain level variation in hosts and viruses, viruses may still affect population wide community structure and abundance. Furthermore, instead of microbial populations changing in response to bottom up effects, chapter three suggested viruses may work in tandem with environmental gradients to shape microbial abundance.

In conclusion, viruses are more than just predators of marine microbes. They are diverse and abundant architects of microbial communities, affecting all life, at all scales. Viruses have adopted a variety of lifestyles to be successful in a diffusive patchy ocean. Ultimately, viral infection is an essential cog that connects the gears of marine microbial ecology and global biogeochemical cycles.

References

Aguirre de Cárcer, D., F. E. Angly, and A. Alcamí. 2014. Evaluation of viral genome assembly and diversity estimation in deep metagenomes. BMC Genomics **15**: 989.

Alexander, H., B. D. Jenkins, T. a. Rynearson, and S. T. Dyhrman. 2015. Metatranscriptome analyses indicate resource partitioning between diatoms in the field. Proc. Natl. Acad. Sci. 201421993.

Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. a Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. a Suttle, and F. Rohwer. 2006. The marine viromes of four oceanic regions. PLoS Biol. **4**: e368.

Avrani, S., O. Wurtzel, I. Sharon, R. Sorek, and D. Lindell. 2011. Genomic island variability facilitates Prochlorococcus-virus coexistence. Nature **474**: 604–8.

Banach-Orlowska, M., I. J. Fijalkowska, R. M. Schaaper, and P. Jonczyk. 2005. DNA polymerase II as a fidelity factor in chromosomal DNA synthesis in Escherichia coli. Mol. Microbiol. **58**: 61–70.

Brum, J. R., J. C. Ignacio-espinoza, S. Roux, et al. 2015. Patterns and ecological drivers of ocean viral communities. Science 348 doi:10.1126/science.1261498.

Brussaard, C. P. D. 2004. The Journal of Eukaryotic Microbiology. **51**: 125–138.

Cassman, N., A. Prieto-Davó, K. Walsh, G. G. Z. Silva, F. Angly, S. Akhter, K. Barott, J. Busch, T. McDole, J. M. Haggerty, D. Willner, G. Alarcón, O. Ulloa, E. F. DeLong, B. E. Dutilh, F. Rohwer, and E. a Dinsdale. 2012. Oxygen minimum zones harbour novel viral communities with low diversity. Environ. Microbiol. **14**: 3043–65.

Culley, A. I., J. A. Mueller, and M. Belcaid. 2014. The Characterization of RNA Viruses in Tropical Seawater Using Targeted PCR and Metagenomics. , doi:10.1128/mBio.01210-14.Editor

Duffy, S., L. a Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. **9**: 267–76.

Hopkins, M., S. Kailasan, A. Cohen, S. Roux, K. P. Tucker, A. Shevenell, M. Agbandje-McKenna, and M. Breitbart. 2014. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. ISME J. 1–11.

Labonté, J. M., and C. a Suttle. 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. ISME J. 1–9.

Lindell, D., J. D. Jaffe, M. L. Coleman, M. E. Futschik, I. M. Axmann, T. Rector, G. Kettler, M. B. Sullivan, R. Steen, W. R. Hess, G. M. Church, and S. W. Chisholm. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature **449**: 83–6.

Lindell, D., J. D. Jaffe, Z. I. Johnson, G. M. Church, and S. W. Chisholm. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. Nature **438**: 86–9.

Rynearson, T. a, and E. V. Armbrust. 2005. Maintenance of clonal diversity during a spring bloom of the centric diatom Ditylum brightwellii. Mol. Ecol. **14**: 1631–40.

Schmidt, H. F., E. G. Sakowski, S. J. Williamson, S. W. Polson, and K. E. Wommack. 2014. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. ISME J. **8**: 103–14.

Weitz, J. S., T. Poisot, J. R. Meyer, C. O. Flores, S. Valverde, M. B. Sullivan, and M. E. Hochberg. 2013. Phage-bacteria infection networks. Trends Microbiol. **21**: 82–91.

