5-2010

The Optimization of Pulsed Field Gel Electrophoresis for Use with Profiling the Freshwater Viral Community

Dustin Robert Glasner
College of William and Mary

The Optimization of Pulsed Field Gel Electrophoresis for Use with Profiling the Freshwater Viral Community

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelor of Science in Biology from
The College of William & Mary

by

Dustin Robert Glasner

Accepted for _____
(Honors, High Honors, Highest Honors)

_____
Kurt E. Williamson, Director

_____
Mark H. Forsyth

_____
Robert J. Hinkle

_____
Oliver P. Kerscher

Williamsburg, VA
May 4, 2010

ABSTRACT

Dustin Glasner: The Optimization of Pulsed Field Gel Electrophoresis
for Use with Profiling the Freshwater Viral Community
(Under the direction of Kurt E. Williamson)

Pulsed field gel electrophoresis (PFGE) has proven to be a useful tool for fingerprinting viral communities in environmental samples. PFGE has the ability to separate larger DNA segments, and it provides sharper resolution and better band separation than standard gel electrophoresis. Since virus genomes are essentially long segments of DNA, the ability to separate larger molecules is vital; as such, PFGE can provide a proxy measure of viral richness through genome size distribution. Despite its documented usefulness, however, PFGE has not been shown to work flawlessly with all samples – especially those from freshwater environments. For samples taken from Lake Matoaka at the College of William & Mary, PFGE has produced non-distinct smearing and unclear banding patterns, limiting its use as a fingerprinting tool. Experiments were run with single phage isolates (species T4, λ, and CrimD) to determine the viral load at which PFGE ceases to produce clear banding. Pulsed field runs with these phage dilutions showed that a minimum of 10^7 viruses must be loaded into a given well of the gel in order to produce a distinguishable band. Artificial phage assemblages were also created using mixtures of T4 and λ. When run on a gel, these mixtures demonstrated that, so long as viruses are loaded at the threshold of detection, distinct bands will be visible, even in the presence of another virus. However, additional problems may arise in interpreting banding patterns due to DNA topology. The work carried out here more clearly illustrates the limitations of PFGE in fingerprinting aquatic viral assemblages,

though further work must be done to gain even deeper insight into the method's usefulness.

ACKNOWLEDGMENTS

Special thanks to my Committee Chair, Dr. Kurt Williamson, and the other members of my committee: Dr. Mark Forsyth, Dr. Rob Hinkle, and Dr. Oliver Kerscher (academic adviser). Their help and encouragement throughout the whole process was paramount to my success. Working in Dr. Williamson's lab for the past two years has been one of the most rewarding and influential experiences of my life, and for that I thank him. This project would not have been possible without the other members of the Williamson Research Lab, especially Dana Hardbower and Daniel Kiernan for their work with viral concentration from water samples and Kimy Javier for her help with CrimD propagation. Thanks are also given to Hilary Whelan of the Forsyth Lab for her assistance with *Mycobacterium smegmatis* growth and CrimD propagation, as well as fellow honors students Chris Givens and Erica Hart for their mutual understanding. Furthermore, thanks to the Forsyth Lab for use of their Kodak Gel Imager and software. Finally, I would like to thank my family and friends for their unwavering love, support, and patience. This experience would not have been a success without the incredible support group I have around me.

This project was supported in part by the DeFontes Fellowship generously extended by the DeFontes family through the College of William & Mary.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

I. INTRODUCTION

Viruses in Aquatic Ecosystems

Viruses are likely the most abundant organisms on the planet (Steward, 2001). Marine abundance has been reported as low as 10^7 and as high as 10^{10} viruses mL^{-1} (Tijdens et al., 2008; Steward & Azam, 1999); freshwater abundance has been reported to be as low as 10^6 and as high as 10^8 viruses mL^{-1} (Maranger & Bird, 1995; Leff et al., 1999). Viruses have been studied fairly extensively in marine ecosystems, and research has determined that viruses play important roles in marine communities, affecting the diversity and abundance of their hosts. It has been estimated that approximately 10-20% of the marine bacteria stock are lost each day as a result of viral infection and reproduction (Suttle, 1994), and the majority of bacteria mortality (estimates of 70% in marine environments, 90-100% in freshwater environments) can be attributed to phage infection (Tijdens et al., 2008) (a bacteriophage, or phage, is a virus that infects bacteria). Through lysis of host cells, viruses affect biogeochemical pathways in aquatic systems by the release of carbon and nitrogen (Wilhelm & Matteson, 2008).

Viruses also drive host evolution and population diversity through selection for resistant hosts and horizontal gene transfer (Wommack et al., 1999). Individuals that are resistant to phage infection are under positive selection, affecting the composition of the host community (Wommack et al., 1999). Horizontal gene transfer can occur via transduction, or the accidental transfer of genetic material from host-to-host as a result of errors in virus reproduction and genome packaging. This type of virus-mediated gene transfer in bacteria has been estimated to occur globally at a rate of 20×10^{15} events per

second (Burke et al., 2001). In addition to transduction, lysogenic interactions can also affect host fitness and evolution by conferring useful genes from the phage during prophage integration (Brüssow et al., 2004). Given the high abundance of viruses in freshwater, it is logical to begin with a null hypothesis that viruses play a comparably important role in the freshwater environment to viruses in the marine environment. However, in terms of viral ecology, freshwater ecosystems are poorly characterized when compared to the body of knowledge concerning the roles of viruses in marine systems. In order to adequately assess the potential differences between marine and freshwater viral assemblages, additional analyses of viral communities in freshwater environments are needed.

Methods for Analyzing Viral Community Richness

The highest-resolution and most complete view of viral diversity can be obtained by metagenomics analysis of a viral assemblage. This technique allows for sequence-based comparisons (Breitbart et al., 2004); however, metagenomics is an expensive method that requires time, computing resources, and computing expertise. Consequently, few labs are able to perform metagenomic analysis on a routine basis, and as a result, several rapid and inexpensive methods have been developed to gain a usable profile of a viral community.

