VIRUS-HOST INTERACTIONS WITHIN GALVESTON BAY, TEXAS

An Undergraduate Research Scholars Thesis

by

CAMERON D JACKSON

Submitted to the Undergraduate Research Scholars program at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by Research Advisor: Dr. Jessica M Labonté

May 2018

Major: Marine Biology

TABLE OF CONTENTS

ABSTRACT

Virus-Host Interactions Within Galveston Bay, Texas

Cameron D Jackson
Department of Marine Biology
Texas A&M University

Research Advisor: Dr. Jessica M Labonté
Department of Marine Biology
Texas A&M University

Viruses are among the most abundant biological entities in the world, numbering approximately $4x10^{30}$ in the ocean alone. Despite the inherent importance of viruses, aspects of host preferences and specificity within the environment remain understudied. This study aimed to connect environmental phages to their environmental bacterial hosts. To this end, two different culture-independent methodologies allowed for the association of viruses to their hosts. First, a new method was tested that allowed for the visualization of virally infected cells through fluorescence microscopy. Secondly, co-occurrence networks were used to analyze PCR amplicons of bacteria and viruses from monthly samples. Together, these approaches allowed for coverage of a much wider range of marine bacteria and their viruses than previous studies. The methods can also lead to flow cytometry sorting and single cell genomics, allowing for a deeper understanding of the infection types present.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jessica Labonté for her guidance and direction over the course of the project. Without her, this project would have remained nothing more than an idea. Thank you to the other members of my lab—Catherine, Abigail, and Jordan—for helping with the daily work.

Thanks also to all my friends from Texas A&M University at Galveston that helped make the experience a great one. Thank you to the many people that acted as sounding boards throughout the scientific process, including the many graduate students in surrounding labs.

Finally, thank you to my father and grandparents, without whose love and support I never could have arrived here today.

NOMENCLATURE

DNA	deoxyribonucleic acid
GF-F	glass fiber grade F
bp	base pairs
kb	kilo base pairs
PCR	polymerase chain reaction
PVDF	polyvinylidene Fluoride
SCG	single-cell genomics
TFF	tangential flow filtration

CHAPTER I

INTRODUCTION

Importance of Viruses

Viruses are the most abundant biological entities within the world (Breitbart, 2012). On average, viral abundance in the ocean is on the scale of 10^7 particles per mL, while bacterial abundance is often an order of magnitude lower at 10^6 (Parikka et al., 2016). Most of these viruses are phages, those that infect prokaryotic organisms. As reservoirs of high genetic diversity and key predators of the ocean's microbial populations, viruses play an important role in the environment (Sullivan et al., 2006). Some phages, called prophages, have the ability to integrate into their hosts, playing a role in the evolution of the host species. Up to one half of all microbes isolated from the ocean contain at least one prophage (Paul, 2008). Additionally, viruses act to release carbon and trace elements locked within the host cells. Despite these important roles, virus–host interactions are poorly understood as most knowledge comes from phage cultivation within the lab. This approach is limited to the 0.1–1% of bacterial cells that are responsive to laboratory conditions (Rappé and Giovannoni, 2003). Specific life strategies, such as lytic infection, are also easier to investigate in the lab through complete lysis assays (Breitbart, 2012). The true answers to "Who infects whom and how within the ocean?" is left mostly answered by the limitations of the studies.

Great Diversity of Viruses

There is no conserved gene across the various families of viruses. The lack of a unifying characteristic such as bacterial 16S ribosomal RNA (rRNA) leads to many difficulties in the study of viral diversity and phylogenetic analyses (Rohwer and Edwards, 2002). Due to the small

genome size (26->200 kb) and physical size of viruses, quantifying the diversity is difficult as well, since most quantifying strategies involve the staining of the genetic material (Steward et al. 2000, Breitbart, 2012). Different methods have uncovered a vast amount of novel diversity (Breitbart, 2012). These methods include polymerase chain reaction (PCR) and single cell genomics (SCG). Degenerate primers used in the PCR amplification target key genes within specific families of viruses, such as *Myoviridae* (Filée et al., 2005), *Podoviridae* (Breitbart et al., 2004), and Cyanophages (Chen et al., 2009). These primers amplify genes such as the major capsid protein (Breitbart and Rohwer, 2004). PCR amplicons have been used to determine viral diversity, genetic relatedness (Rohwer and Edwards, 2002), and biogeography of certain viral families (Breitbart and Rohwer, 2004). SCG is a powerful tool to identify viruses without the need for cultivation. Indeed, when sequencing a single cell, all DNA molecules associated with the cells are sequenced, whether they are from symbionts, chloroplasts and mitochondria in eukaryotes, or viral infections (Stepanauskas, 2015). Viral infections can also be differentiated through SCG results including anomalies in sequence-coverage and genome recovery (Labonté et al. 2015). SCG has revealed novel virus strains (Labonté et al., 2015) and linked a virus to a wide spectrum of hosts (Roux et al., 2014).