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

Appendix I

Infection dynamics of the *Pseudo-nitzschia multiseries* DNA virus

A one step growth curve was performed on the original isolating host, *P. multiseries* CLnn-16 (figure 1), using the methods outlined in chapter 1. The cultures used in this growth curve were not axenic, unlike the *P. multiseries* PC9 and *T. pseudonana* CCMP1335 cultures described in chapter 1. *P. multiseries* CLnn-16 cultures were grown at 0.25 day^{-1} under the same conditions as in chapter 1. Significant changes in the CLnn-16 host cells were noticeable as early as 4 hours post inoculation when declines in host abundance was apparent (figure 1a). The latent period was 12-16 hours post infection, and the burst size was 9,300 viruses $cell^{-1}$ (figure 1b). Bacteria doubled in abundance over the course of the one step growth curve (figure 1c).

Photosynthetic changes were observed in host cells in all three one step growth curves (performed on *P. multiseries* CLnn-16, *P. multiseries* PC9, and *T. pseudonana* CCMP 1335). Chlorophyll-a fluorescence was measured using a Turner AU-10 fluorometer at every time point in all three experiments. Photosynthetic yield of photosystem II (Fv/Fm) was measured with a PhytoPAM fluorometer (Waltz) in the *T. pseudonana* experiment. Declines in the chlorophyll-a concentrations began at 4, 32, and 48 hours post infection in *P. multiseries* CLnn-16, *P. multiseries* PC9, and *T. pseudonana* CCMP 1335 (figure 2). This coincided with declines observed in host cell number. Chlorophyll-*a* fluorescence never reached a value of 0 suggesting that chlorophyll remained in the medium to be detected after the cells had died. Photosynthetic efficiency also declined in concert with chlorophyll-a fluorescence at hour 48 in the *T. pseudonana* one step growth curve (figure 2c). Changes in host cells during infection, primarily shrunken chloroplasts and a change in color from brown to green, were noticeable via light

microscopy (figure 3a). Although cells were lysed which was apparent in the lack of a nucleus compared to control cultures (figure 3b and c), only a few cells were noted to be broken.

Finally, the infectivity of the PmDNAV was assessed under different conditions. The PmDNAV stocks were heated or chilled for 15 minutes at -80, -20, 4, 29, 50, and 100 °C. The viral solutions were allowed to return to room temperature and then inoculated into cultures. For UV treatment, virus aliquots were exposed to a UV lamp (American Ultraviolet Company) with 800μm/cm^2 at 30.5cm distance for 1, 10, and 60 minutes. In each treatment, cultures of *P. multiseries* CLnn-16 were counted, inoculated with treated PmDNAV stocks to equal 13 viruses per cell, and monitored via chlorophyll-a fluorescence. The virus was inactivated by heating to >50°C, while it retained infectivity after 15 minutes of exposure to temperatures from -80°C to 29°C (table 1). Irradiation with UV light inhibited infection of 4 of 5 replicates after 10 minutes, and infection was completely inhibited after 60 minutes of irradiation. No inhibition of infection was seen after 1 minute of UV treatment. Finally, cultures of CLnn-16 showed no difference in susceptibility when inoculated when growing exponentially or 3 days after the onset of stationary phase.

Figure legends

Figure A1. One step growth curve of *P.multiseries* CLnn-16 infected with the *Pseudo-nitzschia multiseries* DNA virus. Solid lines represent control cultures while dashed line represent virus infected cultures. Error bars represent standard error of the mean (n=3).

Figure A2. Photosynthetic changes over the PmDNAV infection cycle for hosts *P. multiseries* CLnn-16 (a), axenic *P. multiseries* PC9 (b), and axenic *T. pseudonana* CCMP1335 (c). Squares

