Physiological and ecological characterization of the *Aureococcus anophagefferens* virus host system

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Eric Robert Gann

May 2020

Copyright © 2020 by Eric Robert Gann
All rights reserved.

Acknowledgments

This would not have been possible without the strong support network of everyone here in Knoxville. I would like to express my deepest gratitude towards both of my advisors, Drs. Steven Wilhelm and Todd Reynolds. The amount of professional and personal guidance each have given to me over the past four and a half years has helped my growth greatly not only as a scientist, but as a person. The great passion for science, teaching, and mentoring both possess helped me maintain my enthusiasm for science through the peaks and valleys of graduate school and life. I strive to be as good of a scientist, and person, as both of the them throughout my career.

I would also like to thank the rest of my dissertation committee, Drs. Erik Zinser, Tim Sparer, and Brad Binder, for the support and guidance. This work would not have been the same without the insight from all of them. Finally, without the support of all of the members, past and present, in the labs, those in the department, and the friends made along the way, I would not have been able to complete this work.

Abstract

Viruses in aquatic systems are a major source of mortality to microbial life, influencing the structure of communities, nutrient cycling, and evolution. Many lineages of eukaryotic algae have been visually confirmed to be infected with viruses, but only a small subset have isolates. Of those where viruses are isolated, an incredible amount of diversity exists, including differences in genome composition, particle sizes, and infection strategies. Under certain conditions specific populations of eukaryotic algae can bloom, which cause millions of dollars in losses. The harmful brown tide bloom forming *Aureococcus anophagefferens* negatively influence sea grass beds and are potentially toxic to shellfish. Viruses have been implicated in the termination of these blooms. A viral isolate, Aureococcus anophagefferens Virus (AaV), has been characterized over the past decade. This work continued characterization of this system through characterization of the particle itself, development of new tools to study the ecology of AaV, and by studying how abiotic factors influence the infection cycle. The particle has the same conserved capsid structure as other related icosahedral viruses and contains at least 43 proteins. Of those packaged, some were found in particles of similar viruses, while others (*i.e.* sugar-degrading proteins) were unique. This study provided novel insight into evolutionarily conserved aspects of these particles, as well as those that are system specific. To understand the role of *Aureococcus* infecting viruses in natural blooms, we developed reliable methods to enumerate total (quantitative PCR) and infectious (plaque and most probable number) particles. Using these tools, the influence of irradiance levels and N:P ratios on the infection cycles were studied. At light levels which reduces growth either caused by acclimation or light shift experiments, an increased latent time and a reduced burst size was observed. Acclimating cultures to N:P ratios

where growth was reduced, negatively influenced the infection cycle, while nutrient shift experiments had not effect. Overall, this work provided novel insights into the *A. anophagefferens* AaV system and developed new tools to study it in the environment. As each algal virus system studied is unique, more research should be conducted to isolate and characterize novel systems.

Table of Contents

Chapter One Introduction: Diversity of cultured algal viruses 1
Introduction 2
Cultured algal-virus systems 4
Trebouxiophyceae 4
Mamiellophyceae 6
Dinophyceae 7
Bacillariophyta 9
Raphidophyceae 12
Phaeophyceae 14
Pelagophyceae 16
Haptophyta 16
Other isolated virus systems 18
Diversity of algal-virus systems requires continued exploration 19
Work in this dissertation 20
Chapter Two Structural and proteomic studies of the Aureococcus anophagefferens Virus demonstrate a global distribution of virus encoded carbohydrate processing 25
Author contributions 26
Abstract 27
Introduction 29
Methods 30
AaV propagation and concentrations 30
Electron microscopy 31
Protein extraction and digestion 32
Protein identification and quantification 33
Bioinformatic comparison of packaged proteins within different viral families 34
Environmental relevance of carbohydrate lyase proteins within the AaV particle 35
Results and Discussion 37
Structural characteristics of the AaV particle 37
Virus particle reconstruction 38
Proteins packaged within the AaV particle 40
Detection of other putative structural proteins 41
Transcriptional Machinery 43
Proteins unique to the AaV particle 44
Comparison of Giant Virus Proteomes 46
Environmental relevance of packaged AaV gene products 48
Examining polysaccharide degrading enzymes in the TARA Oceans datasets 49
Conclusions 52
Acknowledgments 53
Chapter Two Appendix 54
Chapter Three Influence of light on the infection of *Aureococcus anophagefferens* CCMP 1984 by a “giant virus” 99
Publication note and author contributions 100

Publication note 100
Author contributions 100
Abstract 101
Introduction 102
Methods 104
Culture conditions 104
Plaque assay 105
Most probable number assay 106
Enumeration of free AaV particles by quantitative PCR 107
Enumeration of total virus particles 108
Bioinformatic predictions for effects of light on infection 109
Statistics 110
Results 111
Infectivity vs particle abundance of Aureococcus anophagefferens Virus particles 111
Effects of pre-infection light intensities on AaV virus production 112
Effect of reduced light availability post infection 113
Effects of pre-infection light reduction on the progress of AaV infection 115
Informatic examination of the effects of light and virus infection 116
Discussion 117
The infectivity of Aureococcus anophagefferens Virus particles 117
Effects of light acclimation on the progression of infection 118
Effect of growth limiting light on infection 119
Bioinformatic insight into limitations to the virocell at low light 120
Potential ecological impact 121
Acknowledgments 124
Chapter Three Appendix 125
Chapter Four Internal nitrogen pools shape the infection of *Aureococcus anophagefferens* CCMP 1984 by a giant virus 146
Author contributions 147
Abstract 148
Introduction 149
Methods 152
Culture conditions 152
Enumeration of *A. anophagefferens* and AaV 153
Experimental Design 154
Statistical Analyses 156
Results 156
The effects of nitrate concentration acclimation on *A. anophagefferens* growth 156
The effects of nitrate concentration acclimation on the AaV infection cycle 157
The effect of external nitrogen on concentration on the AaV infection cycle 158
Discussion 159
The effects of nitrate acclimation on *Aureococcus* growth and the AaV infection cycle 159
The effects of external nitrogen on the AaV infection cycle 161

Assessing nitrogen requirements for the infection cycle to occur 163
Conclusions 166
Acknowledgments 167
Chapter Four Appendix 168
Chapter Five Conclusions 184
Conclusions 185
Development of new tools 185
Structural and proteomic study of the AaV particle 187
Effects of altering abiotic factors on the infection cycle 188
Future directions 189
List of References 191
Vita 228

List of Tables

Table 2.1. List of sequencing data used in this study. ……54
Table 2.2. Characteristics of structural studies of large icosahedral viruses...…….....56
Table 2.3. Characteristics of the 43 proteins packaged within the AaV particle. ……..57
Table 2.4. Mass spectrometry intensities and normalized intensities to MCP for proteins per particle for all the viral proteins detected within the proteomes...…….....59
Table 2.5. Description of the host proteins detected by proteomics...…….....61
Table 2.6. Mass spectrometry intensities and normalized intensities to MCP for proteins per particle for all the host proteins detected within the proteomes...…...63
Table 2.7. Categories of reciprocal best BLAST hit pairs based on their NCVOG classification (Yutin *et al.,* 2009).……65
Table 2.8. Number of reads mapped to packaged viral proteins from previously published meta-transcriptomes of an A. anophagefferens bloom (Moniruzzaman *et al.,* 2017)...……66
Table 2.9. Metagenomic assembled contigs from TARA Oceans Dataset separated by water depth. …….....69
Table 2.10. Metagenomic assembled contigs from TARA Oceans Dataset separated by size fraction..…….70
Table 2.11. Metagenomic assembled contigs from TARA Oceans Dataset separated by Station. …….....71
Table 2.12. Classification of putative viral contigs containing polysaccharide lyases into viral families by best BLAST hit to NCVOG on contig..……73
Table 2.13. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for percentage of contigs that are viral by water depth...…….....74
Table 2.14. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for percentage of contigs that are viral by water fraction.…...…...……..75
Table 2.15. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for normalized metagenomics reads....…...…….76
Table 2.16. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for normalized metatranscriptomics reads..……77
Table 3.1. Adjusted p-values comparing differences in AaV concentration on Day three (Figure 3.3) based on irradiance level and viral enumeration method as determined by two-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD..…….……125
Table 3.2. Adjusted p-values comparing differences in burst sizes based on irradiance level cells were transferred to after infection as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD (Fig 3.6)..……...126
Table 3.3. Summary of acclimation conditions including host doubling time (Figure 3.4), mean forward scatter (FSU-H) determined by flow cytometry, burst size (MCP copies produced/*A. anophagefferens* cell lysed), and percent of particles determined to be infectious (from Fig 3.3). …….....127
Table 3.4. Forward Scatter (FSC-H) values for high (90 µmol photons m^{-2} s^{-1}) and low (30 µmol photons m^{-2} s^{-1}) light acclimated cultures either maintained at acclimating light or transitioned 5 µmol photons m^{-2} s^{-1} for one day.…….……...….128

Table 3.5. Summary of genes used in analysis for differences between low light conditions (from Frischkorn *et al.* 2014) and the infection cycle (Moniruzzaman *et al.* 2018).....129
Table 3.6. KEGG pathways of *A. anophagefferens* genes differentially expressed in both the CCMP1850 transcriptome (Frischkorn *et al.* 2014) and the CCMP1984 infection cycle transcriptome (Moniruzzaman *et al.* 2018).....130
Table 3.7. Subset of the *Aureococcus anophagefferens* genes predicted to be significantly downregulated in low light (Frischkorn *et al.* 2014) that are related to ribosome biogenesis according to KEGG classifications.....131
Table 4.1. Concentrations of NO_3^- and PO_4^{3-} of the different growth media.....168
Table 4.2. Adjusted p-values comparing differences in end point cell abundance in an uninfected growth curve based on nitrate concentration.....169
Table 4.3. Adjusted p-values comparing differences in doubling time based on nitrate concentration.....170
Table 4.4. Adjusted p-values comparing differences in different flow cytometry parameters based on nitrate concentration.....171
Table 4.5. Adjusted p-values comparing differences in burst size based on nitrate concentrations as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.....172
Table 4.6. Adjusted p-values comparing differences in adsorption rate based on nitrate concentrations as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.....173
Table 4.7. Adjusted p-values comparing differences in end point abundance based on nitrate concentrations added back as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.....174
Table 4.8. Adjusted p-values comparing differences in burst size based on nitrate concentrations added back as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.....175
Table 4.9. Estimated use of atoms within *A. anophagefferens* cell to produce new viruses.....176

List of Figures

Figure 2.1. cryo-EM reconstruction of the AaV virion. …….....78
Figure 2.2. Viral proteins detected in the proteome normalized to the number of MCP per virion by mass spectrometry intensities. ……..79
Figure 2.3. Overview of the proteins found within the AaV particle. ……...80
Figure 2.4. Host proteins detected in multiple runs of the proteome normalized to the number of MCP per virion by mass spectrometry intensities. …….....81
Figure 2.5. Overview of the other proteomes of packaged proteins used in this study.… ……82
Figure 2.6. Comparison of packaged proteins between representative viruses. ……..……..83
Figure 2.7. Circos plot showing translation related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. …….……...85
Figure 2.8. Circos plot showing reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used with unknown functions. …….86
Figure 2.9. Circos plot showing Redox related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. ……...87
Figure 2.10. Circos plot showing MCP reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. ……..88
Figure 2.11. Circos plot showing DEAD/SNF2-like helicases related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. …….....89
Figure 2.12. Circos plot showing transcription/RNA processing-related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. ……..……...90
Figure 2.13. Maximum likelihood phylogenetic relatedness of all characterized polysaccharide lyases within the Carbohydrate-Active Enzyme (CAZy) Database. ……....91
Figure 2.14. Percentage of polysaccharide lyases found on viral contigs to total contigs containing polysaccharide lyases in TARA Oceans metagenomic contigs.…… …..…..….....92
Figure 2.15. Maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral TARA Oceans metagenomic contigs. ……..……..93
Figure 2.16. Subtree of maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral TARA Oceans metagenomic contigs containing the AaV pectate lyases. …….....94
Figure 2.17. Maximum likelihood phylogenetic trees of NCVOG found on TARA Oceans metagenomic contigs within AaV subtree of the maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral contigs. …….....95
Figure 2.18. Contribution of polysaccharide lyases found on viral contigs to total contigs containing polysaccharide lyases by TARA Oceans station. …….....96
Figure 2.19. Normalized mapped reads for TARA Oceans Station 122. ……..97
Figure 2.20. Combined RPKM values of polysaccharide lyases over the 21 hour infection cycle. …….....98
Figure 3.1. Standard curve validating qPCR method to enumerate free AaV particles. ……..…….....134
Figure 3.2. Comparison between total virus particle counts as determined by SYBR green epifluorescence microscopy and infectious particles determined by plaque assay or most probable number. ……..……...…...135

Figure 3.3. Comparison of three different methods to enumerate AaV. ……..…....136
Figure 3.4. Growth curve of *A. anophagefferens* acclimated to three different irradiance levels. …….....137
Figure 3.5. *Aureococcus anophagefferens* virus infection cycle dynamics with acclimation to varying irradiance levels. ……..……..…….138
Figure 3.6. *Aureococcus anophagefferens* virus infection cycle dynamics with acclimation to varying light levels and in lower light. …….……...……139
Figure 3.7. Effects of reducing light levels to 15 and 5 µmol photons m^{-2} s^{-1} after infection on Burst Size (Figure 3.6). ……..…….……..……140
Figure 3.8. Burst Size v. Irradiance Level during the infection from distinct experiments (Table 3.3 and Fig 3.7B). ……...…….……141
Figure 3.9. *Aureococcus anophagefferens* virus infection cycle dynamics with acclimation to varying irradiance levels and then transferred to the dark. …….....142
Figure 3.10. *Aureococcus anophagefferens* virus infection cycle dynamics with pre-acclimation to 5 µmol photons m^{-2} s^{-1} for one day before infection. ……...143
Figure 3.11. Comparison between de novo assembled *Aureococcus anophagefferens* CCMP1850 transcriptome in low light (Frischkorn *et al*. 2014) and *Aureococcus anophagefferens* CCMP1984 transcriptome over the course of the infection cycle (Moniruzzaman *et al.* 2018). …….....144
Figure 3.12. Ecological predictions of *Aureococcus* AaV interactions over the course of a bloom and effects of sinking. ……..145
Figure 4.1. Influence of filtering AaV through a 0.45 µm filter. …..……..177
Figure 4.2. Representative FACS Scatterplot of 0.45 µm filtered lysate stained with SYBR green. …….....178
Figure 4.3. Growth characteristics of *Aureococcus anophagefferens* acclimated to decreasing NO_3^- concentrations. …….…… ….....179
Figure 4.4. AaV infection cycle dynamics of *Aureococcus anophagefferens* acclimated to decreasing NO_3^- concentrations. …… .180
Figure 4.5. Flow cytometry parameters detected over the course of the first 24 hours of the infection cycle. ……....181
Figure 4.6. AaV infection cycle dynamics of *Aureococcus anophagefferens* acclimated to 1.47 mM NO_3^-, with different concentrations of NO_3^- added back after removing external NO_3^-. …….....182
Figure 4.7. Fold change values of transporters found within the *Aureococcus* genome over the course of the infection cycle that are > 1.5 or < -1.5 fold change for infected v. uninfected control cultures with a FDR p-value < 0.05 (Moniruzzaman *et al.,* 2018). ……..183

Chapter One

Introduction: Diversity of cultured algal viruses

Introduction

Estimates of the number of different species of eukaryotic algae have ranged from 10^4 - 10^8, with conservative estimates (~70,000) extrapolated from what has already been described (Guiry, 2012). Algae provide the basis for food webs in aquatic and terrestrial ecosystems and contribute greatly to the amount of oxygen found within the atmosphere (Armbrust, 2009). When conditions are right, algae can form dense blooms with 10^5 - 10^6 cells per mL (Gastrich *et al.*, 2004; Tomaru *et al.*, 2007), which can cause negative side effects, including hypoxia, producing toxic compounds to humans, bivalves, and fish, and decreasing tourism (Berdalet *et al.*, 2015). These events can cause tens of millions of dollars in losses (Tarutani *et al.*, 2001; Gobler and Sunda, 2012). Viruses have been visualized in transmission electron microscopy studies of natural populations of algae since the middle of the last century (as reviewed in Wilhelm *et al.*, 2017), and a few systems had begun to be characterized in the 1980s (Van Etten *et al.*, 1982; Waters and Chan, 1982). When it was determined that viruses were not only abundant in aquatic systems, but were also an important source of mortality (Proctor and Fuhrman, 1990), the field of aquatic virology expanded rapidly. It has been estimated that there are more viruses than any other biological entity in the oceans (Suttle, 2005), contributing to over half of the microbial population lysing each day (Suttle, 2007), although most of these infections are infections of prokaryotes. The high number of infections occurring continuously can influence community structure (Thingstad, 2000), cause nutrients to remain available for the microbial community (Wilhelm and Suttle, 1999), and contribute to the evolution of their hosts (Rohwer and Thurber, 2009).

Of the tens of thousands to millions of species of eukaryotic algae predicted to exist, only about 60 have isolated viruses in culture (Coy *et al.*, 2018). Of those isolated, many of their hosts

cause harmful algal blooms (Sieburth *et al.*, 1988; Tomaru *et al.*, 2004a; Tomaru *et al.*, 2004b), or are easily culturable (Van Etten *et al.*, 1983). Cultured algal viruses are incredibly diverse, having genomes ranging from 5.5 to 668 kbp (Kimura and Tomaru, 2013; Schvarcz and Steward, 2018) and particles ranging from simple self-assembling small capsids (Wu *et al.*, 2012), to those requiring scaffolding proteins predicted to be important for assembly and support (Fang *et al.*, 2019). These viruses possess different types of genomes containing large double-stranded DNA (Schvarcz and Steward, 2018), double-stranded RNA (Brussaard *et al.*, 2004), single-stranded RNA (Tomaru *et al.*, 2004a), and single-stranded DNA (Tomaru *et al.*, 2011a). The same host can be infected by multiple types of viruses including *Micromonas pusilla* (Brussaard *et al.*, 2004; Baudoux and Brussaard, 2005b), *Chaetoceros tenuissimus* (Kimura and Tomarua, 2015), *Heterocapsa circularisquama* (Tarutani *et al.*, 2001; Tomaru *et al.*, 2004a), and *Heterosigma akashiwo* (Nagasaki and Yamaguchi, 1997; Lawrence *et al.*, 2001; Tai *et al.*, 2003). Viruses with large double-stranded DNA genomes belong to multiple families (*i.e. Mimiviridae, Phycodnaviridae*) within the monophyletic group the nucleocytoplasmic large double-stranded DNA viruses (NCLDVs), while those with other genome types do not appear to be related (Attoui *et al.*, 2006; Toyoda *et al.*, 2012; Pagarete *et al.*, 2015). The NCLDV group of viruses contains viruses that infect both photosynthetic and heterotrophic eukaryotes, and some members possess genomes and particle sizes that rival bacteria (Wilhelm *et al.*, 2017). Below describes what has been learned from these isolated systems both from laboratory and field experiments.

Cultured algal-virus systems

Trebouxiophyceae

The most well characterized system of algal viruses are those that infect endosymbiotic algae which live within district lineages of eukaryotes. These zoochlorellae algae can be isolated and grown in liquid culture and on plates (Van Etten *et al.*, 1982). This has allowed for many viruses to be isolated on different strains and species from many freshwater locations (Jeanniard *et al.*, 2013). The most well studied virus is the Paramecium bursaria chlorella virus 1, PBCV-1, which infects *Chlorella* NC64A (Van Etten *et al.*, 1983). Over the past thirty years, much has been learned about the infection cycle (Vanetten *et al.*, 1983; Blanc *et al.*, 2014; Milrot *et al.*, 2017), the particle (Cherrier *et al.*, 2009; Dunigan *et al.*, 2012; De Castro *et al.*, 2013; Fang *et al.*, 2019), and the environmental distribution of chloroviruses (Jeanniard *et al.*, 2013), which belong to the *Phycodaviridae* family. The initiation of the PBCV-1 infection cycle begins when a specialized spike contacts the host cell (Cherrier *et al.*, 2009), then DNA is injected into the cell at least in part due to membrane depolarization that occurs upon infection (Neupärtl *et al.*, 2008; Milrot *et al.*, 2017). Transcription occurs within 7 minutes (Blanc *et al.*, 2014), and DNA replication requires the degradation of the host genome by virally encoded restriction endonucleases (Agarkova *et al.*, 2006). Virus particles assemble in the cytoplasm and are released around 8 hours post infection (Van Etten *et al.*, 1983).

The large (~190 nm) virus particle of PBCV-1 is complex (Dunigan *et al.*, 2012; Fang *et al.*, 2019). The PBCV-1 particle is hypothesized to rely on scaffolding minor structural proteins (Fang *et al.*, 2019), and electrostatic interactions of the major capsid protein (Xian *et al.*, 2019), to assemble in a continuous manner (Xiao *et al.*, 2017). These minor structural proteins are also hypothesized to provide structural support for the major capsid protein (Fang *et al.*, 2019). Also

packaged in the particle are proteins believed to be important in the initiation of the infection cycle, including proteins capable of breaking down the cell wall upon contact (Dunigan *et al.*, 2012; Milrot *et al.*, 2017). Interestingly, the capsid protein of the PBCV-1 particle is glycosylated through proteins that are encoded in the 331 kbp genome (Dunigan *et al.*, 2012; De Castro *et al.*, 2013), which are not necessary for infection, but are hypothesized to improve the stability of the particle (Van Etten *et al.*, 2017).

Although new chloroviruses can be isolated in many freshwater systems (Jeanniard *et al.*, 2013), their role in the environment is not well understood. Naturally, zoochlorellae are found within their host where an infection cannot occur. The concentration of chloroviruses fluctuates over the course of a year and by location but can be continuously found in concentrations from $10^1 – 10^3$ plaque forming units per mL (Quispe *et al.*, 2016). Concentrations and genetic differences of chlorovirus populations are dependent on the type of zoochlorellae used for the plaque assays (Quispe *et al.*, 2016). Chloroviruses have a shared core-genome of ~155 proteins regardless of host they were isolated from and phylogenetically, these viruses cluster based on host they infect, not the location they were isolated from (Jeanniard *et al.*, 2013). Although the concentration of free hosts in the environment may be limited (Quispe *et al.*, 2016), infections by chloroviruses may influence carbon cycling by causing infected cells to clump based on synthesis of cell wall components (Graves *et al.*, 1999). In addition, they may influence movement of the Protozoa hosts zoochlorellae live in, as these viruses can act as chemotactic stimuli (Dunigan *et al.*, 2019).

Mamiellophyceae

Prasinophytes are small algae, distributed worldwide and can make up a large percentage of the eukaryotic fraction of the water column (Waters and Chan, 1982). Viruses for this group were first isolated and characterized from *Micromonas pusilla* (Waters and Chan, 1982; Cottrell and Suttle, 1995). Although large Phycodnaviruses are isolated readily on these multiple hosts, other types of viruses have also been isolated. A double-stranded RNA virus has been isolated on *M. pusilla* and partially characterized (Brussaard *et al.*, 2004; Attoui *et al.*, 2006). These viruses have genomes segmented into 11 parts that total 25 kbp (Attoui *et al.*, 2006) and a 75 nm chloroform resistant capsid (Brussaard *et al.*, 2004), containing at least five major proteins (Brussaard *et al.*, 2004). Large viruses have also been isolated on other prasinophytes including *Micromonas spp*. such as *Micromonas polaris* (Maat *et al.*, 2017), multiple *Ostreococcus spp.* (Bellec *et al.*, 2010), and *Bathycoccus sp.* (Moreau *et al.*, 2010). All of these viruses have dsDNA genomes of sizes between 150 kbp – 205 kbp and belong to the *Phycodnaviridae* (Moreau *et al.*, 2010; Finke *et al.*, 2017). Both globally-distributed *Micromonas pusilla* and *Ostreococcus lucimarinus* have had viruses isolated from distinct locations all over the world (Bellec *et al.*, 2010; Derelle *et al.*, 2015; Finke *et al.*, 2017), while viruses infecting *Ostreococcus tauri* have only been isolated in Mediterranean (Bellec *et al.*, 2010). These viruses can be found in varying concentrations depending on the season (Cottrell and Suttle, 1995). It has been hypothesized from DNA polymerase B amplicons that ecotypes with similar genetic potential are found in similar environments (Finke *et al.*, 2017), however viruses isolated from one location can be more diverse than those isolated from distinct locations (Bellec *et al.*, 2010). Support for this hypothesis has been shown through genomic and metagenomic analyses of prasinoviruses (Monier *et al.*, 2012; Monier *et al.*, 2017). An example of this is a virus encoding

a nitrogen transporter which infects *O. tauri*, normally found in nitrogen deplete conditions (Monier *et al.*, 2017), and this transporter allows for infected cells to take up more nitrogen, which is hypothesized to allow for a more productive infection (Monier *et al.*, 2017).

Resistance develops after the infection of susceptible strains of *Micromonas sp.*, *Bathycoccus sp.*, and *Ostreococcus sp.* These populations can then grow in the presence of their viruses. These strains remain resistant to viruses even after many transfers (Thomas *et al.*, 2011). Prasinophytes contain a unique small chromosome that has a distinct GC content which recombines more readily than the remaining chromosomes (Blanc-Mathieu *et al.*, 2017). Rearrangement of this chromosome allows for one type of resistance for *Ostreococcus spp.* when challenged by viruses (Yau *et al.*, 2016; Yau *et al.*, 2018). This type of resistance does not influence the growth of resistant cells when maintained for many generations in different environments (Heath *et al.*, 2017). This chromosome recombines and exhibits changes in size and expression of the genes present on it. Many of the genes are involved in carbohydrate and glycotransferase synthesis (Yau *et al.*, 2016). This type of resistance is believed to occur at rates between $10^{-3} – 10^{-4}$ of the initial susceptible population (Yau *et al.*, 2018). As this rearrangement does not occur in populations not challenged by viruses (Yau *et al.*, 2016), this allows for an important mechanism of evolution of the host genomes.

Dinophyceae

There have been several viruses isolated from dinoflagellates, all of which are important as they cause harmful blooms. Isolates infecting both *Heterocapsa pygmaea* and *Gymnodinium mikimotoi* have been found to be double-stranded DNA containing viruses that assemble their viruses in the cytoplasm after the nucleus is degraded (Onji *et al.*, 2003; Kim *et al.*, 2012). The

virus infecting *G. mikimotoi* was around 100 nm in diameter and did not possess a lipid membrane (Onji *et al.*, 2003). The virus infecting *H. pygmaea* was larger, 160 nm in diameter, had a latent period less than 36 hours with a burst size between 100 and 250 infectious particles per cell (Kim *et al.*, 2012). The dinoflagellate virus system that is the most characterized is the *Heterocpsa circularisquama* system. This thecate dinoflagellate is important as it annually causes red tides, off the coast of Japan in embayments, which are toxic to bivalves (Horiguchi, 1995). Annually these blooms can produce over 30 million dollars in losses (Tarutani *et al.*, 2001). Both large double-stranded DNA viruses and small single-stranded RNA viruses have been isolated from this system (Tarutani *et al.*, 2001; Tomaru *et al.*, 2004a). The characterized double-stranded DNA virus (HcV) has a 356 bp genome and a particle size of ~200 nm (Tarutani *et al.*, 2001; Ogata *et al.*, 2009). Phylogenetic analysis of the DNA polymerase B gene within the genome clusters this virus within the Asfarviridae, which are classically characterized to infect mammals (Ogata *et al.*, 2009). HcV replicates in the cytoplasm and causes lysis of the host within 32 hours (Nagasaki *et al.*, 2003). The single-stranded RNA viruses that infect *H. circularisquama* have 4.4 kbp genomes encoding two proteins (Tomaru *et al.*, 2004a; Nakayama and Hamaguchi, 2016). The 30 nm particle consists of the RNA genome loosely associating with 180 units of the capsid protein (Miller *et al.*, 2011), and can self-assemble *in vitro* (Wu *et al.*, 2012). These viruses also replicate in the cytoplasm, have a latent period between 24 and 48 hours, and produce between 0.34 – 2.1×10^4 infectious particles per cell lysed (Tomaru *et al.*, 2004a). Regardless of virus type infecting *H. circularisquama,* the protist cells lose their motility, settle to the bottom of their container, and become rounded (Tarutani *et al.*, 2001; Tomaru *et al.*, 2004a). In the case of both virus types, the entire population of susceptible host cells did not lyse after being challenged by viruses, with 10-20% of cells became cyst-like, and

when transferred to fresh virus-free media could grow (and become susceptible to viruses) again (Tarutani *et al.*, 2001; Nagasaki *et al.*, 2003; Tomaru *et al.*, 2004a). Unique to infection by the single-stranded RNA viruses was a resistance mechanism that occurred from a suppression of viral genome replication found in a very small percentage of the population (< 0.5%) (Tomaru *et al.*, 2009a). In the presence of the viruses, this resistance mechanism remains in the majority of cells, but once the pressure is removed, cells revert to being susceptible again (Tomaru *et al.*, 2009a). There appear to be two distinct ecotypes of small RNA viruses present in the environment based on which subset of strains can be infected. This was first hypothesized upon isolation (Tomaru *et al.*, 2004a), but has been seen in environmental studies as well (Tomaru *et al.*, 2007; Nakayama and Hamaguchi, 2016). Viruses similar to the isolated single-stranded RNA viruses appear to have an important role in bloom collapse, as their numbers drastically increase during bloom termination (Nakayama and Hamaguchi, 2016). At the end of the blooms over 80% of the populations are infected with small viruses, even though infections with large viruses are also present (Tomaru *et al.*, 2007). Although the percentage of infectious particles is very low (< 0.01%) (Nakayama and Hamaguchi, 2016), viruses within sediments are hypothesized to be important in modulating bloom severity, and composition from year to year (Tomaru *et al.*, 2007).

Bacillariophyta

Diatoms are believed to have arisen 250 million years ago and currently are hypothesized to contribute ~20% of the photosynthesis on Earth (Armbrust, 2009). Diatoms can proliferate to high abundances in low temperatures. These blooms have the ability to negatively impact the economy through losses in fisheries (Nagasaki *et al.*, 2004; Eissler *et al.*, 2009), and some can

produce compounds toxic to humans (Armbrust, 2009). Although there have been transmission electron microscopy images of diatoms being infected with icosahedral viruses since the early 1990's (Proctor and Fuhrman, 1991), the first viral isolate that was shown to infect diatoms was isolated from surface water and infected the centric diatom, *Rhizosolenia setigera* (Nagasaki *et al.*, 2004). The 32 nm virus replicated in the cytoplasm and contained a single-stranded RNA genome that was 11.2 kbp (Nagasaki *et al.*, 2004). Subsequent studies have isolated small single-stranded RNA viruses from multiple centric diatoms including *Guinardia delicatula* (Arsenieff *et al.*, 2019), and several *Chaetoceros* spp. (Shirai *et al.*, 2008; Tomaru *et al.*, 2009b; Tomaru *et al.*, 2013a; Kimura and Tomarua, 2015), and the pennate diatom *Asterionellopsis glacialis* (Tomaru *et al.*, 2012), all being classified in the genera *Bacillarnavirus* (Toyoda *et al.*, 2012). All of these viruses had genomes around 10 kbp, replicated in the cytoplasm, and were between 22-38 nm in diameter. The genomes for these viruses only encode a few (2-3) open reading frames which are classified as a replication protein and a structural protein (Tomaru *et al.*, 2009b; Kimura and Tomarua, 2015). The structural protein has been hypothesized to be a polyprotein (Kimura and Tomarua, 2015), as the purified particles contain three – four proteins (Tomaru *et al.*, 2009b; Kimura and Tomarua, 2015). Other nucleus-replicating diatom virus isolates have also been shown to contain single-stranded DNA genomes, belonging to the genera *Bacilladnavirus* (Toyoda *et al.*, 2012). These viruses have also been isolated on multiple *Chaetoceros* spp. (Nagasaki *et al.*, 2005; Tomaru *et al.*, 2008; Tomaru *et al.*, 2011a; Tomaru *et al.*, 2011b; Toyoda *et al.*, 2012; Kimura and Tomaru, 2013; Tomaru *et al.*, 2013a), and *Thalassionema nitzchiodes* (Tomaru *et al.*, 2012). These viruses have genomes smaller than their RNA counterparts, ranging from 5.5 kbp (Tomaru *et al.*, 2012) – 8 kbp (Eissler *et al.*, 2009). Interestingly, there is evidence that many of these genomes are circular, with a small (< 1 kbp)

region of double stranded DNA (Tomaru *et al.*, 2008; Tomaru *et al.*, 2012; Toyoda *et al.*, 2012). These viruses also contain genomes with few (2-3) open reading frames (Tomaru *et al.*, 2011a; Kimura and Tomaru, 2013), and particles with 1-4 structural proteins (Nagasaki *et al.*, 2005; Tomaru *et al.*, 2011a; Tomaru *et al.*, 2013b). Within each virus type there was large variability in regard to burst size and length of the infection cycle. In the *Bacilladnaviruses* infecting the genus *Chaetoceros*, the burst sizes ranged from 29 infectious particles per cell (Kimura and Tomaru, 2013) to $1.7x10^3$ infectious particles per cell (Kimura and Tomarua, 2015). A similar range was determined in the *Bacillarnaviruses*, with a maximum burst size of $1.1x10^4$ infectious particles per cell (Shirai *et al.*, 2008). The length of the latent period varied between < 8 hours (Eissler *et al.*, 2009) to > 24 hours (Kimura and Tomarua, 2015) depending on the virus and the host, but all were characterized by a loss of their photosynthetic pigments when infected. Most of the isolated small viruses could only infect a single species, or even specific strains, while a few could infect more than one species (Kimura and Tomarua, 2015).

Future work to study the viruses of *Chaetoceros* spp. could be to compare the differences in the many strains where one or both types of viruses are isolated to determine their differences. As this system was recently able to be genetically modified (Miyagawa-Yamaguchi *et al.*, 2011), more mechanistic studies should be completed. Future studies of diatom viruses will be important to understand their role in modulating diatom populations. This is important in carbon export, as a large amount of the organic matter from diatoms sinks (Armbrust, 2009). Laboratory studies have shown that infecting diatoms with viruses cause an increase in dense aggregates containing acidic polysaccharides and proteins (Yamada *et al.*, 2018). Studies on the residency of the viruses will be important as these lytic agents are not always present in the water column (Bettarel *et al.*, 2005), but can be isolated from sediments readily (Tomaru *et al.*, 2008). Finally,

the influence of the environment on the interactions between viruses and their hosts is necessary. Through metatranscriptomic and laboratory approaches, it was shown that viral infections caused a more rapid loss of biomass in silica stressed populations compared to those in replete conditions (Kranzler *et al.*, 2019). Although this was hypothesized to potentially be related to oxidative stress, which is important in other algal virus systems (Sheyn *et al.*, 2016), this mechanism was not described.

Raphidophyceae

Heterosigma akashiwo is a motile raphidophyte that can bloom to cell densities greater than $1x10^5$ cells/mL in distinct temperate and subarctic waters around the globe (Lawrence *et al.*, 2001; Tomaru *et al.*, 2004b). Blooms initiate from populations that are in a cyst form when conditions are optimal (*i.e.* temperature > 15 °C) (Tobin *et al.*, 2013), maintain high cell densities for weeks, and then within two or three days drop multiple orders of magnitude (Tarutani *et al.*, 2000). Although the first virus isolated that infects *H. akashiwo* was isolated in 1997 (Nagasaki and Yamaguchi, 1997), natural bloom populations were shown to be infected several years earlier (Nagasaki *et al.*, 1994). Viruses were inferred to cause bloom collapse as electron microscopy studies showed low levels of bloom populations were infected during the peak of the bloom, but at the onset of the collapse 20 times more cells were infected (Nagasaki *et al.*, 1994). The collapse of *H. akashiwo* blooms do not appear to be due to dilution (Nagasaki *et al.*, 1994), or lack of nutrients to sustain growth, but correspond to a large increase in *H. akashiwo* lysing viruses (Tarutani *et al.*, 2000; Tomaru *et al.*, 2004b). At bloom collapse the amount of *H. akashiwo* infecting viruses can be one to two orders of magnitude more abundant than their hosts, at concentrations greater than $1x10^6$ mL^{-1} (Nagasaki and Yamaguchi, 1997;

Tarutani *et al.*, 2000). Multiple types of viruses have been isolated using different strains of *H. akashiwo* from waters off the coast of Japan and western Canada including double-stranded DNA viruses with large (>200 nm) particle sizes (Nagasaki and Yamaguchi, 1997), two with smaller particle sizes (30 and 80 nm) and double-stranded DNA genomes (Lawrence and Suttle, 2004), a positive strand single-stranded RNA virus (Tai *et al.*, 2003; Lang *et al.*, 2004), and a virus that replicates exclusively in the nucleus (Lawrence *et al.*, 2001). The large Extended *Mimiviridae* virus has an infection cycle of 33 hours (Lawrence *et al.*, 2006; Maruyama and Ueki, 2016), in which the *H. akashiwo* cells become rounded, lose their photosynthetic pigments and motility and settle to the bottom of their containers (Nagasaki and Yamaguchi, 1997). The loss of photosynthetic pigments corresponds to a gradual decrease in photosynthetic activity, hypothesized to be due to changes in light harvesting complexes and the degradation of photosystem II (Juneau *et al.*, 2003). This loss in pigments and photosynthetic activity occurs with all types of *H. akashiwo* infecting viruses (Juneau *et al.*, 2003), but the rate of decrease is variable. Also, common to all of the viruses isolated, is the loss of motility and sinking (Lawrence and Suttle, 2004). This is potentially important to the collapse of blooms, as it is seen that at bloom termination cells do not travel to the surface of the water column (Nagasaki and Yamaguchi, 1997; Tarutani *et al.*, 2000). Environmental studies of water and sediment where blooms occur reveal these viruses are always present but in varying concentrations (Tarutani *et al.*, 2000; Lawrence *et al.*, 2002; Tomaru *et al.*, 2004b). A study that followed *H. akashiwo* infecting viruses over the course of two blooms using most probable number assays revealed the number of infectious viruses increased at bloom termination. Isolated clonal hosts have different susceptibilities to isolated clonal viruses from the time of isolation as well as at different times in the bloom (Tarutani *et al.*, 2000; Tomaru *et al.*, 2004b). Differences in susceptibility of infection

have been seen by challenging different strains of *H. akashiwo* to those viruses that have been characterized in culture as well (Nagasaki and Yamaguchi, 1997; Lawrence *et al.*, 2001; Tai *et al.*, 2003). These studies lead to the hypotheses that the dominant strain of *H. akashiwo* differs from year to year (Tomaru *et al.*, 2004b), and the dominant strain varies over the course of the bloom (Tarutani *et al.*, 2000; Tomaru *et al.*, 2004b), in part due to viral activity (Tarutani *et al.*, 2000). Unlike bloom waters which appear to be dominated by larger (>200 nm) viruses (Nagasaki *et al.*, 1994; Tomaru *et al.*, 2004b), the sediments appear to be dominated by smaller (~50 nm) viruses (Lawrence *et al.*, 2002). Viruses capable of lysing *H. akashiwo* were detected in a majority of sites along the coast of British Columbia in concentrations > 3000 infectious particles per cm^{-3}. Although the number of infectious particles decreased the further from the water sediment interface, and it was predicted that > 99% would be inactive in 6 months, these viruses could provide seed populations for when *H. akashiwo* cysts become metabolically active again (Lawrence *et al.*, 2002).

Phaeophyceae

The class *Phaeophyceae* contains multicellular brown algae that are present in coastal waters and contain kelp up to 60 feet length (*Macrocystis pyrifera*) (Cock *et al.*, 2011). Viruses have been visualized in this group of algae by transmission electron microscopy since the middle of the twentieth century (reviewed in Wilhelm *et al.*, 2017), but isolation of brown algae containing viruses occurred decades later. Viruses have been isolated infecting multiple *Ectocarpus, Laminaria,* and *Feldmannia* species, *Hincksia hinchsiae, Myriotrichia clavaeformis*, and *Pilayella litoralis* (Henry and Meints, 1992; Kapp *et al.*, 1997; Maier *et al.*, 1998; McKeown *et al.*, 2017). Viruses infecting *Phaeophyceae* members are unique among other algae infecting

viruses as they have been shown to integrate into the genomes of their hosts (Henry and Meints, 1992; Delaroque and Boland, 2008). Viruses can only infect the naked sexual spores / gametes, as the mature cells have a wall that viruses are unable to penetrate (Wolf *et al.*, 1998). In all phaeovirus systems described so far except for viruses infecting *Laminaria*, the infections only occur in sporangia/gametangia cells, while somatic cells continue to replicate normally (Kapp *et al.*, 1997), leading to sterile organisms (Henry and Meints, 1992). The trigger to induce the lytic infection is unknown, but it is hypothesized with work done in *H. hincksiae* to be related to the initiation of sporangia, as the nucleus begins to divide but does not divide into separate cells (Wolf *et al.*, 1998). The replication cycle continues as the nucleus disintegrates and assembly of the virus occurs in cytoplasmic factories (Wolf *et al.*, 1998). Due to the large cell sizes, it is estimated that between $2.5x10^5 – 2x10^6$ viruses are released upon lysis of the cell (Maier *et al.*, 1998; Wolf *et al.*, 1998). All of the characterized phaoviruses have double-stranded DNA genomes between 154– 320 kbps in size (Kapp *et al.*, 1997; Schroeder *et al.*, 2009), with icosahedral particle sizes around 150 nm in diameter with at least 10 proteins particle (Kapp *et al.*, 1997; Maier *et al.*, 1998). The two sequenced genomes categorize the viruses as members of the *Phycodnaviridae* and the contain integrases which are hypothesized to allow for integration into the genomes of their hosts (Schroeder *et al.*, 2009). PCR based approaches to amplify the major capsid protein of phaeoviruses found up to 65% of sporophytes in three common kelp species in the English channel were infected (McKeown *et al.*, 2017), and globally infections were found in 58.5% of *Ectocarpus spp.* sampled, and 4 of 11 kelp sampled including *M. pyrifera* (McKeown *et al.*, 2018), so it appears infections of multicellular algae are ubiquitous.

Pelagophyceae

Although pelagophytes are believed to have originated in deep ocean open waters, and uncultivated members are found throughout the open ocean (Worden *et al.*, 2012), the two members that are most well studied are found in coastal waters (Sieburth *et al.*, 1988; DeYoe *et al.*, 1997). *Aureococcus anophagefferens* and *Aureoumbra lagunensis* cause harmful brown tide blooms in shallow bays which total millions of dollars in losses annually (Gobler and Sunda, 2012). Transmission electron microscopy images of natural blooms have revealed large icosahedral viruses infecting both *A. anophagefferens* and *A. lagunensis* (Sieburth *et al.*, 1988; DeYoe *et al.*, 1997), yet only a single isolate infecting *A. anophagefferens* has been isolated and partially characterized (see below).

Haptophyta

Haptophytes are most commonly found in marine environments, and are unicellular, but can have different asexual and sexual lifecycles. Some have the ability to cause large blooms visible from space that contribute greatly to nutrient cycling of the oceans. Viruses have been isolated on many different haptophyte hosts including multiple *Phaeocystis spp.* (Jacobsen *et al.*, 1996; Baudoux and Brussaard, 2005a), *Emiliania huxleyi* (Castberg *et al.*, 2002), multiple *Chrsochromulina spp.* (Suttle and Chan, 1995; Johannessen *et al.*, 2015; Mirza *et al.*, 2015), and *Prymnesium parva* (Wagstaff *et al.*, 2017). With the exception of the Group II *Phaeocystis globosa* viruses that have ~100 nm capsids and a double-stranded DNA genome of ~175 kbp (Baudoux and Brussaard, 2005a), the haptophyte infecting viruses that have genomic information available all have genomes greater than 400 kbp (Baudoux and Brussaard, 2005a; Johannessen *et al.*, 2015; Stough *et al.*, 2019), with the largest being ~530 kbp (Johannessen *et*

al., 2015). Genomic analysis places all of these large genome containing viruses in the extended *Mimiviridae*, with the exception of the *Emiliania huxleyi* infecting viruses, that are classically characterized as *Phycodnaviridae* (Castberg *et al.*, 2002; Pagarete *et al.*, 2013), but they are phylogenetically divergent (Coy *et al.*, 2018). Most of these viruses have particles < 225 nm in diameter, (Baudoux and Brussaard, 2005a; Wagstaff *et al.*, 2017), but can exceed 300 nm (Johannessen *et al.*, 2015). These particles appear to be complex containing > 10 major and minor proteins (Baudoux and Brussaard, 2005a; Allen *et al.*, 2008), some of which are decorated with regularly repeating fibrils (Yan *et al.*, 2005).

Haptophyte infecting viruses are believed to be important in bloom termination of multiple members including *E. huxleyi* (Vardi *et al.*, 2012), *Phaeocystis pouchetii* (Larsen *et al.*, 2004), and *Phaeocystis globosa* (Brussaard *et al.*, 2005). At the onset of bloom termination, viruses infecting dominant members increase greatly in abundance (Schroeder *et al.*, 2003; Larsen *et al.*, 2004), and can be predicted to be the sole cause of mortality of some blooms (Vardi *et al.*, 2012). Lysis of these bloom formers modulates nutrient cycling. Carbon and other nutrients are then available to the heterotrophic community members causing their increase in abundance (Bratbak *et al.*, 1998), as well as an increase in the abundance of heterotrophic bacteria found within aggregates (Sheik *et al.*, 2014). Viral production alters transparent polysaccharide production and causes clumping of cells. This causes a large number of cells, and their biomass to sink (Vardi *et al.*, 2012; Laber *et al.*, 2018). Although viruses cause bloom termination, there is resistance that occurs throughout these systems. For *E. huxleyi* and *P. pouchetti*, viruses can only infect certain life stages of their hosts. In *E. huxleyi*, viruses can only lyse the diploid cells, where those in the haploid stage of the life cycle cannot be infected (Frada *et al.*, 2008). This morphological switch is hypothesized to be caused by viral infection (Frada *et*

al., 2017). *P. pouchetti* similarly are only able to be infected when cells are free-living, where those found in colonies cannot be infected (Jacobsen *et al.*, 2007), which is hypothesized to occur in *P. globosa* as well (Brussaard *et al.*, 2005). Resistance to different subpopulations of viruses has also been shown to occur in *E. huxleyi* blooms, causing succession of virus and host populations over the course of a bloom (Schroeder *et al.*, 2003), although heterogeneity can also be seen at the single cell level (Rosenwasser *et al.*, 2019).

Other isolated virus systems

Several other systems that have viruses isolated have been described but not characterized in great detail. A virus infecting *Pyraminmonas orientalis* with a large (~560 kbp) genome has been isolated (Sandaa *et al.*, 2001). It was shown that infecting specific times during a light dark cycle can vary the burst size of this virus, although a mechanism was not described (Thyrhaug *et al.*, 2002). The flagellated *Chlorodendrophyceae* genus *Tetraselmis* has had two very different types of viruses isolated. A small 31 kbp double-stranded DNA virus with a 60 nm capsid was isolated on *T. striata.* This virus encodes 33 genes and has a burst size around $4x10^4$ particles per cell (Pagarete *et al.*, 2015). A different *Tetraselmis sp.* was used to isolate the virus (TetV-1) with the largest genome known to infect a eukaryotic algae, of 668 kbp (Schvarcz and Steward, 2018). Although two thirds of the genes in TetV-1 have no putative function, it encodes a large suite of genes related to modulating fermentation pathways in its host, and encodes a hypothetical protein with closest homology to that of the other *Tetraselmis* virus (Pagarete *et al.*, 2015; Schvarcz and Steward, 2018). Finally, double-stranded RNA virus entities have been found in many red algae cultures, with the first being characterized in *Chondrus crispus* (Rousvoal *et al.*, 2016). Although no virus particles have been detected, these entities encode two proteins and are found within the

nuclear genome (Rousvoal *et al.*, 2016). Evidence for these dsRNA viruses has been seen in metatranscriptomes of red algae as well (Lachnit *et al.*, 2016), while there are TEM observations of virus particles found in other red algae species (Lee, 1971; Apt and Gibor, 1991).

Diversity of algal-virus systems requires continued exploration

Currently there is an underrepresentation of viruses isolated compared to the total number of species of algae present (Guiry, 2012; Coy *et al.*, 2018). Even from the subset of possible host virus systems, each host virus system has unique aspects to it, although some similarities exist (*i.e.* continuous reduction of photosynthetic capacity). This can be seen in the large variability in burst sizes (such as four orders of magnitude; Wolf *et al.*, 1998; Maat *et al.*, 2016), differences in lifestyles (lysogeny v. lytic, Henry and Meints, 1992), and particle and genome complexities. All of these differences may influence how the host, virus, and environment interact in natural settings. Factors such as decay rates of free virus particles (Cottrell and Suttle, 1995; Long and Short, 2016) and the influence of abiotic factors on the infection cycle are beginning to be shown to be system specific. Infection cycle dynamics for example, have been shown to be negatively influenced by infections in the dark in some systems (Baudoux and Brussaard, 2008; Thamatrakoln *et al.*, 2019), but not others (Van Etten *et al.*, 1983; Lawrence and Suttle, 2004). Novel trends may be found within the many novel systems available to isolate viruses, as evidence based on transmission electron microscopy observations (reviewed in Wilhelm *et al.*, 2017) and genomic observations of publicly available genomes is present (Blanc *et al.*, 2015; Gallot-Lavallee and Blanc, 2017). With more sequencing and infection cycle dynamics characterized, more information can be gleaned from large datasets such as the TARA Oceans

Dataset (Pesant *et al.*, 2015), and it will be possible to determine functions of unknown proteins within genomes of currently isolated viruses (Fang *et al.*, 2019).

Work in this dissertation

This work focuses on continuing to characterize the *Aureococcus anophagefferens* virus host system. The small (2-3 µm) non-motile spherical *A. anophagefferens* has caused recurrent brown tide blooms in shallow waters off the north eastern coast of the United States since the mid-1980s (Sieburth *et al.*, 1988). Since the initial characterization, brown tide blooms have been found in distinct locations including China (Zhang *et al.*, 2012), South Africa (Probyn *et al.*, 2001), and have spread down the eastern coast of the United States (Popels *et al.*, 2003). This spread is hypothesized to be due in part to transport in ballast water on ships (Doblin *et al.*, 2004). In the summer months, *A. anophagefferens* can obtain cell densities over $2x10^6$ cells per mL (Gastrich *et al.*, 2002). There is no production of compounds that are toxic to humans, but these blooms cause millions of dollars in losses due to the high cell densities which cause intense shading. This decrease in irradiance (less than 10 µmol photons m^{-2} s^{-1}) destroys seagrass beds which in turn affect fisheries. Blooms also negatively influence and are potentially toxic to bivalves and have caused the shellfish industry of New York State to collapse (Bricelj and Lonsdale, 1997; Gobler and Sunda, 2012). *A. anophagefferens* is believed to outcompete other photosynthetic organisms during summer months due to its ability to grow in low irradiance conditions (Milligan and Cosper, 1997), a reduced grazing pressure (Sieracki *et al.*, 2004), and an ability to utilize organic nitrogen and carbon sources (Berg *et al.*, 1997; Berg *et al.*, 2002). Its ability to survive low irradiance and even persist in no light for long (1 month) durations of time allows it to persist when cell densities are high (Popels and Hutchins, 2002; Popels *et al.*, 2007).

A. anophagefferens can utilize urea and other organic nitrogen sources as its nitrogen source (Berg *et al.*, 2002; Fan *et al.*, 2003), allowing for *A. anophagefferens* to proliferate when inorganic nitrogen concentrations decrease and organic nitrogen concentration increase (Gobler and Sunda, 2012). This allows for *A. anophagefferens* populations to be resistant to nitrogen limitations when the rest of the eukaryotic community is limited (Wurch *et al.*, 2011; Wurch *et al.*, 2019). These physiological features are mirrored in the genome. For its size, *A. anophagefferens* has a large genome (>56 MB, 11,501 putative coding sequences) enriched in genes encoding light harvesting complexes and those for organic nutrient utilization (Gobler *et al.*, 2011). Interestingly, the genome has the largest number of selenocysteine containing proteins found in any eukaryote. This causes a high selenium requirement (10 pM) for *A. anophagefferens* to grow (Gobler *et al.*, 2013).

Large icosahedral viruses have been associated with brown tide blooms since the initial characterization of the 1985 brown tide bloom (Sieburth *et al.*, 1988). Transmission electron microscopy studies over the course of the bloom have shown populations continuously infected by large icosahedral viruses at low levels (Gastrich *et al.*, 2002; Gastrich *et al.*, 2004), but at the termination of the bloom over one third of the population was infected (Gastrich *et al.*, 2004). These studies led to the hypothesis that *A. anophagefferens* lytic viruses played a role in bloom collapse, along with other factors such as bacteria capable of lysing *A. anophagefferens* (Frazier *et al.*, 2007). An attempt to isolate viruses infecting *A. anophagefferens* occurred in the early 2000's, and one of these isolates has since been partially characterized. Viruses isolated were similar in size (~140 nm), icosahedral, and could be seen inside *A. anophagefferens* cells (Gobler *et al.*, 2007; Rowe *et al.*, 2008). A telling sign of infection is the loss of the glycocalyx covering the outside of the cell during infection, which has been seen both in the environment (Sieburth *et*

al., 1988), and laboratory studies (Gobler *et al.*, 2007; Rowe *et al.*, 2008). Of the nine clonal viruses that were isolated on the *A. anophagefferens* CCMP1984, all could only lyse specific strains of *A. anophagefferens* (Gobler *et al.*, 2007). Adding lysate from a single virus to natural populations on different days had mixed effects. Some populations decreased in abundance, while others were unchanged (Gobler *et al.*, 2007), suggesting there are subpopulations within the community with different susceptibilities, as seen with other systems (*i.e.* Tomaru *et al.*, 2007). The isolated *A. anophagefferens virus*, AaV, has an infection cycle that is around 24 hours (Rowe *et al.*, 2008; Brown and Bidle, 2014), and a burst size of a few hundred particles per cell (Brown and Bidle, 2014), in ideal conditions. Infection cycle dynamics can be modulated by decreasing temperature or irradiance levels (causing an increase in the time it takes for cultures to clear) (Gobler *et al.*, 2007), or by increasing the multiplicity of infection (reducing the number of particles produced per cell) (Brown and Bidle, 2014).