Within the study of environmental virology, three primary methods exist for analyzing communities: Randomly Amplified Polymorphic DNA-PCR (RAPD-PCR), Transmission Electron Microscopy (TEM), and Pulsed Field Gel Electrophoresis (PFGE). Each of these approaches has its own strengths and weaknesses, and none of the

techniques is capable of providing complete data on the entire virus community.

RAPD-PCR utilizes a 10-mer oligonucleotide primer, whose sequence is generally chosen randomly, to amplify viral DNA. Standard electrophoresis is used to separate the amplicons produced with the primer, and the differences in banding patterns provide a proxy measure for community change across samples (Winget & Wommack, 2008). RAPD-PCR requires relatively little template DNA and no prior knowledge of genome sequences in the sample. Since this method is based on nucleic acid sequences, it also provides deeper insight into the genetic diversity of a community and can be combined with sequencing if desired. The method is, however, not perfect, and banding patterns will not be obtained if the primer fails to a binding site on a genome. DNA fragments of identical or very similar length – but different sequences – will still size at the same location on the gel, making differentiation between the two impossible. Furthermore, only viruses with a double-stranded DNA genome will be amplified. Finally, organic material and other components of environmental samples may also cause problems with primer annealing and genome amplification.

TEM is a direct visualization method for assessing diversity of a viral community. Samples are loaded onto a grid, stained, and then viewed under the microscope. Abundance data may be obtained manually. However, beyond abundance, TEM allows for the observation of morphological differences in viruses, and genome topology does not play a role in detection (Yan-Ming et al, 2006). As a result, a more complete profile of the viral community may be obtained since single and double-stranded DNA and RNA viruses can be observed using TEM. The method is, however, tedious and time consuming. It also relies on the researcher's access to a TEM, which are generally

expensive to both purchase and maintain. Finally, TEM provides little information about genetic differences; even though two viruses may be morphologically similar, their genomes may be completely different.

PFGE is a DNA separation technique that has become popular with viral ecologists for profiling communities in environmental samples (Wommack et al, 1999; Steward, 2001; Filippini & Middelboe, 2007; Tijdens et al, 2008).With PFGE, viral community profiles are generated based on the distribution of viral genome sizes within natural assemblages. In standard gel electrophoresis, a fixed current is run through an agarose gel, separating DNA fragments based upon size. In PFGE, however, alternating homogeneous currents separated by a given angle run for equal time periods in each direction. This is the basis for PFGE's increased resolution and separation of large DNA fragments (whole virus genomes). As the direction of the current switches, larger molecules need more time to reorient their migration to the new current direction; smaller molecules need significantly less time to reorient. Thus, the more time that is provided for each current direction, the larger the DNA molecules that can be separated by PFGE (Birren & Lai, 1993). Gel run conditions can be varied to optimize separation of various size fragments; these conditions include the overall duration of the run, temperature at which the gel runs, voltage for the current, switch time between fields, and angle of separation (Birren & Lai, 1993). Samples are loaded using small agarose plugs that fit into wells in the gel, reducing the frequency of DNA shearing by pipetting. The reduction of DNA shearing is important for environmental virology because PFGE is used to determine the distribution of viral genome sizes within natural assemblages, and shearing of genomic DNA would lead to inaccurate results.

Use of PFGE for Fingerprinting Aquatic Viral Communities

PFGE has been used by viral ecologists because of its reported sensitivity in detecting 10^4-10^6 viruses $lane^{-1}$ (Steward, 2001 ; Filippini & Middelboe, 2007) and its high reproducibility. It has the ability to separate viral genomes from sizes as low as 20 kb up to 300 kb (kilobase pairs) on the same gel run (Steward, 2001). This high range, flexibility, inexpensive nature, and high throughput make PFGE very appealing to environmental virologists who do not necessarily know the composition of the community while sampling.

Using genome size distribution as a proxy for viral diversity in a sample, PFGE is useful in tracking changes in the viral community over time and/or space. By comparing banding patterns produced by different samples, researchers can observe the presence/absence of certain bands and infer how the community has changed. This technique has been used successfully by several researchers investigating viral assemblages in a variety of environments, including rumen (Klieve & Swain, 1993), marine waters (Wommack et al., 1999), freshwaters (Tijdens et al, 2008), and sediments (Filippini & Middelboe, 2007).

In using PFGE to fingerprint viral assemblages in Lake Matoaka, however, we have encountered several issues similar to those observed in other freshwater studies. In most of the gel runs, bands appear to be faint or non-distinct, oftentimes displaying smearing in a lane. Though some data may be extracted from these gels, they are far from optimal, especially for our goal of creating a profile of the community. Despite the fact that PFGE is popular for use with marine samples, it has not been used nearly as extensively to study freshwater systems; as such, its limitations with respect to freshwater

samples are not as well documented (Auguet et al., 2006). By performing simple experiments with PFGE and single virus species, this research seeks to more firmly establish the limitations of pulsed field gel electrophoresis in application towards the fingerprinting of freshwater viral assemblages.