In addition, viruses act as genetic reservoirs and lateral transfer vectors for rate-limiting steps in key biological pathways (Sullivan et al., 2006). For example, the cyanobacterial *psbA* gene that encodes for a protein within photosystem II has been found in many cyanophages. The phage *psbA* gene integrates into the bacterial genome following infection, but can also act in place of the host's gene (Bragg and Chisholm, 2008). Virus–host interactions are important in defining the host's potential for evolution.

Biogeochemical Impact

Through lysis, viruses also contribute to many biogeochemical cycles throughout the ocean. Virus–induced microbial mortality is attributed with releasing approximately 10^9 tons of biologic carbon each day (Suttle, 2007), converting the carbon into dissolved organic matter (DOM). Heterotrophic bacteria then integrate the DOM into the classical ocean food web. This process, known as the viral shunt, is an integral part of the carbon cycle, but it also releases trace elements. Jover et al. (2014) identified the importance that viruses have on the levels of oceanic phosphorus reservoirs. The release of nitrogen from photosynthetic bacteria can also play a role in the environment (Suttle, 2005).

Despite the variety of ecological roles that viruses play, virus–host interactions and specificity are key aspects that remains poorly characterized. Deng et al. (2012) developed a staining process that allows infected bacteria to be seen under fluorescence microscopy. The process, however, was restricted to specific, laboratory cultivated bacteria. So far, the linkage of viruses to their environmental hosts has eluded study. Previous attempts have focused on marine virus isolates infecting cultured host bacterial strains (Sullivan et al., 2003).

Objectives and Hypotheses

This study aimed to link viruses to their environmental hosts through the development of two distinct methods:

Objective 1: A method for the direct visualization of infected bacteria was tested using environmental extracts of bacteria and viruses. This is based on the hypothesis that the first step of viral infection is adsorption and that adsorption is a specific mechanism. The strategy is to infect microbial cells with fluorescently tagged viruses and incubate long enough to allow for

adsorption. Infected cells, which will be tagged by the virus, should be visible under fluorescence microscopy.

Objective 2: Through the comparison of viral gene markers and host 16S rRNA, we can infer viral hosts using co-occurrence network. This is based on the hypothesis that viral infections (virus release, thus virus increase) is following cell lysis (cell death, therefore host decrease). The co-occurrence networks can be used to find trends between host and virus concentration over time.

CHAPTER II

METHODS

Sampling and Concentration

Samples were collected from the boat basin at Texas A&M University at Galveston once per month. In addition, September and October samples were taken weekly throughout the Galveston ship channel following Hurricane Harvey. These samples were used in the place of the monthly boat basin samples for the respective months. Samples are named first by the sampling trip (S1, S3, S4, or S5), then by station number, as listed in Figure 1. At least 20L of seawater was collected and filtered through a nitex membrane (60 µm) into a carboy. The nitex removes all large debris and speed future filtering. The temperature and salinity of the water was recorded using a salinity probe (Seven2Go, Mettler Toledo). A 5mL aliquot was subsampled in order to calculate extraction efficiency.

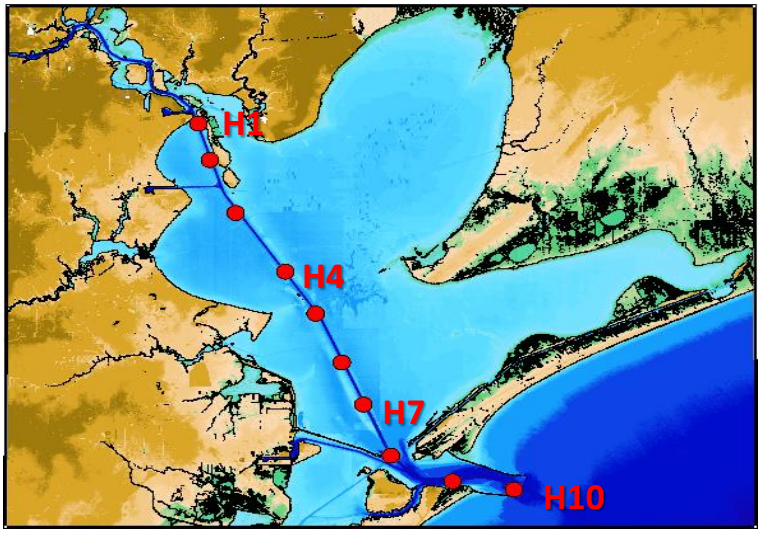

Figure 1: Map of transect used to collect samples following Hurricane Harvey. Only the listed sites were sampled for this project.