The double-stranded genome of AaV is 370,920 bp, encoding 377 coding sequences and 8 tRNAs. Phylogenetic analyses of multiple core NCLDV genes places AaV within the extended *Mimiviridae* (Moniruzzaman *et al.*, 2014). Like other members in the family (La Scola *et al.*, 2003; Fischer *et al.*, 2010; Gallot-Lavallee *et al.*, 2015), AaV has a low GC content (28.7%) (Moniruzzaman *et al.*, 2014), which is worth noting as it is 40.8% less than the GC content of *A. anophagefferens* (Gobler *et al.*, 2011). The genome contains genes for DNA replication and repair as well as transcription, to allow for some independence from the host (Moniruzzaman *et al.*, 2014). Transcriptional profiles of infected cells reveal all but 2 of the predicted coding sequences were active over the infection cycle, with viral transcripts being detected as early as 5 minutes post infection (Moniruzzaman *et al.*, 2018). The infected cells were transcriptionally distinct from healthy cells with up to 42% of genes being differentially regulated at specific

points of the infection (Moniruzzaman *et al.*, 2018). Genes involved light harvesting complexes and the selenocysteine containing proteins notably were downregulated (Moniruzzaman *et al.*, 2018). These studies informed two environmental studies of natural blooms which support that this isolated virus is environmentally relevant (Moniruzzaman *et al.*, 2016; Moniruzzaman *et al.*, 2017). First, degenerate PCR primers to amplify the algae-infecting extended *Mimiviridae* members over the course of 2011 blooms in two locations revealed amplicons 97% similar to AaV present in the third highest abundance. The amplicons that had the high abundance were 87% similar to AaV (Moniruzzaman *et al.*, 2016). Also, transcripts were detected when stringently mapped to the AaV genome of a 2014 bloom metatranscriptome (Moniruzzaman *et al.*, 2017). These studies show that viruses similar to our isolate are present over the course of the bloom, and active infections are occurring.

In this work we continued to characterize the *A. anophagefferens* host-virus system through laboratory experiments with the goal to develop tools to study this system in the environment and understand how abiotic factors influence the infection cycle. To understand to what extent viruses contribute to bloom termination, we need an understanding of how the environment, *A. anophagefferens*, and its viruses all interact. Characterizing the AaV particle provides valuable insight into the initiation of the infection as well as its ability to remain infectious in the environment. Also, this work provides insight into novel functions of proteins that were previously uncharacterized (Fang *et al.*, 2019). Although electron microscopy studies have been used to show natural populations are infected (Gastrich *et al.*, 2002; Gastrich *et al.*, 2004), and total virus counts do not change drastically (Gastrich *et al.*, 2004; Gobler *et al.*, 2007), the dynamics of *A. anophagefferens* infecting virus populations is not known. As in other bloom systems the concentration of viruses infecting blooming organisms changes (Schroeder *et*

al., 2003; Tomaru *et al.*, 2007), so methods to enumerate natural populations are needed. Abiotic factors change considerably over the summer months in these bays (Gobler *et al.*, 2004; Wurch *et al.*, 2019), which can alter the success of the viral infection and bloom collapse. To understand these environmental influences, laboratory experiments were conducted to determine changes in infection cycle dynamics. This work contributes to the growing body of research on this environmentally relevant system and provides new tools to study the viruses infecting *A. anophagefferens* in the environment.

Chapter Two

Structural and proteomic studies of the Aureococcus anophagefferens Virus demonstrate a global distribution of virus encoded carbohydrate processing

Author contributions

This study was developed by Eric. R. Gann, Todd B. Reynolds, Chuan Xiao, and Steven W. Wilhelm. Eric R. Gann generated and concentrated all the viruses used for both microscopy and proteomics. Yeujiao Xian and Chuan Xiao performed the cryo-electron microscopy and reconstructions of the particle. Paul E. Abraham and Robert L. Hettich performed the shotgun proteomics. Bioinformatic analyses were performed by Eric Gann. Eric Gann and Steven Wilhelm wrote the chapter with all authors participating in providing comments in the initial draft and the revision.

*Eric R. Gann and Yeujiao Xian contributed equally to this work.

Abstract

Nucleocytoplasmic large double-stranded DNA viruses (NCLDVs) possess the most complex genomes of known viruses and infect a diverse range of eukaryotic hosts. Some of these viruses play important roles in modulating environmentally relevant populations, such as the harmful bloom forming alga *Aureococcus anophagefferens*. Much is known about the infection cycle of the isolated Aureococcus anophagefferens Virus**,** AaV, but not about the viral particle itself. To study the particle, cryo-electron microscopy was used to reconstruct particle structure while proteomics was performed to determine which proteins are packaged. A 9.5 Å resolution reconstruction of the 1900 Å diameter particle was produced from 570 individual particle movies. The reconstruction demonstrated that AaV possesses a subcapsid organization similar to other large icosahedral viruses determined thus far. In this study, no decorations on the external surface or unique vertices were detected, although there are two pockets between the internal electron dense core and the capsid protein on opposing vertices. Forty-three virus-encoded proteins were detected within purified particles. Functions of these proteins were consistent with some found in other virus encapsulated proteins, such as transcriptional machinery. However, other proteins unique to AaV, such as sugar degradation and binding proteins, were observed and quantified. Packaged proteins were used to determine similarities and differences between members of different NCLDV families: a spectrum of different shared packaged proteins from few (Phycodnaviridae) to many (Pandoraviridae) was observed. Packaged proteins that were shared between viral families were enriched in processes shared by all viruses regardless of host (*i.e.*, transcriptional machinery and redox stress modulating proteins), providing further support that the structure, assembly, and what is packaged within large viruses is evolutionarily conserved. Finally, to determine the extent to which packaged proteins provide metabolic

capabilities to AaV and similar particles, an examination of the TARA Oceans data set revealed that genes of viral origin were a significant component of the carbohydrate lyase genes in the TARA Oceans database. In total, our data demonstrate that the AaV particle itself appears to have prepackaged metabolic capabilities independent of its host. Moreover, our work demonstrates that genes associated with carbohydrate processing by NCLDVs are widespread and abundant in the global oceans.

Introduction

The most biologically complex virus particles are those belonging to the Nucleocytoplasmic Cytoplasmic Large dsDNA Virus (NCLDV) group, which have genomes and particle sizes that rival bacteria in size (Wilhelm *et al.*, 2017). These viruses have been isolated from numerous locations (Legendre *et al.*, 2015, Abrahao *et al.*, 2018), are abundant in the world's oceans (Hingamp *et al.*, 2013, Moniruzzaman *et al.*, 2017), and have been shown to be important in constraining productivity in some ecosystems (Vardi *et al.*, 2012), suggesting this life style strategy is ubiquitous and effective. The genetic repertoire associated with these large genomes allows for more independence from the host's transcriptional (Mutsafi *et al.*, 2010) and translational (Abrahao *et al.*, 2018) machinery during the infection, while the complexity of the virion particle allows for the packaging of functional proteins (Dunigan *et al.*, 2012, Fischer *et al.*, 2014), important in the initial stages of the infection (Zhang *et al.*, 2011, Milrot *et al.*, 2017).

Some NCLDV members play important roles in algal bloom collapse (Gastrich *et al.*, 2004, Vardi *et al.*, 2012). The eukaryotic alga *Aureococcus anophagefferens* causes recurrent algal blooms off the eastern coast of the United States, negatively impacting the economy by influencing tourism and devastating fisheries (Gobler & Sunda, 2012). Natural populations have been shown to be infected by large, icosahedral viruses, with an increased prevalence of infected cells at bloom collapse (Sieburth *et al.*, 1988, Gastrich *et al.*, 2004). The lab model for this system, the Aureococcus anophagefferens virus (AaV), belongs to the extended Mimiviridae family and has a 371 kbp genome that encodes 377 proteins and 8 tRNAs (Moniruzzaman *et al.*, 2014). Many of these proteins encoded by this virus particles have been described as being more consistent with "cellular" life forms than viruses. Previous transcriptomic analyses performed over the course of the infection cycle show transcripts for all but two of these proteins are

expressed over the 21 h infection cycle (Moniruzzaman *et al.*, 2018); however, not much is known about the initial, critical moments of the infection cycle and the role of proteins of virus-particle *vs* host-cell origin. Indeed, given the abundance of these viruses in nature yet paucity of NCLDV model systems and information on their physiological ecology (Needham *et al.*, 2019), complete characterization of these lab models to infer environmental function is a necessity. This is particularly true for the virus particle and its potential to have metabolism outside of the host.

Here we focused specifically on the AaV particle, first reconstructing its capsid structure by cryo-electron microscopy (Cryo-EM) and then using proteomics to determine the protein complement of purified particles. We used this opportunity to compare and contrast both the structure and protein ensemble to other giant viruses to further understand similarities and differences that have occurred since evolution from a common ancestor. Finally, we examined the distribution of one of the packaged functional proteins, a carbohydrate lyase, to approximate virus-particle derived metabolic capabilities (*i.e.* polysaccharide degrading) in the world's oceans. While research most commonly examines the role of viruses in shaping the biogeochemical potential of infected host cells, our work demonstrates that virus particles by themselves may have the unique potential to shape processes in marine systems.

Methods

AaV propagation and concentration

Non-axenic *Aureococcus anophagefferens* CCMP1984 was grown in a modified $ASP_{12}A$ (Gann, 2016), at 19º C with a 14:10 light-dark cycle that included an irradiance level of 90 µmol photons m^{-2} s^{-1}. Cultures were infected with fresh AaV lysate from which cellular debris was removed by a 0.45-µm pore-size syringe filter (Millex-HV 0.45µm nominal pore-size PVDF,

Millipore Sigma, USA). Lysates were pooled (> 80 L) and concentrated by tangential flow filtration (Suttle *et al.*, 1991). First, lysate was concentrated 50x using a Proflux M12 Tangential Flow Filtration System (MilliporeSigma, USA) equipped with a 30 kDa Helicon S10 Ultrafiltration Spiral Cartridge Filter (MilliporeSigma, USA) and then further concentrated using a Lab-scale Tangential Flow Filtration System (Fisher Scientific, USA) equipped with a Durapore 30kDa Pellicon XL Filter (MilliporeSigma, USA). Bacteria were removed by pelleting 50 mL concentrates in serial centrifugations (5000 x g, 5 min) in a Sorvall Lynx 4000 Centrifuge (Thermo Fisher Scientific, USA) with a Fiberlite F14-14x50cy rotor (Thermo Fisher Scientific, USA), and then by filtration using a 0.45 µm pore-size syringe filter (Millex-HV 0.45µm nominal pore-size PVDF, Millipore Sigma, USA). Samples for electron microscopy were stored at 4º C in a final concentration of 1% Triton X-100, while those for proteomics were stored at -20º C without detergent.

Electron microscopy

Purified AaV samples were divided into 1 mL aliquots in 1.5 mL microcentrifuge tubes and centrifuged at 1000 x g for 1 h. To remove the detergent effect, AaV pellets were collected and slowly diluted 200x with Tris-Base buffer A (50 mM TRIS, 500 mM NaCl, 5 mM $MgCl_2$, pH 8.0) using a Fusion 200 syringe pump (Chemyx, USA) at a flow rate of 0.02 mL/min with the sample tube shaking. Next, the diluted AaV sample was concentrated by centrifugation at 1,000 x g for 1 h in 1 mL aliquots and again slowly diluted with Tris-Base buffer B (50 mM TRIS, 5 mM $MgCl_2$, pH 8.0) to reach a final salt concentration of 250 mM. Finally, the AaV sample was concentrated to a volume less than 10 µL by centrifugation. Cryo-EM specimens were prepared by adding 3.5 µL of the final AaV sample on a Quantifoil R 3.5/1 grids (Quantifoil Micro Tools

GmbH, Germany), blotting manually and vitrified using a guillotine style plunging device into liquid ethane. Cryo-EM data was collected at the NIH-funded regional cryo-EM consortia at UCLA and Stanford SLAC on Titan Krios G3 microscopes (Thermo Fischer Scientific, USA) equipped with Gatan BioQuantum Energy filter and K2 detector in EFTEM mode with nanoprobe. The reported magnification was 105,000x, resulting in pixel sizes of 1.36 and 1.34 Å/Pixel in the collected movies from UCLA and Stanford SLAC, respectively. Movies were collected for 7 sec at 5 frame/sec. The total dose level was at approximately 30 e-/$Å^2$. The movies were motion corrected and dose weighted using MotionCor2 (Zheng *et al.*, 2017). A total of 570 particles were boxed from the micrographs using e2boxer.py (Tang *et al.*, 2007) and used for particle reconstruction by the program FREALIGN (Grigorieff, 2007). The resolution of the reconstruction was 9.5Å determined using Fourier Shell correlation (FSC) with threshold of 0.333.

Protein extraction and digestion

Concentrated lysates (4 mL) were further concentrated using 10 kDa molecular weight spin columns (Vivaspin 2, GE Health, USA), and resuspended in sodium deoxycholate (SDC) buffer (2% in 100 mM of NH_4HCO_3, 10 mM dithiothreitol). Cysteines were blocked by adjusting each sample to 30 mM iodoacetamide and incubated in the dark for 15 min at room temperature. Proteins were digested *via* two aliquots of sequencing-grade trypsin (1:75 [w:w], Promega, USA), first overnight followed by a 3 h incubation at 37° C. The peptide flow through was collected *via* centrifugation and the sample adjusted to 1% formic acid to precipitate SDC. Hydrated ethyl acetate was added to each sample at a 1:1 [v:v] ratio three times to effectively remove SDC. Samples were then placed in a SpeedVac Concentrator (Thermo Fischer Scientific, USA) to

remove ethyl acetate and further concentrate the sample. The peptide-enriched flow through was quantified by BCA assay (Pierce Biotechnology, USA), desalted on RP-C18 stage tips (Pierce Biotechnology, USA) and then stored at −80° C.

Protein identification and quantitation

Peptide mixtures were analyzed using an established high-performance peptide sequencing approach (Villalobos Solis *et al.*, 2019), on a Q Exactive Plus mass spectrometer (Thermo Fischer Scientific, USA) coupled with a with a Proxeon EASY-nLC 1200 liquid chromatography (LC) pump (Thermo Fisher Scientific, USA). In brief, peptides were separated on a 75 μm inner diameter microcapillary column packed with 40 cm of Kinetex C18 resin (1.7 μm, 100 Å, Phenomenex). For each sample, a 2 μg aliquot (as determined by BCA assay – above) was loaded in buffer A (0.1% formic acid, 2% acetonitrile) for 10 min, eluted with a linear 90 min gradient of 2 – 20% of buffer B (0.1% formic acid, 80% acetonitrile), followed by an increase in buffer B to 30% for 10 min, another increase to 50% buffer for 10 min, and concluding with a 10 min wash at 98% buffer A. The flow rate was kept at 200 nL min^{-1}. MS data was acquired with the Thermo Xcalibur software version 4.27.19 (Thermo Fischer Scientific, USA) using the data-depending acquisition top10 method where up to 10 different precursor m/z values can be triggered for fragmentation per full scan. Target values for the full scan MS spectra were 1 x 10^6 charges in the 300 – 1,500 m/z range with a maximum injection time of 25 ms. Transient times corresponding to a resolution of 70,000 at m/z 200 were chosen. A 1.6 m/z isolation window and fragmentation of precursor ions was performed by higher-energy C-trap dissociation (HCD) with a normalized collision energy of 27. MS/MS sans were performed at a resolution of 17,500 at m/z 200 with an

ion target value of 1 x 10^5 and a maximum injection time of 50 ms. Dynamic exclusion was set to 20 s to avoid repeated sequencing of peptides.

MS raw data files were processed using the commercial software Proteome Discoverer v2.2 (Thermo Fischer Scientific, USA). Raw data files were searched against both the *Aureococcus anophagefferens* (Gobler *et al.*, 2011) and AaV (Moniruzzaman *et al.*, 2014) reference proteome to which common contaminate proteins had been added. A decoy database, consisting of the reversed sequences of the target database, was appended to discern the false-discovery rate (FDR) at the spectral level. For standard database searching, the peptide fragmentation spectra (MS/MS) were analyzed by Proteome Discoverer v2.2 (Thermo Fischer Scientific, USA). The MS/MS were searched using the MS Amanda v2.0 (Dorfer *et al.*, 2014), and was configured to derive fully tryptic peptides using settings for high-high MS/MS data: MS1 mass tolerance of 10 ppm and MS2 mass tolerance of 0.02 Da. A static modification on cysteines (iodoacetamide; +57.0214 Da) and a dynamic modification on methionine (oxidation; 15.9949) were considered. The results were processed by Percolator (The *et al.*, 2016), to estimate *q* values. Peptide spectrum matches (PSMs) and peptides were considered identified at a *q* value <0.01.

For label-free quantification, MS1-level precursor intensities (area) were derived from the Minora Feature and Precursor ions quantifier nodes using default parameters. Missing values were imputed (Low abundance resampling method) and proteins were normalized by the total peptide amount using Proteome Discoverer (Thermo Fischer Scientific, USA).

Bioinformatic comparison of packaged proteins within different viral families

To compare the protein complement of AaV with other NCLDV members, viral protein sequences and locations of coding regions within the genomes from NCBI were downloaded

(Brister *et al.*, 2015). Viruses with particle proteomes used in this study included Acanthamoeba polyphaga Mimivirus (APMV) (Renesto *et al.*, 2006), Cafeteria roenbergensis Virus (CroV) (Fischer *et al.*, 2014), Emiliania huxleyi Virus 86 (EhV) (Allen *et al.*, 2008), Marseillevirus (Boyer *et al.*, 2009), Melbournevirus (Okamoto *et al.*, 2018), Pithovirus sibericum (Legendre *et al.*, 2014), Mollivirus sibericum (Legendre *et al.*, 2015), Pandorvirus dulcis, Pandorvirus salinus (Legendre *et al.*, 2018), Paramecium bursaria Chlorella Virus 1 (PBCV1) (Dunigan *et al.*, 2012), and Tupanvirus soda lake (Abrahao *et al.*, 2018). Packaged proteins were compared to one another by BLASTp using command line BLAST version 2.8.1+ (Camacho *et al.*, 2009). Reciprocal best BLAST hit (RBH) pairs were determined using bit-score with a maximum e-value cutoff $< 1 \times 10^{-15}$. We focused on RBH pairs where proteins were found to be packaged by both viruses. RBH pairs were grouped based on their Nucleo-cytoplasmic virus orthologous groups (NCVOGs) (Yutin *et al.*, 2009). Clustering of viruses based on the presence/absence of RBH pairs was performed using Bray-Curtis similarity in Primer version 7 (Clarke, 2015). RBH pairs were visualized using Circos (Krzywinski *et al.*, 2009). Statistical analyses were performed using Prism 7.03 (GraphPad, USA).

Environmental relevance of carbohydrate lyase proteins within the AaV particle

To determining the environmental relevance of genes encoding packaged lyases, we used eight publicly available meta-transcriptomes from a 2011 *A. anophagefferens* bloom in Quantuck Bay, New York (Moniruzzaman *et al.*, 2017, Wurch *et al.*, 2019) (Table 2.1) (All figures and tables are located in the Chapter Two appendix). Quality filtered and trimmed reads were mapped to

the AaV genome (0.9 identity fraction, 0.9 length fraction) using CLC Workbench version 12.0 (Qiagen, Germany). Reads that could be mapped to multiple contigs were discarded.

We also explored data from the TARA Oceans expedition to understand the global diversity of polysaccharide lyases encoded by viruses, globally (Pesant *et al.*, 2015). All the assembled TARA Oceans metagenomic assembled contigs (Table 2.1) were screened for polysaccharide lyases by BLASTx using command line BLAST version 2.8.1+ (Camacho *et al.*, 2009), using an e-value cutoff < $1x10^{-10}$. Only biochemically characterized polysaccharide lyases from the Carbohydrate-Active Enzyme (CAZy) Database (Lombard *et al.*, 2014), and those found within the *A. anophagefferens* and AaV genomes were used in our database. To determine if the contigs containing polysaccharide lyases were potentially viral, contigs with polysaccharide lyases were subsequently screened for Nucleo-cytoplasmic virus orthologous groups (NCVOGs) (Yutin *et al.*, 2009), using BLASTx. A contig was termed viral if it had NCVOG hit with an e-value cutoff < $1x10^{-10}$, with the best BLAST hit from the NCVOG BLAST used to assign the contigs putative family. The portion of each of these contigs containing the polysaccharide lyase was using a python script (Gann *et al.*, 2019), and then translated into their amino acid sequence. Polysaccharide lyases from viral contigs which were >80% the length of the smallest biochemically characterized polysaccharide lyase (Hyun-Kyoung *et al.*, 2007), were placed on a reference maximum likelihood phylogenetic base tree using pplacer (Matsen *et al.*, 2010). This reference maximum likelihood tree was constructed using the biochemically characterized polysaccharide lyases from the Carbohydrate-Active Enzyme (CAZy) Database and those found within the *A. anophagefferens* and AaV genomes. The tree was constructed using PhyML 3.0 with the LG model, gamma shape parameter and frequency type estimated from the data. aLRT-SH like statistic was calculated for branch support

(Guindon *et al.*, 2010). Phylogenetic trees were also constructed for the NCVOGs present on some of the contigs to the AaV genome using PhyML with parameters described above. To determine the relative abundance of contigs containing these pectate lyases and their activity a single station (Station 122) was used. Reads from the metagenomes (Table 2.1) for which the contigs were assembled were quality filtered and trimmed before being mapped to the entire contig (0.9 identity fraction, 0.9 length fraction) using CLC Workbench version 12.0 (Qiagen, Germany). Reads from metatranscriptomes (Table 2.1) at the same depth of the metagenomes used to assemble each contig were quality filtered and trimmed before being mapped to the polysaccharide lyase of the contig (0.9 identity fraction, 0.9 length fraction) using CLC Workbench version 12.0 (Qiagen, Germany). Reads that mapped to multiple contigs were discarded.

Results and Discussion

Structural characteristics of the AaV particle

The structure of the external surface of large icosahedral viruses is variable (Colson *et al.*, 2017). Many viruses possess fibrils (*e.g.*, Mimivirus has large (1,250 Å) fibers) and multiples of these can bind to one subunit of the capsid (capsomer) forming a densely packed network (Kuznetsov *et al.*, 2010). In other viruses, fibers are more regularly spaced on specific capsomers within the capsid (Nandhagopal *et al.*, 2002, Yan *et al.*, 2005, Yan *et al.*, 2009). External fibrils have been hypothesized to have different roles, including stabilizing the capsid after adsorption (Zhang *et al.*, 2011), and helping in viral attachment to their host and other organisms to aid in dissemination (Rodrigues *et al.*, 2015). Although fibrils are found of the surface of viruses from diverse families within the NCLDVs, lack of fibrils also occurs (Andreani *et al.*, 2017, Xiao *et*

al., 2017, Okamoto *et al.*, 2018). As shown in the micrograph, the AaV particle does not appear to have fibrils on the outside of its capsid (Figure 2.1B).

Like other icosahedral viruses, with the exception of the Faustovirus (Klose *et al.*, 2016), AaV possesses an electron dense core surrounded by a putative membrane. While the mechanism of entry of AaV into *Aureococcus* is unknown, the putative membrane surrounded core allows for the fusion of the host and virus membrane in other systems (Milrot *et al.*, 2017). Interestingly, this internal membrane does not appear to be maintained at a uniform distance from the capsid, as the two opposing poles of the particle in many images have a larger gap between the capsid and the electron dense core (Figure 2.1B). Thus, while no unique physical structures were detected, these larger pockets could serve to localize proteins essential to the initial stages of the infection. This has previously been observed in the PBCV-1 particle, where it's unique spike vertex contains a pocket below it that is hypothesized to have the enzymes required for the degradation of the cell wall of *Chlorella* NC64A (Cherrier *et al.*, 2009). To this end, it may be that an internal membrane structure is as important as surface decorations with respect to the process of infection.

Virus particle reconstruction

570 particles were used to produce a 9.5 Å-resolution reconstruction of the 1900 Å particle. At this resolution, individual capsomers were resolved with each composed of three copies of the major capsid protein forming a pseudo-hexamer (Figure 2.1A). Viruses assemble into regular icosahedrons using a repeating monomer can be described mathematically by calculating the triangulation number, T number (Caspar & Klug, 1962). The T number can be calculated by determining hexamer capsomers along paths h and k, from one pentamer to the adjacent

pentamer, using the equation $T = h^2 + hk + k^2$. With this approach, AaV has a triangulation number of 169, with h = 7 and k = 8. This is identical to the T number of PBCV-1, which has the same diameter of AaV, ~1900 Å (Fang *et al.*, 2019). The total number of pseudo-hexameric capsomers (N), can be calculated from its T-number using the equation: N = (10*(T – 1)). The AaV capsid therefore contains 1680 pseudo-hexameric capsomers, which are composed of 5040 major capsid proteins, and 12 pentameric capsomers located on the vertexes of the viral capsid, which are made of five copies of an unknown protein (Figure 2.1C). Interestingly, the value of 7 for h is conserved within large icosahedral viruses, with k increasing or decreasing based on the T number and size of individual practices (Table 2.2).

The arrangement of capsomers into distinct subcapsid structures was first described in the late sixties with a structural study of *Sericesthis* iridescent virus (Wrigley, 1969). These substructures were one, three, and five sided, and named disymmetrons, trisymmetrons, and pentasymmetrons, respectively. Structural studies since have only found evidence for trisymmetrons and pentasymmetrons, and mathematically it has been shown that all odd T numbered large icosahedrons can be made by these two symmetrons (Sinkovits & Baker, 2010). AaV possesses 20 trisymmetrons consisting of 66 hexameric capsomers which all face the same orientation, 60º rotated to the adjacent symmetron, and 12 pentasymmetrons (Figure 2.1A). The capsomer orientation within a pentasymmetron has been discussed previously (Xiao *et al.*, 2017, Okamoto *et al.*, 2018), and is no exception to the pentasymmetron size of all large icosahedral viruses at 30 hexameric capsomers and 1 pentameric capsomer (Figure 2.1). Due to the similarities between the substructures of capsids of the various large icosahedral viruses, it is believed that they are assembled in a similar manner. The proposed continuous assembly model proposes that trisymmetron is built continuously from the pentasymmetron. The orientation of

the capsomers within the trisymmetron come from a uniquely oriented capsomer with the pentasymmetron which seeds the position of all the capsomers within the trisymmetron (Xiao *et al.*, 2017). The assembly of these large capsids are hypothesized to be stabilized and guided by minor capsid proteins (Yan *et al.*, 2009, Fang *et al.*, 2019) as well as by electrostatic interactions between the capsomers (Xian *et al.*, 2019).

Proteins packaged within the AaV particle

To further understand how AaV changes the infected cell within the first five min of infection (Moniruzzaman *et al.*, 2018), we determined which proteins are packaged in the purified particle. 43 virus-encoded proteins were detected in at least one of the three proteomic samples run (14 detected in all three runs, 5 detected in only two runs) (Table 2.3). The most abundant protein detected in all three samples was the major capsid protein (AaV_096) (Table 2.4). Using the mass spectrometry intensities (Table 2.4) and the calculated number of MCP copies per virus particle from the cryo-EM reconstruction (Table 2.2), we normalized mass spectrometry intensities to estimate the number of proteins per particle (Figure 2.2, Table 2.4). According to these predictions, sampling was representative of most proteins packaged with the particle as we detected proteins that would be present in low abundances (< 10 copies per particle) (Table 2.4). For proteins found in multiple samples, the Coefficient of Variation (CV) was variable between proteins (Figure 2.2B), so these values can be used to suggest trends. It was noted during the genomic analysis of AaV that most of the genes with a predicted NCLDV origin were found in the central part of the linear genome, while many genes predicted to be acquired horizontally were found within the outer thirds and on either side (Moniruzzaman *et al.*, 2014). Genes for packaged proteins are evenly distributed across the genome, with 12, 13, and 18 of the encoded

genes are found in terminal region A, terminal region B, and the central region, respectively (Figure 2.3A). Eight of the 43 packaged proteins have reciprocal best BLAST hits of proteins packaged by other NCLDV members (e-value cutoff = 1×10^{-15}). Interestingly, 5 of the 8 are found within the central region, with two other proteins are found just outside this region (Figure 2.3A). The first detection of transcripts for genes encoding these proteins were at various times during the infection cycle (Moniruzzaman *et al.*, 2018) (Figure 2.3B). The two time points where the most packaged protein encoding genes were first transcribed were 5 min and 12 h post infection, accounting for 16/43 and 13/43 proteins, respectively (Figure 2.3B). Host proteins (56) were also detected in this analysis (Table 2.5), 20 of which were found in multiple samples (Figure 2.4A). Whether some of the proteins are packaged or contamination is unknown, but as none of the host proteins found in all three samples had a coefficient of variation < 36, it appears these may be contamination (Figure 2.4B, Table 2.6), as predicted previously (Legendre *et al.*, 2018).

Detection of other putative structural proteins

Besides the MCP, only one other putative capsid protein (AaV_247) is found in the genome of AaV (Moniruzzaman *et al.*, 2014), and that was detected in only one sample. The AaV_247 protein appears to play a very minor role in the structure, as there are only 7 copies per particle when normalized by the MCP (Table 2.4, Figure 2.2) in that single run. Cafeteria roenbergensis Virus (CroV) possesses two other minor capsid proteins predicted to be in 60 and 1-2 copies per particle. Due to the low abundance per particle, these proteins are hypothesized to be involved in the assembly of the particle instead of playing an important role in the structure of the particle (Xiao *et al.*, 2017). It is possible that the second capsid protein may be involved in assembly or

some other function that is not solely structural. This is different from PBCV-1, which encodes multiple minor capsid proteins. All of the different capsid proteins within the PBCV-1 particle are believed to contribute to the overall structure of the particle by assembling into the trisymmetrons and pentasymmetrons, instead of playing a minor role (Dunigan *et al.*, 2012).

PBCV-1 also encodes multiple minor capsid proteins that form an intensive network underneath the major capsid protein, which are hypothesized to play roles in stabilizing capsid structure. Through BLASTp searches against minor structural proteins determined to be within the PBCV-1 particle (Fang *et al.*, 2019), three putative minor capsid proteins were found within the AaV particle. AaV_328 has a BLASTp hit (Query Coverage 88%, e-value 6 x 10^{-19}) to PBCV-1 P9 protein (A407L), which is located beneath the major capsid proteins within the pentasymmetrons, and predicted to stabilize the pentasymmetrons and its association with the inner viral membrane (Fang *et al.*, 2019). AaV_214 (detected in all three runs) also had a weak BLASTp hit (Query coverage 26%, e value 1 x10^{-9}) to the minor structural protein P11, which is named the zip protein and is hypothesized to help join neighboring symmetrons (Fang *et al.*, 2019). Finally, the PBCV-1 tape measure protein P2, has weak sequence homology to AaV_232 (Query Coverage 21%, e value 0.006): the predicted tape measure protein in Mimivirus (L454) also has weak homology to AaV_232 (Query Coverage 66%, e value 4 x 10^{-8}). That said, the length of the tape measure protein is hypothesized to be associated with the size of the virus. AaV and PBCV-1 have the same sized particles (1900 Å): AaV's putative tape measure protein consists of 505 amino acids, similar to the PBCV-1 tape measure protein (P2, 576 amino acids). In comparison the Mimivirus tape measure protein is 1257 amino acids in size, consistent with the length predicted by the length of P2 (Fang *et al.*, 2019). Although not all the minor proteins

found in PBCV-1 were detected in AaV particles, these findings suggest that these minor proteins play roles in viral capsid assembly as well as stability.

Transcriptional Machinery

The only other proteins that are found in other NCLDV particles are six proteins that are predicted to be involved in transcription. Transcription related proteins make up 14% of the proteins packaged in AaV, similar to other Mimiviridae members where 10.5% (12/114), 11.1% (16/141), and 10.2% (13/127) packaged in Mimivirus, CroV, and Tupanvirus soda lake respectively (Renesto *et al.*, 2006, Fischer *et al.*, 2014, Abrahao *et al.*, 2018). Unlike Mimivirus, which replicates exclusively in the cytoplasm (Mutsafi *et al.*, 2010), and packages all the necessary subunits of RNA polymerase for transcription (Renesto *et al.*, 2006), AaV appears to be more reliant on the host. We detected only three subunits of RNA polymerase within the AaV particle (AaV_174, AaV_222, AaV_242), which is less than the other larger Mimiviridae members (Renesto *et al.*, 2006, Fischer *et al.*, 2014, Abrahao *et al.*, 2018). AaV_242 and AaV_222 are similar to the two largest and active subunits (J6R and A24R) of the Vaccinia virus RNA polymerase which is capable of transcription of the early genes *in vivo* and *in vitro* (Broyles, 2003). Other transcriptional machinery within the AaV particle included two putative RNA helicases (AaV_180, AaV_269) and a putative mRNA capping enzyme (AaV_211), which are found in the other Mimiviridae particle proteomes (Renesto *et al.*, 2006, Fischer *et al.*, 2014, Abrahao *et al.*, 2018). While how AaV infection initiation remains unknown, it is likely different than the phagocytosis seen with the other larger members of the Mimiviridae (Gonzalez & Suttle, 1993, Mutsafi *et al.*, 2010, Abrahao *et al.*, 2018), as phagotrophy has visualized in *A. anophagefferens*. This may explain why AaV does not need to package all of the components to

begin transcription. This observation may highlight an important difference between the algae-infecting "Extended Mimiviridae" and phagocytic-eukaryote infecting Mimiviridae members. Yet even without packaged transcriptional machinery (Dunigan *et al.*, 2012), virus transcripts are detected within 7 min post infection in a PBCV-1 infection (Blanc *et al.*, 2014), relying on condensed injected viral DNA being translocated into the nucleus (Milrot *et al.*, 2017). More work is needed to understand the beginning of the AaV infection cycle, as viral transcripts were detected 5 min post infection (Moniruzzaman *et al.*, 2018), suggesting the genome can also translocate to the nucleus quickly. Alternatively, some mRNA transcripts may be packaged, as this has been shown in other large dsDNA viruses (Boyer *et al.*, 2009).

Proteins unique to the AaV particle

There are 32 particle associated proteins that are not detected in other NCLDV particles. These include a putative metal dependent hydrolase (AaV_158), and a MdlB domain containing putative ABC transporter (AaV_306), both of which have not been detected in other virus particles. Interestingly these proteins have sequence homology to various families within the genomes of NCLDV members. Also present are 23 proteins with no putative function including 19 which do not have any BLAST hits to NCLDV members (AaV_052, AaV_074, AaV_080, AaV_081, AaV_091, AaV_103, AaV_107, AaV_136, AaV_144, AaV_182, AaV_225, AaV_231, AaV_250, AaV_260, AaV_279, AaV_286, AaV_304, AaV_329, AaV_382). AaV_052 and AaV_382 both contain domain of unknown function (DUF) 285 regions which are noteworthy as the DUF285 region is contained in 13.25% of the AaV coding regions and found in 90 different locations within the *A. anophagefferens* genome (Moniruzzaman *et al.*, 2014). There are genes within viruses infecting *Ostreococcus* similar to AaV_062, although it is

unknown whether these proteins are packaged by their viruses. AaV_175 and AaV_281 have BLAST hits to other NCLDV members, but with no putative function.

The final proteins packaged within the AaV package are hypothesized to be involved in sugar metabolism or sugar-binding. Four of these proteins (AaV_024, AaV_276, AaV_314, AaV_386) are Laminin G domain containing proteins. In total, AaV has 5 paralogous genes within its genome. This group of genes is unique to AaV in the NCLDV group and are believed to have been acquired horizontally from bacteria (Moniruzzaman *et al.*, 2014). Laminin G domains have been shown to interact with various polysaccharides including heparin (Aumailley, 2013). There are no other domains of known function in any of these proteins so their role in the infection cycle is still unknown. AaV also packages a putative beta-1,4 galactosyltransferase (AaV_274). Similar proteins in the chloroviruses are predicted to synthesize the unique glycans decorating the major capsid protein (Van Etten *et al.*, 2017). As it is unknown whether the major capsid protein of AaV is decorated with sugars, the function of this putative galactosyltransferase is unknown, yet AaV_274 has BLAST hits to many members of Iridoviridae and Ascoviridae, suggesting the role of this protein may not be unique. Packaged within the particle are also two proteins with putative functions involved in degrading polysaccharides not found in other NCLDV members. AaV_078 is a putative unsaturated glucuronyl hydrolase which was believed to be acquired from *A. anophagefferens*, which possesses a similar protein (AURANDRAFT_70832). Also packaged is one of three p70%utative pectate lyases found within the genome, AaV_038 (Moniruzzaman *et al.*, 2014). These enzymes may play a role in viral DNA entering *A. anophagefferens*, as has been proposed for PBCV-1 (Van Etten *et al.*, 2017), although it is unknown whether AaV injects its DNA into *A. anophagefferens* in a similar manner to PBCV1 (Milrot *et al.*, 2017). Whether or not this strategy is used by AaV is unknown,

but *A. anophagefferens* cells do lose their outer polysaccharide glycocalyx during the infection cycle with AaV in the laboratory (Rowe *et al.*, 2008). When natural populations are infected with large viruses during blooms the glycocalyx is also lost (Sieburth *et al.*, 1988), suggesting sugar metabolism is important during infection. More work to determine the composition of the glycocalyx and the mechanism by which it is degraded is needed.

Comparison of Giant Virus Proteomes

As virion associated proteins of a subset of NCLDV members have previously been determined (Figure 2.5), we examined similarities and differences of proteins packaged. Of the representative proteomes used, there is a correlation between number of proteins packaged and the diameter of the particle (Figure 2.5A, $Y = 0.086*X + 64.09$, $R^2 = 0.458$) and the genome size (Figure 2.5B, $Y = 0.054 * X + 65.53$, $R^2 = 0.4453$), but the number of packaged proteins varied from 28 (Allen *et al.*, 2008) to ~200 (Legendre *et al.*, 2018). Reciprocal best BLAST hit (RBH) pairs (between representative genomes) were determined, and RBH pairs with both virion associated proteins in the pair were examined (e-value cut off $<1 \times 10^{-15}$). Of the 1,367 proteins packaged, there were a total of 410 RBH pairs (Figure 2.6A), which were classified based on their NCVOG classifications (Table 2.7) (Yutin *et al.*, 2009). Families of viruses appear to package similar proteins (Figure 2.6C). Hierarchical clustering based on similar proteins reveal similar groupings to those determined phylogenetically (Figure 2.6B, 2.6D). Pandoraviridae members have previously been shown to package a large number of similar proteins (Legendre *et al.*, 2018), which can be seen with ~70% of the packaged proteins that are similar between *Pandoravirus salinus* and *Pandoravirus dulcis* (Figure 2.5C). This trend is seen in Marseilleviridae members as well (Fabre *et al.*, 2017) (Figure 2.5C), but not the two

Phycodnaviridae members, with the only RBH pair being the MCP. Mimiviridae members share more packaged proteins, the more similar their hosts are. *Acanthameoba* infecting Mimivirus and Tupanvirus have RBH pairs for ~50% of their packaged proteins. Although CroV infects a different phagocytic host, *Cafeteria roebergensis* (Gonzalez & Suttle, 1993), it shares ~20% of its packaged proteins with Mimivirus and Tupanvirus. This number decreases dramatically to < 10% of shared proteins between AaV and the other three members of the Mimiviridae family (Figure 2.5C, Figure 2.6C). The autotrophic *A. anophagefferens* has the ability to take up organic nutrients in the environment (Berg *et al.*, 2002), but it has never been described to have a phagocytic lifecycle. Differences between AaV and the other members might also be due to differences between Mimiviridae and the algae infecting Extended Mimiviridae, as differences have been described previously (Moniruzzaman *et al.*, 2016, Gallot-Lavallee *et al.*, 2017). These proteins suggest that viruses infecting similar hosts must overcome similar challenges within the virocell, while those that infect drastically different hosts do not.

Of the 410 RBH pairs, 119 were between viruses in different families. With the exception of Mollivirus and Pandoraviridae members sharing ~10% proteins packaged, < 8% of proteins packaged within any of the other representative viruses have RBH pairs in other families (Figure 2.5C), which mirrors the genomic analysis of Mollivirus having the best-matching homologs (15.9%) to Pandoravirus salinus (Legendre *et al.*, 2015). These two families are the only two to have packaged proteins involved in modulating translation of their hosts (Figure 2.7). Besides proteins of unknown function (Figure 2.8), the 26 RBH pairs involved in redox reactions had the most and are found in all of the representative viruses except for AaV and EhV (Figure 2.9). Surviving oxidative stress within the host cell and still being able to produce progeny is a universal task that needs to be overcome by viruses. Increasing oxidative stress increases the

amount of time for viral factory formation in Mimivirus (Yaakov *et al.*, 2019), and EhV requires H_2O_2 production within infected cells to induce cell death to produce progeny (Sheyn *et al.*, 2016). Packaging these proteins can provide an initial counter to the ROS induction that occurs during initial stages of infection. Besides MCP being shared between many viruses (Figure 2.10), the only other group of proteins with more than 10 RBH pairs between families are involved in transcription and RNA processing, including DEAD/SNF2-like helicases. These are shared only between members of the Marseilleviridae, Mimiviridae, and Pithovirus (Figure 2.11, Figure 2.12). Even if the viruses are not completely independent of the host's nucleus (Fabre *et al.*, 2017), there still needs to be a rapid reprogramming of the cell for the infection to proceed, as seen in other systems (Blanc *et al.*, 2014, Moniruzzaman *et al.*, 2018). The similarities and differences of proteins that are packaged between families may provide a glimpse of viruses evolving to survive similar or different harsh environments within their host.

Environmental relevance of packaged AaV gene products

To determine whether proteins packaged within the AaV particle are environmentally relevant we used eight metatranscriptomes (Table 2.1) of a 2011 *A. anophagefferens* bloom that have already shown the reads at a high identity (>97%) mapping to the AaV genome suggesting persistent infections of very similar viruses (Moniruzzaman *et al.*, 2017). We mapped reads less stringently (90% identify) to gain insight into potentially similar viruses infecting cells during the bloom, focusing on proteins packaged in the particle. Of the 8,296 reads that mapped to the AaV genome, 28.71% mapped to those genes which encode packaged proteins (Table 2.8). 33/43 packaged proteins had reads map to them, with 10 having more than 10 reads mapped. Of these highly detected genes, the MCP has the second highest number of reads mapping to it, while all

four of the packaged concanavalin A-like superfamily proteins (AaV_024, AaV_276, AaV_314, AaV_386) have a high number of reads mapped to them, reiterating their potential importance. The remaining five have no putative function annotated, with four (AaV_052, AaV_091, AaV_107, and AaV_206) having no hits within RefSeq.

Examining polysaccharide degrading enzymes in the TARA Oceans dataset

Besides some phage (Latka *et al.*, 2017) and chloroviruses (Van Etten *et al.*, 2017) encode genes to degrade polysaccharide polymers of their host, and AaV packages one of its three putative pectate lyases, we wanted to see how prevalent these genes were in aquatic environments using the TARA Oceans metagenomic assembled contigs (Pesant *et al.*, 2015). Previously, based on phylogeny, it was hypothesized that AaV acquired the three pectate lyases in its genome from either bacteria or *A. anophagefferens* (Moniruzzaman *et al.*, 2014). Using all the biochemically characterized polysaccharide lyases in the Carbohydrate-Active Enzyme (CAZy) Database, which degrade a wide variety of polysaccharide polymers (Figure 2.13), it appears AaV did not acquire these proteins from the host, but instead either a different eukaryote or bacteria. Although the polysaccharide lyases do not tend to cluster based on eukaryotic or prokaryotic origin, or by substrate specificity, the three proteins within the AaV genome all cluster within a group of pectate lyases (Figure 2.13). We searched all the assembled contigs from all metagenomes from the TARA Oceans dataset using these biochemically characterized proteins and found 11,055 contigs with a near full length polysaccharide lyase present (Table 2.9, Table 2.10, and Table 2.11). By searching for Nucleo-cytoplasmic virus orthologous groups (NCVOGs) also on the contigs, 499 of these contigs were classified as potentially of viral origin. Using the best BLAST hit, these contigs were assigned to eight viral families, with the majority

belonging to either Phycodnaviridae (183/499) or Mimiviridae (142/499) (Table 2.12). As NCLDVs range in size from < 0.22 to greater than 1.5 µm in diameter, using all the metagenomes was warranted. Increased confidence that these putative viral contigs were actually viral in origin stemmed from no difference in the fraction of polysaccharide lyases that were viral based on water depth (Figure 2.14A, Table 2.13, Table 2.9), but there was a decrease based on the size fraction of the water used in sample collection (Figure 2.14B, Table 2.14, Table 2.10). Placing the 499 polysaccharide lyases from contigs of viral origin onto the base tree of characterized pectate lyases (Figure 2.13) using pplacer revealed th no grouping of these contigs by assigned family (Figure 2.15). There is also no grouping based on TARA Oceans Station or water depth (data not shown). These data support the hypothesis that these genes are acquired horizontally from many different sources within and between families, and are important to many different types of viruses, potentially infecting a wide range of hosts. The ability to break down large polysaccharide polymers is found in phylogenetically distinct groups of bacteria in aquatic communities (Alderkamp *et al.*, 2007), supporting the potential for acquisition from distinct groups. Five polysaccharide lyases from viral contigs from three stations clustered with the three AaV pectate lyases (Figure 2.16). As three of the NCVOGs used to call those viral contigs were present in the AaV genome, phylogenetic trees were constructed to determine if closely related polysaccharide lyases would be from phylogenetically related viruses (Figure 2.17). Two of the three contigs cluster with AaV (Figure 2.17A and Figure 2.17B), while the other clusters between the Mimiviridae and Extended Mimiviridae (Figure 2.17C), providing support that other viruses in the Extended Mimiviridae possess polysaccharide lyases.

The distribution of these viral contigs containing polysaccharide lyases correlate (Viral lyase contigs = 0.07337*non-viral lyase contigs – 3.992, $R^2 = 0.701$) with non-viral contigs

containing polysaccharide lyases (Figure 2.18 inset). This leads to certain stations where up to 17% of the polysaccharide lyases detected in this analysis were within viral contigs (Figure 2.18). The reason for why certain stations were hotspots for polysaccharide lyase presence (viral or non-viral) is unknown, but it is possible blooms of algae could have occurred. During *Phaeocystis* blooms, high proportions of the community (up to ~33%) have been shown to contain enzymes to degrade its polysaccharide storage molecule, laminarin (Alderkamp *et al.*, 2007). Using a single station (Station 122, Figure 2.18), read mappings of the metagenomes and metatranscriptomes to those contigs were used to look at activity of the polysaccharide lyases and the abundance of the contigs. This station was chosen as it had the highest number of total polysaccharide lyase containing contigs and polysaccharide lyases on viral contigs (Table 2.11, Figure 2.18 inset). In two of the three water depths there was no significant difference between the metagenomic reads mapped back to contigs between viral and non-viral contigs containing polysaccharide lyases, while in surface waters there were slightly (but significantly) more reads mapped to viral contigs containing polysaccharide lyases (Figure 2.19A, Table 2.15). There was no significant differences in the expression between the viral and non-viral polysaccharide lyases at any water depth (Figure 2.19B, Table 2.16).

From this study and others (Latka *et al.*, 2017, Van Etten *et al.*, 2017), it is becoming apparent that enzymes to degrade large polysaccharides are not just found within the cellular fraction of communities. Virus-encoded or virocell-produced enzymes could provide another mechanism by which these large, hard to degrade, polysaccharides can remain in the microbial pool of the oceans. As not all members of the prokaryotic community can degrade high molecular weight polysaccharides (Muhlenbruch *et al.*, 2018), and members differ in their ability to uptake various dissolved organic matter (Elifantz *et al.*, 2007), enzymes being released into

the environment degrade intermediates may benefit, or stimulate the heterotrophic community. As it is hypothesized that these intermediates would only be found when there are large amounts of the substrate (Elifantz *et al.,* 2007), these virocell-produced enzymes may be especially important in bloom conditions. Coupling the viral activity of bloom collapse for algae (Gastrich *et al.,* 2004, Vardi *et al.,* 2012), with the ability of infections able to modulate extracellular polysaccharide production during blooms (Vardi *et al.,* 2012), and an increase in the production of polysaccharide lyases during infection (Figure 2.20), would provide a large amount of substrate for enzymes to aid in the uptake of carbon for the heterotrophic community during bloom collapse.

Conclusions

Although the genetic repertoire of giant viruses is vastly different within and between viral families, this study provides further support that the structure and assembly of large icosahedral viruses is evolutionarily conserved. The 1900 Å particle packages proteins that we hypothesize to be important in the degradation of the host polysaccharides and modulate the host's capacity to transcribe viral genes. In addition the virions contains several proteins which appear to be conserved in the Mimiviridae. Looking at similar proteins packaged across all the giant virus families, it appears that ubiquitous stresses on viral reproduction such as oxidative stress and transcription constrain viruses to package similar proteins even if they infect very different hosts or belong to different families.

Acknowledgments

We thank Gary LeCleir, Mohammad Moniruzzaman, Brittany Zepernick, Samantha Coy, Tim Sparer, Erik Zinser and Brad Binder for discussions about this work. This work was supported by grants from the *National Science Foundation* IOS1922958, the *Gordon & Betty Moore Foundation* (Grant EMS4971) and the *Kenneth & Blaire Mossman Endowment* to the University of Tennessee.

Chapter Two Appendix

Table 2.1. List of sequencing data used in this study. [A] Assembled contigs downloaded from European Nucleotide Archive from https://www.ebi.ac.uk/ena/about/tara-oceans-assemblies. [B] Run accessions downloaded from European Nucleotide Archive. [C] Run name from NCBI Short Reads Archive.

TARA Station 122 Assembled Contigs[A]							TARA Station 122 – Metagenomes[B]
CENF01	CEOO01	CEPX01	CERG01	CESP01	CETY01	CEVH01	ERR594284
CENG01	CEOP01	CEPY01	CERH01	CESQ01	CETZ01	CEVI01	ERR594301
CENH01	CEOQ01	CEPZ01	CERI01	CESR01	CEUA01	CEVJ01	ERR594304
CENI01	CEOR01	CEQA01	CERJ01	CESS01	CEUB01	CEVK01	ERR594305
CENJ01	CEOS01	CEQB01	CERK01	CEST01	CEUC01	CEVL01	ERR598948
CENK01	CEOT01	CEQC01	CERL01	CESU01	CEUD01	CEVM01	ERR594292
CENL01	CEOU01	CEQD01	CERM01	CESV01	CEUE01	CEVN01	ERR594322
CENM01	CEOV01	CEQE01	CERN01	CESW01	CEUF01	CEVO01	ERR594307
CENN01	CEOW01	CEQF01	CERO01	CESX01	CEUG01	CEVP01	ERR598999
CENO01	CEOX01	CEQG01	CERP01	CESY01	CEUH01	CEVQ01	ERR594306
CENP01	CEOY01	CEQH01	CERQ01	CESZ01	CEUI01	CEVR01	ERR598992
CENQ01	CEOZ01	CEQI01	CERR01	CETA01	CEUJ01	CEVS01	ERR594309
CENR01	CEPA01	CEQJ01	CERS01	CETB01	CEUK01	CEVT01	
CENS01	CEPB01	CEQK01	CERT01	CETC01	CEUL01	CEVU01	
CENT01	CEPC01	CEQL01	CERU01	CETD01	CEUM01	CEVV01	TARA Station 122 – Metatranscriptomes[B]
CENU01	CEPD01	CEQM01	CERV01	CETE01	CEUN01	CEVW01	ERR3587201
CENV01	CEPE01	CEQN01	CERW01	CETF01	CEUO01	CEVX01	ERR3587197
CENW01	CEPF01	CEQO01	CERX01	CETG01	CEUP01	CEVY01	ERR3587126
CENX01	CEPG01	CEQP01	CERY01	CETH01	CEUQ01	CEVZ01	ERR3587160
CENY01	CEPH01	CEQQ01	CERZ01	CETI01	CEUR01	CEWA01	ERR3587125
CENZ01	CEPI01	CEQR01	CESA01	CETJ01	CEUS01	CEWB01	ERR3587151
CEOA01	CEPJ01	CEQS01	CESB01	CETK01	CEUT01	CEWC01	ERR3587172
	CEPK01	CEQT01	CESC01	CETL01	CEUU01	CEWE01	

Table 2.1. Continued.

							2011 Bloom Metatranscriptomes[C]
CEOC01	CEPL01	CEQU01	CESD01	CETM01	CEUV01	CEWF01	
CEOD01	CEPM01	CEQV01	CESE01	CETN01	CEUW01	CEWG01	SRR3322145
CEOE01	CEPN01	CEQW01	CESF01	CETO01	CEUX01	CEWH01	SRR3322387
CEOF01	CEPO01	CEQX01	CESG01	CETP01	CEUY01	CEWI01	SRR3322113
CEOG01	CEPP01	CEQY01	CESH01	CETQ01	CEUZ01	CEWJ01	SRR3322397
CEOH01	CEPQ01	CEQZ01	CESI01	CETR01	CEVA01	CEWK01	SRR3322338
CEOI01	CEPR01	CERA01	CESJ01	CETS01	CEVB01	CEWO01	SRR3322357
CEOJ01	CEPS01	CERB01	CESK01	CETT01	CEVC01	CEWP01	SRR3322422
CEOK01	CEPT01	CERC01	CESL01	CETU01	CEVD01	CEWQ01	SRR3322410
CEOL01	CEPU01	CERD01	CESM01	CETV01	CEVE01	CEWR01	
CEOM01	CEPV01	CERE01	CESN01	CETW01	CEVF01	CXWF01	
CEON01	CEPW01	CERF01	CESO01	CETX01	CEVG01		

Table 2.2. Characteristics of structural studies of large icosahedral viruses.