II. MATERIALS AND METHODS

Media Preparation

For growth of *Escherichia coli*, *TSBA + 0.5% NaCl* plates (7.5 g TSB powder; 2.5 g NaCl (s); 7.5 g agar; 500 mL diH_2O) and *TSB + 0.5% NaCl broth* (7.5 g TSB powder; 2.5 g NaCl (s); 500 mL diH_2O) were used, as well as *TSBA soft agar* (3.0 g TSB powder; 1.0 g NaCl (s); 1.5 g agar; 200 mL diH_2O). For growth of *Mycobacterium smegmatis*, *Middlebrook 7H9 Liquid Medium: Neat* (4.7 g 7H9 broth base; 5 mL 40% glycerol stock; 995.3 mL diH_2O) and *AD Supplement* (4.25 g NaCl (s); 25.0 g albumin (Fraction V); 500 mL diH_2O; 10.0 g dextrose) are necessary to prepare *Middlebrook 7H9 Liquid Medium: Complete, No Tween* (89 mL 7H9 liquid medium: neat; 10 mL AD supplement; 100 µL 50 mg mL^{-1} CB stock; 100 µL 10 mg mL^{-1} CHX stock; 1 mL 100 mM $CaCl_2$ stock). For plating of *M.smegmatis*, *Middlebrook 7H10 Agar Plates* (9.5 g 7H10 agar base; 6.25 mL 40% glycerol stock; 443.7 mL diH_2O; 50 mL AD supplement; 500 µL 50 mg mL^{-1} CB stock; 500 µL 10 mg mL^{-1} CHX stock) and *2x TA* (4.7 g 7H9 broth base; 7.0 g agar; 1 L diH_2O) are required. Buffers needed include *SM Buffer* (5.8 g NaCl (s); 2 g $MgSO_4 \cdot 7H_2O$ (s); 50 mL 1.0 M Tris-Cl, pH 7.5; 950 mL diH_2O), *Phage Buffer* (10 mL 1 M Tris stock, pH 7.5; 10 mL 1 M $MgSO_4 \cdot 7H_2O$ stock; 4.0 g NaCl (s); 970 mL diH_2O; 10 mL 100 mM $CaCl_2$ stock) and *1x TE Buffer* (10 mL 1.0 M Tris,

pH 8.0; 2 mL 0.5 M EDTA, pH 8.0; 988 mL diH_2O).

Environmental Sample Preparation

Water samples (8L) were collected at three distinct locations (pier, spillway, inlet) in Lake Matoaka once a month from March 2009 to July 2009. The water was passed through a series of filters (10 μm, 5 μm, 0.22 μm) to remove debris, zoo- and phytoplankton, and bacterioplankton, respectively. Viruses in the cell-free filtrate were then concentrated through tangential flow filtration using a Millipore Prep/Scale filter cartridge (30 kDa) (Suttle et al., 1991). To further concentrate the viruses, samples were then run in a centrifuge (Beckman Coulter Allegra X-15R) through Millipore Centricon devices as described by the manufacturer. The final viral concentrate (ca. 2 mL, ~10^9 viruses mL^{-1}) was stored at 4° C until needed for use. Before sampling in the month of July, the method used for tangential flow filtration was changed to the current system because of availability of materials.

Phage Propagation and Viral Concentrate Preparation

T4: Phage stocks of T4 were obtained from ATCC (11303-B4) and amplified in *Escherichia coli* B (ATCC 11303). To propagate high-titer lysates, flood plating was used. First, a plaque assay was performed to determine the proper dilution of phage to use for confluent lysis; 100 μL of T4 dilution (10^0-10^{-9}) and 100 μL of log phase *E.coli* were mixed with 4.5 mL of TSBA Soft Agar and poured onto TSBA + 0.5% NaCl media plates. After overnight incubation at 37º C, the dilution that displayed confluent lysis was used to infect 10 cultures of *E.coli*; mixtures were plated with TSBA Soft Agar.

Following overnight growth, the plates were flooded with 8 mL of SM Buffer, and incubated for one hour at room temperature. The liquid was aspirated into a 50 mL conical tube using a serological pipette. These plate lysates were then centrifuged at 5250 x g for 20 minutes at 4° C and filtered through a 0.22 μm filter. Lysates were then purified by adding 250 μL of chloroform and inverting several times. After 10 minutes of incubation at 4° C, the tubes were centrifuged at max speed for 20 minutes at 4° C, and the supernatant was passed through a 0.22 μm filter. Lystates were further concentrated using Centricon filters as described by the manufacturer and stored at 4° C until use.

λ: Phage stocks were obtained from ATCC (23724-B2) and amplified in *Escherichia coli* C600 (ATCC 23724). A modified protocol for broth lysis described in *Molecular Cloning* (Sambrook, 2001) was used for propagation of λ. After growing a log-phase culture of *E.coli*, 1 mL of the culture was transferred to a 50 mL conical tube and infected with ~10^7 pfu of λ in 1000 μL SM Buffer. The infected culture was incubated at 37° C for 20 minutes, allowing adsorbtion of phage to host. Then, 40 mL of pre-warmed TSB + 0.5% NaCl liquid medium was added to the tube. After gentle vortexing, the culture was split into two 50 mL tubes (~21mL each) and incubated at 200 rpm for 8-12 hours at 37° C. Following incubation, 250 μL of chloroform was added to each tube, and tubes were placed back on the shaker at 37° C for 15 minutes. The stocks were then centrifuged at 5250 x g for 20 minutes at 4° C, and the supernatant was recovered and passed through a 0.22 μm filter. Lysates were further concentrated by Centricon filtration and stored at 4° C until use.

CrimD: Originally isolated from a soil sample by H. Whelan at the College of William & Mary, stocks of CrimD were amplified from archived concentrates in

Mycobacterium smegmatis (ATCC 700084). *M.smegmatis* was grown for 48 hours at 37° C in Middlebrook 7H9 Liquid Media: Complete, No Tween. As with T4 growth in *E.coli*, CrimD was grown using plate lysis. For each plate in both the initial diagnostic plaque assay and for confluent lysis, 500 µL of host and 100 µL of phage were used. Instead of TSB + 0.5% NaCl plates, Middlebrook 7H10 plates were used. To flood the confluent lysis plates, Phage Buffer was used rather than SM Buffer. Aside from the previously listed exceptions, CrimD propagation was carried out in the same manner as T4 propagation.