A large pressure filtration apparatus was used to efficiently filter the 20L samples. Pre-filtration occurred through a GF-F filter (~0.8 µm pore size; Whatman) to remove eukaryotes. A

second filtration occurred through a Sterivex filter (Millipore) for 1 L and through a PVDF (0.2 μm pore size; Millipore) for the remainder to remove bacteria. The Sterivex filter allows for bacterial resuspension in future work. Finally, the resulting bacteria-free sample that contains the viruses was concentrated using tangential flow filtration (TFF) with pore size of 30 kDa. The permeate, or virus-free seawater, was kept at 4˚C for further use.

Virus Concentrate Efficiency

In order to determine the effectiveness of the filtration, virus counts were compared from before and after the process. For each, 2mL of sample was fixed with formalin to a final volume of 2%. This sample was filtered through a 0.02μm Anodisc filter (22mm diameter). The 0.02μm filter was stained using SYBR Green for 5 minutes in a dark, dry place. The filter was placed onto a slide and viewed under fluorescence microscopy (Lieca; LAS X software). Ten locations on the slide were selected at random to be counted. Viral particles were counted using ImageJ. Each image was converted into binary using the Threshold command. The average number of viruses per location was determined. This can be used to calculate the total virus abundance within the sample through the following equation: Total=RSF*Average/Volume filtered (L), where RSF is the reticle scaling factor of the microscope. RSF must be calculated for the microscope being used by dividing the filterable area of the Anodisc by the exact area of the viewing grid (Patel et al., 2007).

Virus Staining and Visualization

In order to visualize virally infected bacteria, the bacteria and viruses must be stained using dyes that emit at different wavelengths. The protocol used was adapted from Deng et al. (2012).

Bovine Serum Albumin (BSA) was added to a 30 kDa centrifuge TFF cartridge (Millipore) in order to increase the virus recovery efficiency. A volume of 0.5 mL 1% BSA was left in the cartridge for 1 hour at room temperature, then poured out. One mL of the concentrated virus solution was mixed with 100 μL of 10x SYBR Green solution and incubated at 80°C for 10 minutes.

The stained viruses were washed in Tris-EDTA (TE, pH 8) buffer to remove excess dye. A volume of 14 mL of TE and the 1.1 mL of stained viruses were mixed in the prepared centrifuge cartridge. The solution was centrifuged at 4,000 rpm for 15 minutes. The flow through was discarded, then the washing was repeated four more times. The 200 μL of washed virus could then be used to infect environmental bacteria.

The bacteria isolated within the Sterivex filter were resuspended in 500 mL of virus-free seawater (permeate from TFF virus concentration). A volume of 1 mL the resuspended bacteria and 50 μL stained viruses were mixed and incubated at the temperature of collection in order to allow viral adsorption. After 30 minutes, each sample was dyed with 1μL 5mg/mL 4',6-diamidino-2-phenylindole (DAPI) stain for an additional 20 minutes. The samples were filtered through Anodisc filters (0.2 μm pore size; Whatman) and fixed to slides. A fluorescence microscope (Leica; LAS X software) was used to identify the infected bacteria. Infected bacteria were counted using ImageJ parameters.

Nucleic Acid Extractions and PCR

Bacterial and viral DNA needed to be extracted and purified for molecular biology work. Bacterial DNA was extracted from the 0.22 μm filters representing about 4L of filtered seawater. DNEasy Power Soil kits (Qiagen) were used for the regular monthly samples, whereas phenol

chloroform extractions were needed for the post-hurricane samples. DNEasy Power Soil was carried out as per the included instructions.

Once extracted, the DNA from each sample was subsampled to conduct PCR's for identification of common viral families and 16S rRNA bacterial signatures. 1 µL of sample was mixed with a PCR Master Mix containing 5 µL PCR buffer, 1.5 µL MgCl, 1 µL 10 M dNTPs, 2 µL appropriate primer (100 µM), 0.5 µL polymerase, and 39 µL molecular grade (DNAse and RNAse free) H_2O. The primers used and the families they target are summarized in Table 1. The PCR protocol for each primer set is shown in the appendix. All PCR was done in triplicate, then pooled to maximize yield.

Table 1: Primer sets used to target viral families of interest and their bacterial hosts.