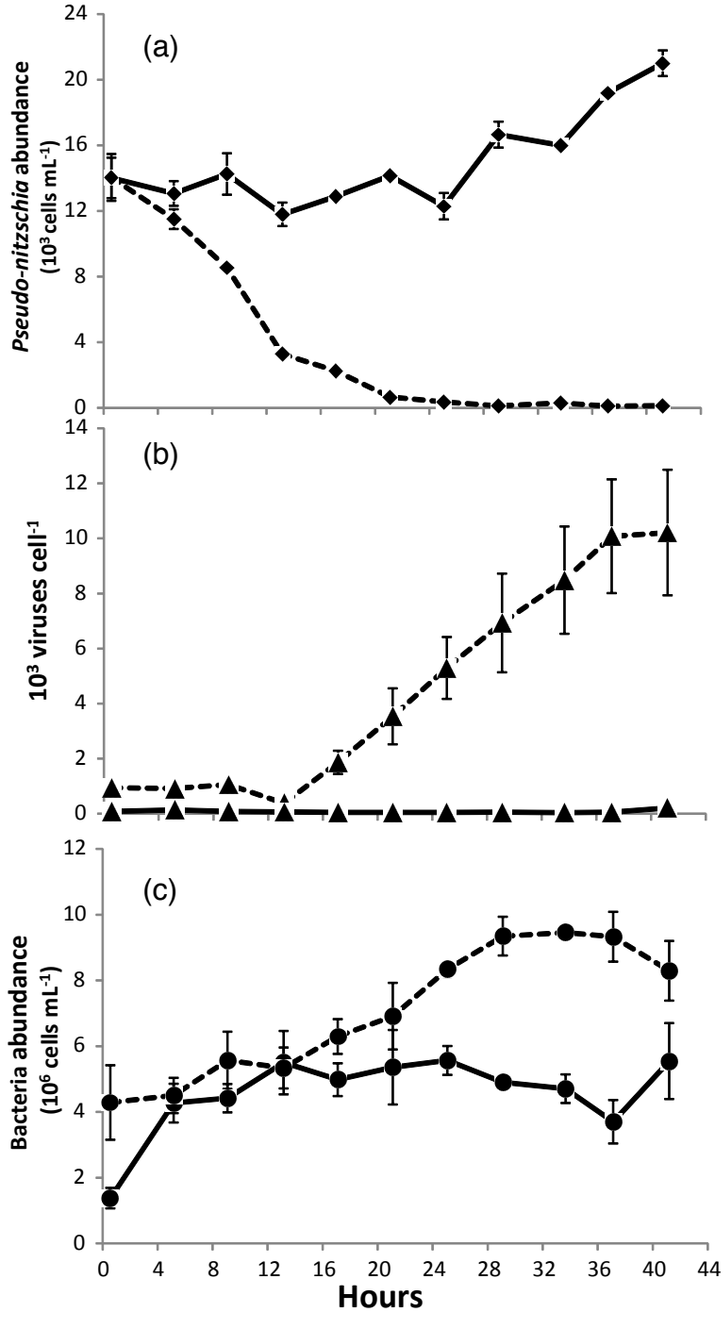
indicate relative fluorescence values; circles Fv/Fm. Dashed line indicates infected cultures; solid lines indicate controls. Error bars are standard deviation (n=3).