Titering Phage Stocks

To determine the titer of both environmental and lab-grown viral concentrates, epifluorescence microscopy was used. Counts were obtained before Centricon filtration of phage stocks in an effort to conserve viral concentrate. 100 µL of phage stock was loaded onto a 13 mm, 0.02 µm pore-size Anodisc (mfg: Whatman) and drawn through the filter under vacuum. The Anodiscs were then stained with 100 µL of 2.5x SYBR Gold (mfg: Invitrogen) in the dark for 15 minutes. After drawing the stain off under vacuum, the Anodiscs were air-dried and mounted onto a slide using a drop of antifade (50 mg *p*-phenylene diamine; 2 mL 1x PBS; 2 mL 1.0 M Tris, pH 8.0; 46 mL 80% glycerol). Virus particles were visualized using an Olympus BX-51 microscope outfitted with a mercury arc lamp and a fluorescein isothiocyanate (FITC) filter. 10 digital pictures were taken for each viral concentrate using a Hamamatsu CCD (charge-coupled device) camera, and virus particles were enumerated in each image using the Metamorph Basic software package (mfg: Molecular Devices). Viral titers (viruses mL^{-1}) were determined

based on the average number of particles in 10 images, the total volume of lysate loaded onto the Anodisc, and area of image observed:

average viruses per image * (π * (filter radius)^2 / image area) * (1 / volume filtered)

Following Centricon filtration, the new titer was mathematically derived based on concentration factor. Final viral titers per mL were as follows: T4, 7.1 x 10^{10}; λ, 2.4 x 10^{10}; CrimD, 4.8 x 10^9.

Community Assembling

Mock-virus assemblages were created using simple mixtures of phages. Assuming a necessary viral load of 10^8 particles, plugs were made with the following mixtures: 50% T4:50%λ; 90% T4:10% λ; and 10% T4:90% λ. Proportions were calculated based upon the viral concentrate titers. If 250 μL of viral concentrate are necessary to make 5 plugs, and 10^8 particles are necessary for each plug, then 6.25 x 10^8 total particles are needed per mixture. 50% of this is 3.125 x 10^8 particles; 90% is 5.625 x 10^8 particles; 10% is 6.25 x 10^7 particles. Dividing these values by the respective phage titer will yield the amount of concentrate to load into each mixture. For example, for a 50% T4:50% λ mixture, 4.40 μL of T4 concentrate and 13.02 μL of λ concentrate are necessary to provide the correct amount of particles; the volume is then brought up to 250 μL using SM Buffer.

Casting Viral Concentrates into Plugs

Agarose plugs were made as previously described by Wommack et al. (http://www.virusecology.org/MOVE/Method%207.html). 250 μL of viral concentrate

and 250 μL of 1.5% InCert agarose (mfr: Lonza) were mixed together in an 2.0mL Eppendorf tube, vortexed, and kept molten in a 65° C heat block. 80 μL of the mixture was pipetted into each well of the plug mold; each plug mold can hold 10 plugs – 2 dilutions per plug mold, 5 plugs per dilution. The plug molds were then stored at 4° C for 20 minutes to allow solidification of each plug. The 5 plugs from each dilution were then displaced from the molds and into 15 mL tubes containing 2 mL of 250 mM EDTA + 1% SDS and 100 μL of 1 mg mL^{-1} Proteinase K solution and stored overnight. Digestion with Proteinase K breaks down the viral capsids, leaving the phage DNA trapped in the agarose plug. After being stored overnight, plugs were rinsed three times with 1x TE Buffer for 30 minutes at a time and stored at 4° C until use.

Pulsed-Field Electrophoresis Conditions

Gels were made using 0.8% PFGE-certified agarose (mfg: Bio-Rad) in 200 mL of 0.5x TBE. All gels were run in a Bio-Rad CHEF-DR II Pulsed Field Electrophoresis System using identical conditions in order to reduce the number of variables that could affect results. Gels were run under the following conditions: run time = 19.5 hours; temperature = 14° C; voltage = 6 V cm^{-1}; initial switch time = 0.3 s; final switch time = 6.5 s. The running buffer was 2.2 L of 0.5x TBE.

Gel Loading

For T4, λ, and CrimD, serial dilutions were carried out on each viral concentrate and loaded into plugs (10^0 – 10^{-9}). The undiluted plug contains approximately 10^9 viruses. For gel loaded with environmental samples from Lake Matoaka, a Bio-Rad DNA ladder

and a λ ladder were used for size markers. With the exception of the λ dilution series, all other gels were run using a New England BioLabs MidRange PFG Marker for size reference; the λ dilution was run with a Bio-Rad λ ladder as a size standard.

Staining and Visualization

Following the run, gels were stained in 400 mL of 0.5x TBE buffer with 60 μL of 10,000x SYBR Gold in the dark at 4° C for 2.5 hours. Gels were visualized on both the GE Healthcare Storm 860 imaging system with ImageQuant software and the Kodak Gel Logic 100 imaging system with Kodak Molecular Imaging Software. Banding patterns generated from environmental viral concentrates were analyzed using ImageQuant software and converted to binary matrix format. Similarity between sample banding patterns was determined by performing cluster analysis on binary data (Dice coefficient) in PAST v. 1.84 (Hammer et al, 2001).

III. RESULTS

Environmental Samples

During July 2009, PFGE was run on 13 viral concentrates obtained from Lake Matoaka (Figure 1). This gel was run in an attempt to track changes in viral diversity in Lake Matoaka over time and space. In most of the lanes, it is possible to distinguish 3-5 bands, and they range from approximately 49 kb in size to 200 kb in size. Most known phage genomes can be found in the observed size range (Ackermann & DuBow, 1987), suggesting that the majority of viruses present in Lake Matoaka infect bacteria. The strongest bands can be seen at approximately 50 kb, 90 kb, and 200 kb.