Virus Family	Primer	Sequence	Reference
Cyanophage	CP-DNAP	F 5′-CCA AAY CTY GCM CAR GT-3′ Ra 5′-CTC GTC RTG SAC RAA SGC-3′ Rb 5′- CTC GTC RTG DAT RAA SGC-3′	Chen et al. 2009
T7 podoviridae	T7-DPOL	F 5'-ARG ARM RIA AYG GIT-3' R 5'-GTR TGD ATR TCI CC-3'	Breitbart et al. 2004
Myoviridae	MZIA	5'-GAT ATT TGI GGI GTT CAG CCI ATG A-3' R 5'-CGC GGT TGA TTT CCA GCA TGA TTT C-3'	Filée et al. 2005
Bacteria	16S	F 5′-GTG YCA GCM GCC GCG GTA A-3' 5′-CCG YCA ATT YMT TTR AGT TT-3'	Parada et al. 2016

CHAPTER III

RESULTS

Virus Concentration

Viral particles were stained with SYBR Green and virus-like particles (VLPs) were counted using fluorescence microscopy (Figure 2). The virus concentrations ranged from $4.8x10^5$–$9.9x10^6$ VLPs per mL of sample, with the highest concentrations observed as Galveston Bay's assemblages returned to normal (Sample 5 taken 09/28, Station H4). The efficiency of the virus concentration protocol was determined for two samples (July and July 2 VC) and the viral recovery was on average 84.35%.

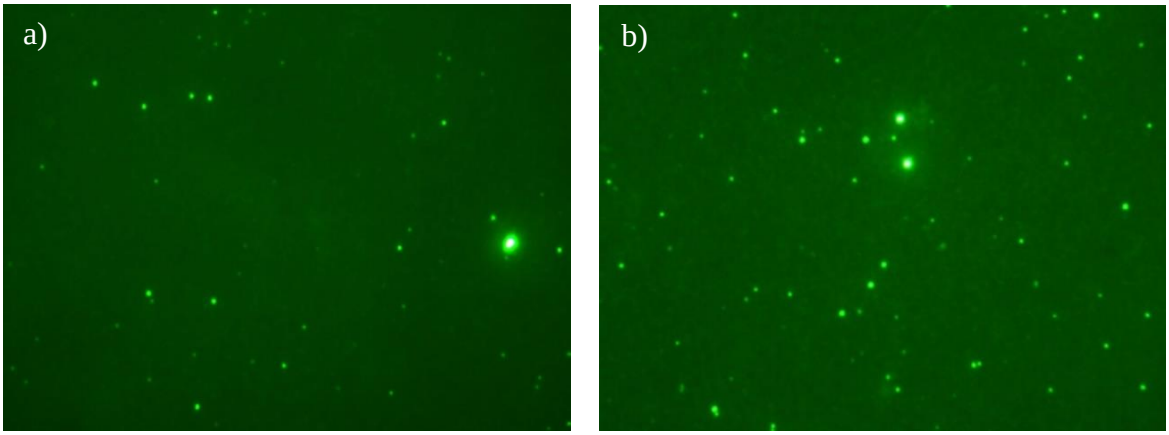

Figure 2: Side-by-side comparison of (a) natural concentrations of viruses and (b) concentrated viruses. Viral particle concentration increased approximately 136 times for this sample (July VC)

Concentration factors were calculated by dividing the starting volume by the final volume of the concentrate (Table 2). In order to maintain relatively consistent concentrations, final volumes were kept to approximately 200mL. The concentration factors vary between 92X and 350X, with an average of 159X.

Table 2: Final concentration factors for each sample tested. Calculations based on volume filtered and final volume of virus concentrate.

Sample	Concentration	Sample	Concentration	Sample	Concentration
July	136.72	S1S1	175.44	S4S1	133.33
July 2	350.00	S1S4	185.19	S4S4	117.65
Oct	112.36	S1S7	289.86	S4S7	106.38
Nov	108.11	S1S10	312.50	S4S10	116.28
Dec	92.59	S3S1	307.69	S5S1	102.56
Jan	163.93	S3S4	134.23	S5S4	102.04
		S3S7	52.36	S5S7	102.56
		S3S10	116.96	S5S10	97.56

Fluorescence Microscopy

We developed a protocol to visualize infected cells, without the need for cultivation, based on the principle that cells will become visible under a fluorescence microscope if infected with a fluorescently labeled virus. We labeled an environmental viral population with SYBR Green (497/520), then incubated a bacterial population from the same sample with the labeled viruses to allow adsorption of the viruses. We counterstained all the cells with DAPI, which has a different excitation than SYBR Green (358/461). Infected cells could be clearly seen through the fluorescent microscopy. Overlaying images taken at each wavelength allowed for distinction between non-infected (green) and infected (blue-green) cells, as seen in Figure 3 below. The difference will allow for single-cell sorting in the future using flow cytometry.