	AaV	PBCV1	PpV-01	Melbournevirus	Pacmanvirus	Faustovirus	CroV
Diameter (nm)	190	190	220	232	250	260	300
h	7	7	7	7	7	7	7
k	8	8	10	13	13	12	18
T number (T)	169	169	219	309	309	277	499
Capsomers	1680	1680	2180	3080	3080	2760	4980
MCP Copies	5040	5040	6540	9240	9240	8280	14940
Citation	This Study	(Fang *et al.*, 2019)	(Yan *et al.*, 2005)	(Okamoto *et al.*, 2018)	(Andreani *et al.*, 2017)	(Klose *et al.*, 2016)	(Xiao *et al.*, 2017)

Table 2.3. Characteristics of the 43 proteins packaged within the AaV particle. [A] Region of AaV is based on third described during the initial characterization of the genome (Moniruzzaman *et al.*, 2014). [B] Data from the published transcriptome of the infection cycle (Moniruzzaman *et al.*, 2018). [C] BLAST results against the minor structural proteins determined by the reconstruction of the PBCV-1 particle (Fang *et al.*, 2019).

Accession Number	Gene Name	Description	Region of AaV[A]	MW (kDa)	First Detected in transcriptome	Number of Samples Found in
Structural						
YP_009052173.1	AaV_096	Capsid Protein	Terminal region A	52.121	5 min	3
YP_009052321.1	AaV_247	putative capsid protein 2	Central Region	64.784	12 h	1
Transcription						
YP_009052248.1	AaV_174	DNA directed RNA polymerase K subunit	Central Region	9.377	6 h	1
YP_009052254.1	AaV_180	putative VV D6R-type helicase	Central Region	86.779	12 h	1
YP_009052285.1	AaV_211	putative mRNA capping enzyme	Central Region	107.644	5 min	1
YP_009052296.1	AaV_222	putative RNA polymerase beta subunit	Central Region	126.625	12 h	1
YP_009052316.1	AaV_242	putative DNA directed RNA polymerase II largest subunit	Central Region	153.787	6 h	1
YP_009052343.1	AaV_269	putative ATP-dependent RNA helicase	Terminal region B	110.004	5 min	1
Polysaccharide Degrading/Binding						
YP_009052116.1	AaV_038	pectate lyase	Terminal region A	43.978	6 h	3
YP_009052102.1	AaV_024	Concanavalin A-like lectin/glucanase superfamily	Terminal region A	231.595	5 min	3
YP_009052156.1	AaV_078	putative unsaturated glucuronyl hydrolase	Terminal region A	55.443	5 min	2
YP_009052348.1	AaV_274	putative beta-1,4 galactosyltranferase	Terminal region B	93.99	5 min	1
YP_009052350.1	AaV_276	Concanavalin A-like lectin/glucanase superfamily	Terminal region B	45.07	6 h	3
YP_009052454.1	AaV_386	Concanavalin A-like lectin/glucanase superfamily	Terminal region B	60.551	12 h	3
YP_009052388.1	AaV_314	Concanavalin A-like lectin/glucanase superfamily	Terminal region B	34.009	6 h	3
Other						
YP_009052233.1	AaV_158	metal dependent hydrolase	Central Region	22.941	12 h	2
YP_009052355.1	AaV_281	putative membrane protein	Terminal region B	20.857	12 h	1
YP_009052380.1	AaV_306	putative ABC transporter family protein	Terminal region B	61.006	5 min	1

Table 2.3. Continued.

Hypothetical						
YP_009052130.1	AaV_052	Hypothetical (DUF285 containing)	Terminal region A	83.579	5 min	3
YP_009052140.1	AaV_062	Hypothetical	Terminal region A	20.844	5 min	3
YP_009052152.1	AaV_074	Hypothetical	Terminal region A	11.423	12 h	1
YP_009052158.1	AaV_080	Hypothetical	Terminal region A	9.594	5 min	3
YP_009052159.1	AaV_081	Hypothetical	Terminal region A	10.239	1 h	1
YP_009052169.1	AaV_091	Hypothetical	Terminal region A	42.516	5 min	2
YP_009052180.1	AaV_103	Hypothetical	Terminal region A	44.076	1 h	1
YP_009052184.1	AaV_107	Hypothetical	Terminal region A	42.877	5 min	1
YP_009052212.1	AaV_136	Hypothetical	Central Region	45.714	12 h	3
YP_009052220.1	AaV_144	Hypothetical	Central Region	30.236	30 min	1
YP_009052241.1	AaV_166	Hypothetical	Central Region	16.066	12 h	1
YP_009052249.1	AaV_175	Hypothetical	Central Region	34.298	30 min	3
YP_009052256.1	AaV_182	Hypothetical	Central Region	13.83	12 h	1
YP_009052288.1	AaV_214	Hypothetical (weak BLAST hit e-9 to P11, zip protein)[C]	Central Region	22.997	30 min	3
YP_009052299.1	AaV_225	Hypothetical	Central Region	20.592	5 min	2
YP_009052305.1	AaV_231	Hypothetical	Central Region	15.689	21 h	3
YP_009052306.1	AaV_232	Hypothetical (weak BLAST hit to P2, tape measure)[C]	Central Region	57.378	1 h	3
YP_009052324.1	AaV_250	Hypothetical	Central Region	17.491	12 h	1
YP_009052334.1	AaV_260	Hypothetical	Central Region	64.878	30 min	1
YP_009052353.1	AaV_279	Hypothetical	Terminal region B	28.328	12 h	1
YP_009052360.1	AaV_286	Hypothetical	Terminal region B	183.451	5 min	1
YP_009052378.1	AaV_304	Hypothetical	Terminal region B	36.271	5 min	1
YP_009052402.1	AaV_328	Hypothetical (BLAST hit to P9, stabilizes pentasymmetrons)[C]	Terminal region B	20.435	12 h	1
YP_009052403.1	AaV_329	Hypothetical	Terminal region B	116.31	5 min	1
YP_009052450.1	AaV_382	Hypothetical (DUF285 containing)	Terminal region B	35.237	5 h	1

Table 2.4. Mass spectrometry intensities and normalized intensities to MCP for proteins per particle for all the viral proteins detected within the proteomes.

Accession Number	Gene Name	Average Sample 1	Average Sample 2	Average Sample 3	Proteins particle^{-1} Sample 1	Proteins particle^{-1} Sample 2	Proteins particle^{-1} Sample 3
YP_009052173.1	AaV_096	6.74E+08	8.15E+08	1.05E+09	5040	5040	5040
YP_009052321.1	AaV_247	955964.7			7.153372		
YP_009052248.1	AaV_174	831595.6			6.222732		
YP_009052254.1	AaV_180	2019714			15.11328		
YP_009052285.1	AaV_211	2475735			18.52564		
YP_009052296.1	AaV_222	1324963			9.914541		
YP_009052316.1	AaV_242	3008596			22.51297		
YP_009052343.1	AaV_269	835972.7			6.255486		
YP_009052102.1	AaV_024	1.24E+08	1.83E+08	3.19E+08	929.4748	1135.224	1526.493
YP_009052116.1	AaV_038	15281427	34508101	26535657	114.3491	213.5093	126.9752
YP_009052156.1	AaV_078		Cannot quant.	Cannot quant.			
YP_009052348.1	AaV_274	1198721			8.969886		
YP_009052350.1	AaV_276	95157139	Cannot quant.	Cannot quant.	712.0497		
YP_009052388.1	AaV_314	73900155	17179584	13631325	552.9862	106.2939	65.22694
YP_009052454.1	AaV_386	1.41E+08	90890892	1.06E+08	1057.345	562.3623	507.5595
YP_009052233.1	AaV_158		3948526	385701.2		24.43041	1.84561
YP_009052355.1	AaV_281	3021310			22.60811		
YP_009052380.1	AaV_306	4644638			34.75529		
YP_009052130.1	AaV_052	31986084	9734407	4370045	239.3481	60.22895	20.911
YP_009052140.1	AaV_062	1026295	5004689	5026787	7.679648	30.96513	24.05356
YP_009052152.1	AaV_074	19734285			147.6693		
YP_009052158.1	AaV_080	81414719	1.07E+08	69972844	609.2168	663.8002	334.8254
YP_009052159.1	AaV_081	2511908			18.79632		
YP_009052169.1	AaV_091		36545243	44470742		226.1136	212.7959
YP_009052180.1	AaV_103	4222648			31.59758		
YP_009052184.1	AaV_107	2407466			18.01478		
YP_009052212.1	AaV_136	3453573	2491742	968582.1	25.84269	15.41696	4.63474
YP_009052220.1	AaV_144	2895241			21.66475		
YP_009052241.1	AaV_166	3197716			23.92813		
YP_009052249.1	AaV_175	13691313	4771932	4593738	102.4505	29.52501	21.98139
YP_009052256.1	AaV_182	3703203			27.71063		
YP_009052288.1	AaV_214	67912933	1879707	274435.4	508.1845	11.63017	1.313195
YP_009052299.1	AaV_225		15758257	11245160		97.49985	53.80895

Table 2.4. Continued.

YP_009052305.1	AaV_231	9261759	21029922	925214.8	69.30466	130.1168	4.427224
YP_009052306.1	AaV_232	8958157	Cannot quant.	Cannot quant.	67.03284		
YP_009052324.1	AaV_250	3585416			26.82924		
YP_009052334.1	AaV_260	6775457			50.69995		
YP_009052353.1	AaV_279	1048100			7.84281		
YP_009052360.1	AaV_286	1247631			9.335879		
YP_009052378.1	AaV_304	1048381			7.844915		
YP_009052402.1	AaV_328	7457721			55.80525		
YP_009052403.1	AaV_329	1453041			10.87293		
YP_009052450.1	AaV_382	958124.8			7.169536		

Table 2.5. Description of the host proteins detected by proteomics. The proteins detected in two of three runs are in blue, while those detected in all three runs are in red.

Description	Genes Detected
ARF1-directed GTPase-activating protein	XP_009034812.1
putative urea active transport protein	XP_009037653.1
anaphase-promoting complex subunit 8-like protein	XP_009037622.1
Cell division protein FtsH	XP_009038582.1
heat shock protein 70	XP_009033273.1
predicted protein	XP_009033220.1, XP_009041980.1, XP_009038981.1, XP_009034221.1, XP_009043494.1, XP_009033043.1, XP_009039271.1, XP_009041703.1, XP_009038784.1, XP_009032377.1, XP_009037560.1, XP_009033628.1, XP_009037597.1, XP_009033155.1, XP_009036072.1, XP_009034946.1, XP_009038429.1, XP_009043268.1, XP_009042844.1, XP_009042952.1, XP_009039958.1, XP_009042897.1, XP_009042766.1, XP_009040236.1, XP_009043278.1, XP_009042681.1
putative ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I	XP_009033030.1
histone	XP_009032689.1, XP_009042081.1, XP_009041779.1, XP_009041608.1, XP_009041154.1, XP_009036131.1, XP_009034585.1, XP_009034430.1, XP_009032767.1, XP_009034195.1
putative hydroxylase	XP_009037422.1
pH domain containing protein	XP_009034850.1
superfamily; cl17037	XP_009042729.1
K homology RNA-binding domain	XP_009042736.1
actinA	XP_009036992.1, XP_009037302.1, XP_009038528.1, XP_009039473.1
Probable 26S proteasome non-ATPase regulatory subunit 3	XP_009034199.1
haemagglutinin	XP_009037968.1

Table 2.5. Continued.

ADP-ribosylation factor	XP_009035447.1
alginate regulatory protein AlgP	XP_009037423.1
photosystem I assembly protein Ycf4	YP_003002020.1
UDP-galactose transporter	XP_009032842.1

Table 2.6. Mass spectrometry intensities and normalized intensities to MCP for proteins per particle for all the host proteins detected within the proteomes.

Accession Number	Average Sample 1	Average Sample 2	Average Sample 3	Proteins particle^{-1} Sample 1	Proteins particle^{-1} Sample 2	Proteins particle^{-1} Sample 3
XP_009042844.1	19004993.4	12203805.56	66886747.91	142.2121399	75.50767189	320.0582282
XP_009042952.1	20613303.19	1322783.858	3162141.936	154.2469337	8.1843593	15.13109214
XP_009039958.1	21007109.44	10429917.65	6843788.991	157.1937397	64.53223104	32.74805618
XP_009042681.1	27989282.88	519221264.5	289437816.6	209.4405259	3212.537982	1384.982192
XP_009037968.1	27989282.88	138193646.3	29105211.77	209.4405259	855.034969	139.2706747
XP_009042897.1	34143139.21	49386483.17	117762344.2	255.4891121	305.5652067	563.5018657
XP_009042766.1	35450172.85	60329924.09	181573436.9	265.2694918	373.2747209	868.8428471
XP_009037423.1	65849442.01	1611798.573	368480.434	492.7436628	9.972557921	1.76320719
XP_009038429.1	90524307.47	46608777.91	49062308.18	677.3827913	288.3789236	234.7669145
XP_009032377.1		2969266.516	311090871.8		18.37151541	1488.593725
XP_009034946.1		43531992.1	101149637.9		269.3421622	484.0087894
XP_009037560.1		35503500.56	50551493.77		219.6680912	241.892782
XP_009033628.1		32798711.36	18020711.95		202.932956	86.23049134
XP_009037597.1		7849962.675	13282309.46		48.56947313	63.55687135
XP_009040236.1		1322783.858	3162141.936		8.1843593	15.13109214
XP_009033155.1		4101278.095	2765595.077		25.37552399	13.23358495
XP_009032842.1		10286886.73	1766691.377		63.64726675	8.453753989
XP_009038784.1		8580336.766	548331.2732		53.08846083	2.623807275
XP_009039271.1		6137258.059	487789.633		37.97258697	2.334110874
XP_009041703.1		6137258.059	487789.633		37.97258697	2.334110874
XP_009034812.1	312755.8438			2.340315351		
XP_009037653.1	598734.2031			4.480257923		
XP_009037622.1	735486.8906			5.503562268		
XP_009036072.1	942204.7656			7.050407918		
XP_009038582.1	1014232.323			7.589381694		
XP_009033273.1	1059002.266			7.924389932		
XP_009033220.1	1264967.948			9.465607012		
XP_009033030.1	1484783.844			11.11046362		
XP_009032689.1	1926888.016			14.41867736		
XP_009032767.1	1926888.016			14.41867736		
XP_009034195.1	1926888.016			14.41867736		
XP_009034430.1	1926888.016			14.41867736		
XP_009034585.1	1926888.016			14.41867736		
XP_009036131.1	1926888.016			14.41867736		

Table 2.6. Continued.

XP_009041154.1	1926888.016			14.41867736		
XP_009041608.1	1926888.016			14.41867736		
XP_009041779.1	1926888.016			14.41867736		
XP_009042081.1	1926888.016			14.41867736		
XP_009037422.1	1963603.805			14.69341731		
XP_009034850.1	3228100.75			24.15549986		
XP_009042729.1	3478711.125			26.03078795		
XP_009042736.1	3979722.958			29.77980082		
XP_009043494.1	5018923.823			37.55601919		
XP_009033043.1	5066075.203			37.90884745		
XP_009043268.1	6312432.688			47.23519454		
XP_009036992.1	6498191.875			48.62520879		
XP_009037302.1	6498191.875			48.62520879		
XP_009038528.1	6498191.875			48.62520879		
XP_009039473.1	6498191.875			48.62520879		
XP_009034199.1	7269192			54.39451243		
XP_009041980.1	7582874.133			56.74175909		
XP_009038981.1	7672332			57.41116184		
XP_009034221.1	8585284.667			64.24267972		
XP_009043278.1	13700260.92			102.517448		
XP_009035447.1	28816906.94			215.6335398		
YP_003002020.1	1375705.75			10.29424502		

.

Table 2.7. Categories of reciprocal best BLAST hit pairs based on their NCVOG classification (Yutin *et al.*, 2009).

Category	Number of RBH pairs	RBH pairs across families
Uncharacterized	215	34
DEAD/SNF2-like helicases	27	14
Transcription/RNA processing (RNA polymerases, Poxvirus early transcription factor, Ribonucease III, XrN 5-'3' exonuclease, mRNA capping enzyme)	39	9
MCP	21	13
Translation (translation initiation inhibitor yjgF family, eukaryotic translation intitation factor 4e)	7	4
Redox (Flavin-containing amine oxidoreductase, Erv1/Alr family oxidoreductase, Thioredoxin)	38	26
Kinases/Phosphatases (Serine/Threonine protein kinase, F10 like kinase, serine threonine phosphatase 2C, Dual specificity phosphatases (DSP); Ser/Thr and Tyr protein phosphatases, acid phosphatase class B)	22	7
Proteases (Ulp1 protease, trypsin-like serine protease, papain-like cysteine peptidase, metallopeptiase WLM, otubain-like protein)	10	2
Other (Patatin phospholipase, uracil-DNA glycosylase, histones, Zn-finger proteins, Poxvirus P4B major core protein, Esterase lipase superfamily, DNA polymerase X, collagen triple helix repeat containing protein, mannose-6P isomerase, glycosyltransferase, Ubiquitin, Nudix hydrolase, lipocalin family protein, FtsJ-like methyltransferase, AAA family ATPase	31	10
Total	410	119

Table 2.8. Number of reads mapped to packaged viral proteins from previously published meta-transcriptomes of an *A. anophagefferens* bloom (Moniruzzaman *et al.*, 2017).

	SRA Accession (SRR)	3322145	3322387	3322113	3322397	3322338	3322357	3322422	3322410	Total
Number of Trimmed Reads		$4.0*10^7$	$4.1*10^7$	$4.4*10^7$	$4.0*10^7$	$2.2*10^7$	$9.6*10^7$	$3.7*10^7$	3.7*107	$3.56*10^8$
Number of reads mapped to AaV		684	1202	1761	1366	189	644	1178	1272	8,296
Reads mapped to genes encoding packaged proteins		231	202	937	303	23	7	333	346	2,382
% of mapped reads that encoding packaged proteins		33.77	16.81	53.21	22.18	12.17	1.09	28.27	27.20	28.71
Gene #	Description									
AaV_096	Capsid Protein	44	65	245	72	11	0	85	70	592
AaV_247	putative capsid protein 2	2	0	4	1	0	0	0	1	8
AaV_174	DNA directed RNA polymerase K subunit	0	0	0	0	0	0	0	0	0
AaV_180	putative VV D6R-type helicase	0	0	0	0	0	0	0	1	1
AaV_211	putative mRNA capping enzyme	0	0	0	2	0	0	0	1	3
AaV_222	putative RNA polymerase beta subunit	0	1	0	0	0	0	0	0	1
AaV_242	DNA directed RNA polymerase II largest subunit	0	1	1	2	0	0	2	3	9
AaV_269	putative ATP-dependent RNA helicase	0	2	0	1	0	0	1	0	4
AaV_024	Concanavalin A-like lectin/glucanase superfamily	104	49	345	92	5	6	95	96	792
AaV_038	pectate lyase	0	1	1	1	0	0	0	2	5

Table 2.8. Continued.

AaV_078	putative unsaturated glucuronyl hydrolase	0	2	3	0	0	0	1	1	7
AaV_274	putative beta-1,4 galactosyltranferase	0	1	0	4	0	0	1	2	8
AaV_276	Concanavalin A-like lectin/glucanase superfamily	10	6	76	11	2	0	13	12	130
AaV_314	Hypothetical	18	16	86	39	1	0	41	31	232
AaV_386	Concanavalin A-like lectin/glucanase superfamily	15	4	33	6	0	0	9	18	85
AaV_158	metal dependent hydrolase	0	0	0	0	0	0	0	0	0
AaV_281	putative membrane protein	0	0	0	0	0	0	0	0	0
AaV_306	putative ABC transporter family protein	0	0	1	1	0	0	1	5	8
AaV_052	Hypothetical	6	5	6	6	0	0	3	10	36
AaV_062	Hypothetical	0	0	0	0	0	1	0	0	1
AaV_074	Hypothetical	0	0	1	1	0	0	0	1	3
AaV_080	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_081	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_091	Hypothetical	32	31	126	47	0	0	71	79	386
AaV_103	Hypothetical	0	3	0	0	0	0	1	0	4
AaV_107	Hypothetical	0	0	1	5	0	0	1	3	10
AaV_136	Hypothetical	0	0	0	0	0	0	1	0	1
AaV_144	Hypothetical	0	0	2	0	0	0	0	0	2
AaV_166	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_175	Hypothetical	0	3	0	0	0	0	1	0	4
AaV_182	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_214	Hypothetical	0	2	0	2	0	0	3	3	10
AaV_225	Hypothetical	0	0	0	0	0	0	0	0	0

Table 2.8. Continued.

AaV_231	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_232	Hypothetical	0	0	1	0	4	0	1	2	8
AaV_250	Hypothetical	0	0	1	1	0	0	0	0	2
AaV_260	Hypothetical	0	2	2	3	0	0	0	3	10
AaV_279	Hypothetical	0	0	0	0	0	0	0	0	0
AaV_286	Hypothetical	0	3	1	1	0	0	0	0	5
AaV_304	Hypothetical	0	2	0	0	0	0	1	0	3
AaV_328	Hypothetical	0	3	0	1	0	0	0	1	5
AaV_329	Hypothetical	0	0	1	4	0	0	1	1	7
AaV_382	Hypothetical	0	0	3	1	0	0	1	0	5

Table 2.9. Metagenomic assembled contigs from TARA Oceans Dataset separated by water depth.

Deep Chlorophyll Maximum	Contig number	Percentage
Total	19388407	100
Non-viral + PL	2789	0.014385
Viral + Polysaccharide Lyases	116	0.000598
mesopelagic	Contig number	Percentage
total	10236193	100
Non-viral + PL	2951	0.028829
Viral + Polysaccharide Lyases	113	0.001104
Epipelagic	Contig number	Percentage
total	32780054	100
Non-viral + PL	4816	0.014692
Viral + Polysaccharide Lyases	270	0.000824

Table 2.10. Metagenomic assembled contigs from TARA Oceans Dataset separated by size fraction.

< 0.22 µm	Contig number	Percentage
total	3942150	100
Non-viral + PL	297	0.007534
Viral + PL	76	0.001928
0.1-0.22 µm	Contig number	Percentage
total	3192414	100
Non-viral + PL	398	0.012467
Viral + PL	35	0.001096
0.22-0.45 µm	Contig number	Percentage
total	5984290	100
Non-viral + PL	753	0.012583
Viral + PL	103	0.001721
0.45-0.8 µm	Contig number	Percentage
total	5986303	100
Non-viral + PL	1707	0.028515
Viral + PL	155	0.002589
0.22-1.6 µm	Contig number	Percentage
total	11381410	100
Non-viral + PL	1521	0.013364
Viral + PL	97	0.000852
0.22-3 µm	Contig number	Percentage
total	31918087	100
Non-viral + PL	5880	0.018422
Viral + PL	33	0.000103

Table 2.11. Metagenomic assembled contigs from TARA Oceans Dataset separated by Station.

Station	Total Contigs	Non-viral + Polysaccharide Lyase	Viral + Polysaccharide Lyase	total	Virus Specific
4	8.00E+05	115	7	122	5.74
7	3.95E+05	127	9	136	6.62
9	7.89E+05	125	15	140	10.71
18	9.92E+05	126	8	134	5.97
22	1.20E+05	14	1	15	6.67
23	4.86E+05	84	1	85	1.18
25	1.04E+06	110	13	123	10.57
30	8.89E+05	131	17	148	11.49
31	4.36E+05	32	2	34	5.88
32	7.96E+05	138	8	146	5.48
33	2.23E+05	17	1	18	5.56
34	1.14E+06	144	11	155	7.10
36	6.97E+05	57	9	66	13.64
37	7.40E+05	116	9	125	7.20
38	1.10E+06	143	13	156	8.33
39	9.64E+05	149	9	158	5.70
41	9.88E+05	90	9	99	9.09
42	7.92E+05	35	1	36	2.78
45	2.52E+05	8	1	9	11.11
46	6.25E+04	5	0	5	0.00
48	4.58E+05	54	3	57	5.26
52	7.76E+05	61	1	62	1.61
56	4.50E+05	87	2	89	2.25
57	2.20E+05	19	0	19	0.00
58	2.69E+05	29	0	29	0.00
62	2.42E+05	21	0	21	0.00
64	1.06E+06	93	1	94	1.06
65	1.24E+06	174	3	177	1.69
66	5.54E+05	56	11	67	16.42
67	9.59E+05	132	10	142	7.04
68	2.16E+06	405	27	432	6.25
70	1.60E+06	401	49	450	10.89

Table 2.11. Continued.

72	1.09E+06	144	8	152	5.26
76	2.25E+06	473	26	499	5.21
78	2.13E+06	288	17	305	5.57
82	5.68E+05	66	0	66	0.00
84	2.87E+05	19	0	19	0.00
85	8.14E+05	146	1	147	0.68
93	6.39E+05	208	3	211	1.42
94	5.08E+05	111	0	111	0.00
96	2.65E+05	44	0	44	0.00
98	7.60E+05	122	0	122	0.00
99	2.89E+05	41	0	41	0.00
100	1.14E+06	235	3	238	1.26
102	1.18E+06	224	0	224	0.00
109	1.02E+06	188	2	190	1.05
110	9.53E+05	133	1	134	0.75
111	1.16E+06	218	0	218	0.00
112	9.62E+05	252	4	256	1.56
122	3.18E+06	720	67	787	8.51
123	2.33E+06	493	25	518	4.83
124	2.82E+06	515	41	556	7.37
125	2.98E+06	616	40	656	6.10
128	5.40E+05	109	3	112	2.68
132	9.96E+05	171	0	171	0.00
133	1.29E+06	225	1	226	0.44
137	9.84E+05	241	2	243	0.82
138	9.61E+05	186	0	186	0.00
140	3.44E+05	51	0	51	0.00
141	3.52E+05	45	0	45	0.00
142	9.37E+05	189	1	190	0.53
145	6.39E+05	113	0	113	0.00
146	5.77E+05	80	0	80	0.00
148	3.60E+05	42	0	42	0.00
148B	3.59E+05	83	1	84	1.19
149	5.91E+05	99	0	99	0.00
150	7.49E+05	91	1	92	1.09
151	7.66E+05	97	1	98	1.02
152	9.55E+05	180	0	180	0.00

Table 2.12. Classification of putative viral contigs containing polysaccharide lyases into viral families by best BLAST hit to NCVOG on contig.

Family	Count	Family/Total Viral
Pithoviridae	4	0.008
Phycodnaviridae	183	0.367
Mimiviridae	142	0.285
Pandoraviridae	50	0.100
Iridoviridae	7	0.014
Ascoviridae	68	0.136
Poxviridae	15	0.030
Extended *Mimiviridae*	30	0.060
Total	499	1.000

Table 2.13. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for percentage of contigs that are viral by water depth.

	DCM	Epipelagic	Mesopelagic
DCM	-	0.9999	0.6103
Epipelagic		-	0.5372
Mesopelagic			-

Table 2.14. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for percentage of contigs that are viral by water fraction.

	0.1-0.22	0.22	0.22-0.45	0.45-0.8	0.22-1.6	0.22-3
0.1--0.22	-	0.0108	1.0000	0.9691	0.5971	0.0078
0.22		-	0.0136	0.0001	<0.0001	<0.0001
0.22-0.45			-	0.9403	0.4877	0.0034
0.45-0.8				-	0.9706	0.0731
0.22-1.6					-	0.2104
0.22-3						-

Table 2.15. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for normalized metagenomics reads.

DCM - cellular vs. DCM - viral	>0.999
DCM - cellular vs. SRF - cellular	0.739
DCM - cellular vs. SRF - viral	<0.001
DCM - cellular vs. MES - cellular	>0.999
DCM - cellular vs. MES - viral	0.08
DCM - viral vs. SRF - cellular	0.999
DCM - viral vs. SRF - viral	0.006
DCM - viral vs. MES - cellular	>0.999
DCM - viral vs. MES - viral	0.161
SRF - cellular vs. SRF - viral	<0.001
SRF - cellular vs. MES - cellular	0.776
SRF - cellular vs. MES - viral	0.023
SRF - viral vs. MES - cellular	<0.001
SRF - viral vs. MES - viral	0.984
MES - cellular vs. MES - viral	0.088

Table 2.16. One-way ANOVA corrected for multiple comparisons using Tukey's multiple comparisons p-values for normalized metatranscriptomics reads.

SRF - cellular vs. SRF - viral	0.908
SRF - cellular vs. DCM - cellular	0.883
SRF - cellular vs. DCM - viral	0.796
SRF - cellular vs. MES - cellular	0.993
SRF - cellular vs. MES - viral	0.997
SRF - cellular vs. all - cellular	0.54
SRF - cellular vs. all - viral	0.504
SRF - viral vs. DCM - cellular	0.996
SRF - viral vs. DCM - viral	>0.999
SRF - viral vs. MES - cellular	0.696
SRF - viral vs. MES - viral	>0.999
SRF - viral vs. all - cellular	>0.999
SRF - viral vs. all - viral	>0.999
DCM - cellular vs. DCM - viral	0.984
DCM - cellular vs. MES - cellular	0.195
DCM - cellular vs. MES - viral	>0.999
DCM - cellular vs. all - cellular	0.999
DCM - cellular vs. all - viral	0.938
DCM - viral vs. MES - cellular	0.491
DCM - viral vs. MES - viral	>0.999
DCM - viral vs. all - cellular	0.996
DCM - viral vs. all - viral	>0.999
MES - cellular vs. MES - viral	0.972
MES - cellular vs. all - cellular	0.019
MES - cellular vs. all - viral	0.131
MES - viral vs. all - cellular	>0.999
MES - viral vs. all - viral	>0.999
all - cellular vs. all - viral	0.984

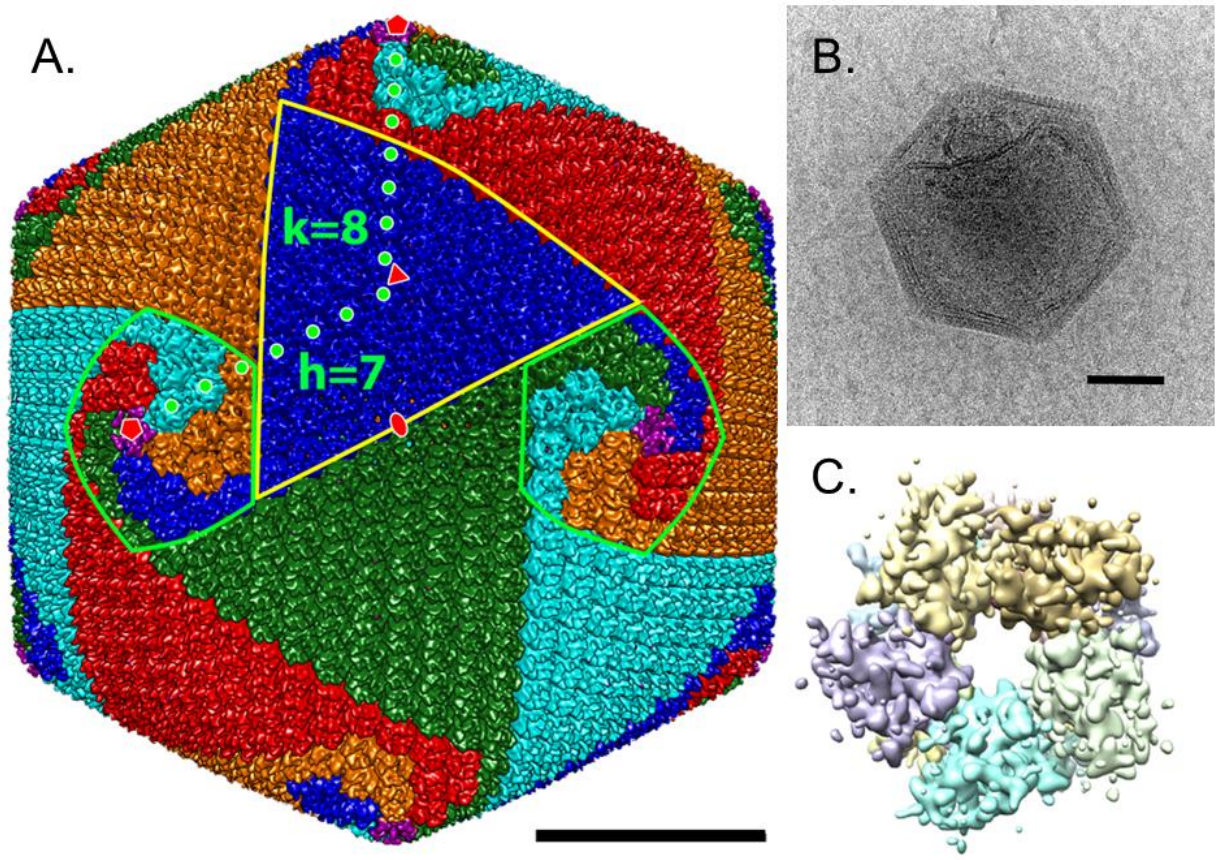
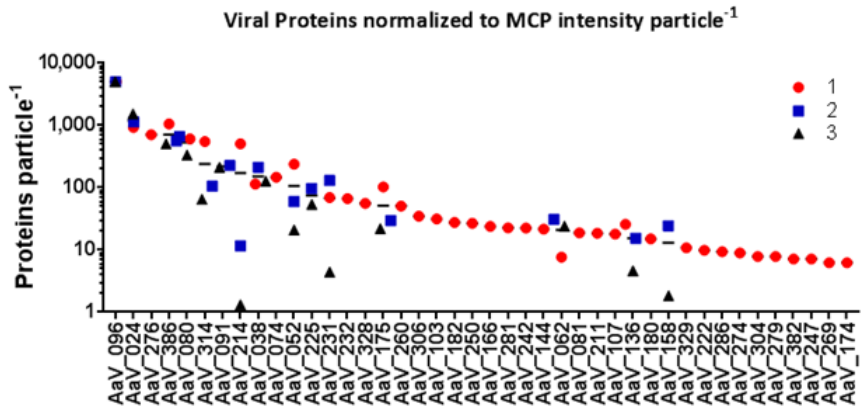


Figure 2.1. cryo-EM reconstruction of the AaV virion. (A) Isosurface of AaV. Capsomers are colored based on their orientation in red, blue, green, cyan, and orange. The pentameric capsomers are colored in purple. The borders of one trisymmetron are highlighted in yellow whereas the borders of two pentasymmetrons are highlighted in light green. The h and k vectors for its T number are indicated in green. 5-fold, 3-fold, and 2-fold symbols are indicated in red. (B) Representative micrograph of an AaV particle. (C) Pentameric capsomer reconstruction. The scale bars in all panels represent 500 Å.

A



B

Gene Name	Mean	Coefficient of Variation
AaV_096	5040.00	0.00
AaV_024	1197.06	20.69
AaV_276	712.05	
AaV_386	709.09	34.87
AaV_080	535.95	26.86
AaV_314	241.50	91.46
AaV_091	219.45	3.03
AaV_214	173.71	136.17
AaV_038	151.61	29.07
AaV_052	106.83	88.99
AaV_225	75.65	28.88
AaV_231	67.95	75.53
AaV_232	67.03	
AaV_175	51.32	70.71
AaV_062	20.90	46.72
AaV_136	15.30	56.60
AaV_158	13.14	85.95
AaV_078		

Figure 2.2. Viral proteins detected in the proteome normalized to the number of MCP per virion by mass spectrometry intensities. A) Normalized viral proteins plotted by the mean of the biological replicates. The black bar indicates the mean of the multiple replicates if detected in multiple samples. B) The mean and coefficient of variation of proteins detected in multiple samples. Bolded gene names indicate those detected in all three samples.

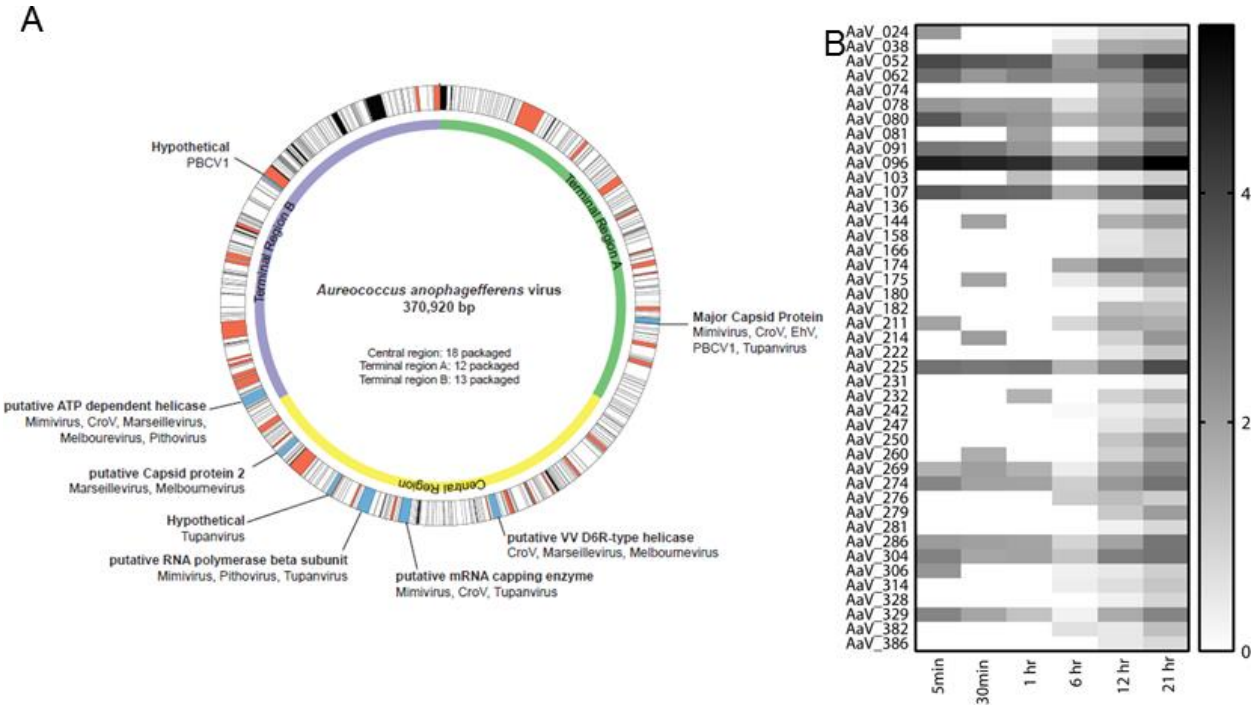
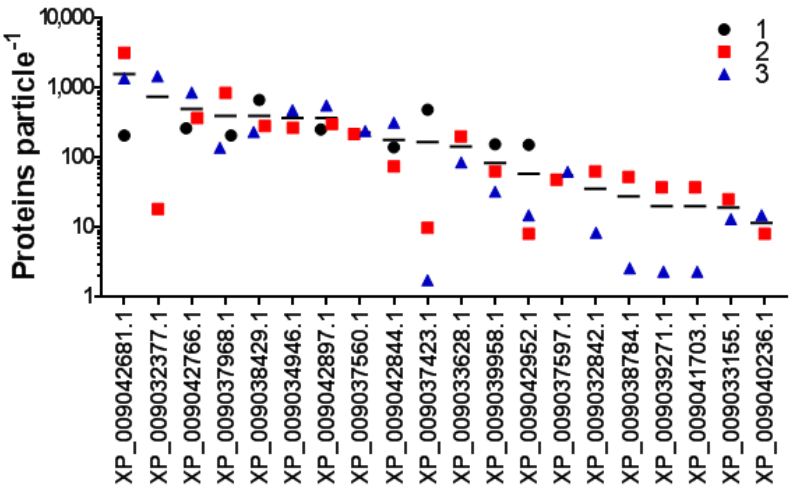


Figure 2.3. Overview of the proteins found within the AaV particle. A) Circos plot of the entire AaV genome. The outer ring is decorated with coding sequences within the genome, where white bars are gene products not packaged, red bars are those packaged but that do not have reciprocal best BLAST hits (e value cutoff < 1 x 10^{-15}), and blue bars are gene products packaged and have reciprocal best BLAST hits in representative proteomes used in this study. Those proteins are labelled and have the viruses for which has an RBH hit. The inner ring shows the three segments of the genome described previously (Moniruzzaman *et al.*, 2014). B) Heatmap of the log-transformed RPKM values + 1 of the genes in which the gene products are found within the particle over the course of the infection cycle (Moniruzzaman *et al.*, 2018).

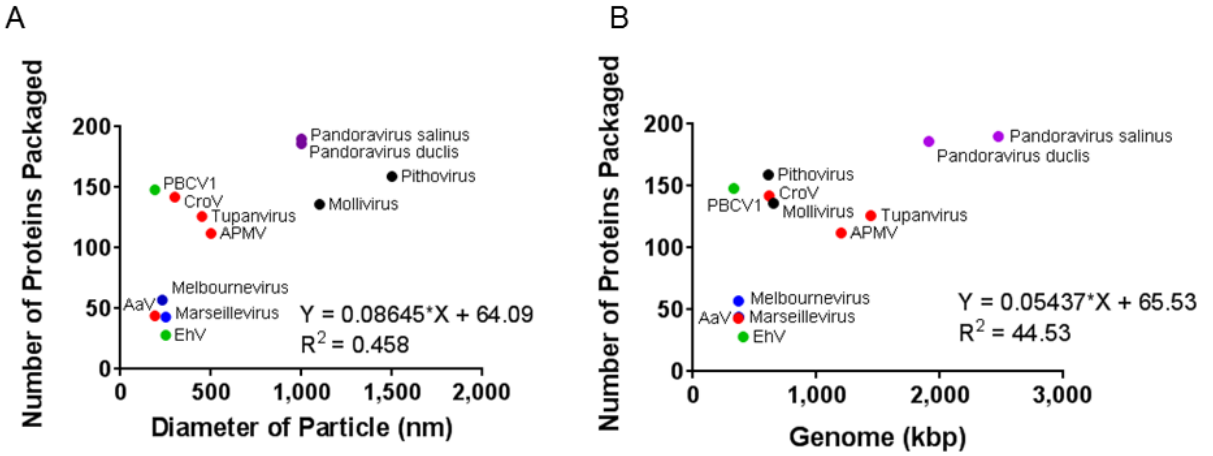
A



B

Accession	Mean	Coefficient of Variation
XP_009042844.1	179.26	57.58
XP_009042952.1	59.19	113.67
XP_009039958.1	84.82	62.24
XP_009042681.1	1602.32	77.11
XP_009037968.1	401.25	80.29
XP_009042897.1	374.85	36.00
XP_009042766.1	502.46	52.30
XP_009037423.1	168.16	136.50
XP_009038429.1	400.18	49.29
XP_009032377.1	753.48	97.56
XP_009034946.1	376.68	28.49
XP_009037560.1	230.78	4.82
XP_009033628.1	144.58	40.36
XP_009037597.1	56.06	13.37
XP_009040236.1	11.66	29.79
XP_009033155.1	19.30	31.45
XP_009032842.1	36.05	76.55
XP_009038784.1	27.86	90.58
XP_009039271.1	20.15	88.42
XP_009041703.1	20.15	88.42

Figure 2.4. Host proteins detected in multiple runs of the proteome normalized to the number of MCP per virion by mass spectrometry intensities. A) Normalized host proteins plotted by the mean of the replicates. The black bar indicates the mean of the multiple replicates. B) The mean and coefficient of variation of proteins detected in multiple samples. Bolded gene names indicate those detected in all three samples.



C

Percentage of Reciprocal Best BLAST hit (RBH) pairs compared to total number of Packaged Proteins

	AaV	APMV	CroV	Tupanvirus	EhV	PBCV1	Marseillevirus	Melbournevirus	P. dulcis	P. salinus	Pithovirus	Mollivirus
AaV	100.00	9.30	9.30	9.30	2.33	4.65	6.98	6.98	0.00	0.00	4.65	0.00
APMV	3.57	100.00	24.11	50.00	0.00	2.68	2.68	1.79	3.57	4.46	3.57	1.79
CroV	2.84	19.15	100.00	18.44	0.71	4.26	1.42	2.13	0.71	0.71	1.42	0.00
Tupanvirus	3.17	44.44	20.63	100.00	0.00	2.38	0.79	0.79	1.59	1.59	3.17	1.59
EhV	3.57	0.00	3.57	0.00	100.00	3.57	0.00	0.00	0.00	0.00	0.00	3.57
PBCV1	1.35	2.03	4.05	2.03	0.68	100.00	0.68	0.68	0.00	0.00	0.00	0.00
Marseillevirus	6.82	6.82	4.55	2.27	0.00	2.27	100.00	75.00	6.82	2.27	4.55	4.55
Melbournevirus	5.26	3.51	5.26	1.75	0.00	1.75	57.89	100.00	7.02	3.51	5.26	3.51
P. dulcis	0.00	2.16	0.54	1.08	0.00	0.00	1.62	2.16	100.00	70.81	0.54	8.65
P. salinus	0.00	2.63	0.53	1.05	0.00	0.00	0.53	1.05	68.95	100.00	0.00	9.47
Pithovirus	1.26	2.52	1.26	2.52	0.00	0.00	1.26	1.89	0.63	0.00	100.00	1.26
Mollivirus	0.00	1.47	0.00	1.47	0.74	0.00	1.47	1.47	11.76	13.24	1.47	100.00

Figure 2.5. Overview of the other proteomes of virion associated proteins used in this study. The number of proteins packaged by each virus was compared to A) Diameter of the particle or B) the genome size. C) Percentage of total proteins packaged by the virus (rows) that share a reciprocal best BLAST hit pairs shared between each other virus (columns). Colored boxes represent the families of viruses. The colors of the icons or boxes indicate viral families with more than one representative: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*.

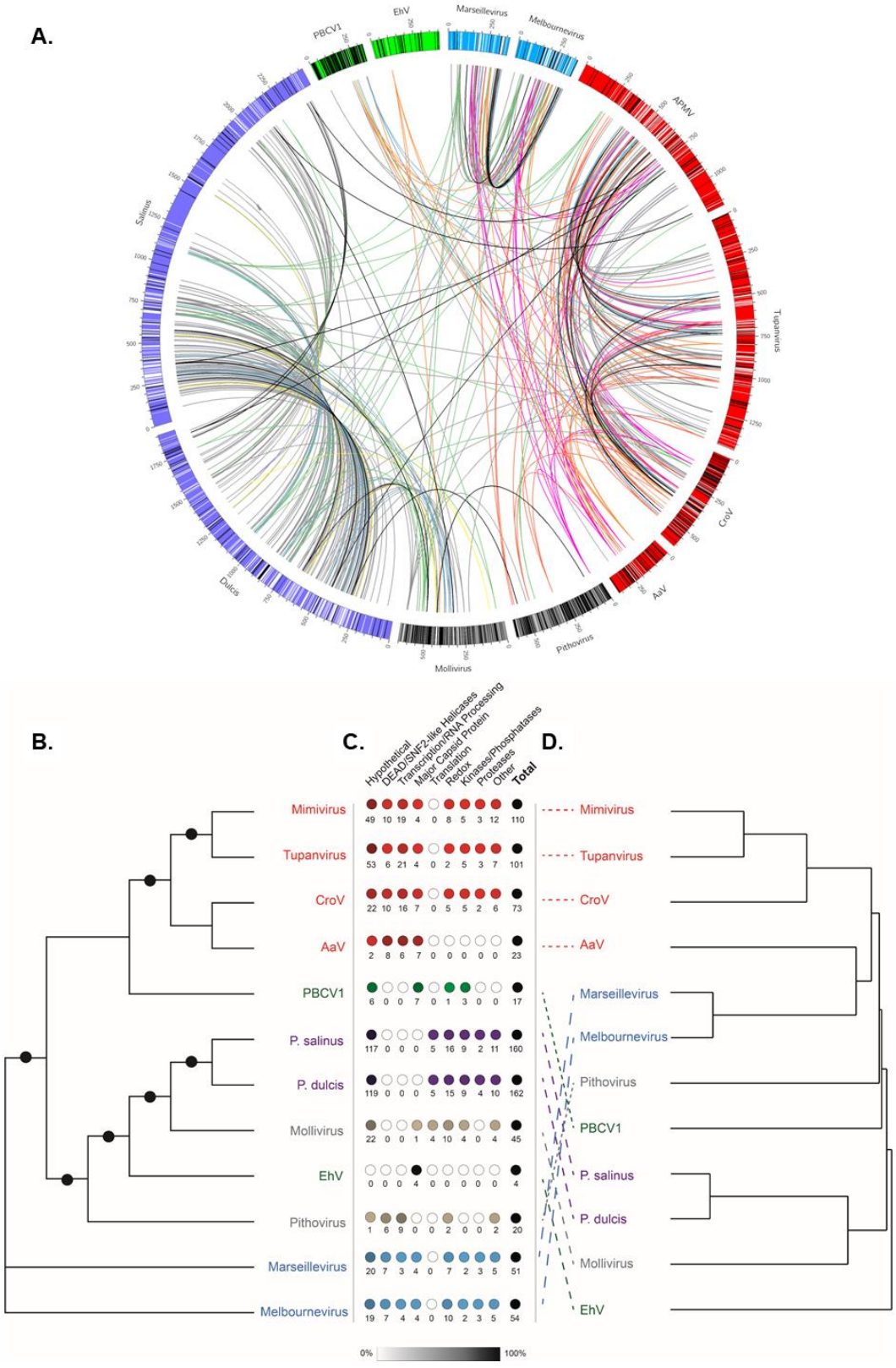


Figure 2.6. Comparison of packaged proteins between representative viruses.

Figure 2.6. Continued. A) The outer bars are the genomes of each of the representative viruses, where there the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs with their color being based on the classification of the NCVOG category (Table 2.7). The colors are: uncharacterized pairs = grey, DEAD/SNF2-like helicases = pink, Redox = green, transcription/RNA processing = red, translation = yellow, capsid proteins = orange, proteases = purple, kinases/phosphatases = blue, other = black. B) Unrooted maximum likelihood DNA Polymerase B (PolB) phylogenetic tree. C) Number of reciprocal best BLAST hit pairs found within each virus proteome by type of NCVOG category (Table 2.7). Numbers below the circles indicate the number of reciprocal best BLAST hit pairs. Shading of the circles represent the percentage of the total number of reciprocal best BLAST hit pairs. D) Hierarchical clustering of resemblance based on Bray-Curtis similarity of presence/absence of reciprocal best BLAST hit pairs for all proteins found within the representative particle proteomes. The colors of the virus names indicate viral families with more than one representative: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*.