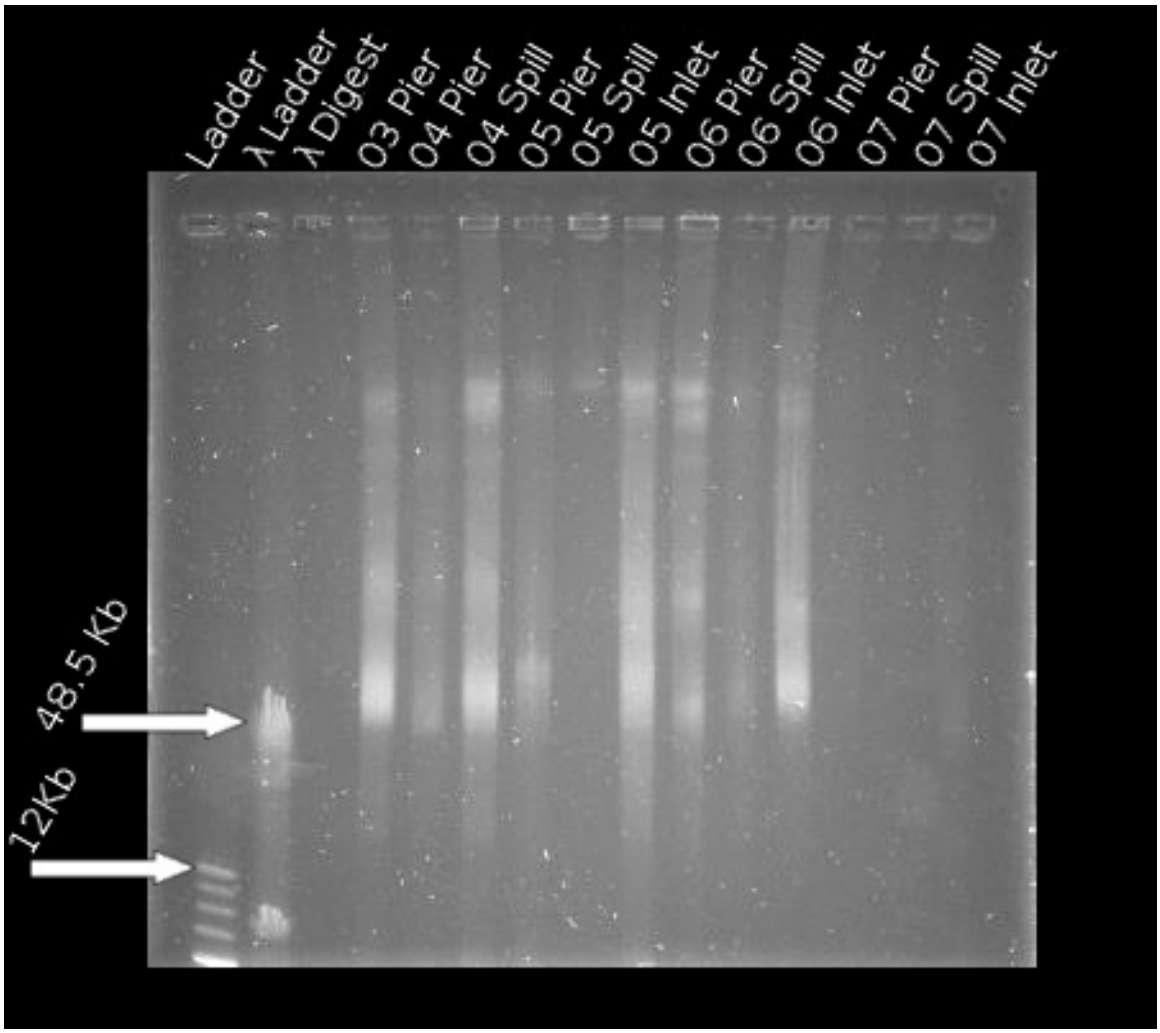

Figure 1 - Lake Matoaka Monthly Samples: Gel from Summer 2009 showing viral concentrates obtained from various locations in Lake Matoaka at monthly time intervals. Distinct bands are hard to determine, and general smearing is the prominent feature in each lane. Bio-Rad DNA Ladder and λ Ladder used as size markers.

Smearing is apparent in all lanes, and as a result, distinct band sizes are hard to delineate. Weak fluorescence can be seen below approximately 50 kb, suggesting that these samples contain viruses with smaller genomes; however, such viruses may not exist in a high enough percentage of the community to create a strong band on the gel.

At the spillway site, fluorescence intensity decreases from April to May, suggesting a decrease in viral abundance; furthermore, bands disappear in May, with the exception of around 200 kb in size. From May to June, the fluorescence signal gets stronger, though bands are not distinguishable. At the pier site, fluorescence intensity seems to remain relatively consistent from May-June. Bands are consistently observed around 50 kb, 90 kb, and 200 kb in size; however, in June, several new bands located around 150 kb-175 kb become apparent. At the inlet site, fluorescence remains constant from May to June, and bands appear at approximately the same sizes. Lanes loaded with viral concentrate from the month of July show little-to-no fluorescence; the most likely cause of this deficiency is a slight methods change from previous months (from a Sartorius Vivaflow 50 ultrafiltration module to a Millipore Prep/Scale Filter).

Cluster Analysis

The data analysis package PAST was used to create a cluster diagram (Figure 2) from the results obtained from Lake Matoaka samples (Figure 1). Band size data was arranged into a presence/absence matrix, and analysis was run using the Dice coefficient, which uses bigrams to compare similarity across two data sets. The formula for the Dice coefficient is: $2C / A + B$, where C is the number of bigrams in common across two data sets, and A & B are the total number of bigrams in a single data set.