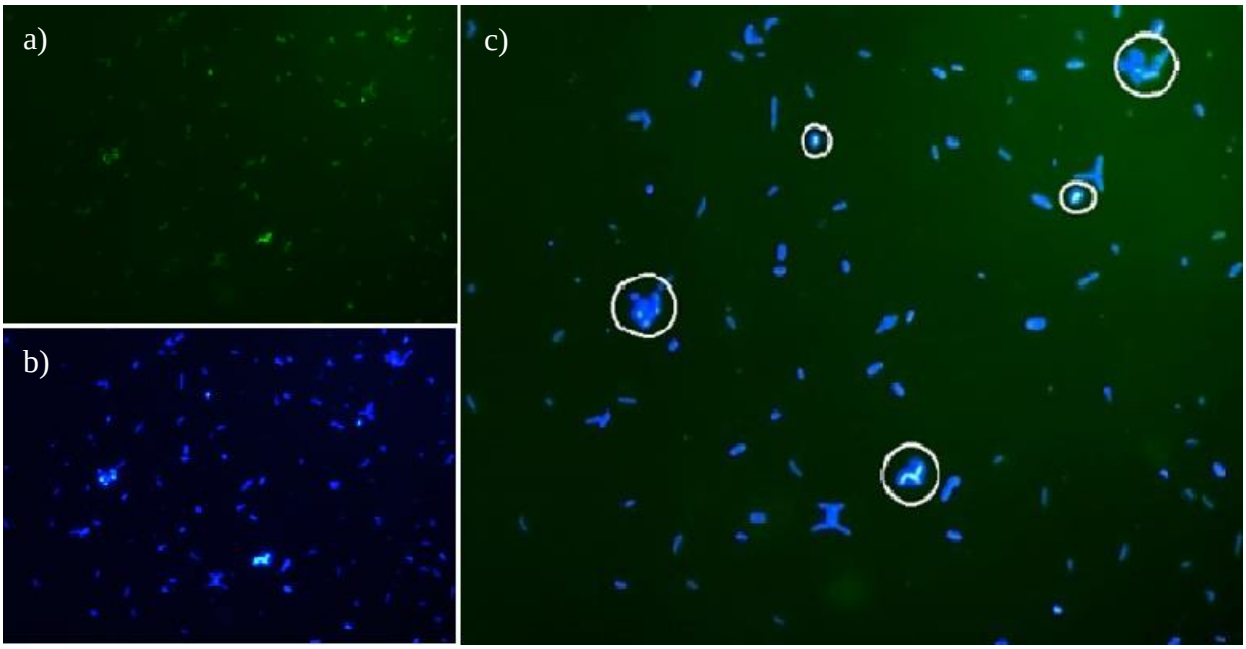

Figure 3: Micrographs of infected cells. (a) SYBR Green stained viruses infecting bacterial cells. (b) DAPI stained bacterial cells (c) Virally-infected bacterial cells. The faint green visible within the blue bacteria represents the stained virus DNA that has infected or adsorbed to the bacteria

PCR Amplification of Viruses and Their Hosts

Amplification of 16S rRNA as well the viral genes went successfully. Strong bands were seen in many of the samples, as seen in Figure 4. The products for the cyanophages, *Podoviridae*, *Myoviridae*, and 16S rRNA yielded amplicon sizes of 600 bp, 500 bp, 380-600 bp, and 300bp respectively. In each trial, the samples from immediately after Hurricane Harvey seem to be returning weak signals, most likely due to the water samples being diluted by the unusually large amount of rain and storm water runoff.