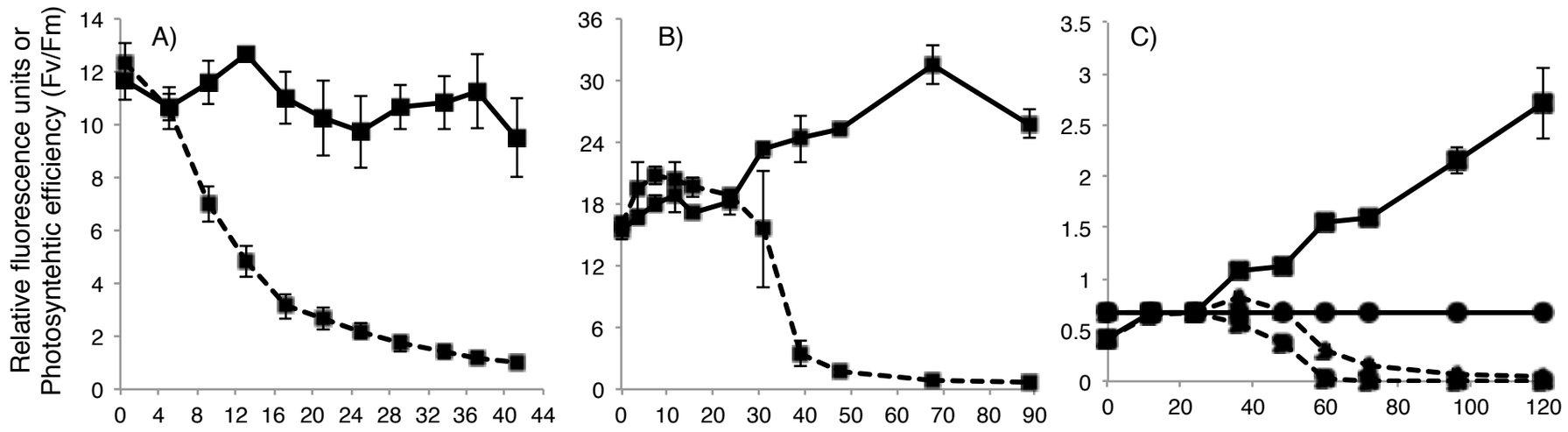
Figure A3. Visual changes in cell characteristics during infection. a) *P. multiseries* CLnn-16 cultures that were treated and lysed with the PmDNAV 24 hours post infection. On the left, a control culture showing healthy brown pigmentation. On the right, an infected culture showing the transition to a green hue. Epifluorescent images of uninfected (b) and infected and lysed (c) *P. multiseries* CLnn-16 cells stained with SYBR Gold viewed at 1000x magnification. Healthy cells show clear nuclei. Lysed cells have no nuclei, but remain intact. Virus like particles can be seen inside the cell.