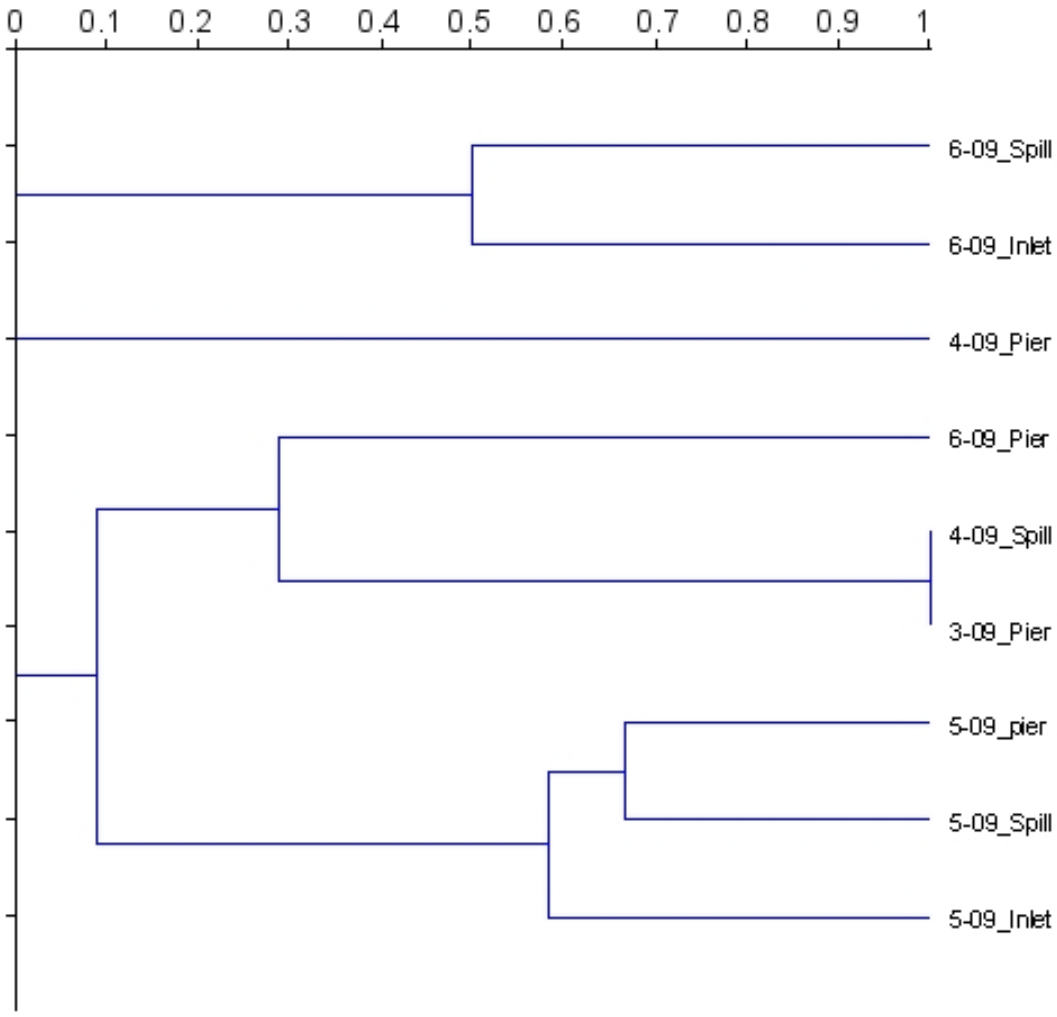

Figure 2 - Cluster Analysis: Cluster diagram created in PAST using the Dice Coefficient to display quantitative, statistical differences between samples.

The resulting cladogram suggests that there are distinguishable differences between samples, but the level at which these differences can be observed is low. For example, several samples (June spillway and inlet, and May pier, spillway and inlet) appear to be about 50% similar by cluster analysis. However, inspection of the gel image (Fig. 1) reveals that these samples typically differ by the position of a single band. Thus, with few identifiable band types in each lane, this type of analysis is highly sensitive to such small differences.

Single Phage Isolates

For all single isolates (T4, λ, and CrimD), bands were observed as low as 10^7 viruses lane^{-1}, but lower dilutions showed no signs of staining (Figures 3, 4, 5, and 6). These effects were observed independently of genome size (T4: 168.9 kb; λ: 48.5 kb; CrimD: 59.8 kb). Though bands were visible at 10^7 viruses lane^{-1}, the clearest band was consistently observed at 10^8, while lanes loaded with 10^9 particles showed signs of overloading and bands were less clear (Figures 3, 4, and 5). In the λ gel (Figure 5), 3 plugs of T4 were loaded to confirm abnormal genome migration of T4 genomes relative to λ genomes.

In the T4 gel (Figure 3), an abnormality was observed. Phage T4 has a reported genome size of 169 kb, but the bands produced on the gel size at approximately 40 kb. In a separate study, PFGE was used to size the genomes of T4-like phages, including T4 (Tetart et al., 2001). In that study, the genome of phage T4 appeared at the expected size of ~170 kb; however, the researchers carried out the digestion of phage capsids under slightly different conditions than in the present study.