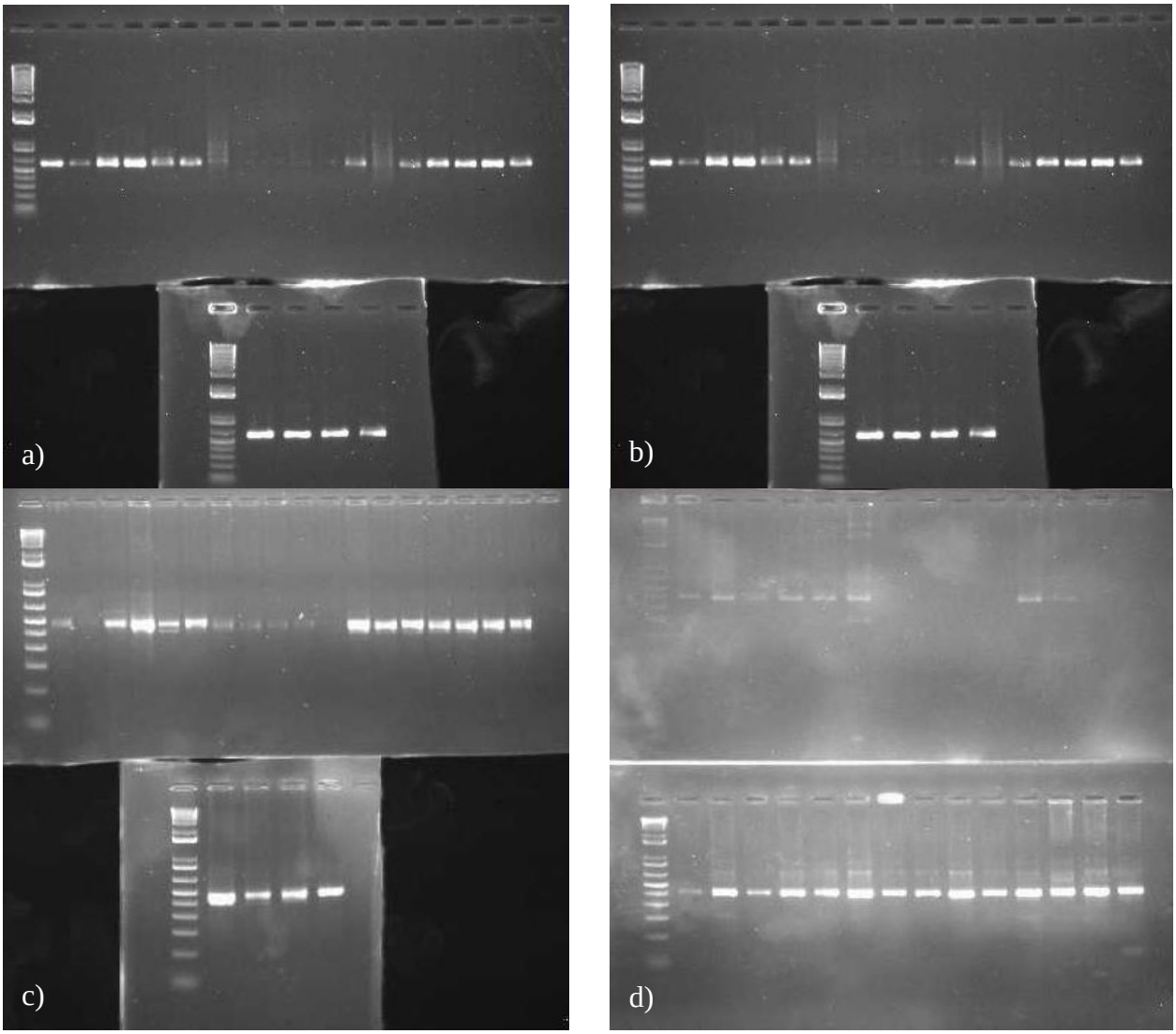

Figure 4: Example PCR Amplicon results following gel electrophoresis. Each gel is a different primer: (a) cyanophages, (b) T7-*Podoviridae*, (c) *Myoviridae*, and (d) bacterial 16S rRNA.

The triplicate PCR products for each viral vamily were pooled to increase the concentration of DNA and reduce PCR amplification bias for sequencing. Unfortunately, contamination fouled the final pooled *Myoviridiae* product. The PCR will be repeated for this primer set. Sequencing of the final PCR amplicons was unable to be carried out during the time frame of this project, but all products will be saved for future molecular work.

CHAPTER IV

DISCUSSION

In this project, we attempted to validate two culture-independent methods of virus–host interactions. Both methods have promising potential following this research. Viral staining and adsorption can lead to isolation of infected cells and single–cell genomics (SCG). PCR amplification and co-occurrence networks can help statistically link viruses to their hosts based on the occurrence of specific sequences. The abundance and dynamics of each viral sequences may lead to the determination of r- and k-type strategy and host specificity among the differing viral populations.

We developed a methodology to visualize virally infected viruses from environmental samples. Previous studies have been limited in their scope. Most have been able to identify viral infections in lab cultures of bacteria (Deng et al., 2007, Breitbart 2012). Genomics approaches applied to microbial samples have found viruses only by chance (Labonté et al., 2015). Microfluidic digital PCR and phageFISH (fluorescence *in situ* hybridization) have also been shown to identify viruses within cultivated samples, but suffer from the lack of knowledge about the diversity of viral genomes within the environment.

The methods presented here can also lead to more information regarding specific virus–host interactions. Bacteria tagged with fluorescent viruses can be separated through flow cytometry according to the emission wavelengths of the SYBR Green stain. Flow cytometry will also allow for the calculation of infection rates, which were too low for microscopy counting to reveal. Cells sorted this way can be analyzed using SCG, which has the power to identify the virus, host, and type of infection (Labonté et al., 2015). Co-occurrence networks can analyze the

metagenomes recovered from the PCR amplicons to determine the abundance and dynamics of virus–host systems over time to establish r- and k-type strategies.