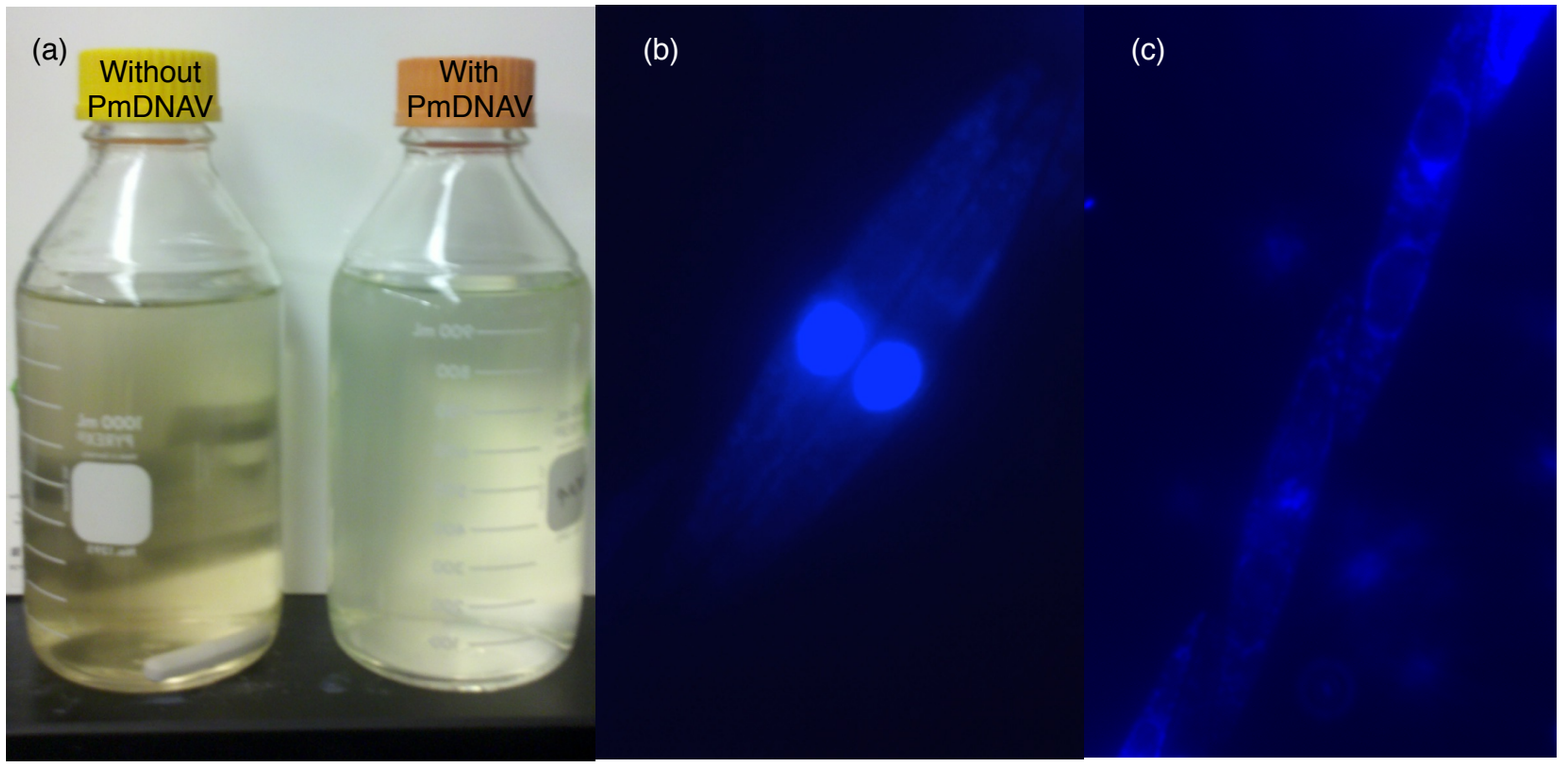
Tables

Table A1. Effects of temperature, UV, and growth phase on the infectivity of the PmDNAV.

Treatment	**Infective**
-80 °C	yes
-20 °C	yes
4 °C	yes
29 °C	yes
50 °C	no
100 °C	no
1 min UV irradiation	yes
10 min UV irradiation	partial
60 min UV irradiation	no
Exponential phase	yes
Stationary phase	yes


(a)
Pseudo-nitzschia abundance
(10³ cells mL⁻¹)
24
20
16
12
8
4
0
(b)
10³ viruses cell⁻¹
14
12
10
8
6
4
2
0
(c)
Bacteria abundance
(10⁶ cells mL⁻¹)
12
10
8
6
4
2
0
0
4
8
12
16
20
24
28
32
36
40
44
Hours


Relative fluorescence units or
Photosyntehtic efficiency (Fv/Fm)
A)
B)
C)
0
2
4
6
8
10
12
14
0
4
8
12
16
20
24
28
32
36
40
44
0
6
12
18
24
30
36
0
10
20
30
40
50
60
70
80
90
0
0.5
1
1.5
2
2.5
3
3.5
0
20
40
60
80
100
120


(a)
Without PmDNAV
With PmDNAV
(b)
(c)

Curriculum Vitae

Michael Curtis Grier Carlson
University of Washington, School of Oceanography
Box 357940, Seattle, WA 98195-7940
Office: Benjamin Hall 323 Phone : (626) 825-2929

EDUCATION

2015 Ph.D., Biological Oceanography
University of Washington, School of Oceanography
Title: The lifestyles of viruses in the sunlit ocean revealed through isolation, genomics, and infection dynamics
Advisor: Gabrielle Rocap

2011 M.Sc., Biological Oceanography
University of Washington, School of Oceanography
Title: Isolation and characterization of a virus infecting the toxigenic diatom *Pseudo-nitzschia multiseries*
Advisor: Gabrielle Rocap

2008 B.A., Biology and History
Pomona College

RESEARCH EXPERIENCE

Present Postdoctoral scholar
Technion University
Adviser: Debbie Lindell

2008-2015 Research Assistant
University of Washington, School of Oceanography
Adviser: Gabrielle Rocap

2007-2008 Research Volunteer
San Diego State University
Prevalence of coral- and seawater-associated adenoviruses in impacted coral reefs
Supervisor: Rebecca Vega Thurber

2007 Edwin W. Pauley Summer Research Program Participant
Hawaii Institute of Marine Biology, University of Hawaii
Impact of viruses on coral health
Mentors: Ruth Gates and Rebecca Vega Thurber