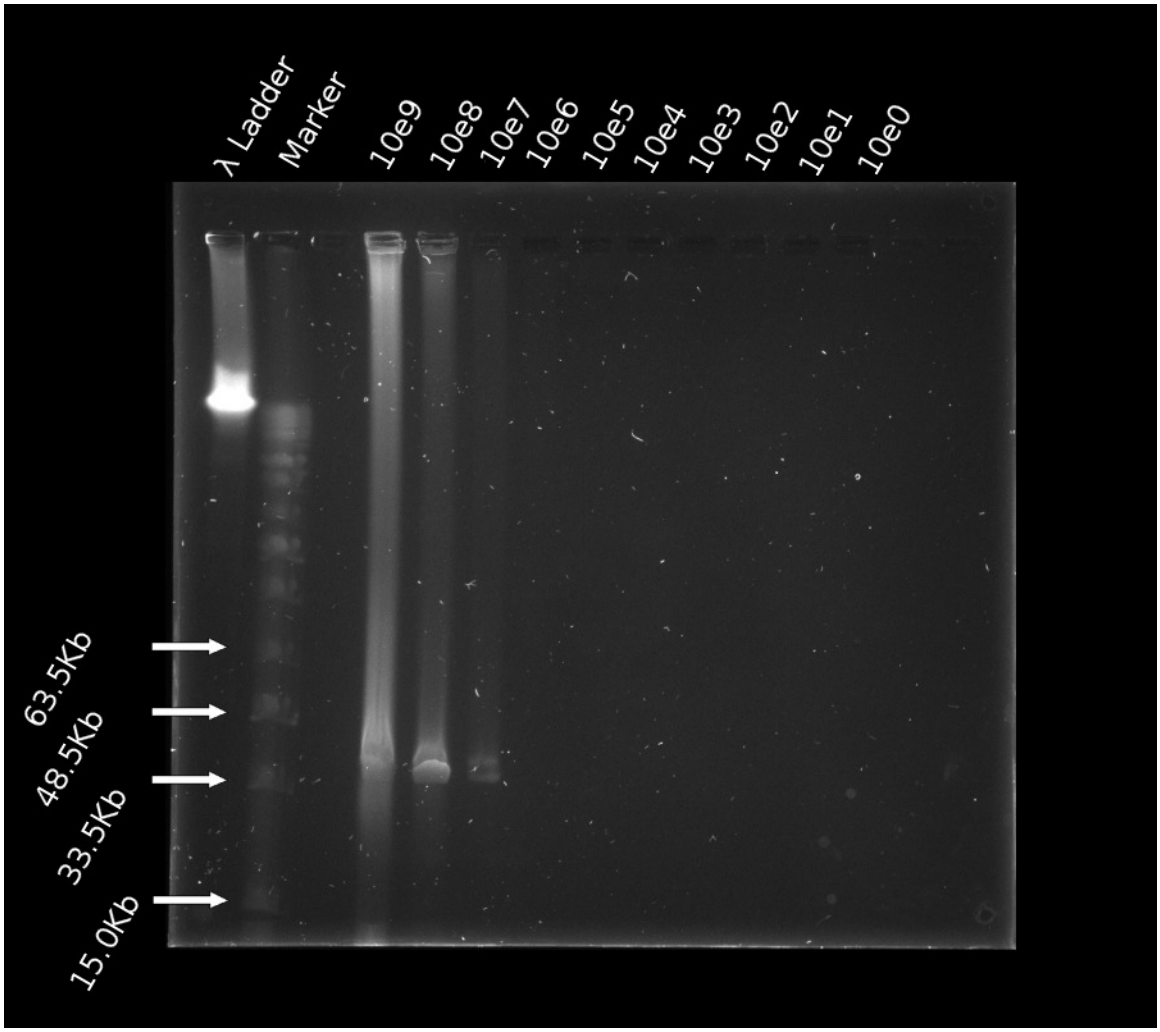

Figure 3 - T4 Dilution Series (Room Temperature): T4 dilution series showing decreasing band intensity with decrease in viral load lane^{-1}. T4 genome should size higher on the gel. BioRad λ Ladder and New England BioLabs MidRange PFG Marker used as size markers.