There is no gene that is conserved among all viruses, which complicates the study of virus–host interactions. Here, we are using available degenerate primers for a few viral groups: *Myoviridae* phages (Filée *et al.*, 2005), *Myoviridae* cyanophages (Fuller *et al.*, 1998), and *Podoviridae* (Labonté *et al.*, 2009). *Myoviridae* phages are the most abundant viruses in marine systems (Breitbart *et al.*, 2002; Bench *et al.*, 2007; Williamson *et al.*, 2008b) and usually infect a broader host range (Sullivan *et al.*, 2003). This abundance is due to a more r-type selection, which relies on high virulence and burst counts following infection (Breitbart, 2012). *Podoviridae* are opportunistic viruses that have a narrower host range. They are abundant, virulent, and replicate rapidly (Suttle, 2007), signs of the k-type selected viruses. In an analysis of surface ocean single amplified genomes looking for viral infections, *Podoviridae* and *Myoviridae* represented the majority of the identified viruses (16/20 viruses) (Labonté *et al.*, 2015). We are aware that the primers may not amplify all the viruses in our sample, but they represent the best currently available proxy for identifying viral infections and confirming the methodology.

Among these two families, host specificity is also highly variant. *Myoviridae* are known to infect a wide range of hosts, often through lytic cycles (Suttle, 2005). This wide range is loosely tied to the morphology of the viruses, owing to the contractile tail they contain. Often, *Myoviridae* target opportunistic hosts that grow in boom-bust cycles during favorable conditions (Suttle, 2007). This leads to the widely accepted theory of virus-predation as a "Kill the Winner" strategy in the most abundant families (Breitbart, 2012). The abundant viral particles that are found represent the progeny generation following the bust of the host organism. Through single-

cell genomics, the fluorescent tagging of viruses presented here could lead to a deeper understanding of the lysogenic tendencies of the less abundant k-selected viruses (Labonté et al., 2015).

PCR amplicon analysis is a powerful tool to link viruses and their hosts over a long span of time (Needham et al., 2017). However, this method is limited to only the families of viruses investigated so far. In order to get a more complete view of the ocean's virome, other families need to be researched in similar manners. *Siphoviridae* are similar to both *Myoviridae* and *Podoviridae* (Rowher and Edwards, 2002). The PCR process only amplifies double stranded (dsDNA), so single-stranded DNA (ssDNA) and RNA viruses are also ignored through the procedures (Breitbart, 2012).

CHAPTER V

CONCLUSION

Here, fluorescent tagging of virally–infected cells has been shown to work on environmental samples through filtration and reintroduction of the viruses. The fluorescence differences will allow for sorting via flow cytometry (Deng et al., 2012). The infectious route can be determined through single cell genomics (SCG) as demonstrated by Labonté et al. (2015). This can lead to finally linking viruses to their hosts in the environment.

Sequencing of the PCR amplicons recovered from the sampling can yield even more in terms of the virus–host linkage. The high yield recovered on the PCR products bodes well for future sequencing. The co-occurrence networks have been shown to link bacteria and viruses over time (Needham et al., 2017). Finding specificity of viruses over an extended time period will help to uncover a previously under-researched field (Breitbart et al., 2011). Together, these methods allow for a better understanding of the currently mysterious virosphere.

REFERENCES

Bench, S., Hanson, T., Williamson, K., Ghosh, D., Radosevich, M., Wang, K., & Wommack, K. E. (2008). Metagenomic Characterization of Chesapeake Bay Virioplankton. *Applied and Environmental Microbiology*, *73*, 7629–7641. https://doi.org/10.1128/AEM.00938-07

Bragg, J. G., & Chisholm, S. W. (2008). Modeling the Fitness Consequences of a Cyanophage-Encoded Photosynthesis Gene. *PLOS ONE*, *3*(10), e3550. https://doi.org/10.1371/journal.pone.0003550

Breitbart, M. (2012). Marine viruses: truth or dare. *Annual Review of Marine Science*, *4*, 425–448. https://doi.org/10.1146/annurev-marine-120709-142805

Breitbart, M., Miyake, J. H., & Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiology Letters*, *236*(2), 249–256. https://doi.org/10.1111/j.1574-6968.2004.tb09654.x

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., … Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*(22), 14250–14255. https://doi.org/10.1073/pnas.202488399

Chen, F., Wang, K., Huang, S., Cai, H., Zhao, M., Jiao, N., & Wommack, K. E. (2009). Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environmental Microbiology*, *11*(11), 2884–2892. https://doi.org/10.1111/j.1462-2920.2009.02033.x

Deng, L., Gregory, A., Yilmaz, S., Poulos, B. T., Hugenholtz, P., & Sullivan, M. B. (2012). Contrasting Life Strategies of Viruses that Infect Photo- and Heterotrophic Bacteria, as Revealed by Viral Tagging. *MBio*, *3*(6), e00373-12. https://doi.org/10.1128/mBio.00373-12

Filée, J., Tétart, F., Suttle, C. A., & Krisch, H. M. (2005). Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12471–12476. https://doi.org/10.1073/pnas.0503404102

Fuller, N. J., Wilson, W. H., Joint, I. R., & Mann, N. H. (1998). Occurrence of a Sequence in Marine Cyanophages Similar to That of T4 g20 and Its Application to PCR-Based Detection and Quantification Techniques. *Applied and Environmental Microbiology*, *64*(6), 2051–2060.

Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W., & Weitz, J. S. (2014). The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature*

Reviews Microbiology, *12*(7), 519–528. https://doi.org/10.1038/nrmicro3289

Labonté, J. M., Swan, B. K., Poulos, B., Luo, H., Koren, S., Hallam, S. J., … Stepanauskas, R. (2015). Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. *The ISME Journal*, *9*(11), 2386–2399. https://doi.org/10.1038/ismej.2015.48

Needham, D. M., Sachdeva, R., & Fuhrman, J. A. (2017). Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *The ISME Journal*, *11*(7), 1614–1629. https://doi.org/10.1038/ismej.2017.29

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parikka, K. J., Le Romancer, M., Wauters, N., & Jacquet, S. (2017). Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. *Biological Reviews of the Cambridge Philosophical Society*, *92*(2), 1081–1100. https://doi.org/10.1111/brv.12271

Patel, A., Noble, R. T., Steele, J. A., Schwalbach, M. S., Hewson, I., & Fuhrman, J. A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nature Protocols*, *2*(2), 269–276. https://doi.org/10.1038/nprot.2007.6

Paul, J. H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME Journal*, *2*(6), 579–589. https://doi.org/10.1038/ismej.2008.35

Rappé, M. S., & Giovannoni, and S. J. (2003). The Uncultured Microbial Majority. *Annual Review of Microbiology*, *57*(1), 369–394. https://doi.org/10.1146/annurev.micro.57.030502.090759

Rohwer, F., & Edwards, R. (2002). The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. *Journal of Bacteriology*, *184*(16), 4529–4535. https://doi.org/10.1128/JB.184.16.4529-4535.2002

Roux, S., Hawley, A. K., Torres Beltran, M., Scofield, M., Schwientek, P., Stepanauskas, R., … Sullivan, M. B. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *ELife*, *3*. https://doi.org/10.7554/eLife.03125

Stepanauskas, R. (2015). Wiretapping into microbial interactions by single cell genomics. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00258

Steward Grieg F., Montiel Janine L., & Azam Farooq. (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography*, *45*(8), 1697–1706. https://doi.org/10.4319/lo.2000.45.8.1697

Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., & Chisholm, S. W. (2006). Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts. *PLOS Biology*, *4*(8), e234. https://doi.org/10.1371/journal.pbio.0040234

Sullivan, M. B., Waterbury, J. B., & Chisholm, S. W. (2003). Cyanophages infecting the oceanic cyanobacterium Prochlorococcus. *Nature*, *424*(6952), 1047–1051. https://doi.org/10.1038/nature01929

Suttle, C. A. (2005, September 14). Viruses in the sea [Special Features]. https://doi.org/10.1038/nature04160

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews. Microbiology*, *5*(10), 801–812. https://doi.org/10.1038/nrmicro1750

Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., … Venter, J. C. (2008). The Sorcerer II Global Ocean Sampling Expedition: Metagenomic Characterization of Viruses within Aquatic Microbial Samples. *PLOS ONE*, *3*(1), e1456. https://doi.org/10.1371/journal.pone.0001456

APPENDIX

PCR Thermocycler Settings for each primer set used during the project.

Cyanophage: CP-DNAP349F, CP-DNAP533Ra, CP-DNAP533Rb

Cycle	Step	Temp	Time
x1	Initial Denaturation	94	3:00
	Denaturation	94	0:30
x35	Annealing	50	0:30
	Extension	72	1:00
x1	Final Extension	72	5:00
x0	Hold	4	-

T7-Podoviridae: T7DPol230F, T7DPol510R

Cycle	Step	Temp	Time
x1	Initial Denaturation	94	5:00
	Denaturation	94	1:00
x35	Annealing	50	1:00
	Extension	72	2:00
x1	Final Extension	72	10:00
x0	Hold	4	-

Myoviridae: T4 MZIA 1 bis, T4 MZIA 6

Cycle	Step	Temp	Time
x1	Initial Denaturation	94	1:30
	Denaturation	94	0:45
x35	Annealing	50	1:00
	Extension	72	0:45
x1	Final Extension	72	5:00
x0	Hold	4	-

Bacterial 16S rRNA: 16S F, 16S R

Cycle	Step	Temp	Time
x1	Initial Denaturation	94	3:00
	Denaturation	94	1:00
x35	Annealing	50	1:00
	Extension	72	1:45
x1	Final Extension	72	10:00
x0	Hold	4	n/a