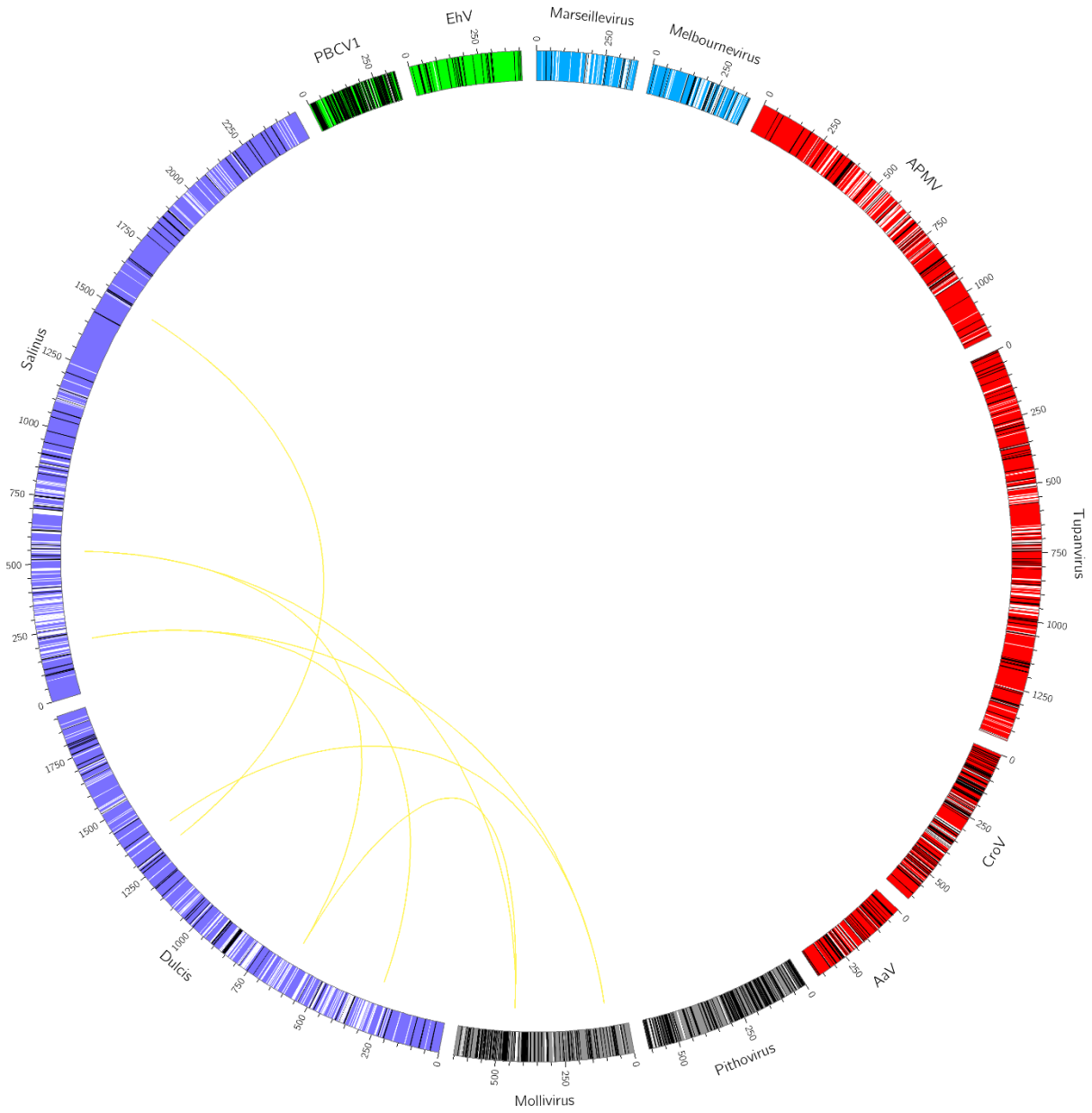


Figure 2.7. Circos plot showing translation related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white lines represent the location of the genes encoding proteins packaged in the particle that have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

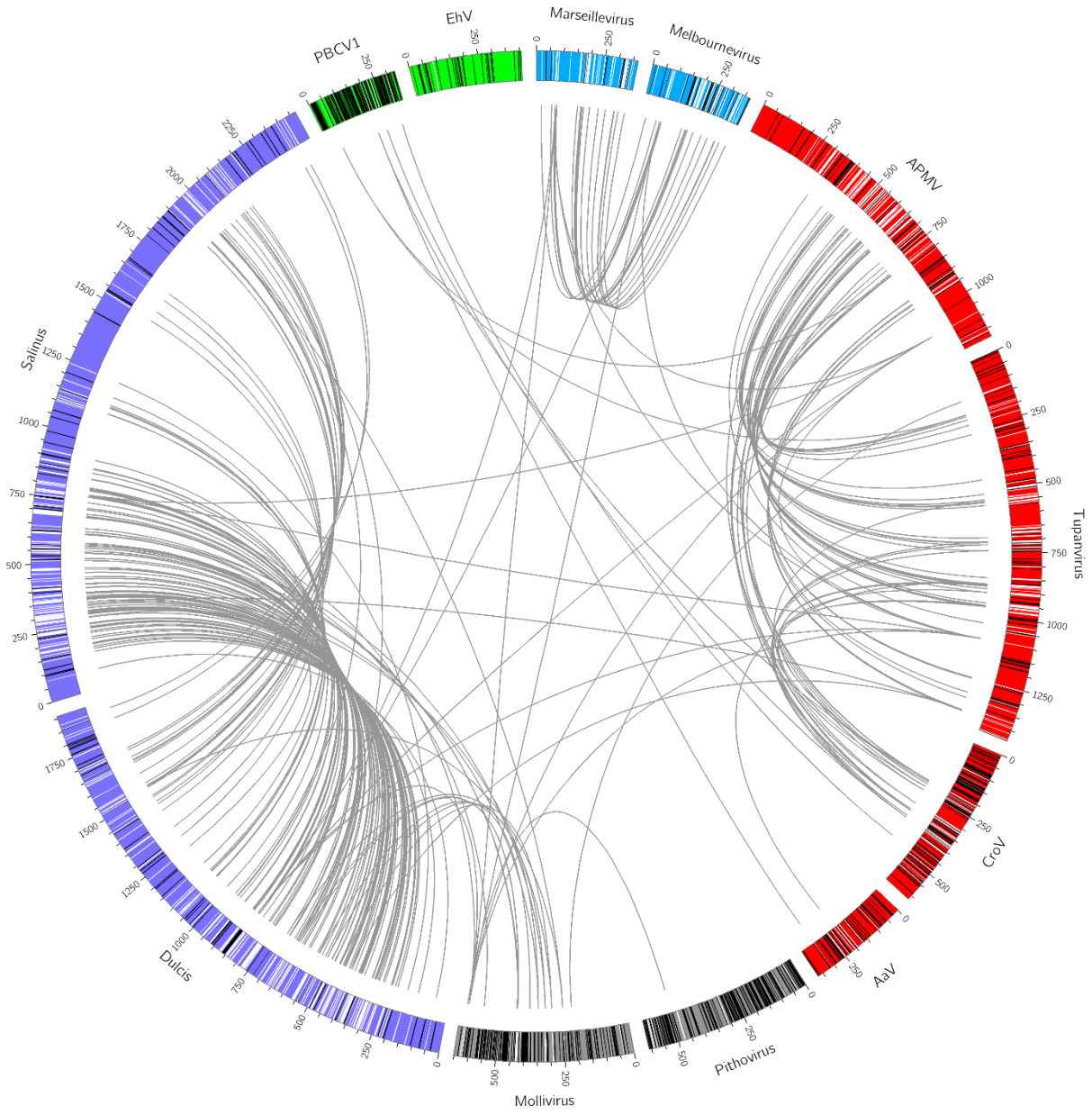


Figure 2.8. Circos plot showing reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used with unknown functions. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white bares lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

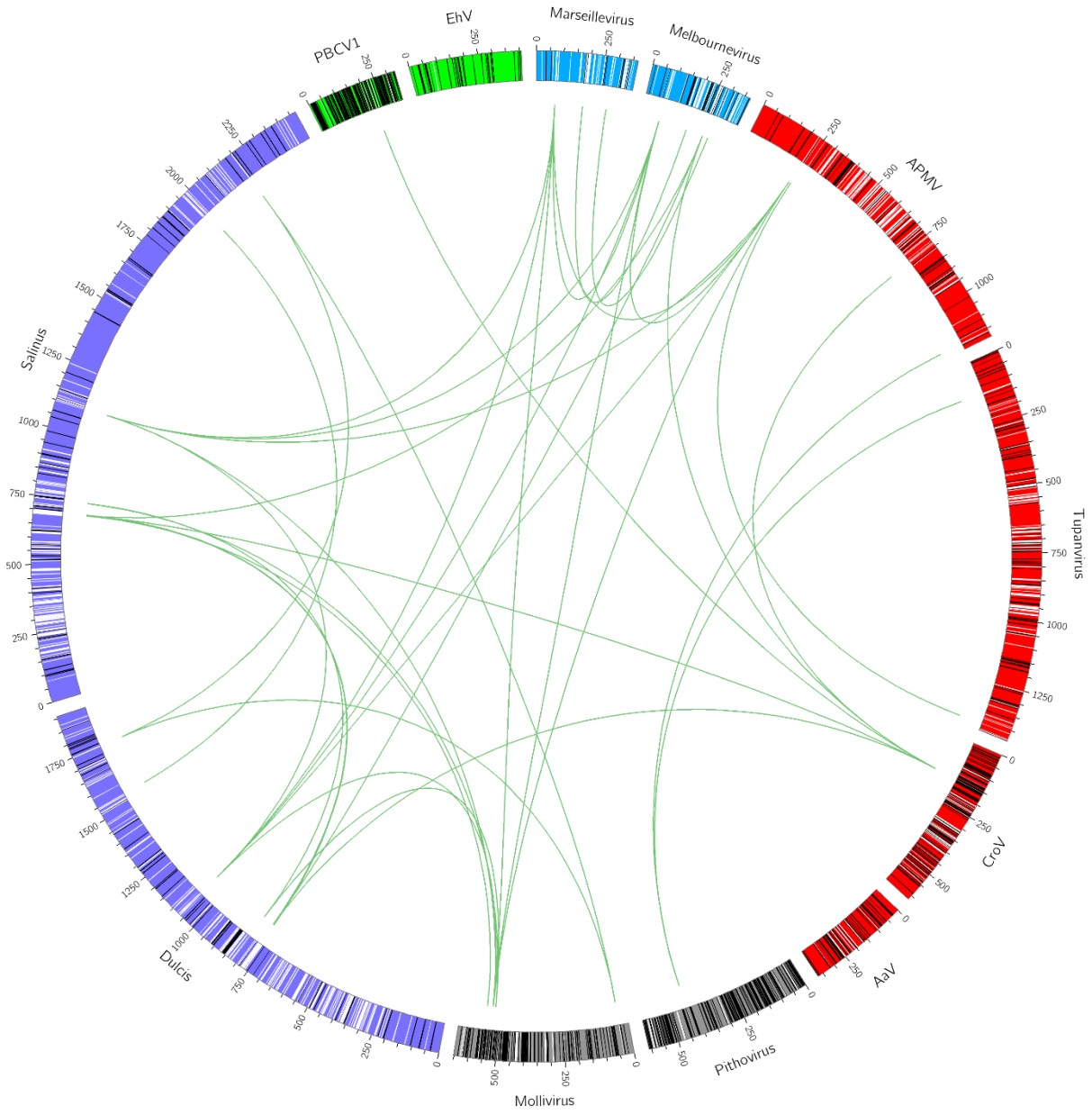


Figure 2.9. Circos plot showing Redox related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white bares lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

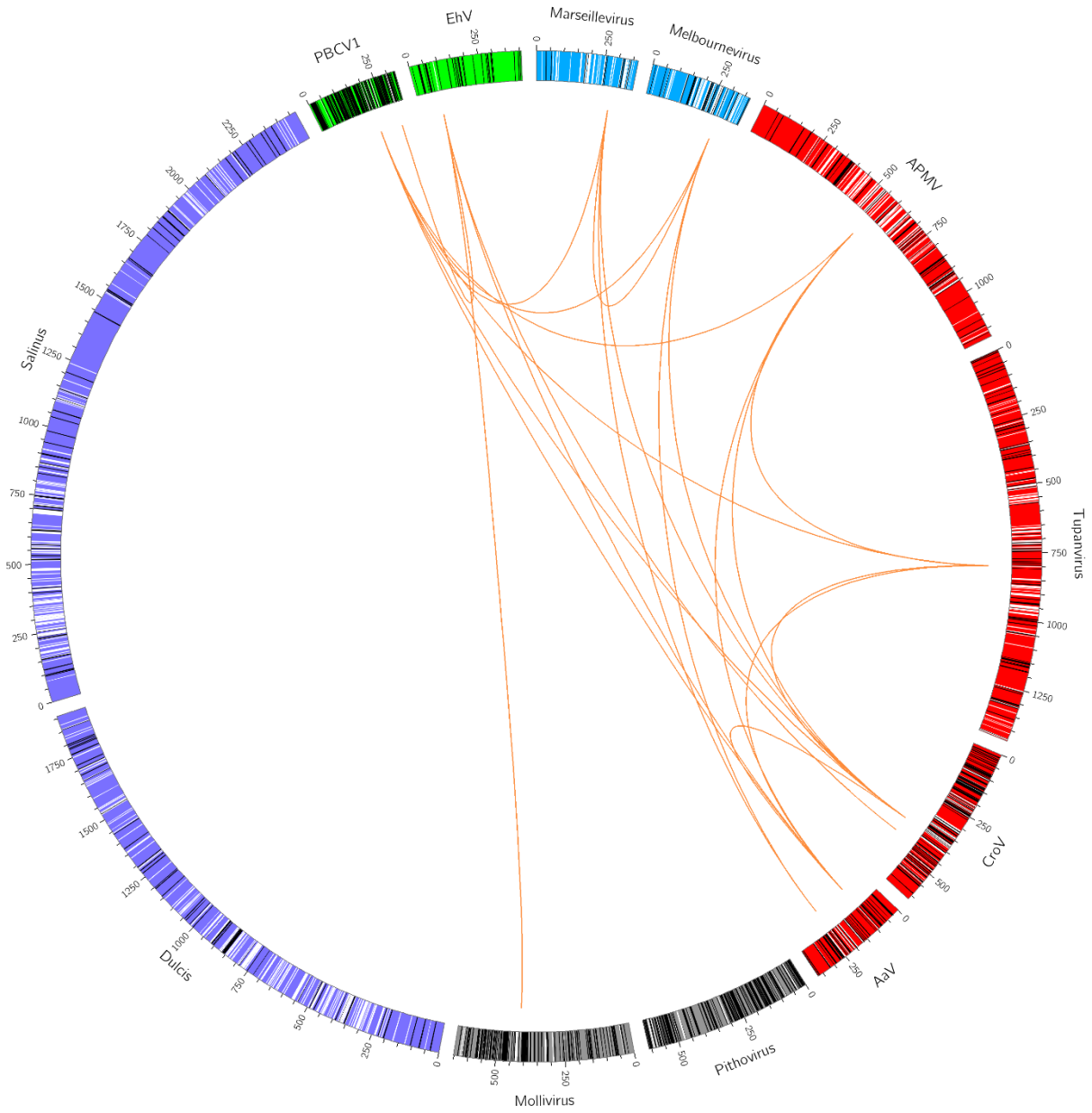


Figure 2.10. Circos plot showing MCP reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white bares lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

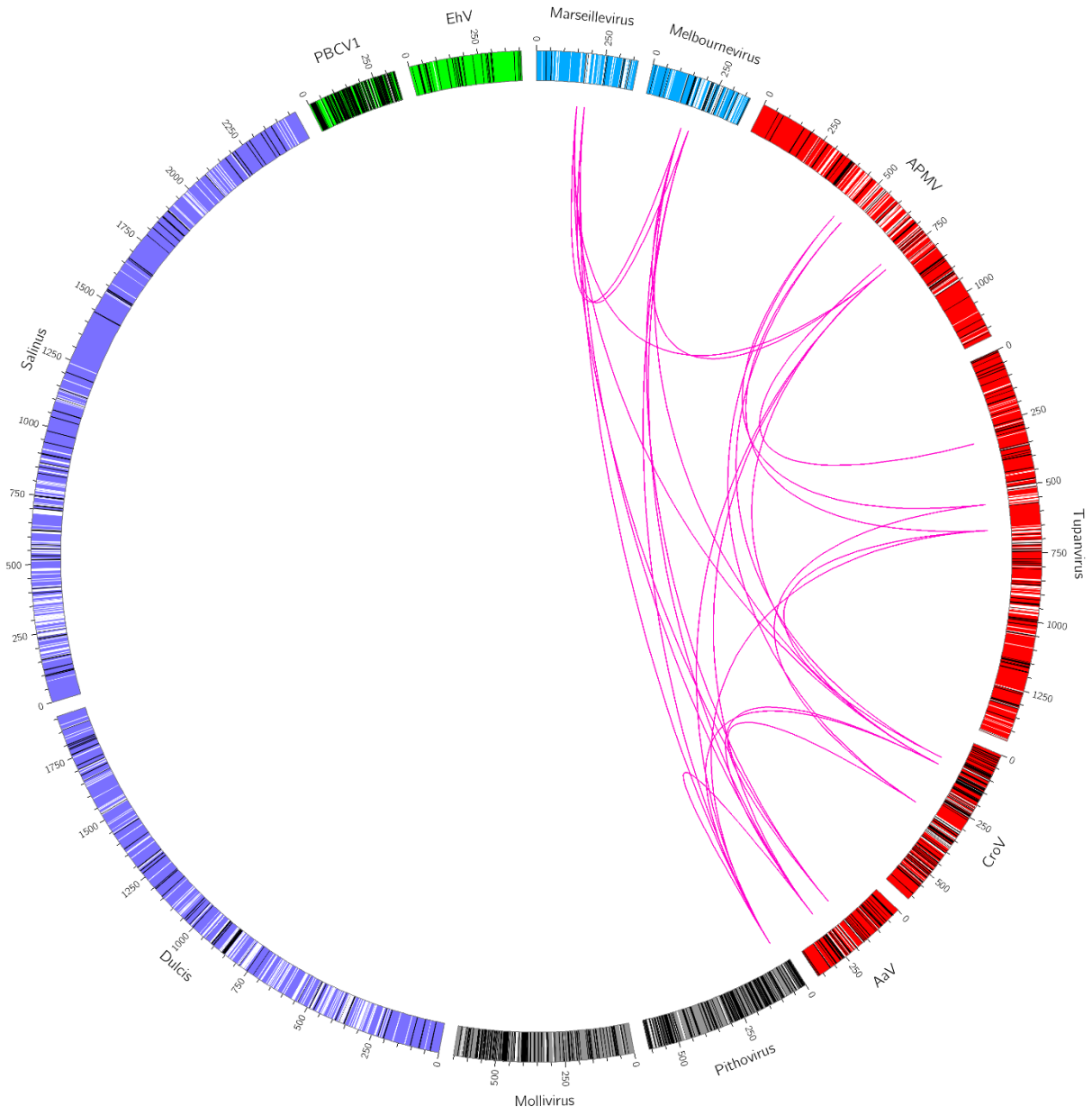


Figure 2.11. Circos plot showing DEAD/SNF2-like helicases related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white bares lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

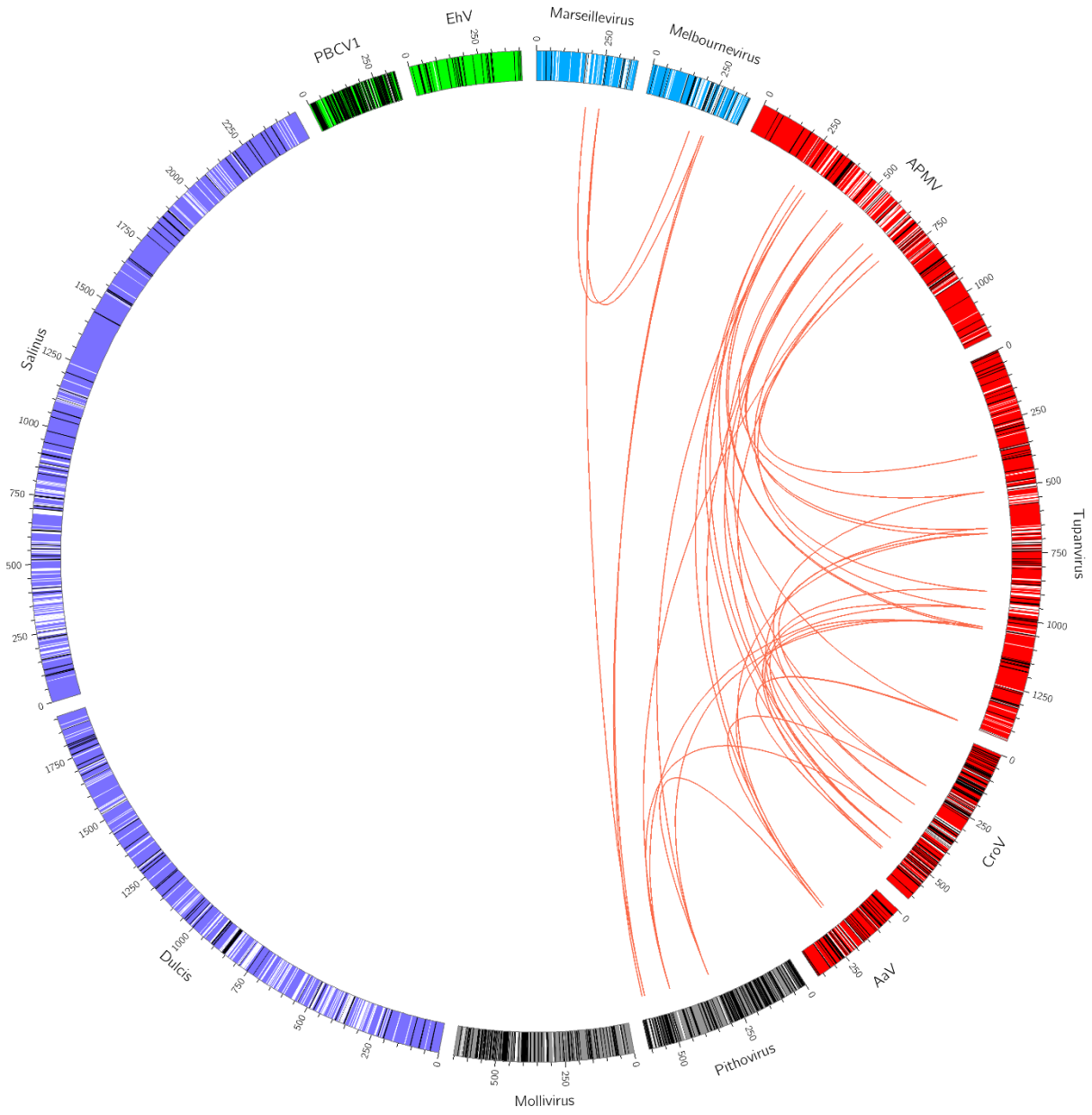


Figure 2.12. Circos plot showing transcription/RNA processing-related reciprocal best BLAST hit pairs found packaged within particles of the representative viruses used. The outer bars are the genomes of each of the representative viruses, where the color is based on the viral family it belongs to: red = *Mimiviridae*, green = *Phycodnaviridae*, Blue = *Marseilleviridae*, Purple = *Pandoraviridae*. The black lines represent the location of the genes encoding proteins packaged in the particle that do not have reciprocal best BLAST hits to any of the other representatives, while white bares lines represent the location of the genes encoding proteins packaged in the particle that do have reciprocal best BLAST hits to any of the other representatives. The lines between viruses are connecting reciprocal best BLAST hit pairs.

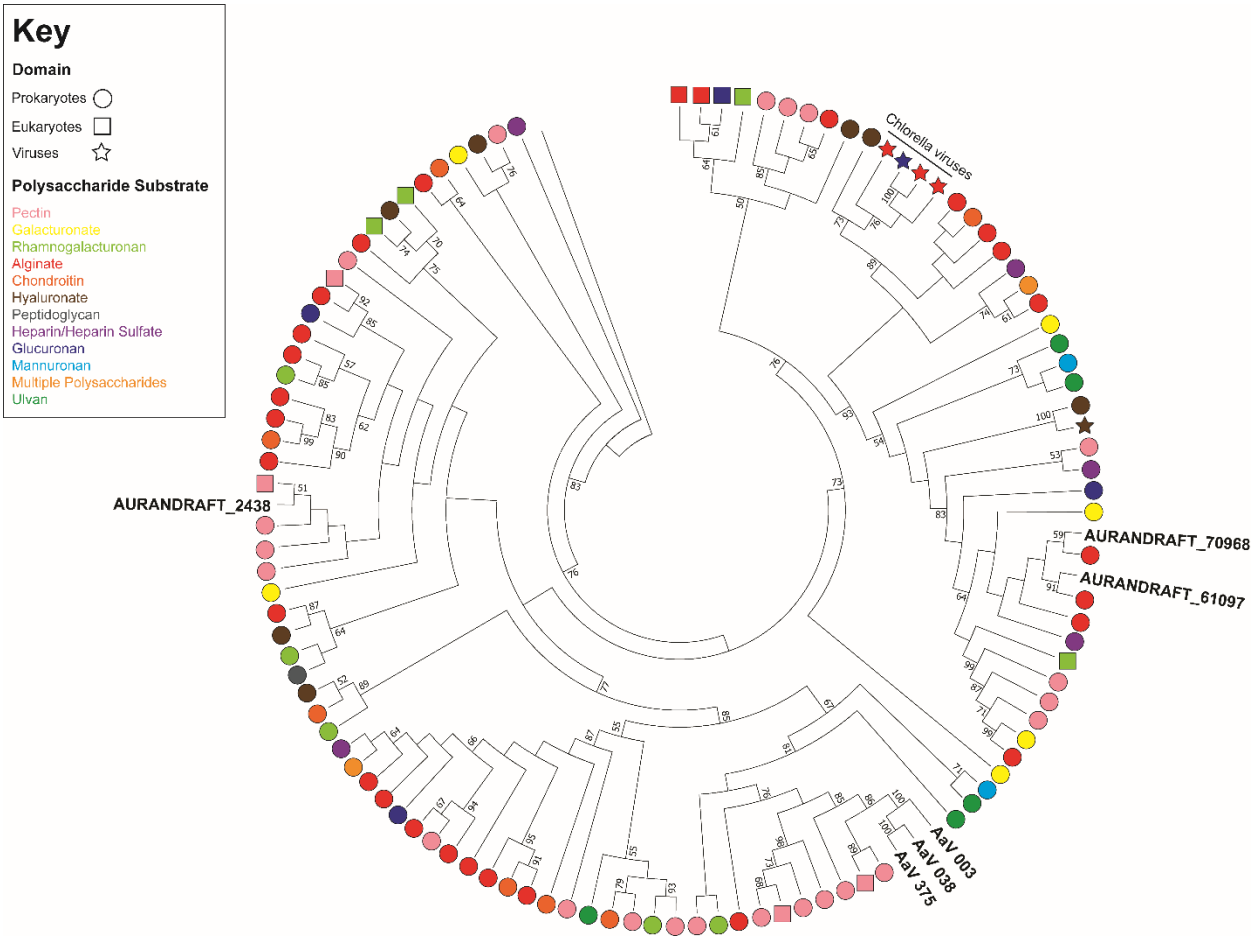


Figure 2.13. Maximum likelihood phylogenetic relatedness of all characterized polysaccharide lyases within the Carbohydrate-Active Enzyme (CAZy) Database. Putative polysaccharide lyases from both *A. anophagefferens* and AaV genomes were also included.

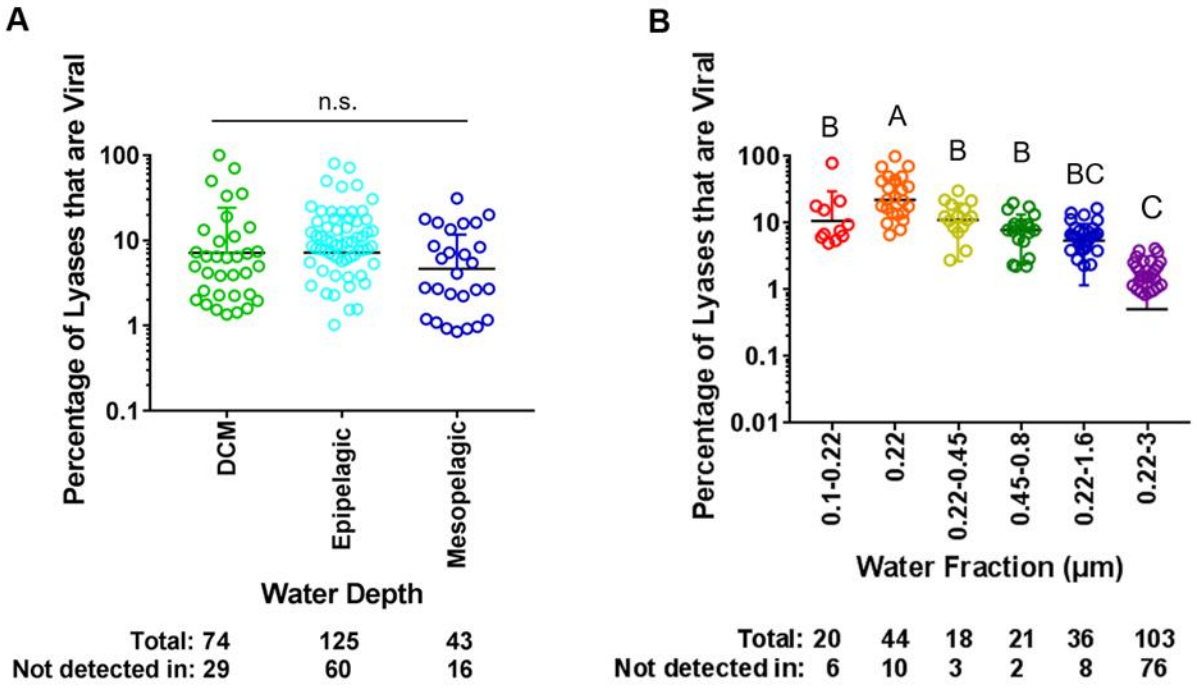


Figure 2.14. Percentage of polysaccharide lyases found on viral contigs to total contigs containing polysaccharide lyases in TARA Oceans metagenomic contigs. A) Contigs separated by water depth at which samples were collected. One-way ANOVA p-values for differences between water depth are in Table 6. B) Contigs separated by size fraction of the sample collected. One-way ANOVA p-values for differences between water depth are in Table 5.

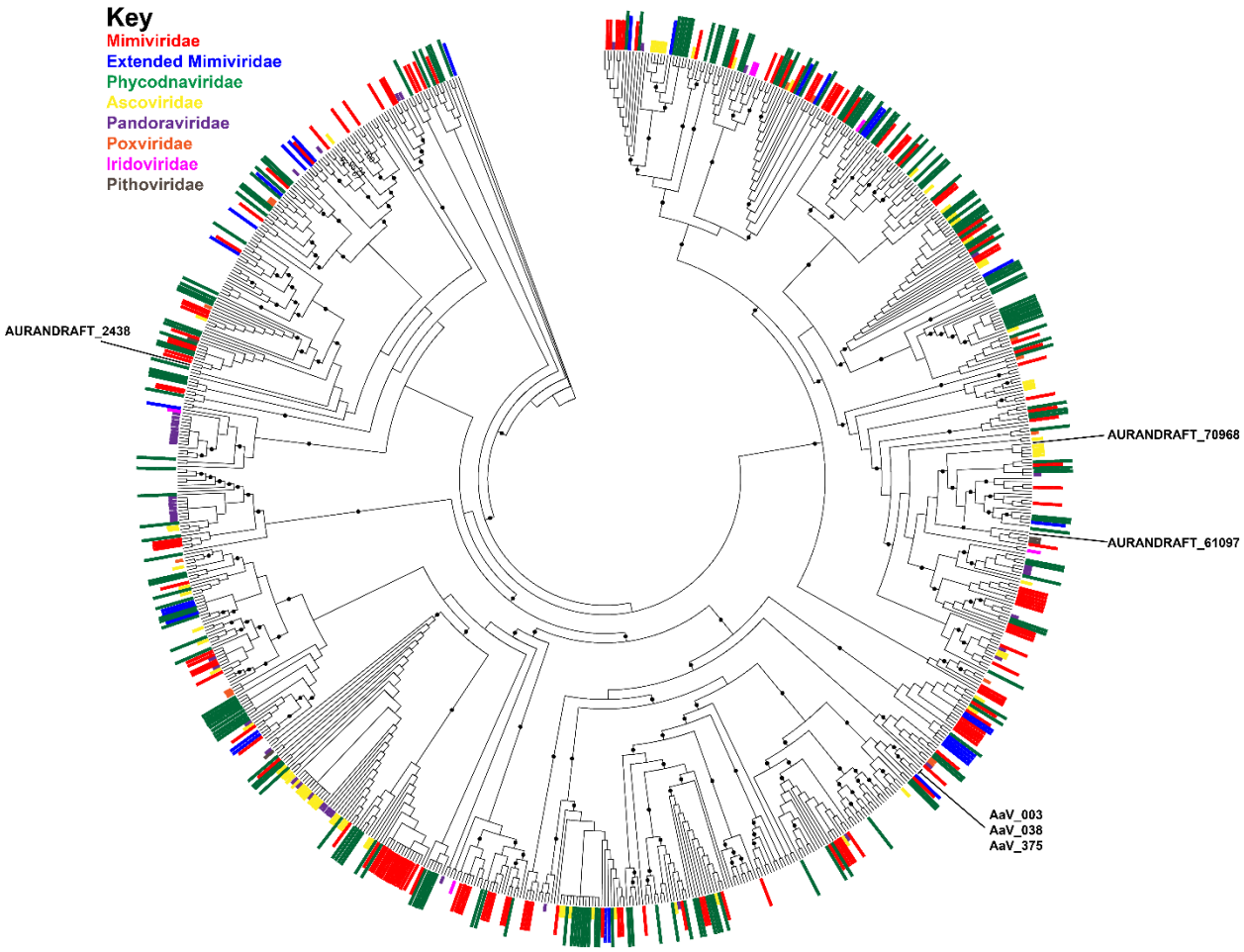


Figure 2.15. Maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral TARA Oceans metagenomic contigs. The classification of the contig by family based on the NCVOG on the contig are denoted by color bar. Reference characterized polysaccharide lyases from the Carbohydrate-Active Enzyme (CAZy) Database do not have a bar at the end of their branch. Black dots represent bootstrap values > 0.5.

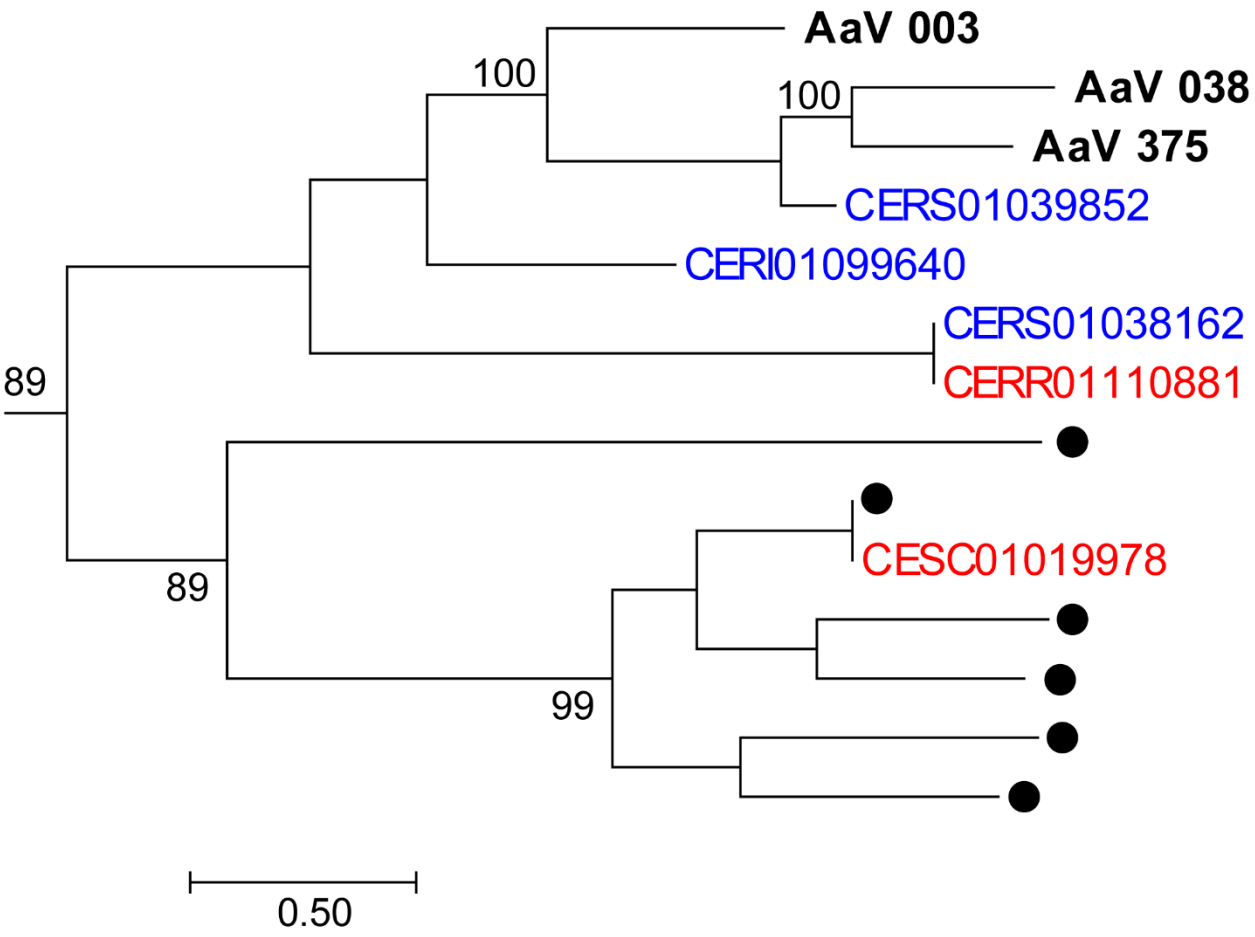


Figure 2.16. Subtree of maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral TARA Oceans metagenomic contigs containing the AaV pectate lyases. Contigs containing a NCVOG in the AaV genome are in blue, while those lacking a NCVOG in AaV are in red. Reference polysaccharide lyases from the Carbohydrate-Active Enzymes (CAZy) Database are denoted by black dots.

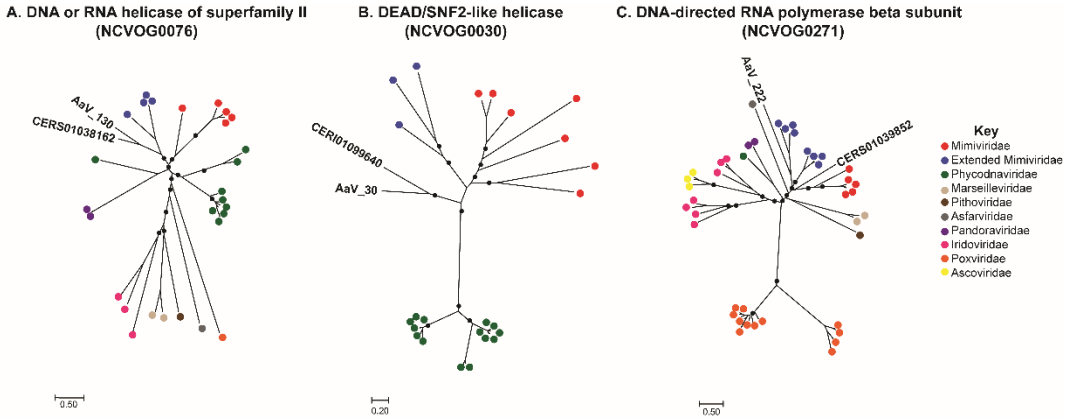


Figure 2.17. Maximum likelihood phylogenetic trees of NCVOG found on TARA Oceans metagenomic contigs within AaV subtree of the maximum-likelihood phylogenetic placement of polysaccharide lyases from putative viral contigs. Viral families are denoted by colored dots. Black dots on branches represent bootstrap values > 0.5.

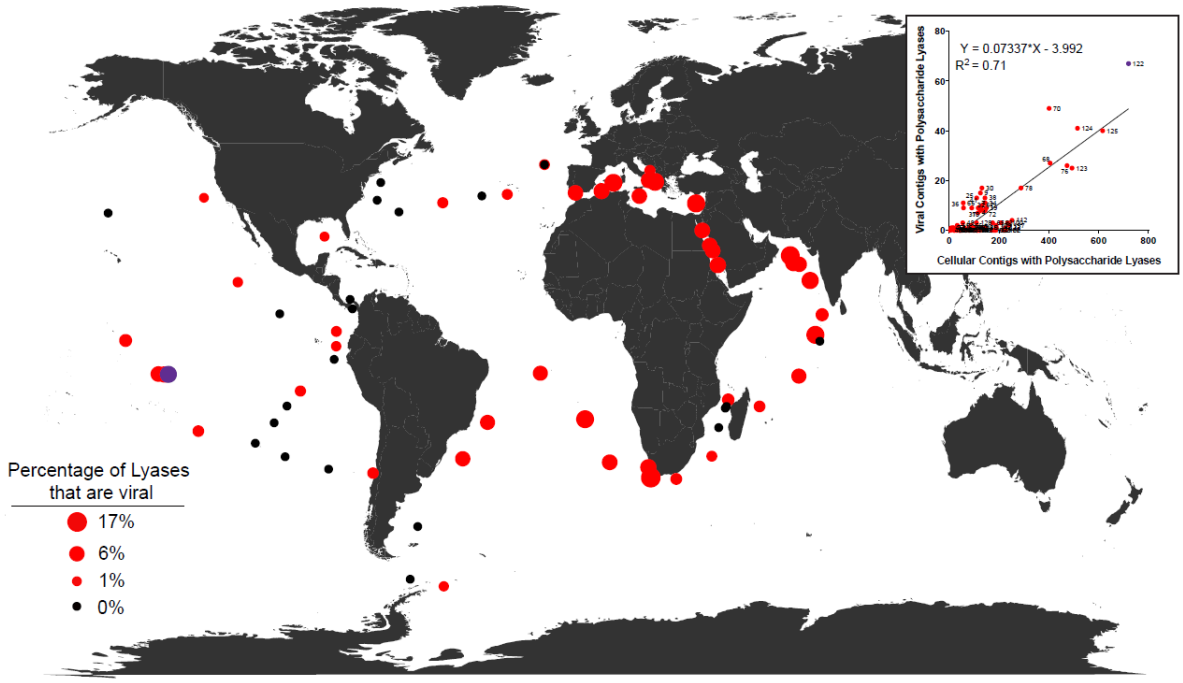


Figure 2.18. Contribution of polysaccharide lyases found on viral contigs to total contigs containing polysaccharide lyases mapped to TARA Oceans stations. Black dots represent stations where no viral contigs containing polysaccharide lyases were found. Inset: TARA Oceans metagenomic contigs containing polysaccharide lyases by station.

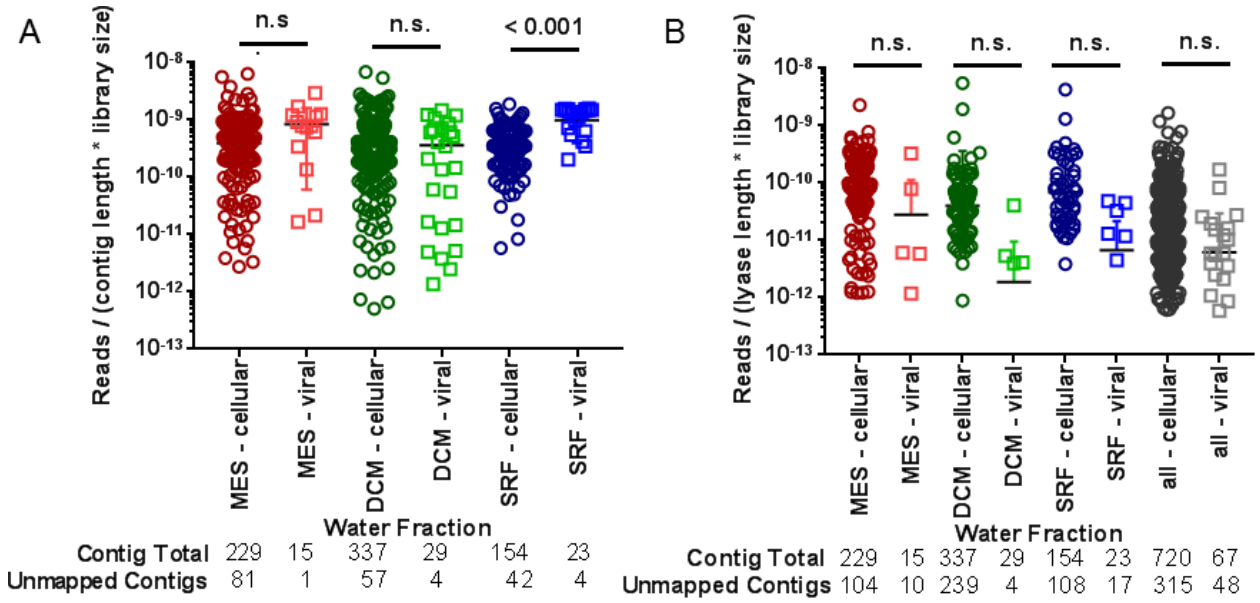


Figure 2.19. Normalized mapped reads for TARA Oceans Station 122. A) Metagenomic reads mapped to the entire contigs. Reads from the metagenome the contig was assembled from were used to map back to each contig. B) Metatranscriptomic reads mapped to the polysaccharide lyases found on each contig. Reads from the metatranscriptomes at the same depth of the water column were used to map back to the polysaccharide lyases found on each contig.

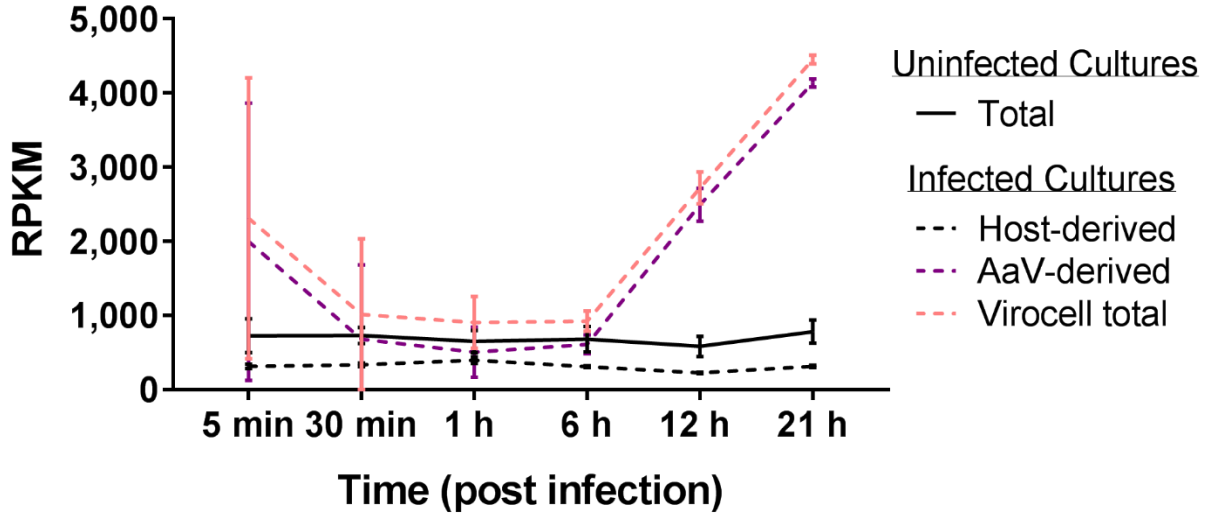


Figure 2.20. Combined RPKM values of polysaccharide lyases over the 21 hour infection cycle. The *A. anophagefferens* proteins used were AURANDRAFT_70968, AURANDRAFT_61097, and AURANDRAFT_2438, and the AaV proteins used were AaV_003, AaV_038, AaV_375.

Chapter Three

Influence of light on the infection of *Aureococcus anophagefferens* CCMP 1984 by a "giant virus"

Publication note and author contributions

Publication note

This chapter has been published in *Plos One* and authored by Eric R. Gann, P. Jackson Gainer, Todd B. Reynolds, and Steven W. Wilhelm. The citation for this work is:

Gann ER, Gainer PJ, Reynolds TB, Wilhelm SW. (2020) Influence of light on the infection of *Aureococcus anophagefferens* CCMP 1984 by a "giant virus". **PLOS ONE**15(1): e0226758.

Author contributions

Experiments were designed by Eric R. Gann, Todd B. Reynolds, and Steven W. Wilhelm. Eric R. Gann and P. Jackson Gainer optimized qPCR assay. All experiments were performed by Eric R. Gann. This chapter was written by Eric R. Gann and Steven W. Wilhelm, with all authors participating in providing comments in the initial and revised drafts.

Abstract

The pelagophyte *Aureococcus anophagefferens* has caused recurrent brown tide blooms along the northeast coast of the United States since the mid-1980's, and more recently spread to other regions of the globe. These blooms, due to the high cell densities, are associated with severe light attenuation that destroys the sea grass beds which provide the basis for many fisheries. Data collected by transmission electron microscopy, PCR, and metatranscriptomic studies of the blooms, support the hypothesis that large dsDNA viruses play a role in bloom dynamics. While a large (~140 nm) icosahedral virus, with a 371 kbp genome, was first isolated more than a decade ago, the constraints imposed by environmental parameters on bloom infection dynamics by Aureococcus anophagefferens Virus, (AaV) remain unknown. To investigate the role light plays in infection by this virus, we acclimated *A. anophagefferens* to light intensities of 30 (low), 60 (medium) or 90 μmol photons m^{-2} s^{-1} (high) and infected cultures at these irradiance levels. Moreover, we completed light shift experiments where acclimated cultures were exposed to even lower light intensities (0, 5, and 15 μmol photons m^{-2} s^{-1}) consistent with irradiance found during the peak of the bloom when cell concentrations are highest. The abundance of viruses produced per lytic event (burst size) was lower in the low irradiance acclimated cultures compared to the medium and high acclimated cultures. Transferring infected cultures to more-limiting light availabilities further decreased burst size and increased the length of time it took for cultures to lyse, regardless of acclimation irradiance level. A hypothetical mechanism for the reduced efficiency of the infection cycle in low light due to ribosome biogenesis was predicted from pre-existing transcriptomes. Overall, these studies provide a framework for understanding light effects on infection dynamics over the course of the summer months when *A. anophagefferens* blooms occur.

Introduction

The pelagophyte *Aureococcus anophagefferens* has caused recurrent brown tide blooms off the eastern coast of the United States since 1985 (Sieburth *et al*., 1988), where blooms can achieve over 10^6 cells mL^{-1} during summer months (Gastrich *et al*., 2002). These blooms typically occur when dissolved inorganic nutrient levels decline, and dissolved organic nutrient levels increase (Gobler *et al*., 2011). Although not known to produce compounds toxic to humans, *A. anophagefferens* blooms cause severe light attenuation that can kill the sea grass beds which provide valuable nurseries and refuge for fish. It is also believed that *A. anophagefferens* may produce compounds toxic to bivalves (Gobler and Sunda, 2012). While blooms were once thought to be constrained to the northeastern seaboard of the USA, they are now becoming a global problem with *A. anophagefferens* spreading down the US eastern coast (Popels *et al*., 2003), as well as to other countries (Probyn *et al*., 2001; Zhang *et al*., 2012), possibly *via* transport in ballast water of ships (Popels and Hutchins, 2002).

During the initial characterization of blooms, natural populations of *A. anophagefferens* were shown, by transmission electron microscopy (TEM), to be infected with large (~140 nm) virus particles (Sieburth *et al*., 1988). Later, TEM studies revealed low percentages of infected cells within natural populations during the onset and peak of the bloom, but during bloom collapse over one-third of the natural *A. anophagefferens* population were infected (Gastrich *et al*., 2004). A virus consistent in both size (~140 nm) and morphology (icosahedral) to both the initial and subsequent studies of the bloom was isolated (Moniruzzaman *et al*., 2014; Moniruzzaman *et al*., 2018). The Aureococcus anophagefferens Virus, AaV, belongs to the algal branch of the *Mimiviridae* family within the Nucleocytoplasmic large dsDNA viruses group (Moniruzzaman *et al*., 2014). AaV is considered to be a giant virus due to its large particle size,

as well as the large genome (371 kbps) it contains (Moniruzzaman *et al*., 2014). Algae-infecting *Mimiviridae* similar to AaV have been detected by PCR over the course of brown tide blooms (Moniruzzaman *et al*., 2016), and transcripts of viral origin are detected within bloom meta-transcriptomes (Moniruzzaman *et al*., 2017), suggesting this virus to be relevant to understanding the natural populations infecting the blooms. As in several other marine algal blooms (e.g., Vardi *et al*., 2012), it is hypothesized that viruses infecting *A. anophagefferens* play an important role in bloom dynamics and collapse.

To constrain the bloom, viruses infecting *A. anophagefferens* must produce enough progeny to remain infectious to lyse the entire population of cells and maintain infectious populations at high enough abundances to encounter susceptible host strains. Changes in productivity (*i.e.* burst size or length of infection cycle) of viral infections and decay rates of free virus particles modulate the dynamics of the host viral system. Many studies in the laboratory and the environment have shown certain environmental factors cause increased viral decay rates although these are virus specific (Wilhelm *et al*., 1998; Long and Short, 2016), but how the changing environment influences the infection cycle within a cell is less understood. Abiotic factors such as irradiance level (Van Etten *et al*., 1983; Baudoux and Brussaard, 2008), temperature (Piedade *et al*., 2018), and nutrients (Clasen and Elser, 2007) have been shown to negatively influence the infection cycle dynamics in various algal virus systems. For *A. anophagefferens*, it is known that decreased temperatures and irradiance increase the time it takes for cultures to lyse, although there was no report of effects on viral abundances (Gobler *et al*., 2007). Moreover, increasing the multiplicity of infection (MOI) causes a reduction in the burst size in cultures (Brown and Bidle, 2014).

Lab experiments of *A. anophagefferens* have shown irradiance levels influence growth (Milligan and Cosper, 1997), transcription of many core metabolic pathways (Frischkorn *et al*., 2014), and alter uptake of nutrients (Pustizzi *et al*., 2004). The current study aimed to understand the effects of irradiance levels on the infection cycle that were caused by changes to the *A. anophagefferens* cell. This is relevant to the *A. anophagefferens* system as there is severe light attenuation over the course of the bloom (Lomas *et al*., 1996), and *A. anophagefferens* is well adapted for low light (Milligan and Cosper, 1997; Gobler *et al*., 2011). These culture-based studies provide a framework for the changing host virus dynamics that occur on a population level over the summer months (as increased cell concentration increases light attenuation). Specifically, we examined how the length of the infection cycle and burst size varied in cultures acclimated to decreased irradiance levels. Moreover, as the light field shifts during the growing season, we completed mock-shift experiments consistent with light attenuation in the bloom to determine how decreases in light availability during infection could shape infection outcome. Finally, we determined differences in infection dynamics in populations transitioning from one physiological state to another due to shifting light compared to pre-acclimated populations. These results illustrate the importance of considering the varying irradiance levels at an individual or population level and the impact on infection cycle dynamics.

Methods

Culture conditions

Non-axenic *Aureococcus anophagefferens* CCMP1984 was grown at 19º C, on a 14:10 light dark cycle in modified $ASP_{12}A$ growth medium (Gann, 2016). Cultures were acclimated to three different irradiance levels (high: 90 µmol photons m^{-2} s^{-1}, medium: 60 µmol photons m^{-2} s^{-1}, and

low: 30 µmol photons m^{-2} s^{-1}) for at least three successive transfers at each irradiance level prior to any experiments. *A. anophagefferens* concentrations were determined using a GUAVA-HT6 flow cytometer (MilliporeSigma, Burlington, MA) with abundances for cells gated on red chlorophyll fluorescence and forward scatter. Mean forward scatter (FSU-H) for the *A. anophagefferens* gated population was also recorded. Cell concentrations were either measured directly after sampling, or first fixed in 0.5% glutaraldehyde (for the one-step Experiment) and stored at 4º C before being measured within 8 h (Moniruzzaman *et al*., 2018). Cultures (20 mL) were infected with fresh AaV lysate that was concentrated using a Lab-scale Tangential Flow Filtration System (Fisher Scientific, Waltham, USA) equipped with a Durapore™ 30kDa Pellicon XL Filter (MilliporeSigma, Burlington, MA). Concentrates were passed through a 0.45-µm pore-size syringe filter (Millex-HV 0.45µm nominal pore-size PVDF, Millipore Sigma, Burlington, USA) before introduction to cultures at a multiplicity of infection (MOI) of ~100 total particles per cell, which is ~1 infectious particle per cell (see below).