AWARDS

2015	Early Career Award- Life of Diatoms Conference
2013-2014	University of Washington School of Oceanography Research Innovation Grant
2012-2015	Environmental Protection Agency STAR Fellowship
2010-2011	Pacific Northwest Center for Human Health and Ocean Studies Traineeship
2008-2010	Pacific Northwest Center for Human Health and Ocean Studies Traineeship
2007	Edwin W. Pauley Summer Research Program fellow
2007	Hart Institute for American History Research Grant

TEACHING EXPERIENCE

Teaching Assistant:

Spring 2012	Oceanography 444: Advanced Field Oceanography
Winter 2012	Oceanography 443: Design of Oceanographic Field Experiments
Spring 2011	Oceanography 201: Introduction to Oceanography Lab
Spring 2010	Oceanography 200: Introduction to Oceanography
2010-2014	Guest Lecturer for Oceanography 535 and Oceanography 497

Student Mentoring:

2014-2015	Andrea Odell, Profiling single stranded DNA viruses in the North Pacific
2014-2015	David Garcia-Prieto, Metagenomic quantification of cyanophage in oxygen minimum zones
2012-2015	Terence Leach, Resolving the phylogeny of the *Pseudo-nitzschia* genus
2011-2014	Nicolette McCary, Seasonality and host range *Pseudo-nitzschia* and their viruses
2012	Matthew Knight, Biogeography of pathogenic bacteria in the coastal waters of southern California and Mexico
2012	Annie Thomson, Distribution of ammonia-oxidizing bacteria in the eastern tropical North Pacific
2012	Sara Geurtsen, Distribution of denitrifying bacteria in the eastern tropical North Pacific
2009-2011	Kyle Frischkorn, Isolation and characterization of *Pseudo-nitzschia* viruses

PRESENTATIONS

Carlson, M., McCary, N., Frischkorn, K., Leach, T., Rocap, G. Viral control of *Pseudo-nitzschia* communities is driven by host specific interactions. Oral presentation at the Life of Diatoms Conference. July 2015. Seattle, WA.

Carlson, M., McCary, N., Rocap, G. Viral host range and permissivity of the diatom Pseudo-nitzschia. March 2015. Oral presentation at Polony Workshop. Haifa, Israel.

Carlson, M., McCary, N., Rocap, G. Viral host range and permissivity of the diatom Pseudo-nitzschia. February 2015. Oral presentation at Aquatic Sciences Meeting. Granada, Spain.

Carlson, M., McCary, N., Rocap, G. Viral host range and permissivity of the diatom Pseudo-nitzschia. June 2014. Poster at Marine Microbes Gordon Research Conference. Boston, MA.

Carlson, M., McCary, N., Rocap, G. Host range of seasonality of Pseudo-nitzschia viruses. February 2014. Oral presentation at Ocean Sciences Meeting. Honolulu, HI.

Carlson, M., Frischkorn, K., Rocap, G. Isolation and characterization of viruses infecting Pseudo-nitzschia. January 2013. Poster at Environmental Virology Workshop. Tucson, AZ.

Carlson, M., Guannel, M., Hubbard, K., Port, J., Turner, J. Connecting researchers in oceans and human health (OHH): the OHHI traineeship program of the Pacific Northwest Consortium. October 2010. Poster at Oceans and Human Health Initiative 2010 National Advisory Panel Meeting. Washington, D.C.

Carlson, M., Frischkorn, K., Rocap, G. Isolation and characterization of viruses infecting Pseudo-nitzschia. June 2010. Poster at Oceans and Human Health Gordon Research Conference. Biddeford, ME.

Carlson, M. and Rocap, G. Isolation and characterization of viruses infecting Pseudo-nitzschia. January 2010. Oral presentation at the UW Oceans and Human Health Seminar Series. Seattle, WA.

Carlson, M. and Rocap, G. Isolation and characterization of viruses infecting Pseudo-nitzschia. December 2009. Oral presentation at the University of Washington for the Department of Oceanography. Seattle, WA.

Carlson M., Van Biber, B., Rocap, G Isolation and characterization of viruses infecting Pseudo-nitzschia. November 2009. Poster at 4th Harmful Algae Symposium. Ocean Shores, WA.