Plaque assay

Assessment of infectious virus particles was determined using a previously described plaque assay (Gann, 2017) (Schroeder *et al*., 2002). The bottom agar was $ASP_{12}A$ with 1% low melting agarose (Fischer Scientific, Waltham, USA), while the top agar was $ASP_{12}A$ with 0.4% low melting agarose. Both were sterilized by Tyndallization. Each plate required ~50 mL of concentrated one-week old *A. anophagefferens* culture. *A. anophagefferens* cultures were pelleted by centrifugation (2000xg, 5 minutes), and resuspended in $ASP_{12}A$ to concentrate the cells fifty-fold. 100 µL of diluted AaV were mixed with 900 µL concentrated *A. anophagefferens* and immediately 3 mL of top agar at 30º C was added and poured onto 18 mL of solidified

bottom agar. Plates were incubated at 19º C on a 14:10 light dark cycle at an irradiance of 90 µmol photons m^{-2} s^{-1} for one to two weeks in a sealed plastic bag with a damp paper towel to prevent the plates from drying out. Plaques were enumerated and the concentration of plaque forming units determined using the following equation:

PFUs/mL = number of plaques / (Dilution * volume plated)

Most probable number assay

A most probable number (MPN) assay for *A. anophagefferens* was developed in parallel with the plaque assay as an independent determination of the number of infectious virus particles (vis a vis Cottrell and Suttle, 1995). Aliquots from one-week old *A. anophagefferens* cultures (150 µL) were dispensed into round bottom 96-well plates (Corning, Corning, USA). AaV was serially diluted (at either 1:4 or 1:5 dilutions) twelve times. The final row for each plate was serially diluted with sterile growth medium as a non-infected control. Plates were incubated at 19º C on a 14:10 light dark cycle at an irradiance of 90 µmol photons m^{-2} s^{-1} for one to two weeks in a sealed plastic bag with a damp paper towel to prevent drying out. *A. anophagefferens* cells in wells were enumerated using in a Cytation 5 plate reader (BioTek, Winooski, USA) set to excitation λ = 614 nm, emission λ = 670 nm as a proxy for biomass within each well. Wells that contained cultures that lysed completely had comparable readings to growth media alone under these settings. The most probable number of infectious particles was calculated by comparing lysed *vs*. unlysed wells at each dilution using previously developed software (Jarvis *et al*., 2010).

Enumeration of free AaV particles by quantitative PCR

A quantitative PCR (qPCR) assay was developed to enumerate gene copy number for the viral major capsid protein (MCP). The forward primer, MCP_F, (5'-TGGATGCACATCTGGAA) was positioned where degenerate primers were designed to amplify all known algal *Mimiviridae* members (Moniruzzaman *et al*., 2016). The forward and reverse primer, MCP_R3 (5'-CAATAAGGGGAAGGGCAAG), amplifies a 196 base pair product that is specific to AaV, as confirmed by sequencing. Reaction mixtures for PCR contained: 2 µL of 0.45m-filtered (Millex PVDF syringe filter, MilliporeSigma, Burlington, USA) AaV lysate, 0.5 µL of 100 mM MCP_F, 0.5 µL of 100 mM MCP_R3, 12.5 µL ABsolute qPCR SYBR Green Mix (Thermo Fisher Scientific, Waltham, USA), and 9.5 µL sterile Milli-Q water. The reaction was performed on a DNA Engine Opticon 2 (Bio-Rad, Hercules, USA) using the following conditions: (i) 95° C for 10 min, (ii) 95° C for 30 sec, (iii) 55° C for 30 sec, (iv) 72° C for 30 sec, (v) repeat II – IV 30 times, (vi) 72° C for 10 min, hold at 4° C. Threshold cycle numbers were determined using Opticon Monitor 3 (Bio-Rad, Hercules, USA).

To convert threshold cycle number to absolute gene copy number the 196 base pair product from the above primers was ligated into the pCR 4-TOPO vector (Thermo Fisher Scientific, Waltham, USA) using the 3' A overhangs generated by the PCR reaction, and transformed *via* heat shock into One Shot TOP10 chemically competent *E. coli* (Thermo Fisher Scientific, Waltham, USA). Plasmids were purified using the QIAprep Spin Mimiprep Kit (Qiagen, Hilden, Germany), and the concentration of DNA was quantified using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA). For each qPCR plate, reactions containing 10-fold serial dilutions of the purified circular plasmid were also run as

standards. The logarithmic trendline of the diluted standards was then used to convert threshold cycle number to copies of MCP.

Burst sizes were calculated from infection experiments by first determining the difference between the stable maximum (time point B) and minimum (time point A) virus densities. In our experiments in section 'Effects of pre-infection light intensities on AaV virus production', this corresponded to days 1 and 3. For our experiments in section 'Effect of reduced light availability post infection' this corresponded to days 1 and 2 for all treatments except the low light acclimated culture maintained at 30 µmol photons m^{-2} s^{-1}, which was days 1 and 3. We then normalized these data to viruses per cell values (i.e., burst size) based on the abundance of cells lysed. In total the following equation was used:

Burst size = (virus concentration B – virus concentration A)/(cell concentration A – cell concentration B)

Enumeration of total virus particles

Virus particle densities were enumerated by epifluorescence microscopy as described previously (Noble and Fuhrman, 1998). Briefly, virus particles were separated from *A. anophagefferens* by filtration (Millex-HV 0.45µm PVDF syringe filter; MilliporeSigma, Burlington, USA). A working solution of SYBR Green DNA stain (Lonza, Basel, Switzerland) and an anti-fade solution were made fresh and stored in the dark before microscopy. To create the working stock of SYBR Green, it was diluted 400-fold in Milli-Q H_2O, and then filter sterilized (Nalgene 0.2µm SFCA Syringe Filter, Thermo Fisher Scientific, Waltham, USA). Anti-fade solution was made by diluting a filter sterilized (Nalgene 0.2µm SFCA Syringe Filter, Thermo Fisher Scientific, Waltham, USA) 1% w/v p-phenylenediamine in an autoclaved 50% - PBS/ 50% -

Glycerol solution (Noble and Fuhrman, 1998). Viruses were collected on a 25-mm diameter, 0.02-µm Anodisc filter (Whatman, Maidstone, UK). Anodiscs were placed on 18 µL of diluted SYBR green and incubated for ten minutes in the dark. The filters were placed between 36 µL of the anti-fade solution on a glass slide and glass cover slip. Random views were counted on a Leica DM5500 microscope at 1000x magnification with a L5 filter cube (Leica, Wetzler, Germany) per slide to determine the number of viruses per grid. The concentration of virus particles was determined by averaging the random views together and accounting for the size of the grid using the following formula:

$$\text{Viruses/mL} = P_f * (A_a/(A_g * V_f)) * D$$

Where P_f is the average viruses counted per field, A_a is the total filterable area of the Anodisc filter, A_g is the area of the eyepiece grid, V_f is the volume filtered, and D is the dilution factor.

Bioinformatic predictions for effects of light on infection

We explored publicly available transcriptomics data to look for a mechanistic link between virus-effects and the results described below. We accessed a previous *de novo* transcriptome which addressed differences between 100 µmol photons m^{-2} s^{-1} and 30 µmol photons m^{-2} s^{-1} acclimated cultures of *Aureococcus anophagefferens* CCMP1850 (Frischkorn *et al*., 2014). Genes that were significantly (Analysis of Sequence Counts (Wu *et al*., 2010), $p < 0.05$) differentially expressed in the low light *vs.* high light conditions and had the best BLAST hit to the *Aureococcus anophagefferens* CCMP1984 genome were compared to the significantly (edgeR (Robinson *et al*., 2010), FDR $p < 0.05$) differentially expressed genes from a transcriptome of *Aureococcus anophagefferens* CCMP1984 over an AaV infection cycle

(Moniruzzaman *et al*., 2018) to determine overlap. Predicted function and cellular processes associated with each gene were assigned using KEGG pathways (Kanehisa *et al*., 2017).

Statistics

Statistical analyses were performed in Prism 7.03 (GraphPad, San Diego, USA) or R software v. 3.4.0 (R Team, 2014). Differences between doubling time and burst sizes in cultures with different acclimation irradiance levels were analyzed using a one-way ANOVA followed by Tukey's HSD post-hoc testing. To determine differences in enumeration methods (qPCR, SYBR green staining, MPN) we utilized a single time point in an experiment (see Effects of pre-infection light intensities on AaV virus production), to enumerate viruses by all three methods. This point was chosen as the epifluorescence counts were not statistically different from one another. The effects of treatment and enumeration method were analyzed by two-way ANOVA. Post-hoc multiple comparisons were adjusted with Tukey's HSD. The adjusted p-values are reported in table 3.1 (All figures and tables are located in the Chapter Three appendix). Differences between treatments in the post-infection light shift experiment were analyzed using a one-way ANOVA followed by Tukey's HSD post-hoc testing. Adjusted p-values are reported in table 3.2.

Results

Infectivity vs particle abundance of Aureococcus anophagefferens Virus particles

As part of this study, we first needed to develop protocols to enumerate both the total concentration of AaV particles, as well as the infectious fraction of the particles or those that are able to produce an active infection once adsorbed to the cell. In the developed qPCR assay, there was a linear relationship between the threshold cycle and the abundance of free viruses determined by epifluorescence microscopy ($C(t) = -1.492 \times \ln(\text{viruses mL}^{-1}) + 36.372$, $R^2 = 0.998$) to a limit of detection between 1.5 – 15 viruses μL^{-1} (Figure 3.1). In our hands, ~1% of AaV particles enumerated by epifluorescence microscopy were infectious as determined by either the most probable number (MPN) (0.77% infectious, SD = 0.12%) or plaque assay (1.76% infectious, SD = 0.28%) (Figure 3.2). We used a single time point in an experiment (see Effects of pre-infection light intensities on AaV virus production) to validate our qPCR assay by also enumerating virus particles by epifluorescence microscopy, as well as by MPN. There was no statistically significant interaction between the effects of the treatment (acclimation irradiance level) and viral enumeration method on the number of viruses counted (Two-way ANOVA; $F(4,30)=0.678$, p-value = 0.613). The treatment did not strongly influence the viral counts (Two-way ANOVA: $F(4,30)=0.678$, p-value = 0.275), while the viral enumeration method did (Two-way ANOVA: $F(4,30)=0.678$, p-value = <0.0001). qPCR counts were higher (average 3.42x, SD = 0.24) than the microscopy counts for each the three different irradiance acclimated cultures (Figure 3.3).

Effects of pre-infection light intensities on AaV virus production

A. anophagefferens cultures were acclimated to three different irradiance levels (high: 90 µmol photons m^{-2} s^{-1}, medium: 60 µmol photons m^{-2} s^{-1}, and low 30 µmol photons m^{-2} s^{-1}). There was no difference in doubling time between medium (1.44 d) and high (1.39 d) irradiance cultures (one-way ANVOA; F=54.55; p = 0.180), while there was a significant difference in doubling time for low irradiance cultures (1.76 d) compared to the other acclimation irradiance levels (one-way ANOVA; F=54.55; p < 0.001) (Table 3.3 and Figure 3.4). Compared to uninfected cultures which continued to increase in cell numbers (Figure 3.5A), cultures infected during late logarithmic growth at these irradiance levels showed no difference in time-to-lysis (Figure 3.5B and 3.5C), or burst size (Table 3.3) for the high light and medium light acclimated cultures (one-way ANOVA; F = 5.3868, p = 0.714). However, there was an increase in the time to complete lysis of the culture (Fig 3.5B), and a reduction in burst size (Table 3.3), for the low irradiance condition compared to both medium light (one-way ANOVA; F = 5.3868, p = 0.085) and high light (one-way ANOVA; F = 5.3868, p = 0.021).

To determine if the different irradiance treatments influenced the percentage of particles produced during lysis that were infectious, MPN assays were performed three days post infection at a time point when the abundance of virus particles per treatment was not significantly different (Figure 3.3). For all three light acclimation levels, the percentage of the total particles determined by epifluorescence microscopy that were infectious as determined by MPN were not significantly different from one another (Figure 3.3, Table 3.3, and Table 3.1). As determined previously (Figure 3.2), a low percentage of the total virus population was infectious, with between 0.48 and 0.98% of particles being infectious at that timepoint (Table 3.3).

Effect of reduced light availability post infection

As there were differences in infection cycle dynamics between the high and low irradiance acclimated cultures, the effect of greater light limitation post-infection (light levels of 5 and 15 μmol photons m^{-2} s^{-1}) was explored. Light levels were chosen as they are relevant to populations late in the blooms when high cell densities occur (Lomas *et al*., 1996), as well as for a reduction in light that occurs while individual cells are sinking. These light levels were also determined to limit growth of uninfected cultures, as cell densities had not increased 4 d after being shifted to the decreased light (Figures 3.6A and 3.6B). When high or low irradiance acclimated cultures were infected and then transferred into either 15 or 5 μmol photons m^{-2} s^{-1}, cultures still maintained a productive infection. Temporally it took two days longer for ~99% of high irradiance acclimated cells to lyse in both limiting light conditions compared to infected cultures maintained at acclimated irradiance levels (Figure 3.6C). Virus production was observed for the first 2 days after the shift, but then virus abundance remained static (2.4 x 10^8 MCP copies mL^{-1}) (Figure 3.6E). Similar trends were seen with low irradiance acclimated cultures (Figures 3.6D and 3.6F). The lytic cycle was extended for cultures transferred into limiting light compared to the infected cultures maintained at acclimated irradiance levels (Figure 3.6D), as well as no further increase in viral abundance 2 days post infection (Figure 3.6F).

The number of viruses produced per lytic event was influenced by the irradiance level at which the infection occurred. When cultures started at high (90 μmol photons m^{-2} s^{-1}) or low (30 μmol photons m^{-2} s^{-1}) irradiance levels were transferred to 15 or 5 μmol photons m^{-2} s^{-1} conditions the burst sizes were not significantly different from one another, despite initially different irradiance levels (Fig 3.7A, Table 3.2). Cultures shifted to limiting light levels had a pronounced reduction in burst size, averaging 19.15% (SD = 16.67%) of the high irradiance

cultures infected and maintained at acclimating irradiance levels (Figure 3.7A, Table 3.2). A linear relationship (burst size = 20.53 x irradiance level + 246.9, $R^2 = 0.600$) between burst size and irradiance level was seen (Figure 3.7B). We note that while between different experiments trends we observed were conserved, similar treatments sometimes gave variable results. One example of this is that there is a significant difference between both 30 µmol photons m^{-2} s^{-1} (unpaired t-test between 30 µmol photons m^{-2} s^{-1} burst sizes from Table 3.3 and Fig 3.7A: $p = 0.019$) and 90 µmol photons m^{-2} s^{-1} (unpaired t-test between 90 µmol photons m^{-2} s^{-1} burst sizes from Table 3.3 and Fig3.7A: $p = 0.0095$). Yet, when burst size is normalized to the 90 µmol photons m^{-2} s^{-1} cultures, the linear relationship between burst size and irradiance level is still maintained (Figure 3.8).

It appeared there were two phases of viral production in cultures transferred to lower light post infection. The first phase was characterized by cells lysing concurrent with an increase in free virus concentration (Day 0-2) (Figures 3.6D and 3.6F), while in the second phase the decline in *A. anophagefferens* cells did not produce more viruses. Over the course of the first two days the mean forward scatter from the flow cytometric estimates of the population decreased before remaining constant for the duration of the experiment for both the uninfected (Figure 3.7C), and infected populations (Figure 3.7D).

Finally, to determine if the infection cycle required light to complete, we infected high, medium, and low irradiance acclimated cultures and transferred them into the dark. There was no increase in host cell concentration in uninfected cultures (Figure 3.9A), unlike cultures kept at the acclimation irradiance (Figure 3.5A), nor was there a large decrease in infected cultures (Figure 3.9B). This lack of cell death was accompanied by no increase in virus concentration in

any of the cultures regardless of light acclimation condition (Figure 3.9C). Virus abundance decreased to 0.16 – 1.61% of starting concentration by day 10.

Effects of pre-infection light reduction on the progress of AaV infection

To determine whether this transition from irradiance levels where the cells could actively grow to where they only persisted (Figures 3.6A and 3.6B) caused the different phases of the infection cycle we observed, cultures acclimated to either high or low irradiance levels were pre-acclimated to 5 µmol photons m^{-2} s^{-1} for one day to allow the population to shift to the altered physiological state, as determined by a decrease in forward scatter (Table 3.4). Cultures were then infected at this limiting irradiance acclimation. *A. anophagefferens* concentrations decreased over the four day experiment in cultures that were kept at their acclimation irradiance (Figure 3.10A), while there was no decrease in host concentration in cultures pre-acclimated and infected at 5 µmol photons m^{-2} s^{-1}. The infection cycle of the limiting irradiance acclimated cultures was determined to be longer (24-30 h) than those cultures maintained at high irradiance (12-18 h) (Fig 3.10C), as determined by an increase in free virus concentration. The cultures pre-acclimated to the 5 µmol photons m^{-2} s^{-1} of light for one day did not show an increase in viruses within the first two days post infection as in the experiment where cultures were not pre-acclimated (Figures 3.6E, 3.6F, and 3.10D). Free virus concentration increased after 30 – 36 h for high irradiance acclimated cultures that were pre-acclimated to limiting irradiance, while low irradiance acclimated cultures that were pre-acclimated to limiting irradiance produced new viruses between 36 - 48 h (Figure 3.10D).

Informatic examination of the effects of light and virus infection

To develop hypotheses concerning processes in *A. anophagefferens* that are different in various acclimated cultures, publicly available transcriptomic data were screened to compare the potential effects of light and the effects of the infection cycle. Published transcriptomes of *Aureococcus anophagefferens* CCMP1850 comparing cultures at two light levels (30 and 100 μmol photons m^{-2} s^{-1}) were examined (Frischkorn *et al*., 2014). Although that study used a different strain, 95% of their reads mapped to *Aureococcus anophagefferens* CCMP1984. To be more stringent, only assembled reads that had top BLAST hits to CCMP1984 were used (Frischkorn *et al*., 2014). Of 1,524 differentially expressed genes detected in low light *vs.* high light cultures, 49.2% of those genes were also differentially expressed in a transcriptome of *Aureococcus anophagefferens* CCMP1984 infected with AaV at 100 μmol photons m^{-2} s^{-1} in at least one time point (Moniruzzaman *et al*., 2018) (Table 3.5). To contrast gene expression between low light the infection cycle transcriptome and infection cycle transcriptomes, KEGG annotations and pathways (Kanehisa *et al*., 2017) were examined (Table 3.6). Pathways included those for metabolism of sugars, nucleic acids, and amino acids. Host nucleic acid scavenging and recycling have been hypothesized to be required for the AT rich AaV to take advantage of the metabolism of the GC rich *A. anophagefferens* (Brown and Bidle, 2014). Interestingly, the pathway that showed the strongest differences between the infection cycle and cells in 30 μmol photons m^{-2} s^{-1} was ribosome biogenesis. 13/15 ribosomal genes detected in the KEGG pathway were downregulated in the 30 μmol photons m^{-2} s^{-1} culture compared to 100 μmol photons m^{-2} s^{-1} (Table 3.6). We then searched for other genes involved in ribosome biogenesis and found 41 genes that were differentially expressed in our subset of both data sets (Table 3.7), most of which

were upregulated at the later time points of the infection cycle, while being down regulated in the 30 µmol photons m^{-2} s^{-1} transcriptome (Figure 3.11).

Discussion

The infectivity of Aureococcus anophagefferens Virus particles

Both most probable number and plaque assays were developed to determine the percentage of particles that were infectious, as this percentage varies by both the biological system (Bratbak *et al*., 1998) and assay (*i.e.,* MPN v. Plaque Assay) employed (Cottrell and Suttle, 1995). From our data, *ca* 1% of the virus particles observed by epifluorescence microscopy were determined to be infectious, a proportion similar to viruses infecting *Heterosigma akashiwo* (Tarutani *et al*., 2006), but much less than other systems (Van Etten *et al*., 1983; Maat *et al*., 2016). In parallel with epifluorescence derived direct counts, a qPCR assay was developed with hopes of eventually deploying it in environmental systems. There was a significant difference between the two types of direct counts. This was not necessarily surprising, as it has previously been reported that circular plasmid standards can overestimate copies within a sample (Hou *et al*., 2010). Yet while the qPCR approach may have produced a relative overestimate, it did provide a lower detection limit relative to epifluorescence. This sensitivity allowed for detection of free virus particle / genome production at 12-18 h, shorter than the 21-24 h infection cycle previously reported (Rowe *et al*., 2008), although we cannot rule out that changes in culture conditions in experiments completed over a decade apart may have played a role.

Effect of light acclimation on the progress of infection

A. anophagefferens thrives in bloom conditions when irradiance levels are reduced due to high cell densities (Lomas *et al*., 1996). The effect of light attenuation on viral production is important from an ecological standpoint as viruses are hypothesized to play a role in bloom collapse (Gastrich *et al*., 2004). To investigate the influence of irradiance, *A. anophagefferens* CCMP1984 was acclimated to high (90 µmol photons m^{-2} s^{-1}), medium (60 µmol photons m^{-2} s^{-1}), and low (30 µmol photons m^{-2} s^{-1}) irradiance levels. The growth rates determined for high and medium acclimated cultures (1.39 and 1.44 d, respectively), were similar (1.61 – 2.04 d) to those seen in several studies growing *A. anophagefferens* at irradiance levels between 90 and 100 µmol photons m^{-2} s^{-1} (Pustizzi *et al*., 2004; Frischkorn *et al*., 2014). In agreement with previous studies (Milligan and Cosper, 1997; Pustizzi *et al*., 2004), there was a reduced growth rate as irradiance level decreased. Doubling times between 1.98 and 2.58 d was reported with cultures growth at similar irradiance levels to the low light acclimation culture in this study (Pustizzi *et al*., 2004). Based on the nitrogen source present the difference between high and low light measured by Pustizzi *et al.* was between 0.10 and 0.53 d (Pustizzi *et al*., 2004), which is like the 0.37 d difference in doubling time between the low and high acclimated cultures. Infecting high and medium irradiance acclimated cultures resulted in > 99% of cultures lysing in 3 days, with statistically indistinguishable burst sizes. In the low irradiance acclimated cultures, delayed lysis of > 99% of the culture with a reduced burst size occurred, as seen in other systems (Maat *et al*., 2016). The increased time for lysis of low irradiance acclimated cultures may partially be explained by the increase in the length of the lytic cycle (see next section).

Effect of growth limiting light on infection

Irradiance during blooms with >5 x 10^5 cells/mL has been reported between 1.1 - 12.8 µmol photons m^{-2} s^{-1} (Lomas *et al*., 1996). To more robustly determine the effects of these limiting light levels, high and low irradiance acclimated cultures were infected and transitioned to 15 µmol photons m^{-2} s^{-1} and 5 µmol photons m^{-2} s^{-1}. Transferring cultures to growth limiting light increased the time of complete lysis of cultures. Burst sizes were reduced during these limiting light shift experiments, and were statistically the same, regardless of acclimation in 30 or 90 µmol photons m^{-2} s^{-1}. In the first two days there was a decrease in mean forward scatter of the *A. anophagefferens* cells dependent on the final light level of the shift. Algae placed low light have previously been shown to have a lower mean forward scatter, which is hypothesized to be due to a decrease in size (Maat *et al*., 2016; Piedade *et al*., 2018). We observed this when decreasing acclimation light levels in our system. It is not clear whether *A. anophagefferens* cells are on average smaller when maintained in low *vs.* high light: we note cell size did not change over the course of a 14 d study in the dark (Popels and Hutchins, 2002). At the least this change in mean forward scatter was indicative of a difference in the physiological state of the entire population, within the first day or two of the experiment compared to the later days. This provided a possible explanation for why free virus abundances do not continue to increase after two days following the shift to lower irradiance levels.

To determine how cells in limiting light respond to viral infection, cultures were transferred to 5 µmol photons m^{-2} s^{-1} for one day before infection. *A. anophagefferens* cell concentrations remained static after infection in contrast to cultures shifted to 5 µmol photons m^{-2} s^{-1} post infection. Production of new viruses was delayed in cultures pre-acclimated to 5 µmol photons m^{-2} s^{-1}, which differed from those not pre-acclimated. The acclimation irradiance level

(30 *vs.* 90 μmol photons m^{-2} s^{-1}) did impact the length of the lytic cycle as cells acclimated in high light produced new viruses 6 – 12 hours earlier, even after pre-acclimation to 5 μmol photons m^{-2} s^{-1}.

A. anophagefferens survives prolonged periods in the dark in culture, but with a previously reported 33% decline in viable cells over time (Popels and Hutchins, 2002). This is similar to the uninfected cell concentration decline over 5 days (22.34 – 35.48%). No productive infection occurred when *A. anophagefferens* was placed in the dark, regardless of acclimation irradiance. Inhibition in the dark has been hypothesized to be due to energy stores within the cell. Infections of the smaller *Micromonas pusilla* are never successful in the dark, but infections of *Phaeocystis globosa* in the dark are successful if cells were acclimated to 250 μmol photons m^{-1} s^{-1}, but not 25 μmol photons m^{-1} s^{-1} (Baudoux and Brussaard, 2008). *A. anophagefferens* is known to utilize its stores when in darkness. For example, during 14 d in darkness, carbohydrates and proteins per cell are reduced by 60% and 89%, respectively (Popels *et al*., 2007). Although we see no increase in free viral abundance in the dark, AaV concentrations do decline 81.6 - 89.6% within the first day, suggesting adsorption can occur in the dark which is in contrast to other systems where adsorption is greatly influenced by light (Thamatrakoln *et al*., 2019). This decrease in AaV may partly be to detritus or heterotrophic bacteria in the culture as well.

Bioinformatic insight into limitations to the virocell at low light

Preexisting *de novo* transcriptomes in a different *A. anophagefferens* strain (CCMP1850) grown in 30 or 100 μmol photons m^{-2} s^{-1} (Frischkorn *et al*., 2014) were used to hypothesize pathways that could be inversely transcribed compared to the transcriptome of an infected cell

(Moniruzzaman *et al*., 2018). We found many KEGG pathways (Kanehisa *et al*., 2017) that were differentially expressed in both 30 μmol photons m^{-2} s^{-1} and at least at one point in the infection cycle. The strongest difference detected from our analysis was ribosome biogenesis, where 41 genes were differentially expressed in both data sets. Under low light conditions, almost all the ribosome biogenesis genes were negatively differentially expressed, and all but two were upregulated at the end (12 and 21 h) of the infection cycle. This down regulation in low light may be due to decreased growth rates, as decreasing growth rate decreases ribosome biogenesis in *Saccharomyces cerevisiae* (Kief and Warner, 1981). In contrast, protein synthesis is essential for new viral progeny, so much so that during Herpes Simplex Virus 1 infections, ribosome production still occurs late in the infection cycle while much of the other cell's machinery has broken down (Simonin *et al*., 1997). This relationship between ribosome biogenesis and translation limits in low light infections has been hypothesized in cyanomyoviruses infecting the cyanobacterium *Synechococcus*. Shifting to higher light levels reduced the length of infections, with the host upregulating genes for ribosome biogenesis and translation (Puxty *et al*., 2018). If there is not enough ribosomal machinery within the infected cell to produce all of the required viral proteins for its infection cycle, the length of the cycle may increase. Alternatively, as the mechanism which triggers cell lysis is unknown in this system, the reduction in burst size could also be explained as not enough complete viruses were synthesized before cell lysis occurred.

Potential ecological impact

Overall, our data contribute to an enhanced understanding of differences in infection dynamics during the course of *A. anophagefferens* blooms as well as highlight potential differences in cells transitioning to lower light while infected, which could also provide a framework for sinking

cells. The role of viruses in sinking cells and carbon export is now being studied more thoroughly. Analyses of environmental and sequencing data collected from the TARA expeditions predicted that viruses may be important drivers of carbon export, potentially more so than organisms classically considered to be (Guidi *et al*., 2016). This is in agreement with previous data that viral infection increases rates of sinking in natural *Emiliania huxleyi* bloom mesocosoms (Vardi *et al*., 2012), as well as in *Heterosigma akashiwo* culture experiments (Lawrence and Suttle, 2004). In the lab, *Aureococcus* settles during growth and phytoplankton populations during *A. anophagefferens* blooms have been shown to settle to the bottom of mesocosms (Kana *et al*., 2004), suggesting understanding the shift from one environment to another in this system is environmentally relevant. We presented three scenarios corresponding to changing environments encountered over the course of an entire bloom event (Figure 3.12). Surface viral infections are most productive during pre-bloom conditions (scenario I) where no reduction in light has occurred. While these cells sink, infections can occur while cells still have light, but not once complete darkness is achieved. As the cell concentration increases, the irradiance begins to decrease due to shading causing surface infections to take longer and reductions in viruses produced per cell to occur (scenario II). At the peak of the bloom (scenario III), light levels are low (Lomas *et al*., 1996), and can severely limit growth. Virus infection dynamics are severely attenuated, but as the cellular concentrations begin to decrease, due to viral infection or other factors, virus infections begin to be more successful due to more light availability (scenario I & II), and with the higher cell densities early in the bloom, more contacts between viruses and the hosts can occur, potentially leading to bloom collapse (Gastrich *et al*., 2004).

Although irradiance levels vary considerably over the course of the bloom (Lomas *et al*., 1996), many other variables that change throughout the bloom (including temperature, nitrogen and phosphorus (Gobler *et al*., 2007)) which have the potential to influence infection success, should be considered. Reducing nutrients, such as nitrogen or phosphorus, have been shown to reduce the burst size and increase the time to lysis in other viruses (Clasen and Elser, 2007; Maat and Brussaard, 2016). An additive effect was shown in the *Phaeocystis globosa* system, where suboptimal light and low phosphorus reduced the PgV burst size to less than ten infectious particles per cell (Maat *et al*., 2016). As late in the bloom, *A. anophagefferens* appears to be phosphorus limited (Wurch *et al*., 2011; Wurch *et al*., 2019), and the amount of light *A. anophagefferens* has is greatly reduced (Lomas *et al*., 1996), could drastically alter the success of an *A. anophagefferens* infecting virus late in the bloom if additive like the *P. globosa* system. As these studies are conducted in the lab, more work is needed to estimate the production rate of viruses over the course of the bloom to determine how the dynamic abiotic factors influence the production of viruses, but these studies do provide evidence for the importance in considering abiotic factors in this bloom system. Also, future work understanding how the interplay between abiotic factors (*i.e.* low DIN:DON ratios), predators (such as viruses), and competing phytoplankton controls *A. anophagefferens* densities in areas where no blooms form will be important as *Aureococcus* spreads globally.

In summary, we showed the AaV infection cycle requires light to proceed, and is influenced by the past and current irradiance the host is in. The length of the infection cycle is increased when light is reduced, and the number of new viruses produced per cell is decreased in lower light, which we hypothesize is due in part to a reduced availability of translational machinery within the cell. We used the data generated in this study to provide an

environmentally relevant framework that describes the potential importance of changing light levels on infections on a population level (shading due to high cell densities) and on an individual level (a sinking cell).

Acknowledgments

We thank Robbie Martin, Gary LeCleir, Mohammad Moniruzzaman, and Samantha Coy for discussions about this work. This work was supported by the National Science Foundation IOS1922958, the *Gordon & Betty Moore Foundation* (Grant EMS4971) and the *Kenneth & Blaire Mossman Endowment* to the University of Tennessee.

Chapter Three Appendix

Table 3.1. Adjusted p-values comparing differences in AaV concentration on Day three (Figure 3.3) based on irradiance level and viral enumeration method as determined by two-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD. High light acclimated are cultured in 90 µmol photons m^{-2} s^{-1}, Medium light acclimated cultures are cultured in 60 µmol photons m^{-2} s^{-1}, low light acclimated cultures are cultured in 30 µmol photons m^{-2} s^{-1}.

Enumeration Method		qPCR			SYBR			MPN		
	Acclimation	High	Med	Low	High	Med	Low	High	Med	Low
qPCR	High	-	0.371	0.939	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Med		-	0.966	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	Low			-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
SYBR	High				-	0.999	0.996	0.14	0.088	0.093
	Med					-	>0.999	0.026	0.012	0.013
	Low						-	0.02	0.001	0.01
MPN	High							-	>0.999	>0.999
	Med								-	>0.999
	Low									-

Table 3.2. Adjusted p-values comparing differences in burst sizes based on irradiance level cells were transferred to after infection as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD (Fig 3.6). Low light corresponds to acclimation irradiance levels of 30 µmol photons m^{-2} s^{-1}, and high light corresponds to 90 µmol photons m^{-2} s^{-1}.

		Low			High		
Acclimated Light	Shifted Light	30 µmol photons m^{-2} s^{-1}	15 µmol photons m^{-2} s^{-1}	5 µmol photons m^{-2} s^{-1}	90 µmol photons m^{-2} s^{-1}	15 µmol photons m^{-2} s^{-1}	5 µmol photons m^{-2} s^{-1}
Low	30 µmol photons m^{-2} s^{-1}	-	0.205	0.617	0.052	0.167	0.394
	15 µmol photons m^{-2} s^{-1}		-	0.966	<0.001	>0.999	>0.999
	5 µmol photons m^{-2} s^{-1}			-	0.001	0.939	0.997
High	90 µmol photons m^{-2} s^{-1}				-	<0.001	<0.001
	15 µmol photons m^{-2} s^{-1}					-	0.999
	5 µmol photons m^{-2} s^{-1}						-

Table 3.3. Summary of acclimation conditions including host doubling time (Figure 3.4), mean forward scatter (FSU-H) determined by flow cytometry, burst size (MCP copies produced/*A. anophagefferens* cell lysed), and percent of particles determined to be infectious (from Fig 3.3). Standard deviation of each value is recorded within the parenthesis.

Acclimating Irradiance	**Low**	**Medium**	**High**
Uninfected Doubling Time (days)	1.763 (0.06)	1.437 (0.008)	1.387 (0.016)
Uninfected Mean FSU-H (Relative Units)	13.23 (0.58)	18.48 (0.99)	18.40 (1.25)
Burst size (Viruses produced / cells lysed)	335 (149)	670 (201)	761 (181)
Percentage of Particles that are infectious	0.982 (0.545)	0.701 (0.481)	0.480 (0.244)

Table 3.4. Forward Scatter (FSC-H) values for high (90 µmol photons m^{-2} s^{-1}) and low (30 µmol photons m^{-2} s^{-1}) light acclimated cultures either maintained at acclimating light or transitioned 5 µmol photons m^{-2} s^{-1} for one day. Standard deviation of each value is recorded within the parenthesis.

Pre-acclimation Light Level	90 µmol photons m^{-2} s^{-1}	90 µmol photons m^{-2} s^{-1}	30 µmol photons m^{-2} s^{-1}	30 µmol photons m^{-2} s^{-1}
Acclimation Light Level	90 µmol photons m^{-2} s^{-1}	5 µmol photons m^{-2} s^{-1}	30 µmol photons m^{-2} s^{-1}	5 µmol photons m^{-2} s^{-1}
Pre-acclimated FSC-H (Relative Units)	15.92 (0.05)	15.92 (0.05)	13.28 (0.06)	13.28 (0.06)
Acclimated FSC-H (Relative Units)	16.62 (0.82)	9.36 (0.22)	12.78 (0.06)	9.52 (0.19)

Table 3.5. Summary of genes used in analysis for differences between low light conditions (from Frischkorn *et al.* 2014) and the infection cycle (Moniruzzaman *et al.* 2018).

		Number of genes with at least 1 significant significantly expressed (>1.5 fold) in Infection				
	# Genes differentially expressed in LowL (Frischkorn et al. 2014)	Upregulated / non-significant in all time points	Mixed / non – significant all time points	Downregulated / non-significant in all time points	# Genes with Kegg IDs	# Genes with Kegg Pathways
LowL upregulated	256	27	16	113	55	37
LowL downregulated	1268	221	22	351	159	39

Table 3.6. KEGG pathways of *A. anophagefferens* genes differentially expressed in both the CCMP1850 transcriptome (Frischkorn *et al.* 2014) and the CCMP1984 infection cycle transcriptome (Moniruzzaman *et al.* 2018). Columns denote whether all of these genes differentially expressed over the course of the infection cycle at every time point are overexpressed (+), underexpressed (-), or a mixed (+/-).

	Down in			Up in				Down in			Up in		
KEGG Pathway	+	+/-	-	+	+/-	+	KEGG Pathway	+	+/-	-	+	+/-	+
ABC transporters			2		1	1	Glyoxylate and dicarboylate			1			
Amino sugar and nucleotide sugar	1		1		1	3	Metabolic pathways	7		7	4	2	14
Aminoacyl-tRNA biosynthesis	1					1	Monobactam biosynthesis			1			
Arachidonic acid metabolism				1			Nitrogen metabolism	1		1			
Arginine and Proline metabolism				1		1	One carbon pool by folate	1					
Ascorbate and aldarate metabolism						1	Pentose and glucuronate			1			1
Biosynthesis of amino acids	2			1		1	Pentose phosphate pathway				1		3
Biosynthesis of antibiotics	1		3	2		7	Peroxisome/Oxidative stress						2
Biosynthesis of secondary metabolites	3		2	2		4	Phagosome			2			1
Biotin metabolism						2	Phosphatidylinositol signaling				1		
Carbon metabolism	1		1	2		3	Porphyrin and chlorophyll	1					
Cyanoamino acid metabolism			1				Protein Export					1	
Cysteine and methioine metabolism	1						Protein processing in	3					
DNA replication			3			1	Purine metabolism			3	1		
Endocytosis	1		3			1	Ribosome				1		
Ether lipid metabolism	1						Ribosome biogenesis in	13		2			
Fatty acid biosynthesis	2					3	RNA degradation			2		1	2
Fatty acid metabolism	2					3	RNA polymerase	1		1			
Fructose and mannose metabolism	1			1			RNA transport	2					2
Galactose metabolism	1					1	Selenocompound metabolism	1		1			
Glutathione metabolism				2			Spliceosome	3					1
Glycerolipid metabolism			1			2	Starch and sucrose metabolism	1		1			
Glycerophospholipid metabolism			2				Sulfur metabolism	1		1			
Glycine, serine and threonine				1			Terpenoid backbone						1
Glycolysis/Gluconeogenesis	1					2	Ubiquitin mediated proteolysis			2			
Glycosaminoglycan degradation	1						Valine, Leucine, and isoleucine				1		

Table 3.7. Subset of the *Aureococcus anophagefferens* genes predicted to be significantly downregulated in low light (Frischkorn *et al.* 2014) that are related to ribosome biogenesis according to KEGG classifications. Significant fold change values over the course of the infection cycle transcriptome are shown. Significantly overexpressed are in blue while those in red are significantly under expressed.

		Significant fold change in Infection Cycle Transcriptome					
Accession	KEGG Gene Description	5 min	30 min	1 h	6 h	12 h	21 h
AURANDRAFT_10453	K13179 DDX18; ATP-dependent RNA helicase DDX18/HAS1 [EC:3.6.4.13]	-	-	-	1.86	2.81	2.73
AURANDRAFT_1296	K14787 MRD1; multiple RNA-binding domain-containing protein 1	-	-	-	2.13	3.87	3.07
AURANDRAFT_14446	K14841 NSA1; ribosome biogenesis protein NSA1	-	1.71	-	2.44	4.99	2.77
AURANDRAFT_1519	K14806 DDX31; ATP-dependent RNA helicase DDX31/DBP7 [EC:3.6.4.13]	-	-	-	-	1.95	-
AURANDRAFT_1767	K14859 SSF1_2; ribosome biogenesis protein SSF1/2	-	-	-	1.65	2.87	2.16
AURANDRAFT_19030	K14768 UTP7; U3 small nucleolar RNA-associated protein 7	-	-	-	1.65	1.98	1.69
AURANDRAFT_20690	K12619 XRN2; 5'-3' exoribonuclease 2 [EC:3.1.13.-]	-	-2.07	-	-	-	-
AURANDRAFT_22083	K06943 NOG1; nucleolar GTP-binding protein	-	-	-	-	2.35	1.52
AURANDRAFT_2416	K14777 DDX47; ATP-dependent RNA helicase DDX47/RRP3 [EC:3.6.4.13]	-	-	-	1.66	2.97	2.17
AURANDRAFT_24989	K12823 DDX5; ATP-dependent RNA helicase DDX5/DBP2 [EC:3.6.4.13]	-	-	-	1.99	2.91	2.67
AURANDRAFT_25894	K14553 UTP18; U3 small nucleolar RNA-associated protein 18	-	-	-	1.66	1.96	1.61
AURANDRAFT_26008	K14809 DDX55; ATP-dependent RNA helicase DDX55/SPB4 [EC:3.6.4.13]	-	-	-	1.52	2.48	2
AURANDRAFT_26477	K14776 DDX10; ATP-dependent RNA helicase DDX10/DBP4 [EC:3.6.4.13]	-	-	-	-	2.1	-
AURANDRAFT_26879	K14847 RPF2; ribosome production factor 2	-	-	-	2.01	2.84	2.48
AURANDRAFT_26912	K14811 DBP3; ATP-dependent RNA helicase DBP3 [EC:3.6.4.13]	-	-	-	1.65	3.18	1.97
AURANDRAFT_27316	K11884 PNO1; RNA-binding protein PNO1	-	-	-	2.35	5	3.02
AURANDRAFT_283	K14569 BMS1; ribosome biogenesis protein BMS1	-	-	-	-	1.64	1.9

Table 3.7. Continued.

AURANDRAFT_28540	K07178 RIOK1; RIO kinase 1 [EC:2.7.11.1]	-	-	-	-	3.44	1.99
AURANDRAFT_31375	K14780 DHX37; ATP-dependent RNA helicase DHX37/DHR1 [EC:3.6.4.13]	-	-	-	-	2.21	1.83
AURANDRAFT_33432	K19306 BUD23; 18S rRNA (guanine1575-N7)-methyltransferase [EC:2.1.1.309]	-	-	-	1.72	2.92	1.59
AURANDRAFT_33949	K07179 RIOK2; RIO kinase 2 [EC:2.7.11.1]	-	-	-	1.58	4.07	2.57
AURANDRAFT_37654	K14775 UTP30; ribosome biogenesis protein UTP30	-	-	-	1.77	3.06	2.64
AURANDRAFT_38045	K14549 UTP15; U3 small nucleolar RNA-associated protein 15	-	-	-	1.9	2.39	2.27
AURANDRAFT_4268	K14843 PES1; pescadillo	-	-	-	1.64	3.98	2.93
AURANDRAFT_4711	K14831 MAK16; protein MAK16	-	-	-	1.67	3.35	2.31
AURANDRAFT_52052	K14857 SPB1; AdoMet-dependent rRNA methyltransferase SPB1 [EC:2.1.1.-]	-	-	-	1.8	2.55	2.41
AURANDRAFT_59367	K14835 NOP2; 25S rRNA (cytosine2870-C5)-methyltransferase [EC:2.1.1.310]	-	-	-	1.61	2.14	2.25
AURANDRAFT_59370	K14824 ERB1; ribosome biogenesis protein ERB1	-	-	-	1.77	2.45	2.09
AURANDRAFT_60066	K14842 NSA2; ribosome biogenesis protein NSA2	-	-	-	1.76	2.45	2.03
AURANDRAFT_60268	K14191 DIM1; 18S rRNA (adenine1779-N6/adenine1780-N6)-dimethyltransferase	-	-	-	2.3	3.45	2.65
AURANDRAFT_62634	K14774 UTP25; U3 small nucleolar RNA-associated protein 25	-	1.5	-	1.78	2.05	1.64
AURANDRAFT_63995	K14557 UTP6; U3 small nucleolar RNA-associated protein 6	-	1.56	-	-	1.85	-
AURANDRAFT_64464	K11883 NOB1; RNA-binding protein NOB1	-	-	-	-	2.59	-
AURANDRAFT_65930	K14554 UTP21; U3 small nucleolar RNA-associated protein 21	-	1.6	-	1.73	1.71	-
AURANDRAFT_68377	K16912 LAS1; ribosomal biogenesis protein LAS1	-	1.71	-	1.63	-	-
AURANDRAFT_68959	K14572 MDN1; midasin	-	-	-	-	-2.12	-2.92
AURANDRAFT_70174	K07562 NMD3; nonsense-mediated mRNA decay protein 3	-	-	-	1.76	3.2	2.25
AURANDRAFT_71183	K14521 NAT10; N-acetyltransferase 10 [EC:2.3.1.-]	-	-	-	-	-	1.58

Table 3.7. Continued.

AURANDRAFT_71380	K14561 IMP4; U3 small nucleolar ribonucleoprotein protein IMP4	-	-	-	-	2.24	3.05
AURANDRAFT_72347	K14537 NUG2; nuclear GTP-binding protein	-	-	-	1.5	2.47	1.63
AURANDRAFT_72521	K14848 RRB1; ribosome assembly protein RRB1	-	-	-	1.68	2.59	1.81

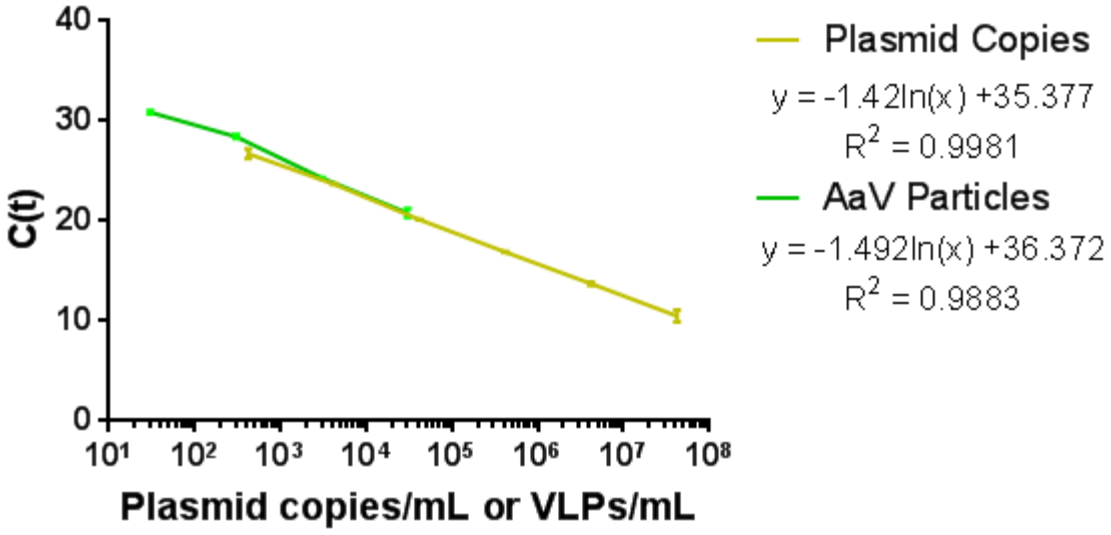


Figure 3.1. Standard curve validating qPCR method to enumerate free AaV particles. Concentration of AaV particles was determined by epifluorescence microscopy, and concentration of plasmids was determined by conversion of DNA concentration to copy number. Points are for n=3 biological replicates ± SD, with technical qPCR reaction replicates.

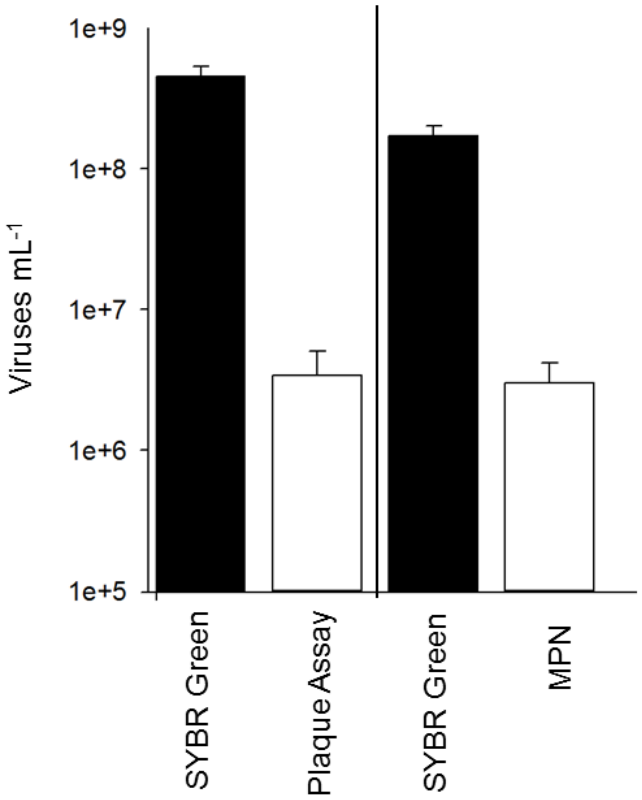


Figure 3.2. Comparison between total virus particle counts as determined by SYBR green epifluorescence microscopy and infectious particles determined by plaque assay or most probable number. For each, triplicate replicate cultures were infected, and once the cultures cleared, aliquots were taken for either type of enumeration. Plaque assays for each culture were done at two dilutions in duplicate. MPN assays had duplicate plates with 7 replicates per plate. There was a significant difference between the microscopy counts and the infectious particle counts (SYBR Green v. Plaque assay, pair t-test, $p < 0.01$; SYBR Green v. MPN, paired t-test, $p < 0.001$).

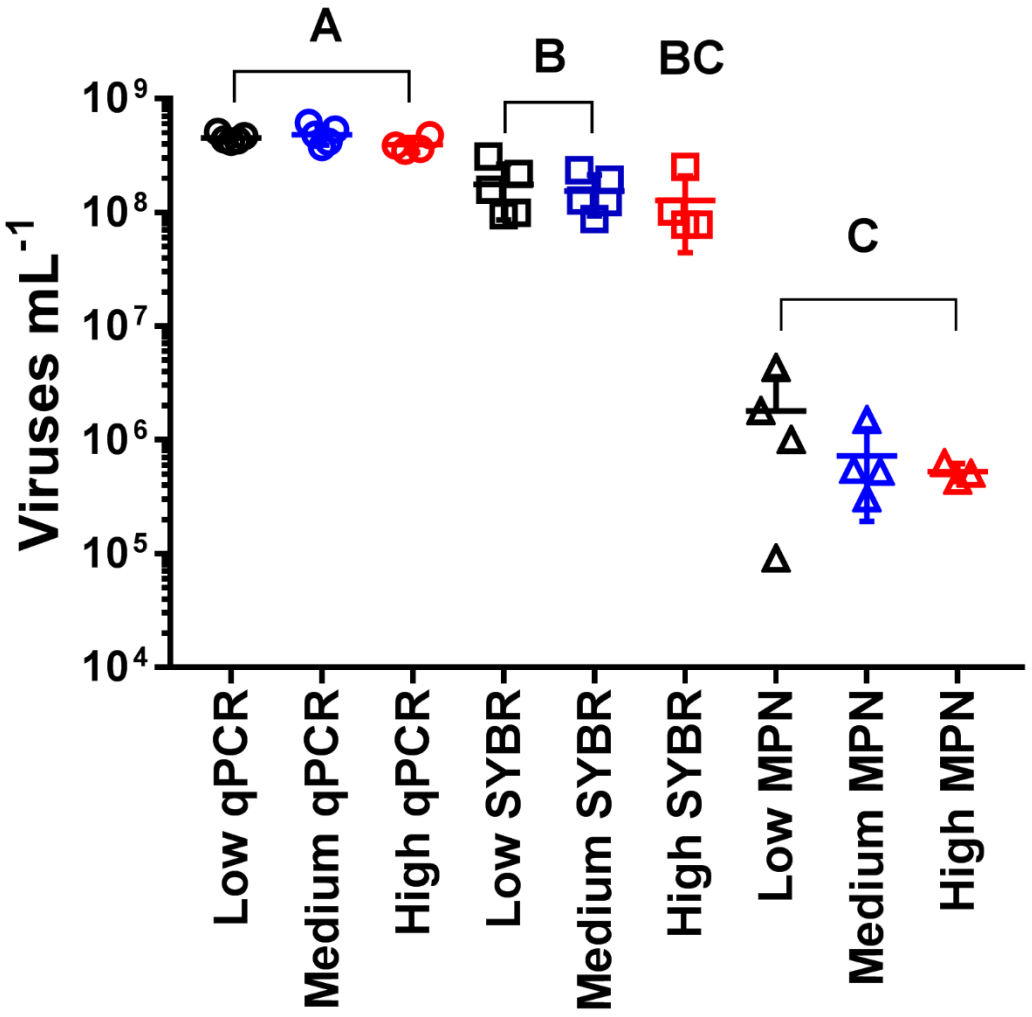


Figure 3.3. Comparison of three different methods to enumerate AaV. Epifluorescence microscopy with SYBR green, qPCR, and most probable number counts were used to enumerate virus concentration 3 days post infection of cultures acclimated to three different irradiance levels (high: 90 µmol photons m^{-2} s^{-1}, medium: 60 µmol photons m^{-2} s^{-1}, and low: µmol photons m^{-2} s^{-1}). Letters represent statistically indistinguishable concentrations (two-way ANOVA; see Methods and Table 3.1).

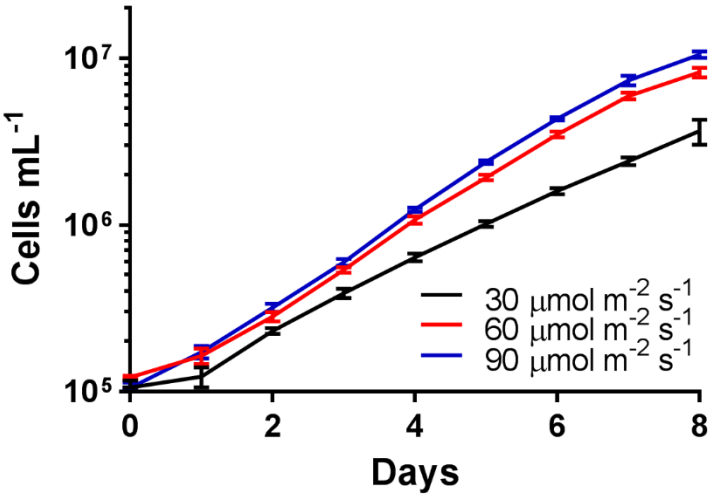


Figure 3.4. Growth curve of *A. anophagefferens* acclimated to three different irradiance levels. to the three irradiance levels: black line - low (30 µmol photons m^{-2} s^{-1}), blue line – medium (60 µmol photons m^{-2} s^{-1}), red line – high (90 µmol photons m^{-2} s^{-1}). Points are for n=5 five biological replicates ± SD.

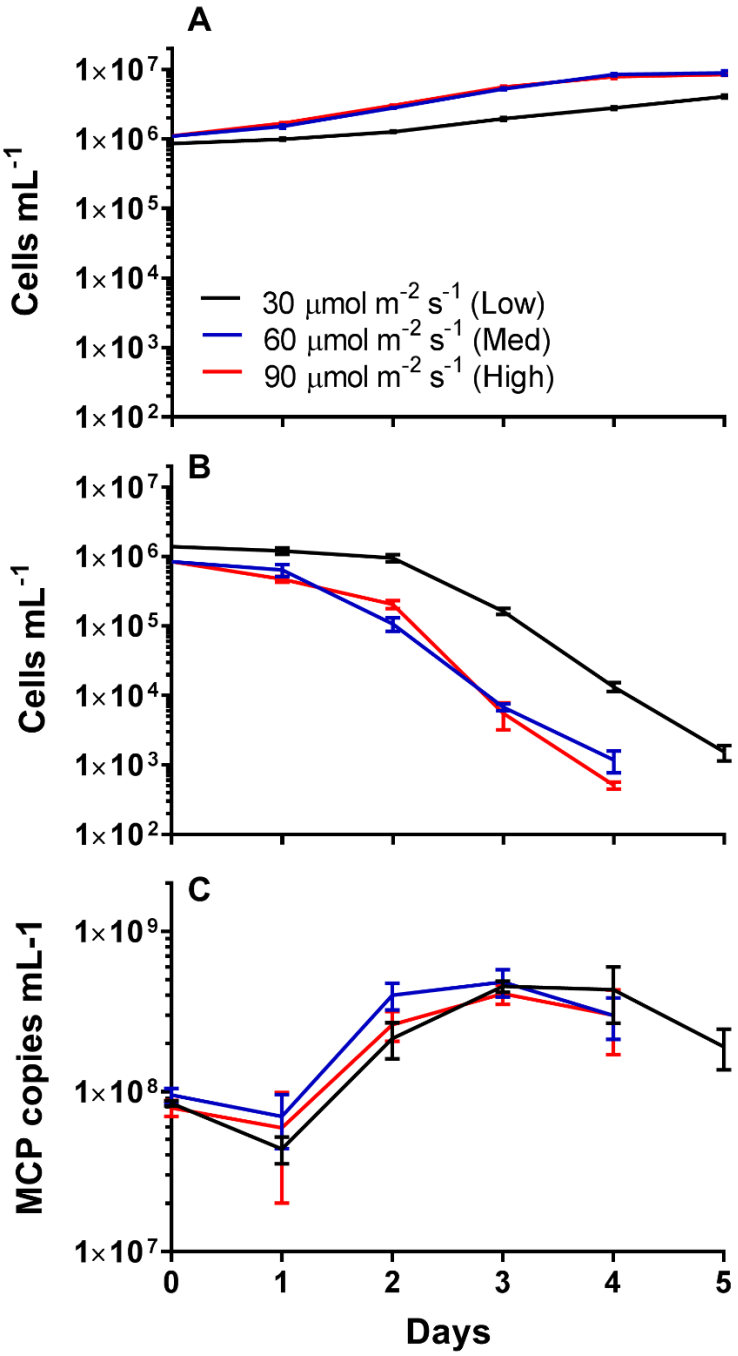


Figure 3.5. *Aureococcus anophagefferens* virus infection cycle dynamics with acclimation to varying irradiance levels. *A. anophagefferens* host concentrations either uninfected (A) or infected (B), and C) major capsid protein (MCP) copies mL^{-1} over the course of the 5-day experiment. Cultures were infected on day 0. Red lines are high irradiance acclimated cultures (90 µmol photons m^{-2} s^{-1}), blue lines are medium irradiance acclimated cultures (60 µmol photons m^{-2} s^{-1}), and black lines are low irradiance acclimated cultures (30 µmol photons m^{-2} s^{-1}). All symbols are for n=5 five biological replicates ± SD.

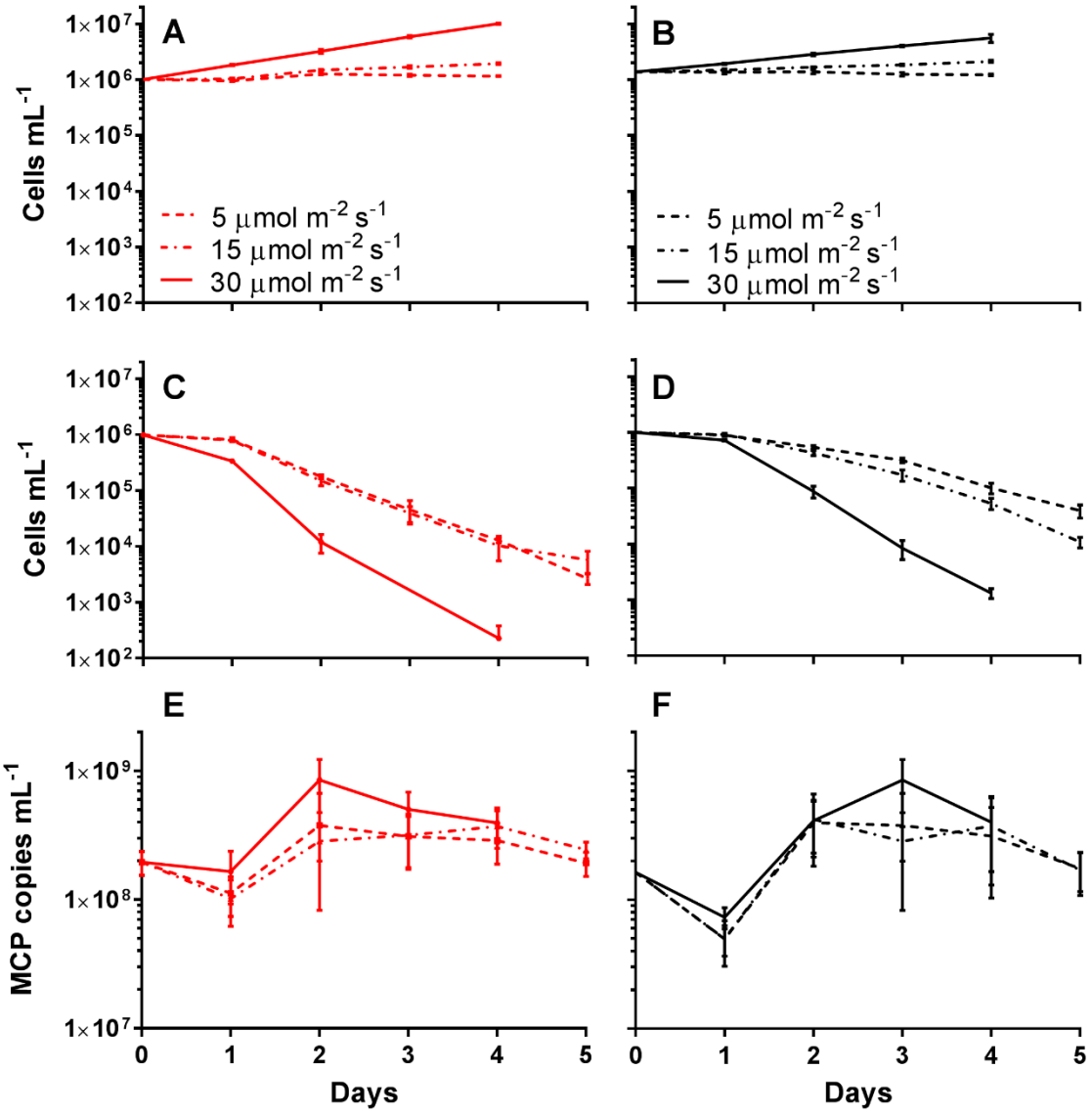


Figure 3.6. ***Aureococcus anophagefferens*** **virus infection cycle dynamics with acclimation to varying light levels and in lower light**. *A. anophagefferens* host concentrations for uninfected cultures acclimated to A) 90 µmol photons m^{-2} s^{-1} and B) 30 µmol photons m^{-2} s^{-1}, and infected cultures acclimated to C) 90 µmol photons m^{-2} s^{-1} and D) 30 µmol photons m^{-2} s^{-1} over the 5-day experiment. MCP copies mL^{-1} for cultures acclimated to E) 90 µmol photons m^{-2} s^{-1} and F) 30 µmol photons m^{-2} s^{-1} over the 5-day experiment. Cultures were infected on day 0. Solid lines indicate acclimated cultures maintained at acclimated irradiance levels after infection, dotted and dashed lines indicate acclimated cultures transferred into 15 µmol photons m^{-2} s^{-1} light after infection, and dashed lines indicate acclimated cultures transferred into 5 µmol photons m^{-2} s^{-1} light after infection. All symbols are for n=5 five biological replicates ± SD.

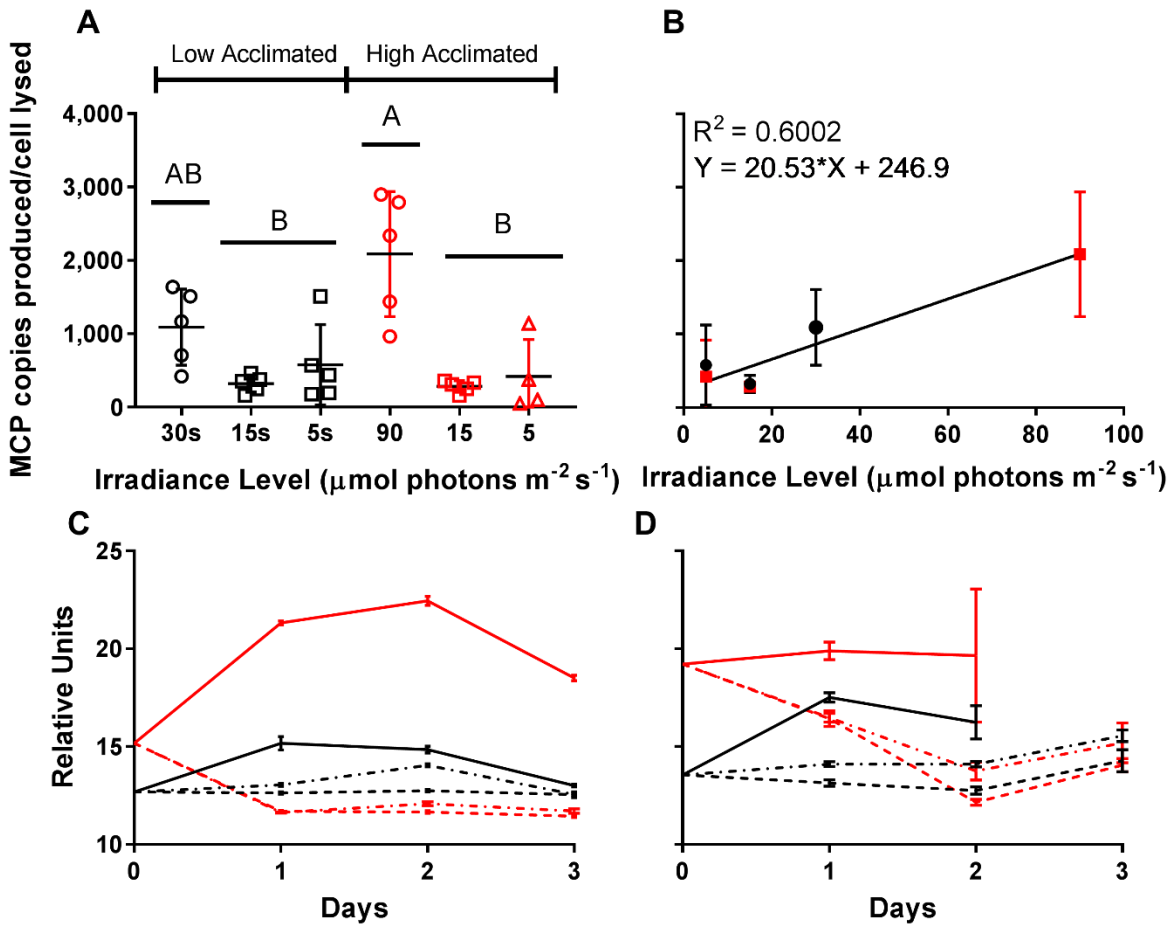


Figure 3.7. Effects of reducing light levels to 15 and 5 µmol photons m^{-2} s^{-1} after infection on Burst Size (Figure 3.6). Red Lines and symbols are high irradiance acclimated (90 µmol photons m^{-2} s^{-1}), while black lines are low irradiance acclimated (30 µmol photons m^{-2} s^{-1}). A) MCP copies produced per cell lysed for each treatment. Letters represent statistically indistinguishable burst sizes (one-way ANOVA; see Methods and Table 3.2). B) Burst Size v. Irradiance Level during the infection. C) Mean forward scatter of uninfected *A. anophagefferens* cultures over the first three days of the 5-day experiment (Figure 3.5C, Figure 3.5D) as determined by flow cytometry. D) Mean forward scatter of infected *A. anophagefferens* cultures over the course of the 5-day experiment (Figure 3.5C, Figure 3.5D) as determined by flow cytometry. Plotted are only populations >10,000 cells/mL. Solid lines indicate acclimated cultures maintained at acclimated irradiance levels after infection, dotted and dashed lines indicate acclimated cultures transferred into 15 µmol photons m^{-2} s^{-1} light after infection, and dashed lines indicate acclimated cultures transferred into 5 µmol photons m^{-2} s^{-1} light after infection. Error is plotted as standard deviation.

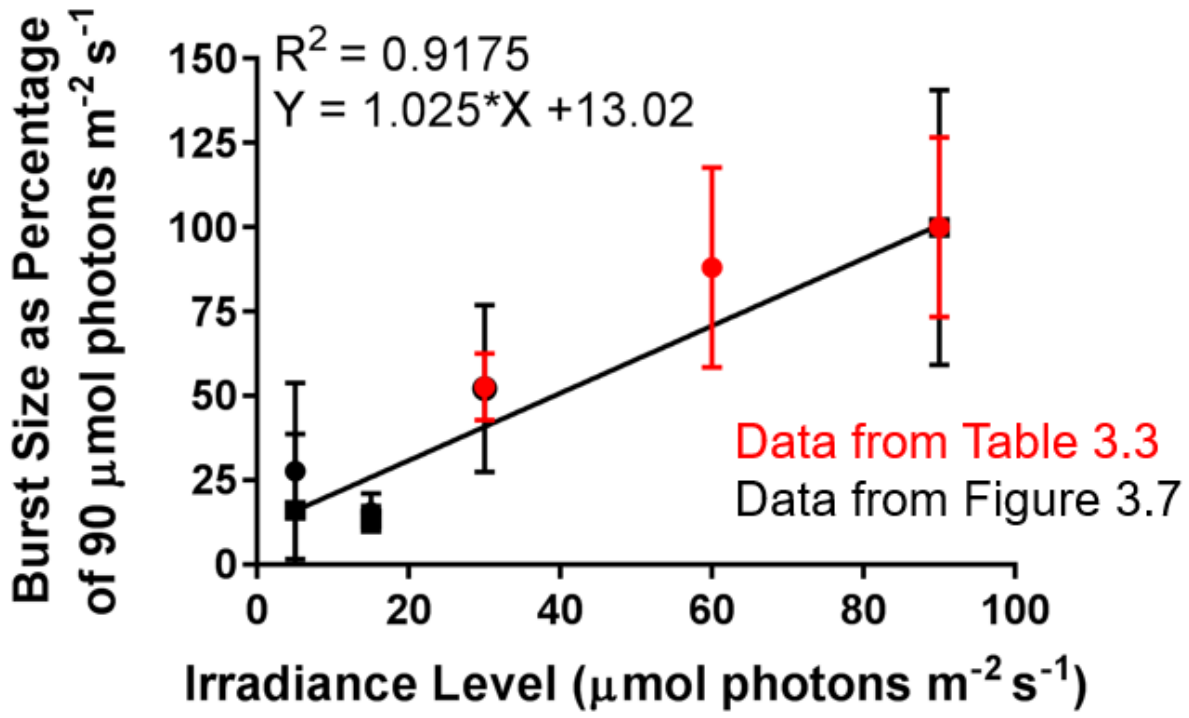


Figure 3.8. Burst Size v. Irradiance Level during the infection from distinct experiments (Table 3.3 and Fig 3.7B). Burst sizes were normalized to that of the 90 µmol photons m^{-2} s^{-1} acclimated cultures from each experiment.

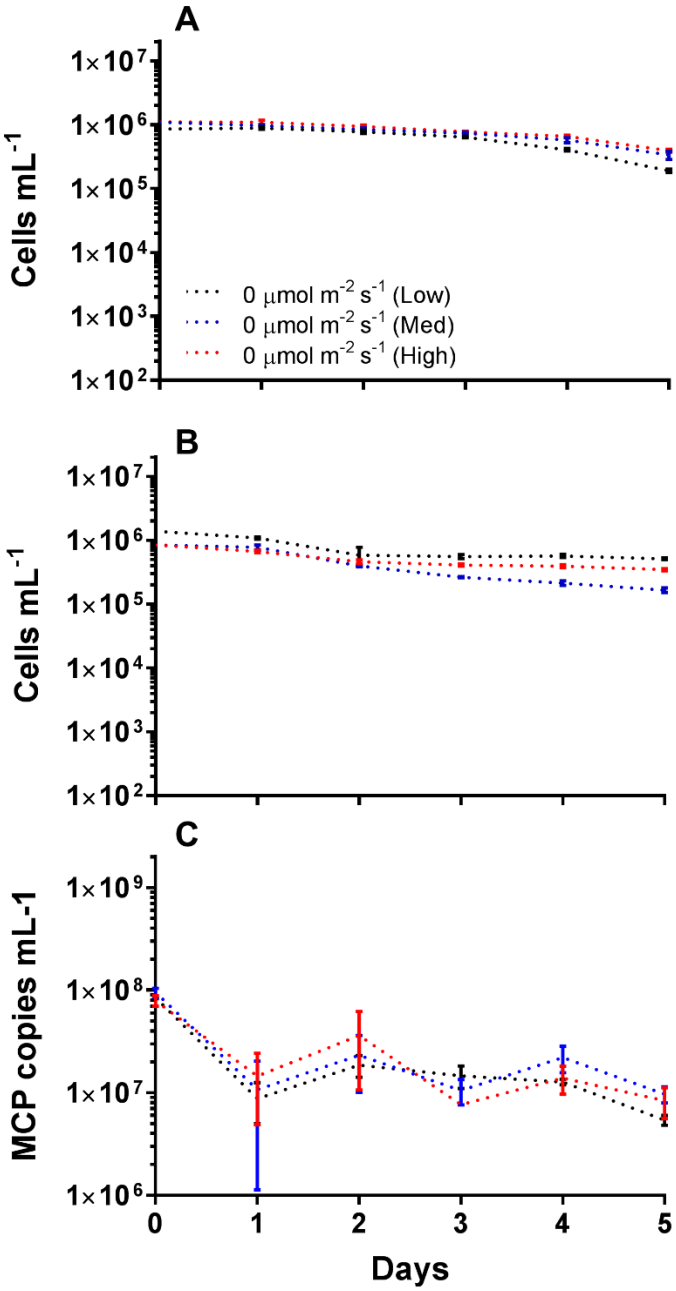


Figure 3.9. *Aureococcus anophagefferens* virus infection cycle dynamics with acclimation to varying irradiance levels and then transferred to the dark. *A. anophagefferens* host concentrations either uninfected (A) or infected (B), and C) major capsid protein (MCP) copies mL^{-1} over the course of the 5-day experiment. Cultures were infected on day 0, and transferred to the dark. Red lines are high irradiance acclimated cultures (90 µmol photons m^{-2} s^{-1}), blue lines are medium irradiance acclimated cultures (60 µmol photons m^{-2} s^{-1}), and black lines are low irradiance acclimated cultures (30 µmol photons m^{-2} s^{-1}).All symbols are for n=5 biological replicates ± SD. Control cultures not placed in the dark are shown in Fig3.5.

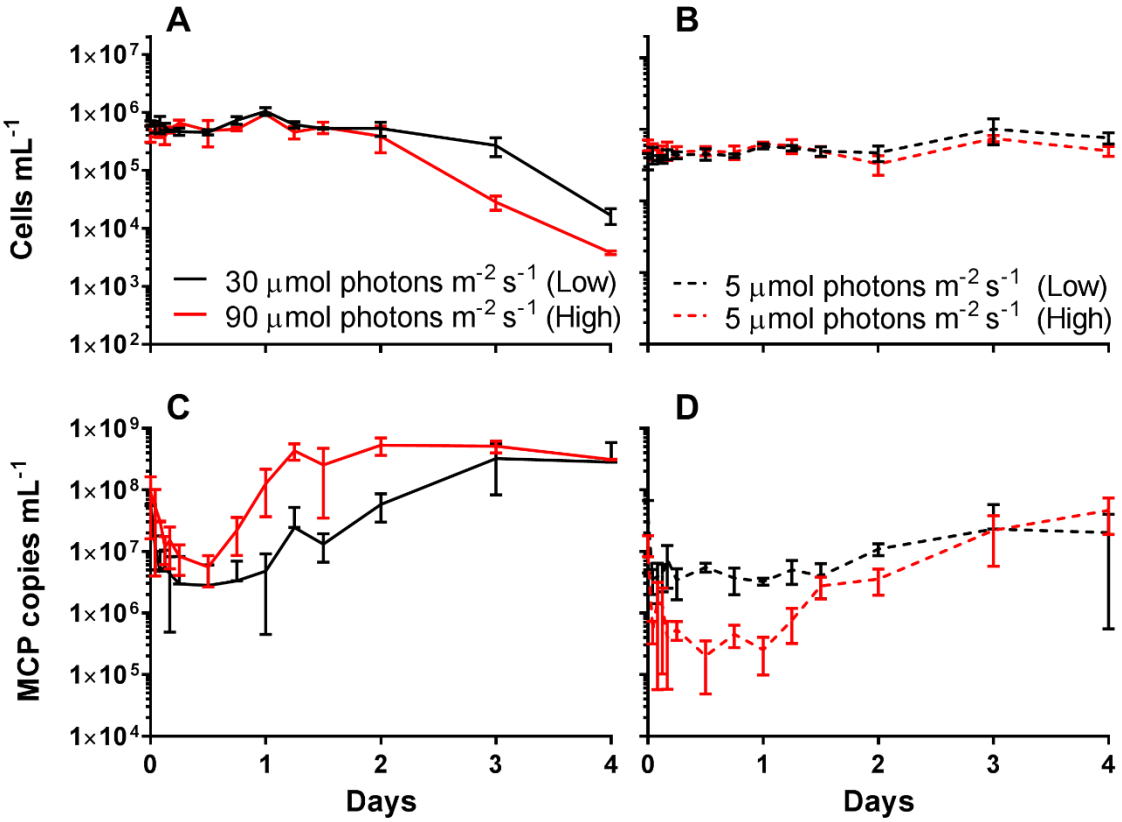


Figure 3.10. *Aureococcus anophagefferens* virus infection cycle dynamics with pre-acclimation to 5 µmol photons m^{-2} s^{-1} for one day before infection. Red lines are high irradiance acclimated (90 µmol photons m^{-2} s^{-1}) and black lines are low irradiance acclimated (30 µmol photons m^{-2} s^{-1}). *A. anophagefferens* host concentrations for acclimated cultures A) maintained at acclimated irradiance levels after infection or B) infected after one day at 5 µmol photons m^{-2} s^{-1}. MCP copies mL^{-1} for acclimated cultures C) maintained at acclimated irradiance levels after infection or D) infected after one day at 5 µmol photons m^{-2} s^{-1}. Day 0 on the graphs are when cultures were infected. Points are for n=3 three biological replicates ± SD.

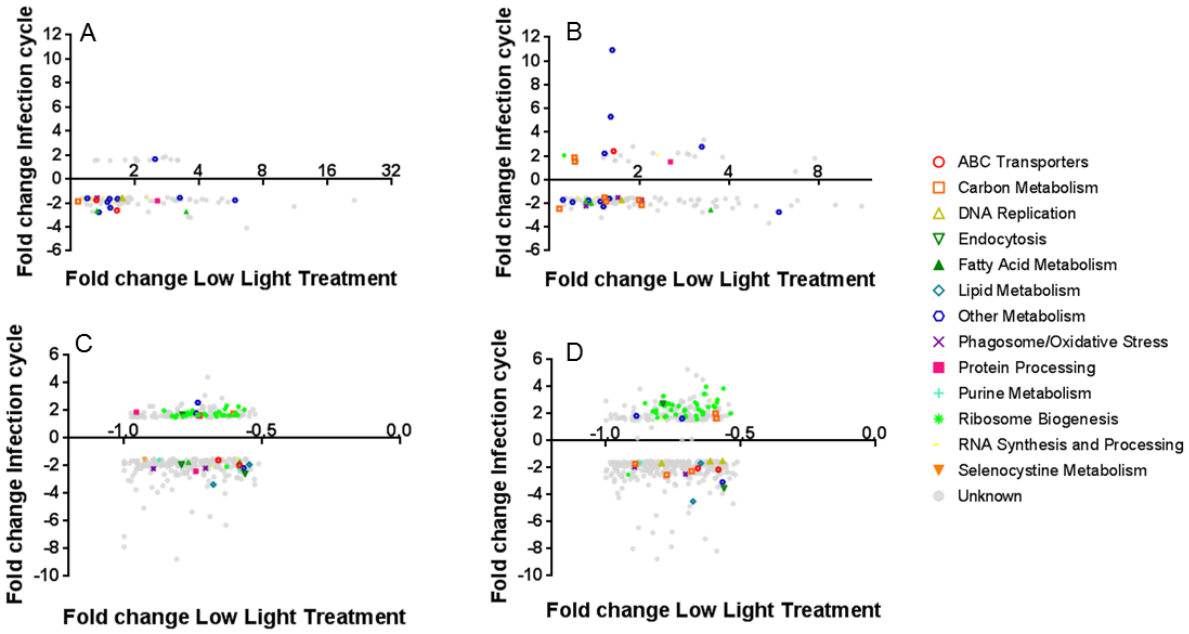


Figure 3.11. Comparison between *de novo* assembled *Aureococcus anophagefferens* CCMP1850 transcriptome in low light (Frischkorn *et al.* 2014) and *Aureococcus anophagefferens* CCMP1984 transcriptome over the course of the infection cycle (Moniruzzaman *et al.* 2018). A) Significantly overexpressed genes detected in Frischkorn *et al.* Transcriptome compared to genes significantly over expressed during the early stages of the infection cycle (5 min, 30 min, 1 hr, 6 hr). B) Significantly overexpressed genes detected in Frischkorn *et al.* Transcriptome compared to genes significantly over expressed during the late stages of infection cycle (12 hr and 21 hr). C) Significantly underexpressed genes detected in Frischkorn *et al.* Transcriptome compared to genes significantly over expressed during the beginning of infection cycle (5 min, 30 min, 1 hr, 6 hr). D) Significantly underexpressed genes detected in Frischkorn *et al.* Transcriptome compared to genes significantly over expressed during the late stages of infection cycle (12 hr and 21 hr). The value for the fold change over the course of the infection cycle is the average of significantly differentially expressed fold change for either the early stages of infection (A, C) or the late stages of infection (B, D).

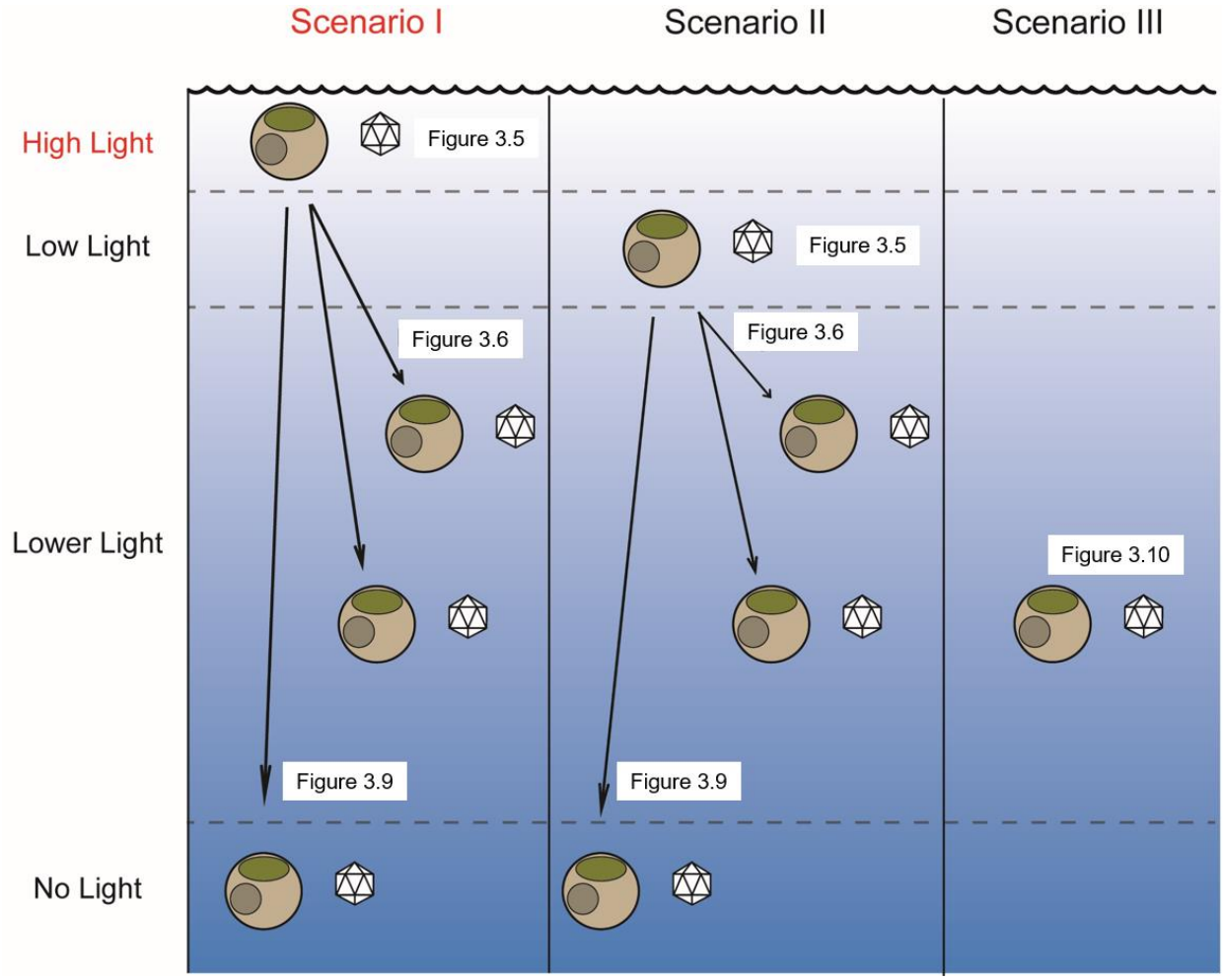


Figure 3.12. Ecological predictions of *Aureococcus* AaV interactions over the course of a bloom and effects of sinking. High light corresponds to 90 µmol photons m^{-2} s^{-1}, low light corresponds to 30 µmol photons m^{-2} s^{-1}, lower light corresponds to 15 and 5 µmol photons m^{-2} s^{-1}.

Chapter Four

Internal nitrogen pools shape the infection of *Aureococcus anophagefferens* CCMP 1984 by a giant virus

Author contributions

Experiments were designed by Eric R. Gann, Todd B. Reynolds, and Steven W. Wilhelm. Eric R. Gann and Brennan J. Hughes performed uninfected growth curves and one-step experiments. Eric R. Gann performed nutrient shift experiments and bioinformatic analyses. This chapter was written by Eric R. Gann and Steven W. Wilhelm, with all authors participating in providing comments in the initial and revised drafts.

Abstract

The pelagophyte *Aureococcus anophagefferens* blooms annually in shallow bays around the world, where it is hypothesized to outcompete other phytoplankton in part by using alternative nitrogen sources. The high proportion of natural populations that are infected during the late stages of the bloom suggest viruses cause bloom collapse. We hypothesized that the Aureococcus anophagefferens Virus (AaV) infection cycle would be negatively influenced in cultures acclimated to decreasing N:P ratios, but N:P ratio at the time of the infection cycle will have no effect. Cultures acclimated in the NO_3^- concentration (0.0147 mM; N:P = 0.1225) that showed reduced end point cell abundances, forward scatter (a proxy for size), and red fluorescence (a proxy for chlorophyll), produced less viruses per cell at a slower rate. Decreasing the external concentration of nitrogen after infection did not influence burst size or time to lysis. These data suggest that all the nitrogen required for new progeny is found within the host cells before infection and is recycled. Flow cytometric data of the infection cycle showed a reduction in red fluorescence around twelve hours post infection, consistent with degradation of nitrogen-rich chloroplasts during the infection cycle. Using cell and virus quota estimates, we determined that *A. anophagefferens* cells had enough nitrogen and carbon for the estimated lower burst sizes but did not contain enough phosphorous. Consistent with this observation nitrate and sugar transporters were not over expressed in the publicly available transcriptome data of the infection cycle, while several phosphorus transporters were. Our data demonstrate that the dynamics of viruses infecting *A. anophagefferens* over the course of a bloom is dictated by the host cell state upon infection, which is set *a priori* by external nutrient supplies.

Introduction

Since the accidental discovery of the ‘giant’ Mimivirus, viruses with particle sizes and genome sizes larger than many prokaryotic organisms have been isolated and are believed to belong to a monophyletic group termed the Nucleocytoplasmic Large dsDNA Viruses (NCLDVs) (Wilhelm et al., 2017). The increased genetic potential within these genomes allows for less dependence on the host during the infection cycle. This is evident in the production of new viruses by cytoplasmic virus factories created after infection that do not rely on the host nucleus to begin transcription (Fabre et al., 2017), as well as by the presence of viral proteins that modulate translation (Abrahao et al., 2018; Mizuno et al., 2019). This shifting paradigm, that viruses are more independent than originally believed, has led to the idea that the infected cell, sometimes termed “the virocell”, is an independent entity different from an uninfected host (Forterre, 2011). This is perhaps best illustrated by the novel products created through a combination of host and virally encoded genes (Rosenwasser et al., 2014).

The environment of the virocell can drastically alter the phenotype and replication parameters of the virus. Culturing amoeba hosts without bacteria for 150 generations caused the production of Mimivirus without fibrils and with a reduced genome (Boyer et al., 2011). Quenching oxidative stress produced during infection of *Emiliania huxleyi* prevents new viruses from being produced (Sheyn et al., 2016). To this end, the growth environment of the cell before infection is important as it shapes the physiological state of the host. Most viruses accumulate the building blocks for new progeny by some combination of degrading current host stores (*i.e.*, the nuclear genome for nucleotides) (Agarkova et al., 2006) and driving continued uptake of nutrients from the external environment (Waldbauer et al., 2019).

Changing extracellular conditions could play a crucial role(s) in environments where viruses are hypothesized to be important in the control of host populations. An important example of this is the blooms of eukaryotic algae in which viruses can be the sole cause of bloom collapse (Brussaard et al., 2005; Vardi et al., 2012) and are implicated in bloom prevention (Haaber and Middelboe, 2009). The pelagophyte *Aureococcus anophagefferens* blooms annually in shallow, well-mixed bays in several locations around the world, resulting in severe light attenuation from high cell densities ($> 10^6$ cells mL^{-1}) and apparent toxicity to bivalves that result in millions of dollars in economic losses (Gobler and Sunda, 2012). A major driver of the bloom's demise has been hypothesized to be large (~190 nm) viruses, because a high proportion of cells are visibly infected at the time of the collapse (Gastrich et al., 2004). One of these viruses, Aureococcus anophagefferens virus (AaV), has been isolated and partially characterized. It belongs to the *Mimiviridae* family of viruses which house some of the largest viruses known to science (Wilhelm et al., 2016). Although infected cells within natural populations have been seen since the initial description of the species (Sieburth et al., 1988), how the changing environment over the course of the bloom influences the infection cycle is not known. Variable abiotic or biotic conditions (*i.e.* low temperature, low light) that may reduce the efficiency of the AaV infection (Gobler et al., 2007), could influence the succession of bacteria or other eukaryotic algae via altered release of dissolved organic matter (Gobler et al., 1997), and the nutrients contained within free virus particles (Jover et al., 2014).

To understand ecosystem constraints on the virus-mediated lysis of *A. anophagefferens* blooms, we examined the effect of decreasing nitrogen concentrations on the infection cycle of AaV. *A. anophagefferens* blooms form when inorganic nitrogen concentrations are drawn down and organic nitrogen concentrations are elevated (Gobler and Sunda, 2012). *A. anophagefferens*

can utilize many different nitrogen compounds (Berg et al., 2002; Berg et al., 2008), and the assimilation of alternative organic nitrogen sources allows *A. anophagefferens* to avoid nitrogen-limitation when other organisms become growth-limited (Gobler et al., 2004; Wurch et al., 2019). The continued use of this alternative nitrogen pool decreases total extracellular N concentrations (Lomas et al., 1996; Gobler et al., 2011), thus altering the N:P ratio in the environment. In lab experiments with other algal virus systems, infection during low nitrogen conditions reduced burst sizes and increased the length of the infection cycle (Cheng et al., 2015; Maat and Brussaard, 2016), though viral production is not always stimulated by nitrogen additions in natural populations (Bratbak et al., 1993; Motegi and Nagata, 2007). In this study, the influence of acclimation to decreasing nitrate concentrations (therefore decreasing extracellular N:P ratios) was studied as cells infected at different points of the bloom are exposed to differing concentrations of nitrogen. Also, we were interested in determining whether the infected cell would be influenced by the external nitrogen concentration, as this appears to be system specific (Fowler and Cohen, 1948; Wikner et al., 1993). We have determined that the AaV infection cycle was negatively influenced by *A. anophagefferens* acclimated to NO_3^- concentrations that negatively influence growth and flow cytometry parameters, while altering the extracellular N:P ratio at the time of the infection had no effect. This suggested the internal pools of nitrogen were more important than external ones after the onset of infection in this system.

Methods

Culture conditions

Non-axenic *Aureococcus anophagefferens* CCMP1984 cultures were grown at 19 ºC, on a 14:10 light dark cycle at 100 µmol photons m^{-2} s^{-1} in modified $ASP_{12}A$ growth medium (Gann, 2016). Cultures were acclimated to four different NO_3^- concentrations (1.47 mM, 0.735 mM, 0.147 mM, 0.0147 mM) (Table 4.1) (all tables and figures in this chapter are found in the Chapter 4 Appendix) by at least three successive transfers at each NO_3^- concentration prior to experimentation. Attempting to acclimate cultures to NO_3^- concentrations less than 0.0147 mM would not grow, therefore the 0.0147 mM NO_3^- was the lowest concentration used. Previous studies have also demonstrated that culturing *A. anophagefferens* in higher concentrations of a sole nitrogen source (0.030 mM NH_4^+) cause physiologically and transcriptional differences compared to replete conditions (Frischkorn et al., 2014). To prevent nutrient carry over, cultures were pelleted (2000xg, 5 minutes), and resuspended in fresh $ASP_{12}A$ with the respective NO_3^- concentration. The AaV lysate that was used to infect cultures was concentrated 50-fold from freshly lysed cultures using a Lab-scale Tangential Flow Filtration System (Fisher Scientific, Waltham, USA) equipped with a Durapore™ 30 kDa Pellicon XL Filter (MilliporeSigma, Burlington, MA). To deplete lysate of nitrate, the concentrate was diluted 20-fold with $ASP_{12}A$ without added nitrate, and then re-concentrated, before being passed through a 0.45-µm pore-size syringe filter (Millex-HV 0.45µm nominal pore-size PVDF, Millipore Sigma, Burlington, USA). Less than 0.5 mL was added to cultures (20 mL) for each experiment.

Enumeration of *A. anophagefferens* and AaV

Before enumerating concentrations, 1 mL of cultures were either directly fixed in 0.5% glutaraldehyde (for *A. anophagefferens* enumeration) or first passed through a 0.45-µm pore-size syringe filter (Millex-HV 0.45µm nominal pore-size PVDF, Millipore Sigma, Burlington, USA) before fixation (for AaV enumeration). Although filtering cultures for enumeration of viruses reduced the concentration of viruses (~43%) (Figure 4.1), we wanted to reduce any clumping or adsorption of viruses to cellular material. Fixed samples were stored at 4º C for at least 30 min before being measured on a FACSClibur flow cytometer (Becton, Dickinson and Company, Franklin Lakes, NJ). *A. anophagefferens* populations were gated on red chlorophyll a fluorescence and forward scatter (Moniruzzaman et al., 2018). Mean forward scatter (FSC-H), side scatter (SSC-H), and red fluorescence (FL3-H) for the gated *A. anophagefferens* populations was also recorded.

AaV was enumerated as described previously (Brussaard, 2004; Brown and Bidle, 2014). Briefly, the fixed AaV samples were diluted 100-fold in $ASP_{12}A$ and then inoculated with a 10,000-fold diluted SYBR Green DNA stain (Lonza, Basel, Switzerland), before being incubated at 80 º C for 10 min. AaV populations were gated on SYBR green emission (excitation: 520 nm) and side scatter (Supplemental Figure 4.2). To determine host and virus concentrations a known concentration of diluted FluoSphere Carboxylate-Modified Microspheres, 1.0 µm, yellow-green fluorescent (505/515) (Invitrogen, Waltham, MA) was added to each sample before being ran for one minute.

Experimental Design

To determine the influence of nitrate concentration on growth of *A. anophagefferens*, multiple cultures at late logarithmic growth that were acclimated to the four different nitrate concentrations (Table 4.1) and pelleted by centrifugation (2000xg, 5 minutes). Cell pellets were resuspended in media containing the respective NO_3- concentration and pooled to concentrate the cultures ~50 fold. To observe the entire growth curve, four 20 mL biological replicates were inoculated to a final cell density between $5x10^4$ and $1x10^5$. To assess differences in the physiological state of the cells due to the different NO_3^- concentrations, the mean of various flow cytometry parameters (FSC-H, SSC-H, FL3-H) of the gated *A. anophagefferens* cells were averaged on two days during the middle of logarithmic growth (days 6 and 7). Doubling time of the cultures was calculated using the following equation, where days three and eight were used as $time_N$ and $time_{No}$, respectively:

$$\text{Doubling time} = (time_N - time_{No}) / ((\log(concentration_N) - \log(concentration_{No}))/\log(2))$$

To determine the influence of acclimation to different nitrate concentration on the AaV infection cycle, multiple cultures at late logarithmic growth that were acclimated to the four different nitrate concentrations (Table 4.1) were pelleted by centrifugation (2000xg, 5 minutes). Cell pellets were resuspended in media containing the respective NO_3- concentration and pooled to concentrate the cultures ~50 fold. Five biological replicates were inoculated to a final cell density at 10^6 and infected at a multiplicity of infection (m.o.i) ~100 particles per cell in 20 mL. Uninfected controls were also done in five biological replicates. A high m.o.i. is used as a small fraction of AaV are infectious, and infecting cultures with m.o.i's between ~20-100 allows for lysis of the majority of cells (>99%) within three days. Cultures were infected at the start of the

light cycle. Samples were taken immediately upon infection, and then every four hours for 24 hours, and 48 and 72 hours post infection. Sampling throughout the first twenty-four hours allowed for the adsorption rate of AaV to be determined as described previously (Cottrell and Suttle, 1995) using the following formula:

$$\text{Adsorption Rate} = a/N$$

Where a = slope of the linear regression of the natural log of free viruses over the course of the first twenty-four hours and N is equal to the concentration of *A. anophagefferens*. The number of viruses produced per cell lyses, or burst size, was calculated using the following formula:

$$\text{Burst size} = (\text{virus concentration}_N - \text{virus concentration}_{No}) / (\text{cell concentration}_N - \text{cell concentration}_{N0})$$

Host and virus concentrations used for the calculation were the last point when free viruses did not decline more (24 hours for 1.47 mM, 0.735 mM, and 0.147 mM NO_3^- cultures and 48 hours for the 0.0147 mM NO_3^- cultures) and 72 hours for times N and N_o, respectively. Changes in cellular physiology during the infection cycle was determined through various flow cytometry parameters (FSC-H, SSC-H, FL3-H) of the gated *A. anophagefferens* cells.

To determine the influence of external nitrate concentration on the AaV infection cycle, multiple 1.47 mM NO_3^- concentration acclimated cultures at late logarithmic growth that were first pelleted by centrifugation (2000xg, 5 minutes). Cell pellets were resuspended in $ASP_{12}A$ with no added nitrate and pooled to concentrate the cultures ~50 fold. Immediately before infection, NO_3^- was added to the cultures (final cell density = $1x10^6$) at three different final

concentrations (1.47 mM, 0.0147 mM, and no added NO_3^-). Five biological replicates were infected at a m.o.i ~20 particles per cell in 20 mL, where five uninfected controls were brought up to 20 mL with $ASP_{12}A$ with no nitrate. Samples were taken each day for four days. Burst sizes were calculated as described above, using time points of 1 and 4 days for times N and N_o, respectively.

Statistical Analyses

Statistical analyses were performed using Prism 7.03 (GraphPad, San Diego, USA). To compare differences between treatments, a one-way ANOVA was used followed by Tukey's HSD post-hoc testing.

Results

The effects of nitrate concentration acclimation on *A. anophagefferens* growth

Cultures were acclimated to four different NO_3^- concentrations (1.47 mM, 0.735 mM, 0.147 mM, 0.0147 mM), with a constant PO_4^{3-} concentration (0.12 mM), producing environments with variable N:P ratios (Table 4.1). There was a difference in the length of the logarithmic phase of growth as well as the final abundance of cells reached between the 0.0147 mM acclimated cultures and the other three. For the 1.47 mM, 0.735 mM, and 0.147 mM NO_3^- acclimated cultures, the length of the logarithmic phase of growth was 9-10 days (Figure 4.3A), and there was no significant difference in end point abundances which were: 1.7×10^7 cells mL^{-1} (SD = 9.7×10^5), 1.9×10^7 cells mL^{-1} (SD = 8.7×10^5), 1.6×10^7 cells mL^{-1} (SD = 1.4×10^5), respectively (Table 4.2). The 0.0147 mM NO_3^- acclimated culture had a final cell abundance of half (8.5×10^6 cells mL^{-1} SD = 9.7×10^5) the three higher concentrations at the end of the 14-day

experiment, with the logarithmic phase being 1-2 days shorter (Figure 4.3A). The four different acclimating NO_3^- concentrations did not greatly influence the doubling time of the cultures, with all cultures having doubling times between 0.99 and 1.09 days (Figure 4.3B, Table 4.3). To further understand how decreasing N:P ratios influences *A. anophagefferens* cells, we used parameters measured during flow cytometry (Figure 4.3C). The relative values of forward scatter (FSC-H), side scatter (SSC-H), and red fluorescence (FL3-H) of all cells gated as *A. anophagefferens* were averaged during two consecutive days in mid-logarithmic growth (Figure 4.3C). The 0.735 mM and 0.147 mM mM NO_3^- acclimated cultures had FSC-H, SSC-H, and FL3-H values that were not significantly different from one another (Table 4.4). Cultures acclimated to 1.47 mM NO_3^- had a slight, but significant increase for FSC-H and SSC-H, whereas there was statistically no difference with FL3-H compared to the 0.735 mM and 0.147 mM NO_3^- acclimated cultures (Table 4.4). There was a significant decrease in all three parameters for the 0.0147 mM NO_3^- acclimated cultures compared to the other three (Table 4.4).

The effects of nitrate concentration acclimation on the AaV infection cycle

When cultures acclimated to the four NO_3^- concentrations were infected with AaV, only the 0.0147 mM NO_3^- acclimated cultures showed a difference in infection dynamics. The 1.47 mM, 0.735 mM, and 0.147 mM NO_3^- acclimated cultures all had similar trends as previously reported (Rowe et al., 2008; Brown and Bidle, 2014), where new viruses were produced and cells lysed after 24 hours (Figure 4.4A, Figure 4.4B), while cell lysis and virus production was delayed one day for the 0.0147 mM acclimated culture. The burst sizes for the 1.47 mM (465.8 viruses produced cell lysed $^{-1}$, SD = 32.2), 0.735 mM (427.1 viruses produced cell lysed $^{-1}$, SD = 93.7), and 0.147 mM (465.8 viruses produced cell lysed $^{-1}$, SD = 32.2) NO_3^- acclimated cultures were

not significantly different from one another but had a burst size three times greater than the 0.0147 mM (142.6 viruses produced cell lysed $^{-1}$, SD = 40.4) NO_3^- acclimated cultures (Figure 4.4C, Table 4.5). There were no differences in the adsorption rates due to NO_3^- concentration (Figure 4.4D, Table 4.6). We flow cytometry parameters to look at changes between infected and uninfected cells over the first 24 hours of the infection cycle (Figure 4.5). All uninfected populations increased in FSC-H, and FL3-H, while all but the 0.0147 mM NO_3^- acclimated culture increased in SSC-H. The FSC-H and SSC-H of infected cells either remained the same or decreased slightly (< 5%) (Figure 4.5A, 4.5B). The FL3-H values remained the same for the first twelve hours of the infection cycle but then decreased between 6.8 - 9.5% for all the NO_3^- concentrations (Figure 4.5C).

The effect of external nitrogen on concentration on the AaV infection cycle

To determine whether the external nitrate concentration affects viral production, we pelleted *A. anophagefferens* cultures grown at 1.47 mM and added back decreasing NO_3^- concentrations (1.47 mM, 0.0147 mM, 0 mM) and measured growth of control and infected cultures (Figure 4.6). In control cultures, the end point abundance after the four days significantly decreased with reduced NO_3^- concentrations (Table 4.7). The final cell abundances were $6.2x10^6$ cells mL^{-1} (SD = $3.9x10^5$), $5.0x10^6$ cells mL^{-1} (SD = $5.9x10^5$), and $3.4x10^6$ cells mL^{-1} (SD = $3.1x10^5$), for cultures that had 1.47 mM, 0.0147 mM, and 0 mM NO_3^- added back, respectively. Although there was a hinderance to the growth of control cultures, infecting cultures with decreasing NO_3^- concentrations added back did not change the length of time for cells to lyse (Figure 4.6A), new viruses to be produced (Figure 3B), or a decrease the burst size (Figure 4.6C, Table 4.8).

Discussion

The effects of nitrate acclimation on *Aureococcus* growth and the AaV infection cycle

An effect of acclimation of *A. anophagefferens* cultures to decreasing NO_3^- concentrations was only seen in the lowest concentration (0.0147 mM), where there was a decreased biomass accumulation and end point cell abundance. These acclimated cultures appeared to be physiologically different than the other cultures as there was a reduction in forward scatter, side scatter, and red fluorescence, as detected by flow cytometry. These values suggest that the cells in the lowest NO_3^- concentrations are smaller and have less chlorophyll a per cell. In previous studies of various *A. anophagefferens* strains the maximum efficiency of photosystem II (F_v/F_m) is reduced in lower or limiting nitrogen concentrations (Berg et al., 2008; Dong et al., 2014; Frischkorn et al., 2014), and *in vivo* fluorescence per cell is also reduced (Dong et al., 2014). This data is also supported by transcriptomes of *A. anophagefferens* CCMP1850 in low nitrogen environments, where there is a down regulation of light harvesting complexes and an increase in proteases and peptidases (Frischkorn et al., 2014). The reduction of red fluorescence or chlorophyll a per cell was observed in other species of algae as well, leading to a reduced F_v/F_m (Berges and Falkowski, 1998; Maat et al., 2016a; Maat and Brussaard, 2016). It was hypothesized that this reduction in chlorophyll a content decreases the nitrogen requirement within the cell (Maat and Brussaard, 2016).

Similarly only infections of the 0.0147 mM NO_3^- acclimated cultures did negatively influence burst size with a prolonged production of new viruses. A prolonged infection cycle has been seen in multiple systems due to growth in lower nitrogen concentrations (Bratbak et al., 1998; Maat and Brussaard, 2016), as has the reduced burst size (Bratbak et al., 1998; Cheng et

al., 2015; Maat and Brussaard, 2016). We sampled multiple times throughout the first 24 hours of the experiment to detect differences in infected populations by flow cytometry and to determine adsorption rates. There was no difference in adsorption rates between any of the acclimated NO_3^- concentration treatments. The average adsorption rate determined in this study (5.9×10^{-10} mL min^{-1}) agrees with that previously reported in this system (5.8×10^{-10} mL min^{-1}) (Brown and Bidle, 2014). The *Chlorella variabilis* NC64A infecting Paramecium bursaria Chlorella Virus, PBCV-1, adsorption rate was also not influenced by altered less optimum C:P ratios (Clasen and Elser, 2007).