Carlson M., Van Biber, B., Rocap, G. Isolation and characterization of viruses infecting Pseudo-nitzschia. October 2009. Oral presentation at Oceans and Human Health Initiative Principal Investigators Meeting. Seattle, WA.

Carlson, M. A Pomona perspective on the Edwin W. Pauley Program. July 2009. Invited oral presentation at the 25th Anniversary of the Edwin W. Pauley Program. Kane'ohe, Hawaii.

Carlson, M., Van Biber, B., Rocap, G. Viral-phytoplankton interactions: Isolation, characterization, and ecology of viruses infecting the diatom Pseudo-nitzschia. May 2009. Poster presented at the 8th SCOR Marine Virus Working Group Meeting. Newark, DE.

PRESENTATIONS BY UNDERGRADUATE MENTEES

Leach, T, **Carlson, M.**, Guannel, M., Rocap, G. Uncovering the evolutionary history of *Pseudo-nitzschia*. Poster presentation at at the Life of Diatoms Conference. July 2015. Seattle, WA.

McCary, N., **Carlson, M.**, Rocap, G. Seasonal and strain specific variation in *Pseudo-nitzschia* virus susceptibility in the Pacific Northwest. Oral presentation at Joint Aquatic Sciences Meeting. June 2014. Portland, OR.

Leach. T., Guannel, M., **Carlson**, **M.**, Rocap, G. Uncovering the evolutionary history of Pseudo-nitzschia. Oral presentation at Ocean Sciences Meeting. Oral presentation at Ocean Sciences Meeting. February 2014. Honolulu, HI.

Frischkorn, K. **Carlson, M.**, Rocap. G. Characterization of viruses that infect *Pseudo-nitzschia*. Oral presentation at Oceans and Human Health Seminar. November 2010. Seattle, WA.

CRUISES AND FIELD WORK

April 2013-April 2014 Monthly sampling at Penn Cove and Grayland Beach State Park

March 16- 27, 2012 UW Oceanography Senior Thesis Cruise. San Diego, CA to Manzanillo, Mexico. R.V. Thompson

October 8-19, 2011 Oceans and Human Health Cruise. R.V. Thompson

October 29-30, 2010 Oceans and Human Health Cruise. Puget Sound, WA. R.V. Thompson

November 17-19, 2009 Oceans and Human Health Cruise. Puget Sound, WA. R.V. Thompson

Summers 2007-2010 Shore based sampling expeditions in Puget Sound, WA, Washington, Oregon, and California coasts, and Kane'ohe Bay, HI.

OUTREACH AND BROADER IMPACTS

2015 Graduate Recruitment, Retention and Diversity (GRRAD) committee member

2012-2015 Mentor for students in the Genomics Outreach for Minorities, Alliances for Learning and Vision for Underrepresented Americans (GenOM ALVA)

Biosketch

Michael Curtis Grier Carlson was born in 1986 in Claremont, CA. His parents, Catharine Grier Carlson and Mark Carlson, were both Presbyterian ministers. He attended Pomona College in Claremont, CA, after which he attended the University of Washington, in Seattle, WA.

Michael's interest and passion for the ocean began early, spending summers at the beach with his grandparents swimming, sailing, lifeguarding, and working at an aquarium shop. He was inspired to understand the ocean during trips to Hawaii where he spent his time on coral reefs. In 6th grade, he forayed into marine biology during a summer course at the Polytechnic School. There, with friends, the life-long dream to pursue marine science was crystalized. In the summer his junior year at Pomona College, Michael interned at the Hawaii Institute of Marine Biology at the University of Hawaii where he research viruses on coral in Kane'ohe Bay. Being excited enough to wake up at 7 every morning (a feat for him), Michael decided to purse marine science research. He continued his passions at the University of Washington studying viruses that infect marine phytoplankton.

Michael's passions have always been dual, driven by the need for balance. He competitively swam and played tennis. He plays the harp and paints. He majored in biology and history at Pomona College. He studied viruses and their hosts of his Ph.D. He continues this theme as a postdoctoral scholar at the Technion University in Israel studying marine cyanobacteria and their viruses.