Forward scatter and side scatter remain constant over the course of the infection cycle, while red fluorescence begins to decrease around twelve hours post infection. A reduction in red fluorescence has been detected in other systems (Lawrence et al., 2006; Lonborg et al., 2013) specific points in the infection cycle (Lawrence et al., 2006). This correlates with a reduction in chlorophyll a and F_v/F_m over the course of the infection cycle (Schatz et al., 2014). Measurements for F_v/F_m have not been determined during the *Aureococcus* infection cycle, but daily measurements for control and infected cultures have shown an initial increase compared to control cultures followed by a drastic decrease (Gobler et al., 2007). In both natural populations and lab cultures, when many viruses are present in the cells, there is either no chloroplast or it is smaller than in uninfected cells (Sieburth et al., 1988; Gastrich et al., 2004; Rowe et al., 2008). More work is needed to determine whether this reduction in chloroplast size corresponds to the decrease in red fluorescence beginning around twelve hours post infection.

It was hypothesized that the differences in burst sizes in cells grown in low nitrogen environments were caused by a limitation of resources available within the host (Cheng et al., 2015), due to a decreased nitrogen quota within cells grown in low nitrogen (Cheng et al., 2015;

Maat and Brussaard, 2016). Altered internal quota has also been hypothesized to cause a reduced burst size in cells grown at altered C:P ratios (Clasen and Elser, 2007). In *Thalassiosira weissflogii, Dunaliella tertiolecta* (Berges and Falkowski, 1998), and *Chlorella variabilis* NC64A (Cheng et al., 2015), as cells became more limited for nitrogen, their internal C:N ratio increased, hypothesized to be due to the reduction in nitrogen stores or non-essential proteins (Berges and Falkowski, 1998). This reduced chlorophyll a content within the cells and decreased F_v/F_m as discussed above may prevent viruses from acquiring all of the energy or nitrogen required to produce the large number of viruses per cell. Algae also have increased starch and lipid production (Cheng et al., 2015), and protease activity (Berges and Falkowski, 1998; Frischkorn et al., 2014) under nitrogen stress, which may reduce the efficiency of infection. More work is needed to understand how stresses influence nutrient and metabolite pools of virocells.

The effects of external nitrogen on the AaV infection cycle

The dependence on the external nutrient pools is variable depending on the system and the nutrient. Transitioning *Escherichia coli* grown in a nutrient rich broth to minimal media without various components greatly reduced T2 phage produced, and the effects of specific nutrient(s) added back partially restored viral production (Fowler and Cohen, 1948), while phosphorus starved eukaryotic alga *Micromonas pusilla* did not decrease the length of time to produce new viruses but did significantly increase the number of viruses produced (Maat et al., 2016). In contrast, in three marine phage systems, phage genomes were recycled from the host instead of being derived from the external environment (Wikner et al., 1993). This was also seen in the *Synechococcus* WH8102- cyanophage S-SM1 system where most of the nitrogen within phage

proteins were derived from its host. Although, the amount varied depending on light availability (Waldbauer et al., 2019). To test this in our system, cultures acclimated to a high external NO_3^- concentration were shifted to environments with decreasing concentrations of external NO_3^- concentrations. There was no delay in virus production or decreased burst size when cells were moved from a high NO_3^- external concentration (1.47 mM) to media where there was little (0.0147 mM) or no NO_3^- added back compared to cultures with a high NO_3^- external concentration added back. We acknowledge that this method may cause some carry-over of nutrients, but this method allows for us to see differences caused by external NO_3^- concentrations. There is support for this with the control cultures, as decreasing external NO_3^- concentrations caused a significantly reduced end point cell abundance. Cultures with no added NO_3^- only achieved 55% of the cell abundance with 1.47 mM NO_3^- added back. This is worth noting as acclimating cultures to 0.147 mM NO_3^- (10% that of normal $ASP_{12}A$), had no effect on growth or the infection cycle.

One caveat of this work is the cultures were non-axenic. Although there are axenic cultures, and methods for the production axenic cultures exist (Berg et al., 2002), we were unable to remove the heterotrophic bacteria when *A. anophagefferens* is cultured on a defined growth medium. It appears there is a compound required by *A. anophagefferens* not present in $ASP_{12}A$. The use of a defined medium, despite the requirement of bacteria, allowed us to control the NO_3^- added, and any variability that would exist in a medium with a natural sea water base would not. Previous estimates of the number of heterotrophic bacteria in our cultures during late logarithmic growth of *A. anophagefferens* are between 1.8 and $8.1x10^5$ colony forming units mL^{-1}. The amount of nitrogen found within the heterotrophic community associated with *A. anophagefferens* therefore could reduce the amount of NO_3^- available to *A. anophefferens*. Using

estimates of nitrogen found within cells in natural and culture studies (between 1.6 – 100 fg per cell) (Fagerbakke et al., 1996), we can estimate there would be $2.88x10^5$ – $8.1x10^8$ fg of nitrogen sequestered in the heterotrophic community in our cultures. This would account for $1.4x10^{-4}$ – 0.4% of the nitrogen found within the 0.0147 mM NO_3^- acclimated cultures. Although this is a small amount of the total external nitrogen available, heterotrophic bacteria may influence the growth of *A. anophagefferens* in other ways. More work is needed to understand the interactions between *A. anophagefferens* and heterotrophic bacteria, as well as if these bacteria influence the infection cycle in any way.

Assessing nitrogen requirements for the infection cycle to occur

From the experiments performed in this study, the data did not reject the hypothesis that there is enough nitrogen present within an *A. anophagefferens* cell for the infection cycle to proceed. Or that the infected cell does not utilize external nitrogen pools. To further support this hypothesis, we performed theoretical calculations from available data about *A. anophagefferens* nutrient pools, C:N:P amounts within the algal virus, PBCV-1, and burst size ranges to assess whether enough nitrogen exists in an infected cell (Table 4.9).

The amount of carbon and nitrogen in an *A. anophagefferens* cell has been calculated to be $2.2x10^3$ fg per cell and $3.4x10^2$ fg per cell, respectively. The phosphorus concentration is estimated to be $2.1x10^1$ fg per cell based on estimates using the Redfield ratio (Gobler et al., 1997). The amount of carbon, nitrogen, and phosphorus within individual AaV particles remains unknown, but similar values have been estimated for PBCV-1. Given PBCV-1 and AaV have similar diameters of ~1900Å (AaV: Chapter 2 of this work, PBCV-1: (Fang et al., 2019), similar genome sizes (AaV: ~371 kbp Moniruzzaman et al., 2014 vs. PBCV-1: ~331 kbp Dunigan et al.,

2012), both contain internal membranes (Rowe et al., 2008; Milrot et al., 2017), and similar capsid structures (AaV: data not shown; PBCV-1: Fang et al., 2019) we feel PBCV-1 serves as an appropriate proxy. The C:N:P ratio for PBCV-1 was estimated to be 17:5:1. This was determined by totaling the C, N, and P found in each macromolecule in the particle (Clasen and Elser, 2007). With knowledge of the macromolecules in a particle, and weight of a single particle ($1x10^9$ Da) (Van Etten et al., 1983), the amount of C, N, and P per particle was determined.

The burst size of AaV has been suggested to be dependent on multiplicity of infection where larger m.o.i's produce less viruses per cell (Brown and Bidle, 2014). This is not unique, as it has been shown in other NCLDV systems (*i.e.* Van Etten et al., 1982). In this study two multiplicities of infection were used (~20 and ~100 viruses per cell). Comparing the burst sizes with the 1.47 mM NO_3^- acclimated cultures in both experiments, our results add further evidence that multiplicity of infection influences burst size. This can be seen as the low multiplicity of infection produced a burst size 5.2 times greater than that of the high multiplicity of infection. Our estimated burst sizes are greater than those determined previously (191 viruses per cell), (2.4 – 12.6 times less than this study) (Brown and Bidle, 2014). Although the same strain of *A. anophagefferens* and AaV were used, we note the growth medium was different. The previous study utilized f/2 media (Guillard 1975), which has a natural sea water base, and different nutrients and metals. Therefore it is possible this could account for some of the differences.. For our calculations we used the burst size calculated from a high and low multiplicity of infection of 1.47 mM NO_3^- cultures: 466 viruses per cell and 2,415 viruses per cell, respectively. Although these values are much greater than those determined previously (191 viruses per cell, (Brown and Bidle, 2014)), they are still less than the theoretical maximum number of viruses that could be produced in an *A. anophagefferens* cell. *A. anophagefferens* is between 2 – 3 µm (Gobler and

Sunda, 2012), so its volume, assuming 2.5 µm, is $8.19x10^9$ nm^3, and the volume of an icosahedral with a diameter of 190 nm is $2.174x10^6$ nm^3. Therefore, the theoretical maximum number of particles that can be produced is 3,767 particles. Using these values, around one quarter of carbon and one half of nitrogen within an *Aureococcus* cell would be required to produce a burst size of 466 particles, while over 150% of the phosphorus within the cells would be required. At a high burst size, the quotas of C, N, and P within the cell would not be enough to produce the number of particles (Table 4.9). These numbers do not consider any virally encoded proteins that are required for production of new progeny but are not packaged. The PBCV-1 genome, for instance, encodes 416 proteins, yet only 148 of them are packaged (Dunigan et al., 2012). Also, these calculations do not consider whether cellular machinery required for the infection needs to be continually synthesized.

Although these estimates are somewhat crude, it suggests that the virocell has a strong requirement for external phosphorus but not nitrogen to produce a burst size similar to those determined in this study. Using the previously published transcriptome of the infected *A. anophagefferens* culture, transporters in the host were evaluated (Moniruzzaman et al., 2018). There were no NO_3^- transporters over expressed at any point sampled, and most sugar transporters were either under expressed or did not change (Figure 4.7). In contrast, several inorganic phosphate transporters were over expressed late in the infection cycle, as previously has been reported (Moniruzzaman et al., 2018). These transporters are upregulated in phosphorus stressed cultures (Wurch et al., 2011).

Differences in uptake of external nitrogen pools in eukaryotic algae seem to depend on the evolution the host virus pair, and not necessarily on their laboratory growth rate as previously hypothesized (Fowler and Cohen, 1948; Waldbauer et al., 2019). Both *Ostreococcus tauri* and *A.*

anophagefferens have similar growth rates when grown on nitrate (Worden et al., 2004), but the virocell's reliance on the external environment for nitrogen appears to be different. The lack of change in infection cycle dynamics due to external nitrogen concentration, and the presence of virus-encoded nucleotide recycling genes (Brown and Bidle, 2014; Moniruzzaman et al., 2014), suggest AaV has evolved to utilize the internal stores within *A. anophagefferens.* In contrast, viruses that infect *O. tauri* encode their own nitrogen transporters, which aid host transporters to assimilate more external nitrogen than uninfected cells, suggesting this system evolved to utilize external sources (Monier et al., 2017). AaV does not possess any putative nitrogen or phosphorus transporters.

Conclusions

Many harmful algal blooms are terminated by lysis of viruses (Brussaard et al., 2005; Vardi et al., 2012), and this process likely plays important roles in nutrient cycling. The interplay between nutrient availability for the hosts and maximized virus production is not well characterized and depends on the specific virus-host system (Maat and Brussaard, 2016). Here we have contributed to the growing data about the influence of nitrogen on the infection cycle for an ecologically important virus-host system. Acclimation of cultures to NO_3^- concentrations that negatively influence growth caused a decreased efficiency of the AaV infection cycle. Our observation suggests that nitrogen-pollution in the early stages of the bloom may influence natural virus-mediated constraints on the bloom but once infection has set in then nitrogen-loading would have no influence on this control (but may enhance bloom growth). Determining how the infection cycle of AaV responds to the many other nitrogenous species (Berg et al., 2008) may provide

further insight into how the host's metabolism shapes the outcome of interactions with this viral giant.

Acknowledgments

We thank Gary LeCleir, Mohammad Moniruzzaman, Brittany Zepernick, and Samantha Coy for discussions about this work. This work was supported by the National Science Foundation IOS1922958, the Gordon & Betty Moore Foundation (Grant EMS4971) and the Kenneth & Blaire Mossman Endowment to the University of Tennessee.

Chapter 4 Appendix

Table 4.1. Concentrations of NO_3^- and PO_4^{3-} of the different growth media.

Percent NO_3^- of $ASP_{12}A$	$[NO_3^-]$	$[PO_4^{3-}]$	N:P
100%	1.47 mM	0.12 mM	12.25
50%	0.735 mM	0.12 mM	6.125
10%	0.147 mM	0.12 mM	1.225
1%	0.0147 mM	0.12 mM	0.1225

Table 4.2. Adjusted p-values comparing differences in end point cell abundance in an uninfected growth curve based on nitrate concentration. Adjusted p-values were determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

End Point Abundances Adjusted p-values (One-way ANOVA; F = 47.31)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.145	0.901	<0.001
0.735 mM	0.145	-	0.054	<0.001
0.147 mM	0.901	0.054	-	<0.001
0.0147 mM	<0.001	<0.001	<0.001	-

Table 4.3. Adjusted p-values comparing differences in doubling time based on nitrate concentration. Adjusted p-values were determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

Doubling Times Adjusted p-values (One-way ANOVA; F = 12.1)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.941	0.053	0.036
0.735 mM	0.941	-	0.021	0.085
0.147 mM	0.053	0.021	-	<0.001
0.0147 mM	0.036	0.085	<0.001	-

Table 4.4. Adjusted p-values comparing differences in different flow cytometry parameters based on nitrate concentration. Adjusted p-values were determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

FSC-H Adjusted p-values (One-way ANOVA; F = 92.98)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.018	0.012	<0.001
0.735 mM	0.018	-	0.992	<0.001
0.147 mM	0.012	0.992	-	<0.001
0.0147 mM	<0.001	<0.001	<0.001	-
SSC-H Adjusted p-values (One-way ANOVA; F = 40.16)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.034	0.034	<0.001
0.735 mM	0.034	-	0.869	<0.001
0.147 mM	0.034	0.869	-	<0.001
0.0147 mM	<0.001	<0.001	<0.001	-
FL3-H Adjusted p-values (One-way ANOVA; F = 133.9)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.145	0.079	<0.001
0.735 mM	0.145	-	0.812	<0.001
0.147 mM	0.079	0.812	-	<0.001
0.0147 mM	<0.001	<0.001	<0.001	-

Table 4.5. Adjusted p-values comparing differences in burst size based on nitrate concentrations as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

Burst Size adjusted p-values (One-way ANOVA; F = 18.89)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.827	0.678	<0.001
0.735 mM	0.827	-	0.240	<0.001
0.147 mM	0.678	0.240	-	<0.001
0.0147 mM	<0.001	<0.001	<0.001	-

Table 4.6. Adjusted p-values comparing differences in adsorption rate based on nitrate concentrations as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

Adsorption Rates adjusted p-values (One-way ANOVA; F = 0.3886)				
[NO_3^{-1}]	**1.47 mM**	**0.735 mM**	**0.147 mM**	**0.0147 mM**
1.47 mM	-	0.813	0.900	0.776
0.735 mM	0.813	-	0.997	>0.999
0.147 mM	0.900	0.997	-	0.990
0.0147 mM	0.776	>0.999	0.990	-

Table 4.7. Adjusted p-values comparing differences in end point abundance based on nitrate concentrations added back as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

End Point adjusted p-values (One-way ANOVA; F = (40.45)			
$[NO_3^{-1}]$	**1.47 mM**	**0.0147 mM**	**0.0 mM**
1.47 mM	-	0.004	<0.001
0.0147 mM	0.004	-	<0.001
0.0 mM	<0.001	<0.001	-

Table 4.8. Adjusted p-values comparing differences in burst size based on nitrate concentrations added back as determined by one-way ANOVA with post-hoc multiple comparisons being adjusted with Tukey's HSD.

Burst Size adjusted p-values (One-way ANOVA; F = 0.6372)			
[NO_3^{-1}]	**1.47 mM**	**0.0147 mM**	**0.0 mM**
1.47 mM	-	0.566	0.988
0.0147 mM	0.566	-	0.653
0.0 mM	0.988	0.653	-

Table 4.9. Estimated use of atoms within *A. anophagefferens* cell to produce new viruses. [A]Estimates of mass are based on those of PBCV 1 which a mass of $1x10^9$ Da, and a C:N:P ratio of 17:5:1 (Clasen and Elser, 2007). [B]Based on estimated or calculated masses per *A. anophagefferens* cell (Gobler *et al*., 1997).

	1 Virus	466 Viruses	2415 Viruses
	Mass in Daltons[A]		
C	7.4×10^8 Da	3.45×10^{11} Da	1.79×10^{12} Da
N	2.2×10^8 Da	1.02×10^{11} Da	5.31×10^{11} Da
P	4.4×10^7 Da	2.05×10^{11} Da	1.06×10^{11} Da
	Mass in Grams		
C	1.23×10^{-15} g	5.73×10^{-13} g	2.97×10^{-12} g
N	3.65×10^{-16} g	1.70×10^{-13} g	8.81×10^{-13} g
P	7.31×10^{-17} g	3.41×10^{-14} g	1.77×10^{-13} g
	Percentage of atom mass in *A. anophagefferens*[B]		
C	0.055%	25.63%	132.83%
N	0.11%	51.26%	265.65%
P	0.35%	163.10%	845.25%

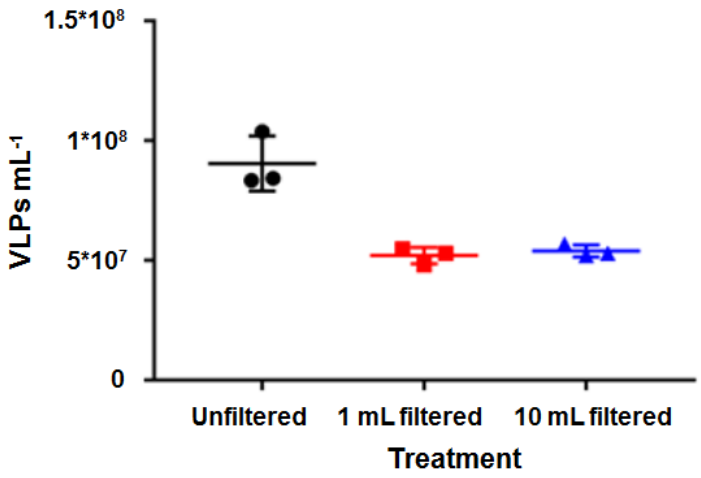


Figure 4.1. Influence of filtering AaV through a 0.45 µm filter. The different treatments were: unfiltered; 1 mL of culture passed through the filter, or 10 mL of culture passed through the filter. Results are from technical triplicates.

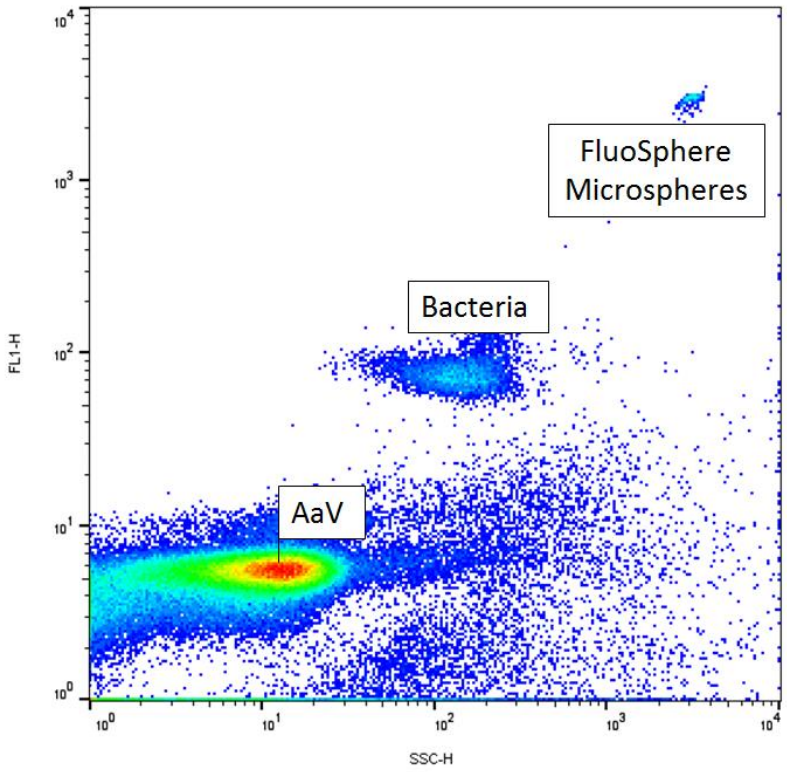


Figure 4.2. Representative FACS Scatterplot of 0.45 µm filtered lysate stained with SYBR green.

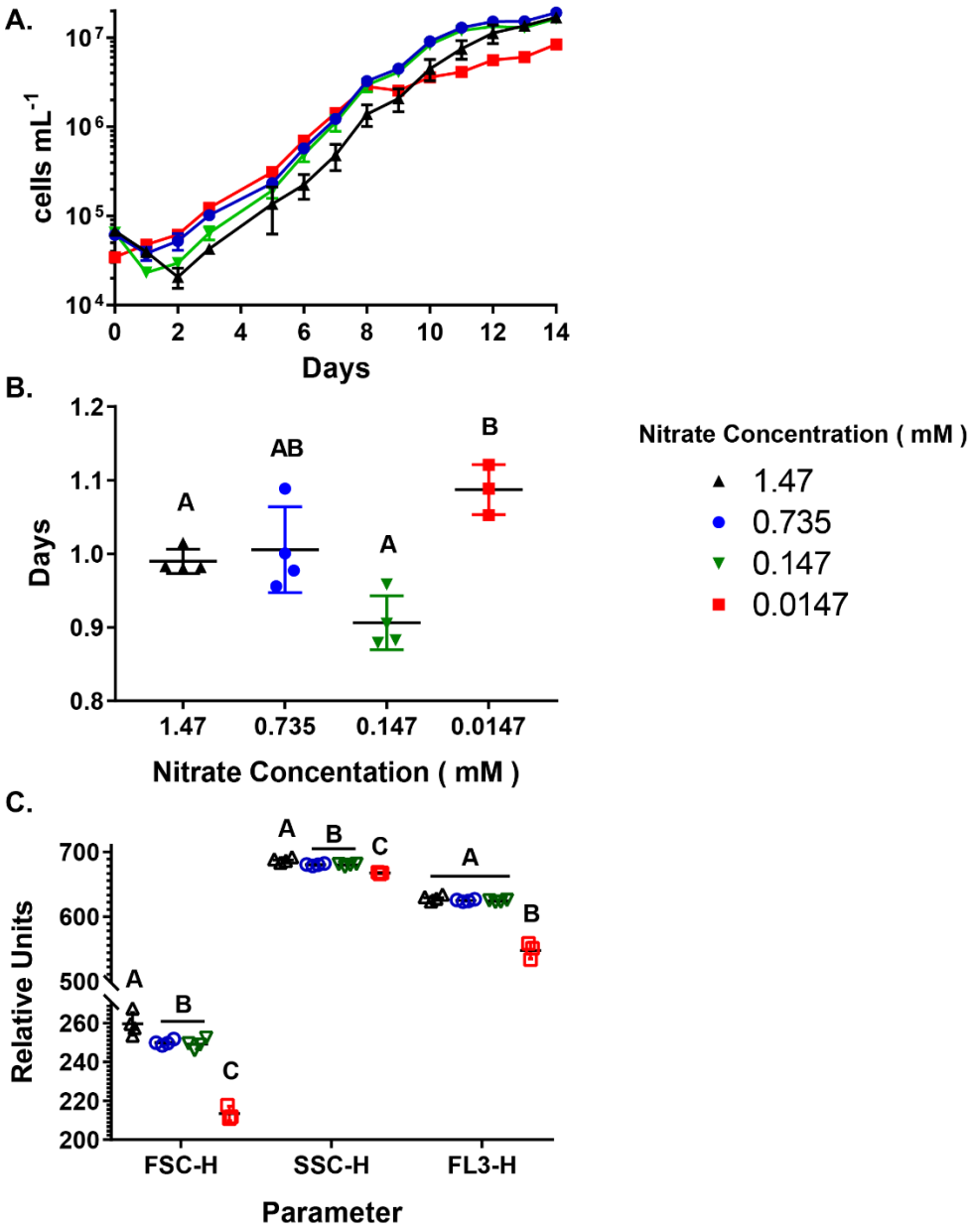


Figure 4.3. Growth characteristics of *Aureococcus anophagefferens* acclimated to decreasing NO_3^- concentrations. A) *A. anophagefferens* concentration over time. B) Doubling times of *A. anophagefferens*. C) Flow cytometry averages of two consecutive days of *A. anophagefferens* populations during logarithmic growth. FSC-H = Forward Scatter, SSC-H = Side Scatter, and FL3-H = Chlorophyll a fluorescence. The different nitrate conditions are represented by the following colors and shapes: 1.47 mM - black upward triangles, 0.735 mM - blue circles, 0.147 mM - green downward triangles, and 0.0147 mM – red squares. Letters indicate that the two values being compared are not significant ($p > 0.05$). Tables with significance values are Tables 4.2 and 4.3 for data for figures 4.3B and 4.3C, respectively. Points are for n=4 biological replicates ± SD.

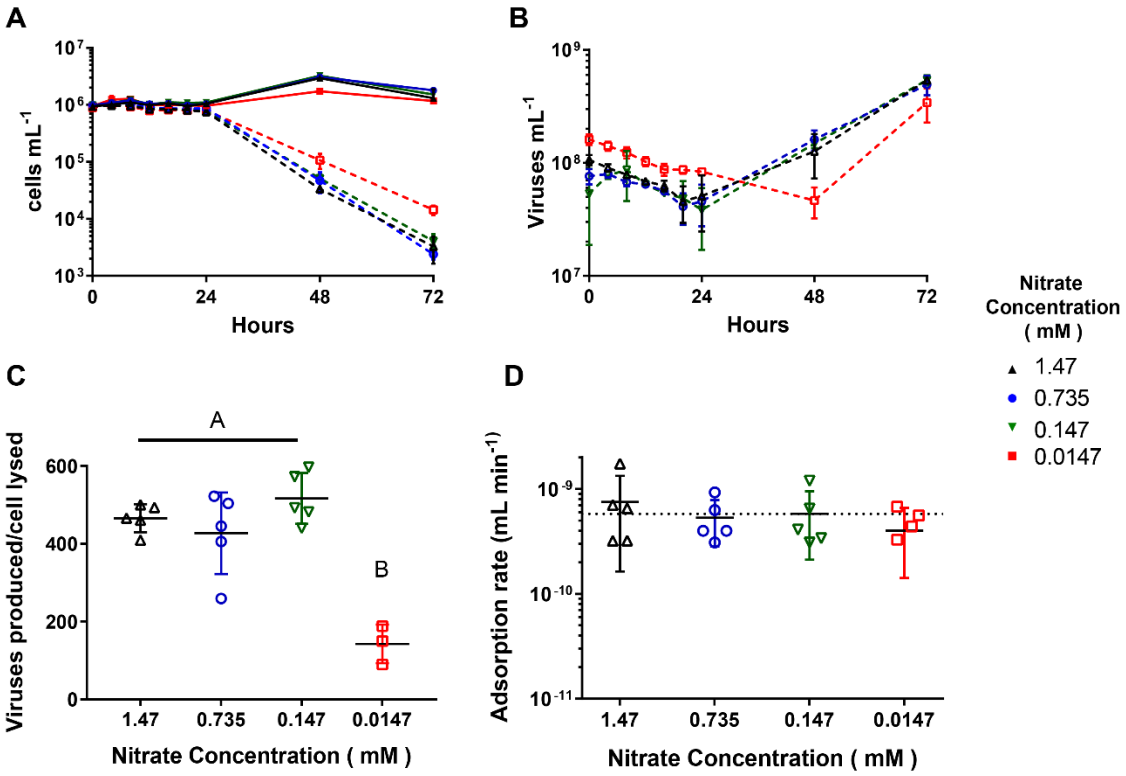


Figure 4.4. AaV infection cycle dynamics of *Aureococcus anophagefferens* acclimated to decreasing NO_3^- concentrations. A) *A. anophagefferens* and B) AaV concentrations over the three day experiment. C) AaV produced per host cell lysed. D) Adsorption rates of AaV as calculated as previously reported (Cottrell and Suttle, 1995). The dotted line represents the adsorption rate (-5.8 x 10^{-10} mL min^{-1}) previously reported for AaV (Brown and Bidle, 2014). The different nitrate conditions are represented by the following colors and symbols: 1.47 mM - black upward triangles, 0.735 mM - blue circles, 0.147 mM - green downward triangles, and 0.0147 mM – red squares. Filled in symbols with solid connecting lines represent uninfected controls while open symbols with dashed lines represent infected cultures. Letters indicate that the two values being compared are not significant ($p > 0.05$). Tables with significance values are Tables 4.5 and 4.6 for data for figures 4.4C and 4.4D, respectively. Points are for n=5 biological replicates ± SD.

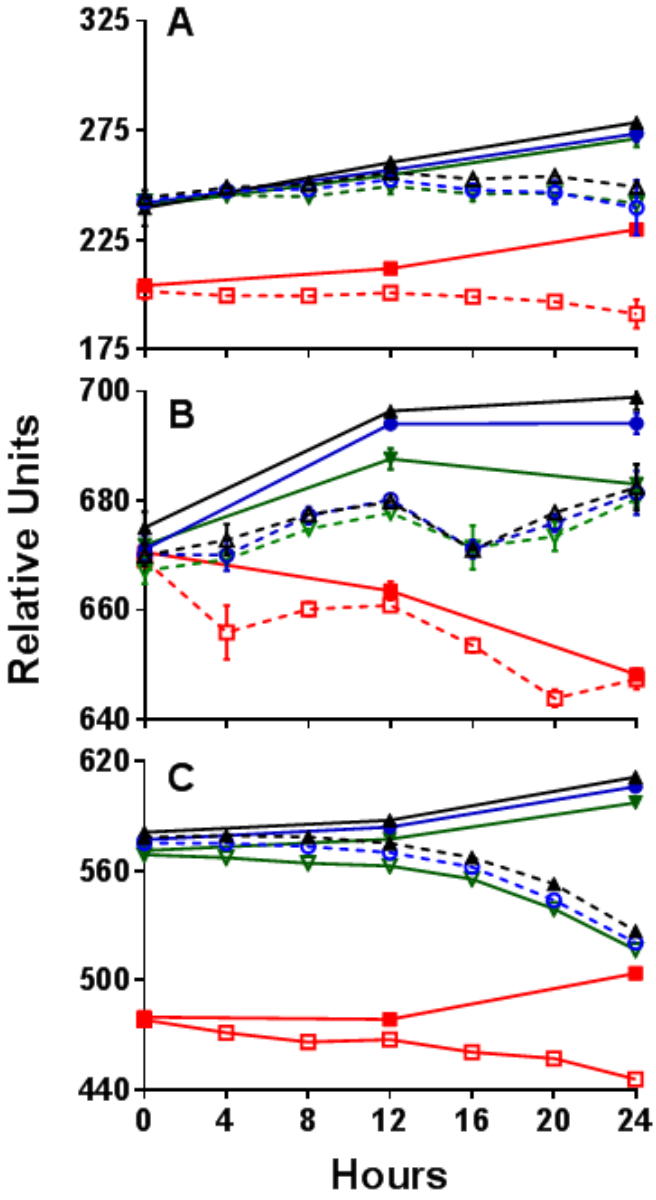


Figure 4.5. Flow cytometry parameters detected over the course of the first 24 hours of the infection cycle. The average of A) FSC-H, B) SSC-H, and C) FL3-H of each sample's gated *Aureococcus* population over time. The different nitrate conditions are represented by the following colors and symbols: 1.47 mM - black upward triangles, 0.735 mM - blue circles, 0.147 mM - green downward triangles, and 0.0147 mM – red squares. Filled in symbols with solid connecting lines represent uninfected controls while open symbols with dashed lines represent infected cultures. Points are for n=5 biological replicates ± SD.

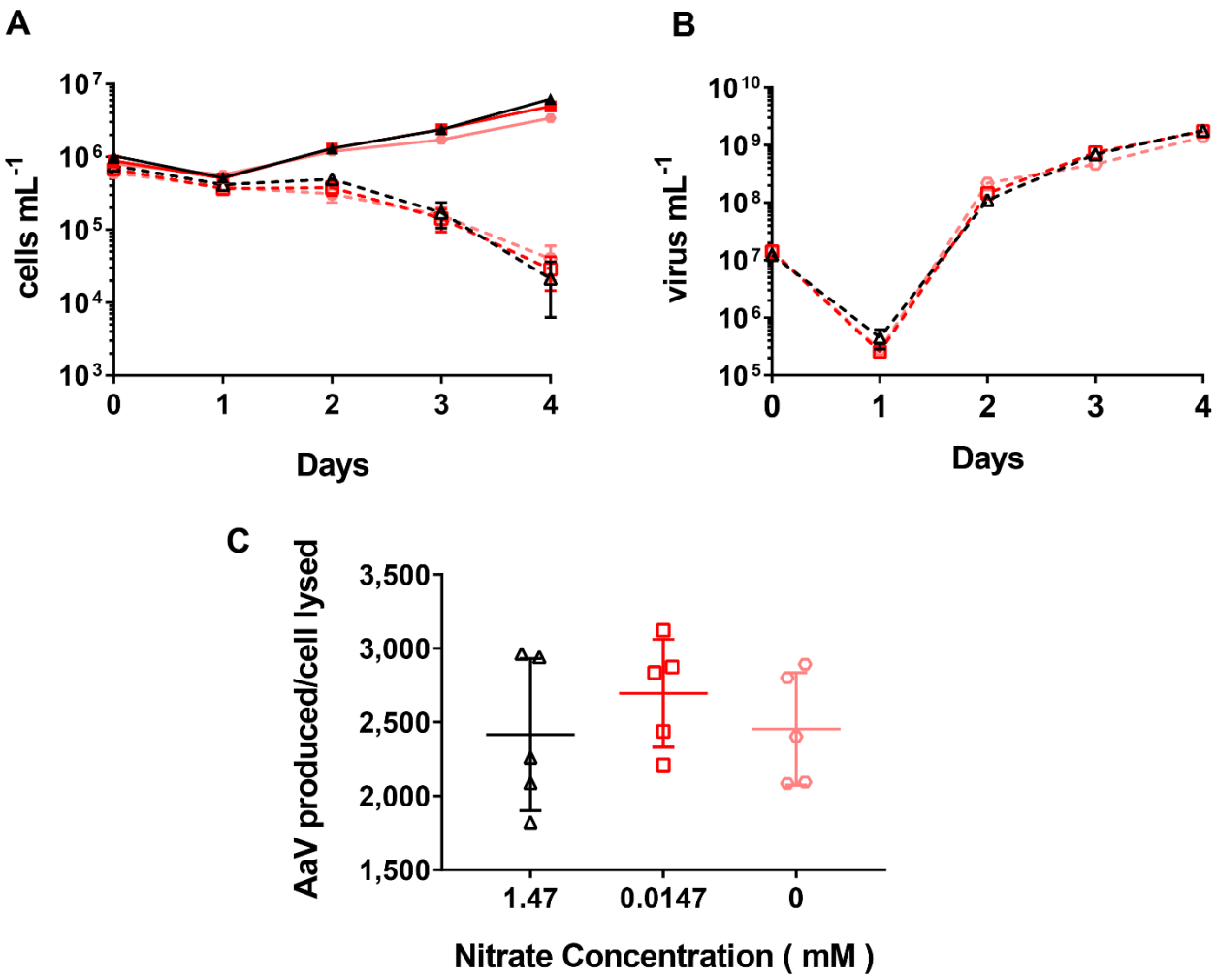


Figure 4.6. AaV infection cycle dynamics of *Aureococcus anophagefferens* acclimated to 1.47 mM NO_3^-, with different concentrations of NO_3^- added back after removing external NO_3^-. A) *A. anophagefferens* and B) AaV concentrations over the four day experiment. C) AaV produced per host cell lysed. The different nitrate conditions added back are represented by the following colors and symbols: 1.47 mM - black upward triangles, 0.0147 – red squares, and 0.0 mM = light red hexagons. Filled in symbols with solid connecting lines represent uninfected controls while open symbols with dashed lines represent infected cultures. Letters indicate that the two values being compared are not significant ($p > 0.05$). Table 4.8 has significance values data for figure 4.6C. Points are for n=5 biological replicates ± SD.

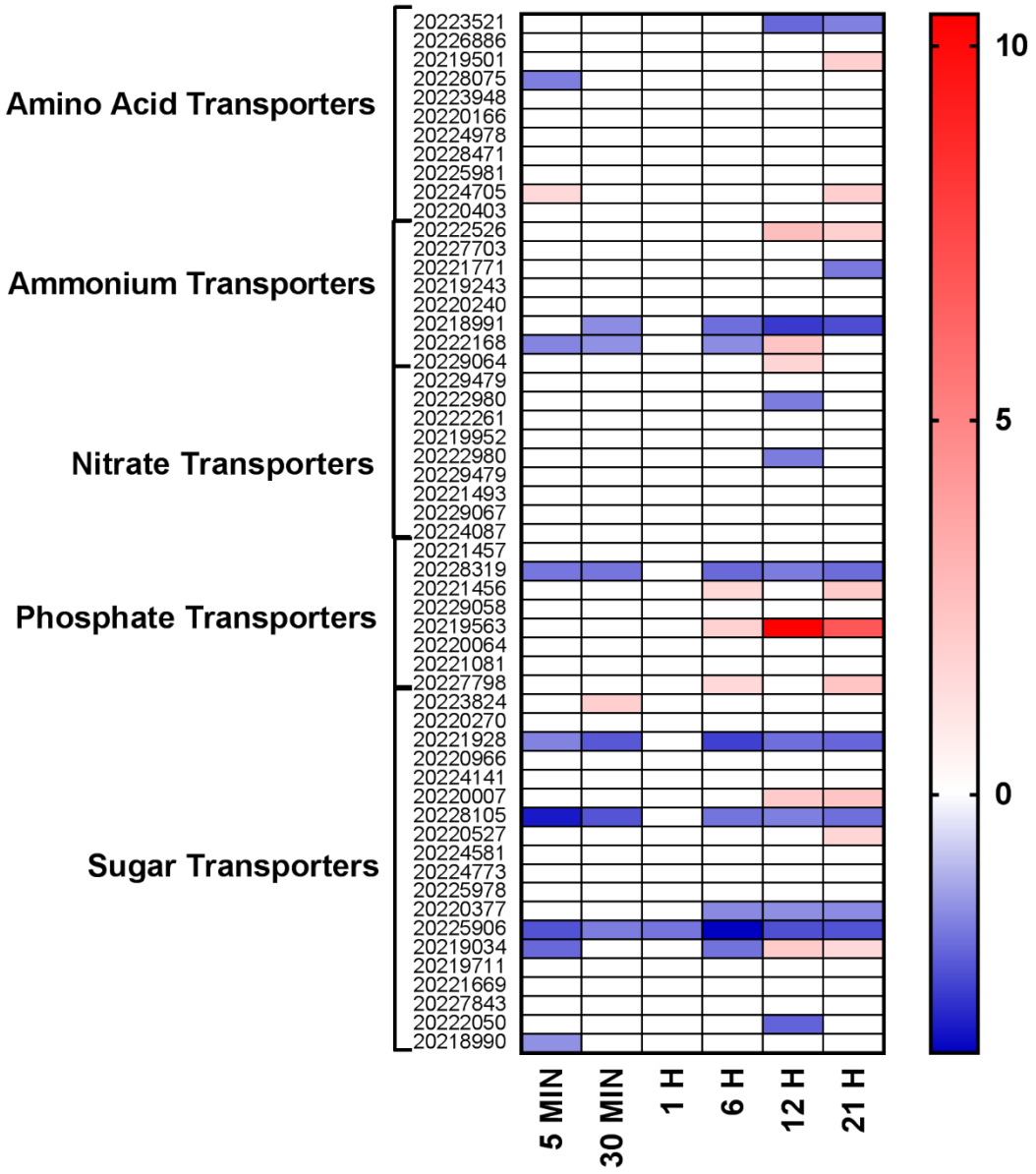


Figure 4.7. Fold change values of transporters found within the *Aureococcus* genome over the course of the infection cycle that are > 1.5 or < -1.5 fold change for infected v. uninfected control cultures with a FDR p-value < 0.05 (Moniruzzaman et al., 2018).

Chapter Five

Conclusions

Conclusions

Our work continued studying the isolated *Aureococcus anophagefferens* infecting virus, AaV. Research into this virus-host system has occurred over more than a decade, and previous work in the laboratory settings translated into understanding more about this virus host system in the environment and the evolution of the *Mimiviridae* family. Assembling the genome of AaV (Moniruzzaman *et al.*, 2014), allowed for the basis of an amplicon-based study looking at the diversity of algae infecting *Mimiviridae* in natural *A. anophagefferens* blooms (Moniruzzaman *et al.*, 2016). Performance of a transcriptomics-based study over the course of the infection cycle (Moniruzzaman *et al.*, 2018a) revealed the conserved early promoter motif found in heterotrophic eukaryote infecting Mimiviridae members (Fischer *et al.*, 2010; Legendre *et al.*, 2010) is also found in viruses that infect algae. In this current work we characterized the particle of AaV, developed new tools that can be used to study *A. anophagefferens* infecting viruses in the environment, and determined that environmentally relevant abiotic factors that can negatively influence the infection cycle.

Development of new tools

The scientific contributions of this work would not have been possible without the development of novel tools to study with this system. Transitioning from a natural seawater base (Moniruzzaman *et al.*, 2018b), to a defined growth medium (Gann, 2016), allowed for a more robust and consistent culturing of *A. anophagefferens* and its virus. The large amount of biomass required for the particle study and the ability to change nutrients would not have been possible without this. There are limitations though. We are unable to culture *A. anophagefferens* without the heterotrophic bacteria (that we perceive as “contaminants”) present in our cultures, while

previously axenic cultures have been reported to be maintained (Berg *et al.*, 2002). More work to improve this medium could allow for studies of the importance of bacteria in this system, both pertaining to the virus and not. Studies on other algal hosts should transition away from seawater-based media to allow for more control of the consistency of culturing, as well as to allow for specific questions to be asked beyond initial characterization of algal virus systems. Also developed within this dissertation were assays to enumerate total (quantitative PCR) and infectious (plaque and most probable number) AaV particles. These tools determined that a low percentage of particles (~1%) are infectious on this strain of *A. anophagefferens*. The variability in percentage of particles that are infectious varies system to system (Van Etten *et al.*, 1983; Tarutani *et al.*, 2006). It would be interesting to determine infectivity of particles cultivated on different permissible strains in multiple hosts. This would allow for a more comprehensive understanding of the differences in the number of particles produced versus those that are infectious. Determining the cause(es) for the loss of infectivity within the cell is also an interesting and an environmentally relevant problem that could be addressed. These assays allow for a robust study of the *A. anophagefferens* infecting viruses in the environment to be completed in the future. The only information about these viruses comes from transmission electron microscopy studies (Gastrich *et al.*, 2002; Gastrich *et al.*, 2004), but the concentrations of free viruses in the water column and sediment below remains unknown. Understanding fluxes of these populations is important as viruses infecting the blooming organisms can change drastically over the course of the bloom (Schroeder *et al.*, 2003; Tomaru *et al.*, 2004). This should be done with multiple strains, through multiple techniques, to understand these dynamics of viral succession. Finally, these tools can be used to look at the degradation of particles over time in the water column and the sediment to determine how *A. anophagefferens* infecting

viruses can continue to influence blooms from year to year, and how similar viruses remain present years later (Moniruzzaman *et al.*, 2016).

Structural and proteomic study of the AaV particle

Although large viruses can have distinct morphologies (Legendre *et al.*, 2015; Abrahao *et al.*, 2018), the structure and assembly of large icosahedral viruses appears to be evolutionarily conserved. Like the other reconstructions of large viruses, the reconstruction of AaV supports the continuous assembly of the trisymmetron from the pentasymmetron (Xiao *et al.*, 2017). Also, conserved is the fixed size of the pentasymmetron (Xiao *et al.*, 2017; Okamoto *et al.*, 2018; Fang *et al.*, 2019). Continuing to determine the structure of novel viruses will provide more support for the continuous assembly hypothesis and allow for more data to be available for modelling interactions between specific components of the particle (Xian *et al.*, 2019). As *A. anophagefferens* grows optimally in salinities close to sea water, the processing of AaV for cyro-electron microscopy provides a challenge for imaging with optimal resolution. As this hurdle has been overcome for other systems as well (Yan *et al.*, 2005; Fischer *et al.*, 2010), there should be more attempts with novel marine systems by adapting similar protocols. At least 43 proteins were packaged within the AaV particle. It appears that proteins packaged within the particle are somewhat conserved within families. As an example, all *Mimiviridae* family members, package machinery for transcription (Renesto *et al.*, 2006; Fischer *et al.*, 2014; Abrahao *et al.*, 2018), while in other families it is lacking (Dunigan *et al.*, 2012, Allen *et al.*, 2008). Proteins packaged within each virus also appear to provide unique interactions between the virus and its host. AaV contains proteins involved in sugar binding and break down. Whether or not these enzymes play a role in entry, the degradation of the glycocalyx which causes the balding of infected *A.*

anophagefferens cells (Sieburth *et al.*, 1988; Rowe *et al.*, 2008), or a different unknown process is to be determined. More work to characterize the activity of these proteins, their localization, and the changing in polysaccharide pools of the host during infection is needed. As with many algae-infecting viruses (Schvarcz and Steward, 2018; Stough *et al.*, 2019), the genome of AaV has many proteins with unknown function (Moniruzzaman *et al.*, 2014). Proteomic and structural studies can be an avenue to determine novel structural proteins within the genome. This has been done extensively with PBCV-1 (Fang *et al.*, 2019), but also with other large viruses (Andres *et al.*, 2019). Continuing to further characterize this system and others will allow for the discovery of the function of many unknown proteins.

Effects of altering abiotic factors on the infection cycle

Over the course of the bloom, *A. anophagefferens* is exposed to changing environmental conditions from light availability to fluctuation in nutrients (Lomas *et al.*, 1996; Wurch *et al.*, 2019). Altering abiotic factors can influence *A. anophagefferens* (Frischkorn *et al.*, 2014), and in turn can modulate the success of an AaV infection (Gobler *et al*, 2007). In this work, it was shown that the acclimation of cultures to irradiance levels or N:P ratios that reduce growth rates and biomass accumulation also reduces the efficiency of the infection cycle of AaV (*i.e.* a reduced burst size and an increased length of infection cycle). Shifting to low or no irradiance had an increased effect, where in no light, the infection could not proceed. The ability of viruses to lyse cultures in the dark is specific to the system (Baudoux and Brussaard, 2008; Thamatrakoln *et al.*, 2019), and hypotheses about nutrient stores allowing for infections to proceed should be experimental tested (Baudoux and Brussaard, 2008). Performing these experiments with sources of carbon added could provide insight into the continued infections

occurring at peak conditions, as *A. anophagefferens* can utilize organic compounds for growth (Berg *et al.*, 1997). Estimates of the amount of nitrogen in an *A. anophagefferens* cell allowed for the hypothesis that there is not a large requirement for external nitrogen to produce the number of viruses each cell does in optimal conditions. This was supported by our nutrient shift experiments. This work should be followed up on in other conditions and with the use of proteomics to determine the amount of nitrogen from the environment versus the host cell used in proteins and nucleic acids under varying conditions (Waldbauer *et al.*, 2019). Overall other -omics techniques could provide more information about the differences in cells grown in suboptimal conditions, which could then provide information to perform studies of bloom communities (Rosenwasser *et al.*, 2019), and how efficient infections of these communities would be. These interactions are critical in understanding how AaV contributes to the termination of a bloom, as well as persistence until host numbers increase again, whether it be in sediments or in the water column.

Future directions

This work was performed with a single viral isolate that has been continuously passaged on a single host strain. Without using multiple strains of the host or virus, important differences or trends may not be observed in the laboratory (Zimmerman *et al.*, 2019), or in the environment (Tomaru *et al.*, 2007). Future work in this system, and others, needs to focus on obtaining more isolates on multiple hosts. For *A. anophagefferens*, there are blooms worldwide, and some research is being conducted on strains not isolated from the northeastern coast of the United States (Dong *et al.*, 2014). Obtaining different strains may allow for the isolation of novel viruses that could not infect the United States strains. As network analyses (Moniruzzaman *et al.*, 2017)

and amplicon libraries (Moniruzzaman *et al.*, 2016) have provided evidence that viruses both similar and different than AaV occur in the blooms at high abundances, more work to understand their role in the bloom is important. Understanding differences in types of viruses (*i.e.* double-stranded DNA versus single-stranded RNA genomes) isolated on the same host should be explored, due to the interesting differences in infection cycle dynamics and persistence in nature. Finally, as a field, more systematic approaches to improving isolation, cultivation, cryo-preservation, and -omics, of current and novel systems are important to understand the many roles of algal viruses globally.

List of References

Abrahao, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., Arantes, T., *et al*. (2018). Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. **Nature Communications** 9.

Agarkova, I.V., Dunigan, D.D., and Van Etten, J.L. (2006). Virion-associated restriction endonucleases of chloroviruses. **Journal of Virology** 80(16), 8114-8123.

Alderkamp, A.C., van Rijssel, M., and Bolhuis, H. (2007). Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin. **Fems Microbiology Ecology** 59(1), 108-117.

Allen, M.J., Howard, J.A., Lilley, K.S., and Wilson, W.H. (2008). Proteomic analysis of the EhV-86 virion. **Proteome Science** 6.

Andreani, J., Khalil, J.Y.B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., *et al*. (2017). Pacmanvirus, a New Giant Icosahedral Virus at the Crossroads between Asfarviridae and Faustoviruses. **Journal of Virology** 91(14).

Andres, G., Charro, D., Matamoros, T., Dillard, R.S., and Abrescia, N.G.A. (2019). The cryo-EM structure of African swine fever virus unravels a unique architecture comprising two icosahedral protein capsids and two lipoprotein membranes. **Journal of Biological Chemistry**.

Apt, K.E., and Gibor, A. (1991). The ultrastructure of gall on the red alga *Gracilaria ephihippisora*. **Journal of Phycology** 27(3), 409-413.

Armbrust, E.V. (2009). The life of diatoms in the world's oceans. **Nature** 459(7244), 185-192.

Arsenieff, L., Simon, N., Rigaut-Jalabert, F., Le Gall, F., Chaffron, S., Corre, E., *et al*. (2019). First viruses Infecting the marine diatom *Guinardia delicatula*. **Frontiers in Microbiology** 9.

Attoui, H., Jaafar, F.M., Belhouchet, M., de Micco, P., de Lamballerie, X., and Brussaard, C.P.D. (2006). *Micromonas pusilla* reovirus: a new member of the family Reoviridae assigned to a novel proposed genus (*Mimoreovirus*). **Journal of General Virology** 87, 1375-1383.

Aumailley, M. (2013). The laminin family. **Cell Adhesion & Migration** 7(1), 48-55.

Baudoux, A.C., and Brussaard, C.P. (2005). Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. **Virology** 341(1), 80-90.

Baudoux, A.C., and Brussaard, C.P.D. (2008). Influence of irradiance on virus-algal host interactions. **Journal of Phycology** 44(4), 902-908.

Bellec, L., Grimsley, N., and Desdevises, Y. (2010). Isolation of prasinoviruses of the green unicellular algae *Ostreococcus spp.* on a worldwide geographical scale. **Applied and Environmental Microbiology** 76(1), 96-101.

Berdalet, E., Fleming, L.E., Gowen, R., Davidson, K., Hess, P., Backer, L.C., *et al*. (2015). Marine harmful algal blooms, human health and wellbeing: challenges and opportunities in the 21st century. **Journal of the Marine Biological Association of the United Kingdom**.

Berg, G.M., Glibert, P.M., Lomas, M.W., and Burford, M.A. (1997). Organic nitrogen uptake and growth by the chrysophyte *Aureococcus anophagefferens* during a brown tide event. **Marine Biology** 129(2), 377-387.

Berg, G.M., Repeta, D.J., and Laroche, J. (2002). Dissolved organic nitrogen hydrolysis rates in axenic cultures of *Aureococcus anophagefferens* (*Pelagophyceae*): comparison with heterotrophic bacteria. **Applied and Environmental Microbiology** 68(1), 401-404.

Berg, G.M., Shrager, J., Glockner, G., Arrigo, K.R., and Grossman, A.R. (2008). Understanding nitrogen limitation in *Aureococcus anophagefferens* (*Pelagophyceae*) through cDNA and qRT-PCR analysis. **Journal of Phycology** 44(5), 1235-1249.

Berges, J.A., and Falkowski, P.G. (1998). Physiological stress and cell death in marine phytoplankton: Induction of proteases in response to nitrogen or light limitation. **Limnology and Oceanography** 43(1), 129-135.

Bettarel, Y., Kan, J., Wang, K., Williamson, K.E., Cooney, S., Ribblett, S., *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.

Blanc, G., Gallot-Lavallee, L., and Maumus, F. (2015). Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. **Proceedings of the National Academy of Sciences of the United States of America** 112(38), E5318-E5326.

Blanc, G., Mozar, M., Agarkova, I.V., Gurnon, J.R., Yanai-Balser, G., Rowe, J.M., *et al.* (2014). Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in Paramecium bursaria Chlorella Virus 1. **Plos One** 9(3).

Blanc-Mathieu, R., Krasovec, M., Hebrard, M., Yau, S., Desgranges, E., Martin, J., *et al.* (2017). Population genomics of picophytoplankton unveils novel chromosome hypervariability. **Science Advances** 3(7).

Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., *et al.* (2011). Mimivirus shows dramatic genome reduction after intraamoebal culture. **Proceedings of the National Academy of Sciences of the United States of America** 108(41), 17234-17234.

Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., *et al.* (2009). Giant *Marseillevirus* highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. **Proceedings of the National Academy of Sciences of the United States of America** 106(51), 21848-21853.

Bratbak, G., Egge, J.K., and Heldal, M. (1993). Viral mortality of the marine alga *Emiliania huxleyi* (*Haptophyceae*) and termination of algal blooms. **Marine Ecology Progress Series** 93(1-2), 39-48.

Bratbak, G., Jacobsen, A., and Heldal, M. (1998). Viral lysis of *Phaeocystis pouchetii* and bacterial secondary production. **Aquatic Microbial Ecology** 16(1), 11-16.

Bratbak, G., Jacobsen, A., Heldal, M., Nagasaki, K., and Thingstad, F. (1998). Virus production in *Phaeocystis pouchetii* and its relation to host cell growth and nutrition. **Aquatic Microbial Ecology** 16(1), 1-9.

Bricelj, V.M., and Lonsdale, D.J. (1997). *Aureococcus anophagefferens*: Causes and ecological consequences of brown tides in US mid-Atlantic coastal waters. **Limnology and Oceanography** 42(5), 1023-1038.

Brister, J.R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI viral genomes resource. **Nucleic Acids Research** 43(Database issue), D571-577.

Brown, C.M., and Bidle, K.D. (2014). Attenuation of virus production at high multiplicities of infection in *Aureococcus anophagefferens*. **Virology** 466-467, 71-81.

Broyles, S.S. (2003). Vaccinia virus transcription. **Journal of General Virology** 84, 2293-2303.

Brussaard, C.P. (2004). Optimization of procedures for counting viruses by flow cytometry. **Applied and Environmental Microbiology** 70(3), 1506-1513.

Brussaard, C.P.D., Mari, X., Van Bleijswijk, J.D.L., and Veldhuis, M.J.W. (2005). A mesocosm study of *Phaeocystis globosa* (*Prymnesiophyceae*) population dynamics - II. Significance for the microbial community. **Harmful Algae** 4(5), 875-893.

Brussaard, C.P.D., Noordeloos, A.A.M., Sandaa, R.A., Heldal, M., and Bratbak, G. (2004). Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. **Virology** 319(2), 280-291.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., *et al.* (2009). BLAST+: architecture and applications. **BMC Bioinformatics** 10, 421.

Caspar, D.L.D., and Klug, A. (1962). Physical Principles in Construction of Regular Viruses. **Cold Spring Harbor Symposia on Quantitative Biology** 27, 1-&.

Castberg, T., Thyrhaug, R., Larsen, A., Sandaa, R.A., Heldal, M., Van Etten, J.L., *et al.* (2002). Isolation and characterization of a virus that infects *Emiliania huxleyi* (*Haptophyta*). **Journal of Phycology** 38(4), 767-774.

Cheng, Y.S., Labavitch, J., and VanderGheynst, J.S. (2015). Organic and inorganic nitrogen impact *Chlorella variabilis* productivity and host quality for viral production and cell lysis. **Applied Biochemistry and Biotechnology** 176(2), 467-479.

Cherrier, M.V., Kostyuchenko, V.A., Xiao, C.A., Bowman, V.D., Battisti, A.J., Yan, X.D., *et al.* (2009). An icosahedral algal virus has a complex unique vertex decorated by a spike. P**roceedings of the National Academy of Sciences of the United States of America** 106(27), 11085-11089.

Clarke, K.R. and Gorley, R.N. (2015). PRIMER v7: User Manual/Tutorial. Plymouth: PRIMER-EPIymouth.

Clasen, J.L., and Elser, J.J. (2007). The effect of host *Chlorella* NC64A carbon : phosphorus ratio on the production of Paramecium bursaria Chlorella Virus-1. **Freshwater Biology** 52(1), 112-122.

Cock, J.M., Peters, A.F., and Coelho, S.M. (2011). Brown algae. **Current Biology** 21(15), R573-R575.

Colson, P., La Scola, B., Levasseur, A., Caetano-Anolles, G., and Raoult, D. (2017). Mimivirus: leading the way in the discovery of giant viruses of amoebae. **Nature Reviews Microbiology** 15(4), 243-254

Cottrell, M.T., and Suttle, C.A. (1995). Dynamics of a lytic virus infecting the photosynthetic marine picoflagellate *Micromonas pusilla*. **Limnology and Oceanography** 40(4), 730-739.

Coy, S.R., Gann, E.R., Pound, H.L., Short, S.M., and Wilhelm, S.W. (2018). Viruses of eukaryotic algae: diversity, methods for detection, and future directions. **Viruses-Basel** 10(9).

De Castro, C., Molinaro, A., Piacente, F., Gurnon, J.R., Sturiale, L., Palmigiano, A., *et al.* (2013). Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. **Proceedings of the National Academy of Sciences of the United States of America** 110(34), 13956-13960.

Delaroque, N., and Boland, W. (2008). The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. **BMC Evolutionary Biology** 8.

Derelle, E., Monier, A., Cooke, R., Worden, A.Z., Grimsley, N.H., and Moreau, H. (2015). Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*. **Journal of Virology** 89(11), 5812-5821.

DeYoe, H.R., Stockwell, D.A., Bidigare, R.R., Latasa, M., Johnson, P.W., Hargraves, P.E., *et al.* (1997). Description and characterization of the algal species *Aureoumbra lagunensis gen. et sp. nov.* and referral of *Aureoumbra* and *Aureococcus* to the *Pelagophyceae*. **Journal of Phycology** 33(6), 1042-1048.

Doblin, M.A., Popels, L.C., Coyne, K.J., Hutchins, D.A., Cary, S.C., and Dobbs, F.C. (2004). Transport of the harmful bloom alga *Aureococcus anophagefferens* by oceangoing ships and coastal boats. **Applied and Environmental Microbiology** 70(11), 6495-6500.

Dong, H.P., Huang, K.X., Wang, H.L., Lu, S.H., Cen, J.Y., and Dong, Y.L. (2014). Understanding strategy of nitrate and urea assimilation in a Chinese strain of *Aureococcus anophagefferens* through RNA-Seq Analysis. **Plos One** 9(10).

Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., *et al.* (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. **Journal of Proteome Research** 13(8), 3679-3684.

Dunigan, D.D., Al-Sammak, M., Al-Ameeli, Z., Agarkovao, I.V., DeLong, J.P., and Van Etten, J.L. (2019). Chloroviruses lure hosts through long-distance chemical signaling. **Journal of Virology** 93(7).

Dunigan, D.D., Cerny, R.L., Bauman, A.T., Roach, J.C., Lane, L.C., Agarkova, I.V., *et al.* (2012). Paramecium bursaria Chlorella Virus 1 proteome reveals novel architectural and regulatory features of a giant virus. **Journal of Virology** 86(16), 8821-8834.

Eissler, Y., Wang, K., Chen, F., Wommack, K.E., and Coats, D.W. (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.

Elifantz, H., Dittell, A.I., Cottrell, M.T., and Kirchman, D.L. (2007). Dissolved organic matter assimilation by heterotrophic bacterial groups in the western Arctic Ocean. **Aquatic Microbial Ecology** 50(1), 39-49.

Fabre, E., Jeudy, S., Santini, S., Legendre, M., Trauchessec, M., Coute, Y., *et al.* (2017). Noumeavirus replication relies on a transient remote control of the host nucleus. **Nature Communications** 8, 15087.

Fagerbakke, K.M., Heldal, M., and Norland, S. (1996). Content of carbon, nitrogen, oxygen, sulfur and phosphorus in native aquatic and cultured bacteria. **Aquatic Microbial Ecology** 10(1), 15-27.

Fan, C., Glibert, P.M., Alexander, J., and Lomas, M.W. (2003). Characterization of urease activity in three marine phytoplankton species, *Aureococcus anophagefferens*, *Prorocentrum minimum*, and *Thalassiosira weissflogii*. **Marine Biology** 142(5), 949-958.

Fang, Q.L., Zhu, D.J., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., *et al.* (2019). Near-atomic structure of a giant virus. **Nature Communications** 10.

Finke, J.F., Winget, D.M., Chan, A.M., and Suttle, C.A. (2017). Variation in the genetic repertoire of viruses infecting *Micromonas pusilla* reflects horizontal gene transfer and links to their environmental distribution. **Viruses-Basel** 9(5).

Fischer, M.G., Allen, M.J., Wilson, W.H., and Suttle, C.A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. **Proceedings of the National Academy of Sciences of the United States of America** 107(45), 19508-19513.

Fischer, M.G., Kelly, I., Foster, L.J., and Suttle, C.A. (2014). The virion of Cafeteria roenbergensis virus (CroV) contains a complex suite of proteins for transcription and DNA repair. **Virology** 466, 82-94.

Forterre, P. (2011). Manipulation of cellular syntheses and the nature of viruses: the virocell concept. **Comptes Rendus Chimie** 14(4), 392-399.

Fowler, C.B., and Cohen, S.S. (1948). Chemical studies in host-virus interactions; a method of determining nutritional requirements for bacterial virus multiplication. **Journal of Experimental Medicine** 87(4), 259-274.

Frada, M., Probert, I., Allen, M.J., Wilson, W.H., and de Vargas, C. (2008). The "Cheshire Cat" escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. **Proceedings of the National Academy of Sciences of the United States of America** 105(41), 15944-15949.

Frada, M.J., Rosenwasser, S., Ben-Dor, S., Shemi, A., Sabanay, H., and Vardi, A. (2017). Morphological switch to a resistant subpopulation in response to viral infection in the bloom-forming coccolithophore *Emiliania huxleyi.* **Plos Pathogens** 13(12).

Frazier, A.D., Rowe, J.M., Rentz, C.A., Gobler, C.J., and Wilhelm, S.W. (2007). Bacterial lysis of *Aureococcus anophagefferens* CCMP 1784 (*Pelagophyceae*). **Journal of Phycology** 43(3), 461-465.

Frischkorn, K.R., Harke, M.J., Gobler, C.J., and Dyhrman, S.T. (2014). *De novo* assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. **Frontiers in Microbiology** 5, 375.

Gallot-Lavallee, L., and Blanc, G. (2017). A glimpse of nucleo-cytoplasmic large DNA virus biodiversity through the eukaryotic genomics window. **Viruses** 9(1).

Gallot-Lavallee, L., Blanc, G., and Claverie, J.M. (2017). Comparative Genomics of Chrysochromulina ericina Virus and other microalga-infecting large DNA viruses highlights their intricate evolutionary relationship with the established *Mimiviridae* family. **Journal of Virology** 91(14).

Gallot-Lavallee, L., Pagarete, A., Legendre, M., Santini, S., Sandaa, R.A., Himmelbauer, H., *et al.* (2015). The 474-kilobase-pair complete genome sequence of CeV-01B, a virus infecting *Haptolina* (*Chrysochromulina*) *ericina* (*Prymnesiophyceae*). **Microbiology Resource Announcements** 3(6).

Gann, E. (2017). Plaque Assay for the Aureococcus anophagefferens Virus (AaV) [Online]. dx.doi.org/10.17504/protocols.io.g2nbyde: https://www.protocols.io/view/plaque-assay-for-the-aureococcus-anophagefferens-v-g2nbyde.

Gann, E.R. (2016). ASP12A Recipe for culturing *Aureococcus anophagefferens* [Online]. Available: https://www.protocols.io/view/asp12a-recipe-for-culturing-aureococcus-anophageff-f3ybqpw.

Gann, E.R., Pound, H.L., and Wilhelm, S.W. (2019). Python Script: Extracting aligned portion of viral hallmark genes from blastx sequence. [Online]. Available: https://github.com/Wilhelmlab/general-scripts/blob/master/Pound2019_Extract_aligned.py.

Gastrich, M.D., Anderson, O.R., and Cosper, E.M. (2002). Viral-like particles (VLPS) in the alga, *Aureococcus anophagefferens* (*Pelagophyceae*), during 1999-2000 brown tide blooms in Little Egg Harbor, New Jersey. **Estuaries** 25(5), 938-943.

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

Gobler, C.J., Anderson, O.R., Gastrich, M.D., and Wilhelm, S.W. (2007). Ecological aspects of viral infection and lysis in the harmful brown tide alga *Aureococcus anophagefferens*. **Aquatic Microbial Ecology** 47(1), 25-36.

Gobler, C.J., Berry, D.L., Dyhrman, S.T., Wilhelm, S.W., Salamov, A., Lobanov, A.V., *et al.* (2011). Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. **Proceedings of the National Academy of Sciences of the United States of America** 108(11), 4352-4357.

Gobler, C.J., Boneillo, G.E., Debenham, C.J., and Caron, D.A. (2004). Nutrient limitation, organic matter cycling, and plankton dynamics during an *Aureococcus anophagefferens* bloom. **Aquatic Microbial Ecology** 35(1), 31-43.

Gobler, C.J., Hutchins, D.A., Fisher, N.S., Cosper, E.M., and Sãnudo-Wilhelmy, S.A. (1997). Release and bioavailability of C, N, P, Se, and Fe following viral lysis of a marine chrysophyte. **Limnology and Oceanography** 42(7), 1492-1504.

Gobler, C.J., Lobanov, A.V., Tang, Y.Z., Turanov, A.A., Zhang, Y., Doblin, M., *et al.* (2013). The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, *Aureococcus anophagefferens*. **ISME Journal** 7(7), 1333-1343.

Gobler, C.J., and Sunda, W.G. (2012). Ecosystem disruptive algal blooms of the brown tide species, *Aureococcus anophagefferens* and *Aureoumbra lagunensis*. **Harmful Algae** 14, 36-45.

Gonzalez, J.M., and Suttle, C.A. (1993). Grazing by marine nanoflagellates on viruses and virus-sized particles - ingestion and digestion. **Marine Ecology Progress Series** 94(1), 1-10.

Graves, M.V., Burbank, D.E., Roth, R., Heuser, J., DeAngelis, P.L., and Van Etten, J.L. (1999). Hyaluronan synthesis in virus PBCV-1-infected *Chlorella*-like Green Algae. **Virology** 257(1), 15-23.

Grigorieff, N. (2007). FREALIGN: High-resolution refinement of single particle structures. **Journal of Structural Biology** 157(1), 117-125.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., *et al.* (2016). Plankton networks driving carbon export in the oligotrophic ocean. **Nature** 532(7600), 465-470.

Guillard, R.R.L. 1975. Culture of phytoplankton for feeding marine invertebrates. pp 26- 60. In Smith W.L. and Chanley M.H (Eds.) Culture of Marine Invertebrate Animals. Plenum Press, New York, USA

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic Biology** 59(3), 307-321.

Guiry, M.D. (2012). How many species of algae are there? **Journal of Phycology** 48(5), 1057-1063.

Haaber, J., and Middelboe, M. (2009). Viral lysis of *Phaeocystis pouchetii*: implications for algal population dynamics and heterotrophic C, N and P cycling. **ISME Journal** 3(4), 430-441.

Heath, S.E., Knox, K., Vale, P.F., and Collins, S. (2017). Virus resistance is not costly in a marine alga evolving under multiple environmental stressors. **Viruses-Basel** 9(3).

Henry, E.C., and Meints, R.H. (1992). A persistent virus-infection in *Feldmannia* (*Phaeophyceae*). **Journal of Phycology** 28(4), 517-526.

Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J., *et al.* (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. **ISME Journal** 7(9), 1678-1695.

Horiguchi, T. (1995). Heterocapsa circularisquama sp. nov. (Peridiniales, Dinophyceae): A new marine dinoflagellate causing mass mortality of bivalves in Japan. **Phycological Research** 43(3), 129-136.

Hou, Y., Zhang, H., Miranda, L., and Lin, S. (2010). Serious overestimation in quantitative PCR by circular (supercoiled) plasmid standard: microalgal *pcna* as the model gene. **PLoS One** 5(3), e9545.

Hyun-Kyoung, K., Jae-Chang, L., Nam-Hyun, K., Song-Hui, K., Guk, K.J., and Chul, C.K. (2007). Purification and characterization of the extracellular alginate lyase from Streptomyces sp. MET 0515. **Journal of Life Science** 17(5).

Jacobsen, A., Bratbak, G., and Heldal, M. (1996). Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (*Prymnesiophyceae*). **Journal of Phycology** 32(6), 923-927.

Jacobsen, A., Larsen, A., Martinez-Martinez, J., Verity, P.G., and Frischer, M.E. (2007). Susceptibility of colonies and colonial cells of *Phaeocystis pouchetii* (*Haptophyta*) to viral infection. **Aquatic Microbial Ecology** 48(2), 105-112.

Jarvis, B., Wilrich, C., and Wilrich, P.T. (2010). Reconsideration of the derivation of Most Probable Numbers, their standard deviations, confidence bounds and rarity values. **Journal of Applied Microbiology** 109(5), 1660-1667.

Jeanniard, A., Dunigan, D.D., Gurnon, J.R., Agarkova, I.V., Kang, M., Vitek, J., *et al.* (2013). Towards defining the chloroviruses: a genomic journey through a genus of large DNA viruses. **BMC Genomics** 14.

Johannessen, T.V., Bratbak, G., Larsen, A., Ogata, H., Egge, E.S., Edvardsen, B., *et al.* (2015). Characterisation of three novel giant viruses reveals huge diversity among viruses infecting *Prymnesiales* (*Haptophyta*). **Virology** 476, 180-188.

Jover, L.F., Effler, T.C., Buchan, A., Wilhelm, S.W., and Weitz, J.S. (2014). The elemental composition of virus particles: implications for marine biogeochemical cycles. **Nature Reviews Microbiology** 12(7), 519-528.

Juneau, P., Lawrence, J.E., Suttle, C.A., and Harrison, P.J. (2003). Effects of viral infection on photosynthetic processes in the bloom-forming alga *Heterosigma akashiwo*. **Aquatic Microbial Ecology** 31(1), 9-17.

Kana, T.M., Lomas, M.W., MacIntyre, H.L., Cornwell, J.C., and Gobler, C.J. (2004). Stimulation of the brown tide organism, *Aureococcus anophagefferens*, by selective nutrient additions to *in situ* mesocosms. **Harmful Algae** 3(4), 377-388.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. **Nucleic Acids Research** 45(D1), D353-D361.

Kapp, M., Knippers, R., and Mueller, D.G. (1997). New members of a group of DNA viruses infecting brown algae. **Phycological Research** 45(2), 85-90.

Kief, D.R., and Warner, J.R. (1981). Coordinate control of syntheses of ribosomal ribonucleic-acid and ribosomal-proteins during nutritional shift-up in *Saccharomyces cerevisiae*. **Molecular and Cellular Biology** 1(11), 1007-1015.

Kim, J., Kim, C.H., Takano, Y., Jang, I.K., Kim, S.W., and Choi, T.J. (2012). Isolation and physiological characterization of a new algicidal virus infecting the harmful dinoflagellate *Heterocapsa pygmaea*. ***Plant Pathology Journal*** *28*(4), 433-438.

Kimura, K., and Tomaru, Y. (2013). Isolation and characterization of a single-stranded DNA virus infecting the marine diatom *Chaetoceros sp* Strain SS628-11 Isolated from Western Japan. **Plos One** 8(12).

Kimura, K., and Tomarua, Y. (2015). Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. **Applied and Environmental Microbiology** 81(3), 1120-1131.

Klose, T., Reteno, D.G., Benamar, S., Hollerbach, A., Colson, P., La Scola, B., *et al.* (2016). Structure of faustovirus, a large dsDNA virus. **Proceedings of the National Academy of Sciences of the United States of America** 113(22), 6206-6211.

Kranzler, C.F., Krause, J.W., Brzezinski, M.A., Edwards, B.R., Biggs, W.P., Maniscalco, M., *et al.* (2019). Silicon limitation facilitates virus infection and mortality of marine diatoms. **Nature Microbiology** 4(11), 1790-1797.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al.* (2009). Circos: an information aesthetic for comparative genomics. **Genome Research** 19(9), 1639-1645.

Kuznetsov, Y.G., Xiao, C.A., Sun, S.Y., Raoult, D., Rossmann, M., and McPherson, A. (2010). Atomic force microscopy investigation of the giant mimivirus. **Virology** 404(1), 127-137.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., *et al.* (2003). A giant virus in amoebae. **Science** 299(5615), 2033.

Laber, C.P., Hunter, J.E., Carvalho, F., Collins, J.R., Hunter, E.J., Schieler, B.M., *et al.* (2018). Coccolithovirus facilitation of carbon export in the North Atlantic. **Nature Microbiology** 3(5), 537-547.

Lachnit, T., Thomas, T., and Steinberg, P. (2016). Expanding our understanding of the seaweed holobiont: RNA viruses of the red alga *Delisea pulchra*. **Frontiers in Microbiology** 6.

Lang, A.S., Culley, A.I., and Suttle, C.A. (2004). Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. **Virology** 320(2), 206-217.

Larsen, A., Flaten, G.A.F., Sandaa, R.A., Castberg, T., Thyrhaug, R., Erga, S.R., *et al.* (2004). Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. **Limnology and Oceanography** 49(1), 180-190.

Latka, A., Maciejewska, B., Majkowska-Skrobek, G., Briers, Y., and Drulis-Kawa, Z. (2017). Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. **Applied Microbiology Biotechnology** 101(8), 3103-3119.

Lawrence, J.E., Brussaard, C.P.D., and Suttle, C.A. (2006). Virus-specific responses of *Heterosigma akashiwo* to infection. **Applied and Environmental Microbiology** 72(12), 7829-7834.

Lawrence, J.E., Chan, A.M., and Suttle, C.A. (2001). A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (*Raphidophyceae*). **Journal of Phycology** 37(2), 216-222.

Lawrence, J.E., Chan, A.M., and Suttle, C.A. (2002). Viruses causing lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (*Raphidophyceae*) are widespread in coastal sediments of British Columbia, Canada. **Limnology and Oceanography** 47(2), 545-550.

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

Lee, R.E. (1971). Systemic viral material in the cells of the freshwater red alga *Sirodotia tenuissima* (Holden) Skuja. **Journal of Cell Science** 8(3), 623.

Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., *et al.* (2010). mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. **Genome research** 20(5), 664-674.

Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., *et al.* (2014). Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. **Proceedings of the National Academy of Sciences of the United States of America** 111(11), 4274-4279.

Legendre, M., Fabre, E., Poirot, O., Jeudy, S., Lartigue, A., Alempic, J.M., *et al.* (2018). Diversity and evolution of the emerging Pandoraviridae family. **Nature Communications** 9(1), 2285.

Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., *et al.* (2015). In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. **Proceedings of the National Academy of Sciences of the United States of America** 112(38), E5327-5335.

Lomas, M.W., Glibert, P.M., Berg, G.M., and Burford, M. (1996). Characterization of nitrogen uptake by natural populations of *Aureococcus anophagefferens* (*Chrysophyceae*) as a function of incubation duration, substrate concentration, light, and temperature. **Journal of Phycology** 32(6), 907-916.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. **Nucleic Acids Research** 42(Database issue), D490-495.

Lonborg, C., Middelboe, M., and Brussaard, C.P.D. (2013). Viral lysis of *Micromonas pusilla*: impacts on dissolved organic matter production and composition. **Biogeochemistry** 116(1-3), 231-240.

Long, A.M., and Short, S.M. (2016). Seasonal determinations of algal virus decay rates reveal overwintering in a temperate freshwater pond. **ISME Journal** 10(7), 1602-1612.

Maat, D.S., Bale, N.J., Hopmans, E.C., Damste, J.S.S., Schouten, S., and Brussaard, C.P.D. (2016). Increasing P limitation and viral infection impact lipid remodeling of the picophytoplankter *Micromonas pusilla*. **Biogeosciences** 13(5), 1667-1676.

Maat, D.S., Biggs, T., Evans, C., van Bleijswijk, J.D.L., van der Wel, N.N., Dutilh, B.E., *et al*. (2017). Characterization and temperature dependence of arctic *Micromonas polaris* viruses. **Viruses-Basel** 9(6).

Maat, D.S., and Brussaard, C.P.D. (2016). Both phosphorus- and nitrogen limitation constrain viral proliferation in marine phytoplankton. **Aquatic Microbial Ecology** 77(2), 87-97.

Maat, D.S., de Blok, R., and Brussaard, C.P.D. (2016). Combined phosphorus limitation and light stress prevent viral proliferation in the phytoplankton species *Phaeocystis globosa*, but not in *Micromonas pusilla*. **Frontiers in Marine Science** 3.

Maat, D.S., van Bleijswijk, J.D.L., Witte, H.J., and Brussaard, C.P.D. (2016). Virus production in phosphorus-limited *Micromonas pusilla* stimulated by a supply of naturally low concentrations of different phosphorus sources, far into the lytic cycle. **FEMS Microbiology Ecology** 92(9).

Maier, I., Wolf, S., Delaroque, N., Muller, D.G., and Kawai, H. (1998). A DNA virus infecting the marine brown alga *Pilayella littoralis* (*Ectocarpales*, *Phaeophyceae*) in culture. **European Journal of Phycology** 33(3), 213-220.

Maruyama, F., and Ueki, S. (2016). Evolution and phylogeny of large DNA viruses, *Mimiviridae* and *Phycodnaviridae* including newly characterized Heterosigma akashiwo Virus. **Frontiers in Microbiology** 7.

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

McKeown, D.A., Schroeder, J.L., Stevens, K., Peters, A.F., Saez, C.A., Park, J., *et al.* (2018). Phaeoviral infections are present in *Macrocystis*, *Ecklonia* and *Undaria* (*Laminariales*) and are influenced by wave exposure in *Ectocarpales*. **Viruses-Basel** 10(8).

McKeown, D.A., Stevens, K., Peters, A.F., Bond, P., Harper, G.M., Brownlee, C., *et al.* (2017). Phaeoviruses discovered in kelp (*Laminariales*). **ISME Journal** 11(12), 2869-2873.

Miller, J.L., Woodward, J., Chen, S., Jaffer, M., Weber, B., Nagasaki, K., *et al.* (2011). Three-dimensional reconstruction of Heterocapsa circularisquama RNA virus by electron cryo-microscopy. **Journal of General Virology** 92(8), 1960-1970.

Milligan, A.J., and Cosper, E.M. (1997). Growth and photosynthesis of the 'brown tide' microalga *Aureococcus anophagefferens* in subsaturating constant and fluctuating irradiance. **Marine Ecology Progress Series** 153, 67-75.

Milligan, K.L.D., and Cosper, E.M. (1994). Isolation of virus capable of lysing the brown tide microalga, *Aureococcus anophagefferens*. **Science** 266(5186), 805-807.

Milrot, E., Shimoni, E., Dadosh, T., Rechav, K., Unger, T., Van Etten, J.L., *et al.* (2017). Structural studies demonstrating a bacteriophage-like replication cycle of the eukaryote-infecting Paramecium bursaria chlorella virus-1. **Plos Pathogens** 13(8).

Mirza, S.F., Staniewski, M.A., Short, C.M., Long, A.M., Chaban, Y.V., and Short, S.M. (2015). Isolation and characterization of a virus infecting the freshwater algae *Chrysochromulina parva*. **Virology** 486, 105-115.

Miyagawa-Yamaguchi, A., Okami, T., Kira, N., Yamaguchi, H., Ohnishi, K., and Adachi, M. (2011). Stable nuclear transformation of the diatom *Chaetoceros sp*. **Phycological Research** 59(2), 113-119.

Mizuno, C.M., Guyomar, C., Roux, S., Lavigne, R., Rodriguez-Valera, F., Sullivan, M.B., *et al.* (2019). Numerous cultivated and uncultivated viruses encode ribosomal proteins. **Nature Communications** 10.

Monier, A., Chambouvet, A., Milner, D.S., Attah, V., Terrado, R., Lovejoy, C., *et al.* (2017). Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. **Proceedings of the National Academy of Sciences of the United States of America** 114(36), E7489-E7498.

Monier, A., Welsh, R.M., Gentemann, C., Weinstock, G., Sodergren, E., Armbrust, E.V., *et al.* (2012). Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. **Environmental Microbiology** 14(1), 162-176.

Moniruzzaman, M., Gann, E.R., LeCleir, G.R., Kang, Y., Gobler, C.J., and Wilhelm, S.W. (2016). Diversity and dynamics of algal Megaviridae members during a harmful brown tide caused by the pelagophyte, *Aureococcus anophagefferens*. **FEMS Microbiology Ecology** 92(5).

Moniruzzaman, M., Gann, E.R., and Wilhelm, S.W. (2018). Infection by a giant virus (AaV) induces widespread physiological reprogramming in *Aureococcus anophagefferens* CCMP1984 - a harmful bloom algae. **Frontiers in Microbiology** 9.

Moniruzzaman, M., LeCleir, G.R., Brown, C.M., Gobler, C.J., Bidle, K.D., Wilson, W.H., *et al.* (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.

Moniruzzaman, M., Wurch, L.L., Alexander, H., Dyhrman, S.T., Gobler, C.J., and Wilhelm, S.W. (2017). Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. **Nature Communications** 8.

Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010). Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. **Journal of Virology** 84(24), 12555-12563.

Motegi, C., and Nagata, T. (2007). Enhancement of viral production by addition of nitrogen or nitrogen plus carbon in subtropical surface waters of the South Pacific. **Aquatic Microbial Ecology** 48(1), 27-34.

Muhlenbruch, M., Grossart, H.P., Eigemann, F., and Voss, M. (2018). Mini-review: Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. **Environmental Microbiology** 20(8), 2671-2685.

Mutsafi, Y., Zauberman, N., Sabanay, I., and Minsky, A. (2010). Vaccinia-like cytoplasmic replication of the giant Mimivirus. **Proceedings of the National Academy of Sciences of the United States of America** 107(13), 5978-5982.

Nagasaki, K., Ando, M., Itakura, S., Imai, I., and Ishida, Y. (1994). Viral mortality in the final stage of *Heterosigma akashiwo* (*Raphidophyceae*) red tide. **Journal of Plankton Research** 16(11), 1595-1599.

Nagasaki, K., Tomaru, Y., Katanozaka, N., Shirai, Y., Nishida, K., Itakura, S., *et al.* (2004). Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. **Applied and Environmental Microbiology** 70(2), 704-711.

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

Nagasaki, K., Tomaru, Y., Tarutani, K., Katanozaka, N., Yamanaka, S., Tanabe, H., *et al.* (2003). Growth characteristics and intraspecies host specificity of a large virus infecting the dinoflagellate He*terocapsa circulatisquama*. **Applied and Environmental Microbiology** 69(5), 2580-2586.

Nagasaki, K., and Yamaguchi, M. (1997). Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (*Raphidophyceae*). **Aquatic Microbial Ecology** 13(2), 135-140.

Nakayama, N., and Hamaguchi, M. (2016). Multiplex reverse transcription quantitative PCR detection of a single-stranded RNA virus HcRNAV infecting the bloom-forming dinoflagellate *Heterocapsa circularisquama*. **Limnology and Oceanography: Methods** 14(6), 370-380.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X.D., Baker, T.S., Graves, M.V., *et al.* (2002). The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. **Proceedings of the National Academy of Sciences of the United States of America** 99(23), 14758-14763.

Needham, D.M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C.J., Hehenberger, E., et al. (2019). A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. **Proceedings of the National Academy of Sciences of the United States of America** 116(41), 20574.

Neupärtl, M., Meyer, C., Woll, I., Frohns, F., Kang, M., Van Etten, J.L., *et al.* (2008). Chlorella viruses evoke a rapid release of K+ from host cells during the early phase of infection. **Virology** 372(2), 340-348.

Noble, R.T., and Fuhrman, J.A. (1998). Use of SYBR Green I for rapid epifluorescence counts of marine viruses and bacteria. **Aquatic Microbial Ecology** 14(2), 113-118.

Ogata, H., Toyoda, K., Tomaru, Y., Nakayama, N., Shirai, Y., Claverie, J.M., *et al.* (2009). Remarkable sequence similarity between the dinoflagellate-infecting marine girus and the terrestrial pathogen African swine fever virus. **Virology Journal** 6.

Okamoto, K., Miyazaki, N., Reddy, H.K.N., Hantke, M.F., Maia, F.R.N.C., Larsson, D.S.D., *et al.* (2018). Cryo-EM structure of a *Marseilleviridae* virus particle reveals a large internal microassembly. **Virology** 516, 239-245.

Onji, M., Nakano, S., and Suzuki, S. (2003). Virus-like particles suppress growth of the red-tide-forming marine dinoflagellate *Gymnodinium mikimotoi*. **Marine Biotechnology** 5(5), 435-442.

Pagarete, A., Grebert, T., Stepanova, O., Sandaa, R.A., and Bratbak, G. (2015). Tsv-N1: A Novel DNA algal virus that Infects *Tetraselmis striata*. **Viruses-Basel** 7(7), 3937-3953.

Pagarete, A., Lanzen, A., Puntervoll, P., Sandaa, R.A., Larsen, A., Larsen, J.B., *et al.* (2013). Genomic sequence and analysis of EhV-99B1, a new coccolithovirus from the Norwegian fjords. **Intervirology** 56(1), 60-66.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., *et al.* (2015). Open science resources for the discovery and analysis of Tara Oceans data. **Scientific Data** 2, 150023.

Piedade, G.J., Wesdorp, E.M., Montenegro-Borbolla, E., Maat, D.S., and Brussaard, C.P.D. (2018). Influence of irradiance and temperature on the virus MpoV-45T infecting the arctic picophytoplankter *Micromonas polaris*. **Viruses-Basel** 10(12).

Popels, L.C., Cary, S.C., Hutchins, D.A., Forbes, R., Pustizzi, F., Gobler, C.J., *et al.* (2003). The use of quantitative polymerase chain reaction for the detection and enumeration of the harmful alga *Aureococcus anophagefferens* in environmental samples along the United States East Coast. **Limnology and Oceanography-Methods** 1, 92-102.

Popels, L.C., and Hutchins, D.A. (2002). Factors affecting dark survival of the brown tide alga *Aureococcus anophagefferens* (*Pelagophyceae*). **Journal of Phycology** 38(4), 738-744.

Popels, L.C., MacIntyre, H.L., Warner, M.E., Zhang, Y.H., and Hutchins, D.A. (2007). Physiological responses during dark survival and recovery in *Aureococcus anophagefferens* (*Pelagophyceae*). **Journal of Phycology** 43(1), 32-42.

Probyn, T., Pitcher, G., Pienaar, R., and Nuzzi, R. (2001). Brown tides and mariculture in Saldanha Bay, South Africa. **Marine Pollution Bulletin** 42(5), 405-408.

Proctor, L.M., and Fuhrman, J.A. (1990). Viral mortality of marine bacteria and cyanobacteria. **Nature** 343(6253), 60-62.

Proctor, L.M., and Fuhrman, J.A. (1991). Roles of viral-infection in organic particle-flux. **Marine Ecology Progress Series** 69(1-2), 133-142.

Pustizzi, F., MacIntyre, H., Warner, M.E., and Hutchins, D.A. (2004). Interaction of nitrogen source and light intensity on the growth and photosynthesis of the brown tide alga *Aureococcus anophagefferens*. **Harmful Algae** 3(4), 343-360.

Puxty, R.J., Evans, D.J., Millard, A.D., and Scanlan, D.J. (2018). Energy limitation of cyanophage development: implications for marine carbon cycling. **ISME Journal** 12(5), 1273-1286.

Quispe, C.F., Sonderman, O., Seng, A., Rasmussen, B., Weber, G., Mueller, C., *et al.* (2016). Three-year survey of abundance, prevalence and genetic diversity of chlorovirus populations in a small urban lake. **Archives of Virology** 161(7), 1839-1847.

Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., *et al.* (2006). Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. **Journal of Virology** 80(23), 11678-11685.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics** 26(1), 139-140.

Rodrigues, R.A., dos Santos Silva, L.K., Dornas, F.P., de Oliveira, D.B., Magalhaes, T.F., Santos, D.A., *et al.* (2015). Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. **Journal of Virology** 89(23), 11812-11819.

Rohwer, F., and Thurber, R.V. (2009). Viruses manipulate the marine environment. **Nature** 459(7244), 207-212.

Rosenwasser, S., Mausz, M.A., Schatz, D., Sheyn, U., Malitsky, S., Aharoni, A., *et al.* (2014). Rewiring host lipid metabolism by large viruses determines the fate of *Emiliania huxleyi*, a bloom-forming alga in the Ocean. **Plant Cell** 26(6), 2689-2707.

Rosenwasser, S., Sheyn, U., Frada, M.J., Pilzer, D., Rotkopf, R., and Vardi, A. (2019). Unmasking cellular response of a bloom-forming alga to viral infection by resolving expression profiles at a single-cell level. **Plos Pathogens** 15(4).

Rousvoal, S., Bouyer, B., Lopez-Cristoffanini, C., Boyen, C., and Collen, J. (2016). Mutant swarms of a totivirus-like entities are present in the red macroalga *Chondrus crispus* and have been partially transferred to the nuclear genome. **Journal of Phycology** 52(4), 493-504.

Rowe, J.M., Dunlap, J.R., Gobler, C.J., Anderson, O.R., Gastrich, M.D., and Wilhelm, S.W. (2008). Isolation of a non-phage-like lytic virus infecting *Aureococcus anophagefferens*. **Journal of Phycology** 44(1), 71-76.

Sandaa, R.A., Heldal, M., Castberg, T., Thyrhaug, R., and Bratbak, G. (2001). Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (*Prymnesiophyceae*) and *Pyramimonas orientalis* (*Prasinophyceae*). **Virology** 290(2), 272-280.

Schatz, D., Shemi, A., Rosenwasser, S., Sabanay, H., Wolf, S.G., Ben-Dor, S., *et al.* (2014). Hijacking of an autophagy-like process is critical for the life cycle of a DNA virus infecting oceanic algal blooms. **New Phytologist** 204(4), 854-863.

Schroeder, D.C., Oke, J., Hall, M., Malin, G., and Wilson, W.H. (2003). Virus succession observed during an *Emiliania huxleyi* bloom. **Applied and Environmental Microbiology** 69(5), 2484-2490.

Schroeder, D.C., Oke, J., Malin, G., and Wilson, W.H. (2002). Coccolithovirus (Phycodnaviridae): Characterisation of a new large dsDNA algal virus that infects *Emiliania huxleyi*. **Archives of Virology** 147(9), 1685-1698.

Schroeder, D.C., Park, Y., Yoon, H.M., Lee, Y.S., Kang, S.W., Meints, R.H., *et al.* (2009). Genomic analysis of the smallest giant virus--*Feldmannia sp.* virus 158. **Virology** 384(1), 223-232.

Schvarcz, C.R., and Steward, G.F. (2018). A giant virus infecting green algae encodes key fermentation genes. **Virology** 518, 423-433.

Sheik, A.R., Brussaard, C.P., Lavik, G., Lam, P., Musat, N., Krupke, A., *et al.* (2014). Responses of the coastal bacterial community to viral infection of the algae *Phaeocystis globosa*. **ISME Journal** 8(1), 212-225.

Sheyn, U., Rosenwasser, S., Ben-Dor, S., Porat, Z., and Vardi, A. (2016). Modulation of host ROS metabolism is essential for viral infection of a bloom-forming coccolithophore in the ocean. **ISME Journal** 10(7), 1742-1754.

Shirai, Y., Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., and Nagasaki, K. (2008). Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. **Applied and Environmental Microbiology** 74(13), 4022-4027.

Sieburth, J.M., Johnson, P.W., and Hargraves, P.E. (1988). Ultrastructure and ecology of *Aureococcus anophagefferens gen. et. sp. nov* (*Chrysophyceae*) - the dominant picoplankter during a bloom in Narragansett Bay, Rhode Island, Summer 1985. **Journal of Phycology** 24(3), 416-425.

Sieracki, M.E., Gobler, C.J., Cucci, T.L., Thier, E.C., Gilg, I.C., and Keller, M.D. (2004). Pico- and nanoplankton dynamics during bloom initiation of *Aureococcus* in a Long Island, NY bay. **Harmful Algae** 3(4), 459-470.

Simonin, D., Diaz, J.J., Masse, T., and Madjar, J.J. (1997). Persistence of ribosomal protein synthesis after infection of HeLa cells by herpes simplex virus type 1. **Journal of General Virology** 78, 435-443.

Sinkovits, R.S., and Baker, T.S. (2010). A tale of two symmetrons: Rules for construction of icosahedral capsids from trisymmetrons and pentasymmetrons. **Journal of Structural Biology** 170(1), 109-116.

Stough, J.M.A., Yutin, N., Chaban, Y.V., Moniruzzaman, M., Gann, E.R., Pound, H.L., *et al.* (2019). Genome and environmental activity of a Chrysochromulina parva Virus and Its virophages. **Frontiers in Microbiology** 10.

Suttle, C.A. (2005). Viruses in the sea. **Nature** 437(7057), 356-361.

Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. **Nature Reviews Microbiology** 5(10), 801-812.

Suttle, C.A., and Chan, A.M. (1995). Viruses infecting the marine prymnesiophyte *Chrysochromulina spp*. - isolation, preliminary characterization, and natural-abundance. **Marine Ecology Progress Series** 118(1-3), 275-282.

Suttle, C.A., Chan, A.M., and Cottrell, M.T. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. **Applied and Environmental Microbiology** 57, 721-726.

Tai, V., Lawrence, J.E., Lang, A.S., Chan, A.M., Culley, A.I., and Suttle, C.A. (2003). Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (*Raphidophyceae*). **Journal of Phycology** 39(2), 343-352.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., *et al.* (2007). EMAN2: An extensible image processing suite for electron microscopy. **Journal of Structural Biology** 157(1), 38-46.

Tarutani, K., Nagasaki, K., Itakura, S., and Yamaguchi, M. (2001). Isolation of a virus infecting the novel shellfish-killing dinoflagellate *Heterocapsa circularisquama*. **Aquatic Microbial Ecology** 23(2), 103-111.

Tarutani, K., Nagasaki, K., and Yamaguchi, M. (2000). Viral impacts on total abundance and clonal composition of the harmful bloom-forming phytoplankton *Heterosigma akashiwo*. **Applied and Environmental Microbiology** 66(11), 4916.

Tarutani, K., Nagasaki, K., and Yamaguchi, M. (2006). Virus adsorption process determines virus susceptibility in *Heterosigma akashiwo* (*Raphidophyceae*). **Aquatic Microbial Ecology** 42(3), 209-213.

Team, R.C. (2014). "A language and environment for statistical computing". (Vienna, Austria).

Thamatrakoln, K., Talmy, D., Haramaty, L., Maniscalco, C., Latham, J.R., Knowles, B., *et al.* (2019). Light regulation of coccolithophore host-virus interactions. **New Phytologist** 221(3), 1289-1302.

The, M., MacCoss, M.J., Noble, W.S., and Kall, L. (2016). Fast and accurate protein false discovery rates on large-scale proteomics data sets with Percolator 3.0. **The Journal of the American Society for Mass Spectrometry** 27(11), 1719-1727.

Thingstad, T.F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. **Limnology and Oceanography** 45(6), 1320-1328.

Thomas, R., Grimsley, N., Escande, M.L., Subirana, L., Derelle, E., and Moreau, H. (2011). Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. En**vironmental Microbiology** 13(6), 1412-1420.

Thyrhaug, R., Larsen, A., Brussaard, C.P.D., and Bratbak, G. (2002). Cell cycle dependent virus production in marine phytoplankton. **Journal of Phycology** 38(2), 338-343.

Tobin, E.D., Grunbaum, D., Patterson, J., and Cattolico, R.A. (2013). Behavioral and physiological changes during benthic-pelagic transition in the harmful alga, *Heterosigma akashiwo*: potential for rapid bloom formation. **PLoS One** 8(10), e76663.

Tomaru, Y., Hata, N., Masuda, T., Tsuji, M., Igata, K., Masuda, Y., *et al.* (2007). Ecological dynamics of the bivalve-killing dinoflagellate *Heterocapsa circularisquama* and its infectious viruses in different locations of western Japan. **Environmental Microbiology** 9(6), 1376-1383.

Tomaru, Y., Katanozaka, N., Nishida, K., Shirai, Y., Tarutani, K., Yamaguchi, M., *et al.* (2004). Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. **Aquatic Microbial Ecology** 34(3), 207-218.

Tomaru, Y., Mizumoto, H., and Nagasaki, K. (2009). Virus resistance in the toxic bloom-forming dinoflagellate *Heterocapsa circularisquama* to single-stranded RNA virus infection. **Environmental Microbiology** 11(11), 2915-2923.

Tomaru, Y., Shirai, Y., Suzuki, H., Nagumo, T., and Nagasaki, K. (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., Shirai, Y., Toyoda, K., and Nagasaki, K. (2011). Isolation and characterisation of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus*. **Aquatic Microbial Ecology** 64(2), 175-184.

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

Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., and Nagasaki, K. (2009). Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming Diatom *Chaetoceros socialis*. **Applied and Environmental Microbiology** 75(8), 2375-2381.

Tomaru, Y., Tarutani, K., Yamaguchi, M., and Nagasaki, K. (2004). Quantitative and qualitative impacts of viral infection on a *Heterosigma akashiwo* (*Raphidophyceae*) bloom in Hiroshima Bay, Japan. **Aquatic Microbial Ecology** 34(3), 227-238.

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

Tomaru, Y., Toyoda, K., Kimura, K., Takao, Y., Sakurada, K., Nakayama, N., *et al.* (2013). Isolation and characterization of a single-stranded RNA virus that infects the marine planktonic diatom *Chaetoceros sp* (SS08-C03). **Phycological Research** 61(1), 27-36.

Tomaru, Y., Toyoda, K., Suzuki, H., Nagumo, T., Kimura, K., and Takao, Y. (2013). New single-stranded DNA virus with a unique genomic structure that infects marine diatom *Chaetoceros setoensis*. **Scientific Reports** 3. doi: 10.1038/srep03337.

Toyoda, K., Kimura, K., Hata, N., Nakayama, N., Nagasaki, K., and Tomaru, Y. (2012). Isolation and characterization of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros sp* (strain TG07-C28). **Plankton & Benthos Research** 7(1), 20-28.

Van Etten, J.L., Agarkova, I., Dunigan, D.D., Tonetti, M., De Castro, C., and Duncan, G.A. (2017). Chloroviruses Have a Sweet Tooth. **Viruses-Basel** 9(4).

Van Etten, J.L., Burbank, D.E., Xia, Y., and Meints, R.H. (1983). Growth-cycle of a virus, PBCV-1, that Infects *Chlorella*-like algae. **Virology** 126(1), 117-125.

Van Etten, J.L., Meints, R.H., Kuczmarski, D., Burbank, D.E., and Lee, K. (1982). Viruses of symbiotic *Chlorella*-like algae isolated from *Paramecium bursaria* and *Hydra viridis*. **Proceedings of the National Academy of Sciences of the United States of America** 79(12), 3867-3871.

Vardi, A., Haramaty, L., Van Mooy, B.A.S., Fredricks, H.F., Kimmance, S.A., Larsen, A., *et al.* (2012). Host-virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. **Proceedings of the National Academy of Sciences of the United States of America** 109(47), 19327-19332.

Villalobos Solis, M.I., Giannone, R.J., Hettich, R.L., and Abraham, P.E. (2019). Exploiting the dynamic relationship between peptide separation quality and peptide coisolation in a multiple-peptide matches-per-spectrum approach offers a strategy to optimize bottom-up proteomics throughput and depth. **Analytical Chemistry** 91(11), 7273-7279.

Wagstaff, B.A., Vladu, I.C., Barclay, J.E., Schroeder, D.C., Malin, G., and Field, R.A. (2017). Isolation and characterization of a double stranded DNA Megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. **Viruses-Basel** 9(3).

Waldbauer, J.R., Coleman, M.L., Rizzo, A.I., Campbell, K.L., Lotus, J., and Zhang, L. (2019). Nitrogen sourcing during viral infection of marine cyanobacteria. **Proceedings of the National Academy of Sciences of the United States of America** 116(31), 15590-15595.

Waters, R.E., and Chan, A.T. (1982). *Micromonas pusilla* Virus - the virus growth-cycle and associated physiological events within the host-cells - host range mutation. **Journal of General Virology** 63(Nov), 199-206.

Wikner, J., Vallino, J.J., Steward, G.F., Smith, D.C., and Azam, F. (1993). Nucleic-acids from the host bacterium as a major source of nucleotides for 3 marine bacteriophages. **FEMS Microbiology Ecology** 12(4), 237-248.

Wilhelm, S.W., Bird, J.T., Bonifer, K.S., Calfee, B.C., Chen, T., Coy, S.R., *et al.* (2017). A Student's Guide to Giant Viruses Infecting Small Eukaryotes: From Acanthamoeba to Zooxanthellae. **Viruses-Basel** 9(3).

Wilhelm, S.W., Coy, S.R., Gann, E.R., Moniruzzaman, M., and Stough, J.M.A. (2016). Standing on the shoulders of giant viruses: 5 lessons learned about large viruses infecting small eukaryotes and the opportunities they create. **PLoS Pathogens** 12, e1005752.

Wilhelm, S.W., and Suttle, C.A. (1999). Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. **BioScience** 49(10), 781-788.

Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A., and Jeffrey, W.H. (1998). The role of sunlight in the removal and repair of viruses in the sea. **Limnology and Oceanography** 43(4), 586-592.

Wolf, S., Maier, I., Katsaros, C., and Muller, D.G. (1998). Virus assembly in *Hincksia hincksiae* (*Ectocarpales*, *Phaeophyceae*) an electron and fluorescence microscopic study. **Protoplasma** 203(3-4), 153-167.

Worden, A.Z., Janouskovec, J., McRose, D., Engman, A., Welsh, R.M., Malfatti, S., *et al.* (2012). Global distribution of a wild alga revealed by targeted metagenomics. **Current Biology** 22(17), R675-677.

Worden, A.Z., Nolan, J.K., and Palenik, B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. **Limnology and Oceanography** 49(1), 168-179.

Wrigley, N.G. (1969). An Electron Microscope Study of Structure of *Sericesthis* Iridescent Virus. **Journal of General Virology** 5, 123-&.

Wu, Y., Kim, W., Kim, S.-W., Eom, C.Y., Yang, H., and Shin, H.J. (2012). Expression and self-assembly of *Heterocapsa circularisquama* RNA virus-like particles synthesized in *Pichia pastoris*. **Chinese Science Bulletin** 57(25), 3288-3293.

Wu, Z.J., Jenkins, B.D., Rynearson, T.A., Dyhrman, S.T., Saito, M.A., Mercier, M., *et al.* (2010). Empirical bayes analysis of sequencing-based transcriptional profiling without replicates. **BMC Bioinformatics** 11.

Wurch, L.L., Alexander, H., Frischkorn, K.R., Haley, S.T., Gobler, C.J., and Dyhrman, S.T. (2019). Transcriptional shifts highlight the role of nutrients in harmful brown tide dynamics. **Frontiers in Microbiology** 10.

Wurch, L.L., Haley, S.T., Orchard, E.D., Gobler, C.J., and Dyhrman, S.T. (2011). Nutrient-regulated transcriptional responses in the brown tide-forming alga *Aureococcus anophagefferens*. **Environmental Microbiology** 13(2), 468-481.

Xian, Y.J., Karki, C.B., Silva, S.M., Li, L., and Xiao, C. (2019). The roles of electrostatic interactions in capsid assembly mechanisms of giant viruses. **International Journal of Molecular Sciences** 20(8).

Xiao, C., Fischer, M.G., Bolotaulo, D.M., Ulloa-Rondeau, N., Avila, G.A., and Suttle, C.A. (2017). Cryo-EM reconstruction of the *Cafeteria roenbergensis* virus capsid suggests novel assembly pathway for giant viruses. **Scientific Reports** 7.

Yaakov, L.B., Mutsafi, Y., Porat, Z., Dadosh, T., and Minsky, A. (2019). Kinetics of Mimivirus infection stages quantified using image flow cytometry. **Cytometry Part A** 95(5), 534-548.

Yamada, Y., Tomaru, Y., Fukuda, H., and Nagata, T. (2018). Aggregate formation during the viral lysis of a marine diatom. **Frontiers in Marine Science** 5.

Yan, X.D., Chipman, P.R., Castberg, T., Bratbak, G., and Baker, T.S. (2005). The marine algal virus PpV01 has an icosahedral capsid with T=219 quasisymmetry. **Journal of Virology** 79(14), 9236-9243.

Yan, X.D., Yu, Z.Y., Zhang, P., Battisti, A.J., Holdaway, H.A., Chipman, P.R., *et al.* (2009). The capsid proteins of a large, icosahedral dsDNA virus. **Journal of Molecular Biology** 385(4), 1287-1299.

Yau, S., Caravello, G., Fonvieille, N., Desgranges, E., Moreau, H., and Grimsley, N. (2018). Rapidity of genomic adaptations to prasinovirus infection in a marine microalga. **Viruses-Basel** 10(8).

Yau, S., Hemon, C., Derelle, E., Moreau, H., Piganeau, G., and Grimsley, N. (2016). A viral immunity chromosome in the marine picoeukaryote, *Ostreococcus tauri*. **Plos Pathogens** 12(10).

Yutin, N., Wolf, Y.I., Raoult, D., and Koonin, E.V. (2009). Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution. **Virology Journal** 6.

Zhang, Q.C., Qiu, L.M., Yu, R.C., Kong, F.Z., Wang, Y.F., Yan, T., *et al.* (2012). Emergence of brown tides caused by *Aureococcus anophagefferens* Hargraves et Sieburth in China. **Harmful Algae** 19, 117-124.

Zhang, X.Z., Xiang, Y., Dunigan, D.D., Klose, T., Chipman, P.R., Van Etten, J.L., *et al.* (2011). Three-dimensional structure and function of the Paramecium bursaria chlorella virus capsid. **Proceedings of the National Academy of Sciences of the United States of America** 108(36), 14837-14842.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. **Nature Methods** 14(4), 331-332.

Zimmerman, A.E., Bachy, C., Ma, X.F., Roux, S., Jang, H.B., Sullivan, M.B., *et al.* (2019). Closely related viruses of the marine picoeukaryotic alga *Ostreococcus lucimarinus* exhibit different ecological strategies. **Environmental Microbiology** 21(6), 2148-2170.

Vita

Eric R. Gann was born in Syracuse, New York, U.S.A, in 1993. He was raised in Oneida, New York. After graduating Oneida High School, he attended the University of Massachusetts Amherst where he received a B.S degree in Microbiology in 2015. He moved from Amherst to Knoxville to pursue this degree.