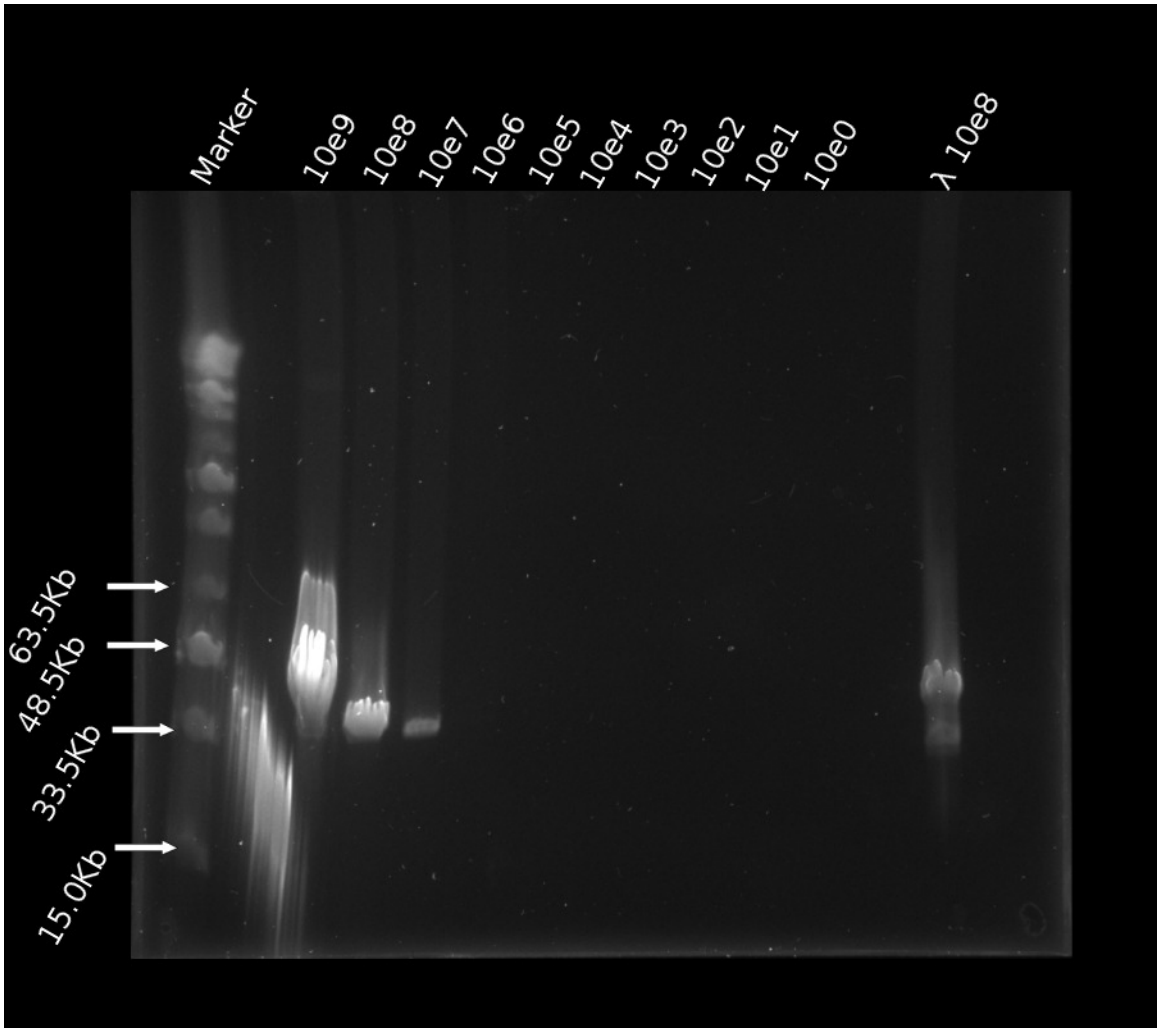

Figure 4 - T4 Dilution Series (55° C): Diagnostic gel of T4 dilution series digested at 55° C overnight. No significant size change in band position was observed relative to room temperature incubation. λ plug loaded for size reference, along with NEB MidRange PFG Marker. Artifactual staining appears in lane 2.

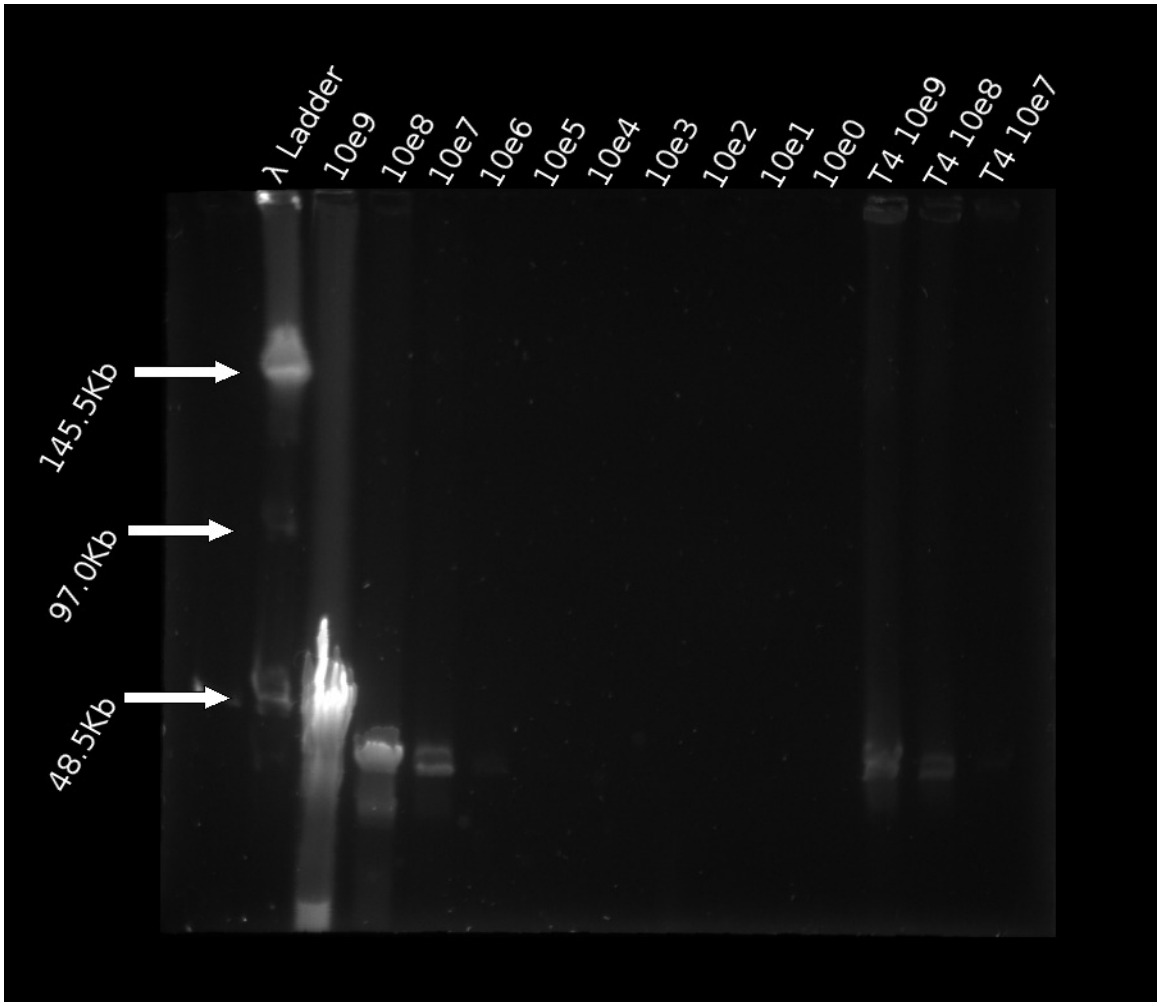

Figure 5 - λ Dilution Series: Dilution series of λ showing decreasing band intensity. T4 plugs loaded to confirm abnormal genome migration. Bio-Rad λ Ladder used as size marker.

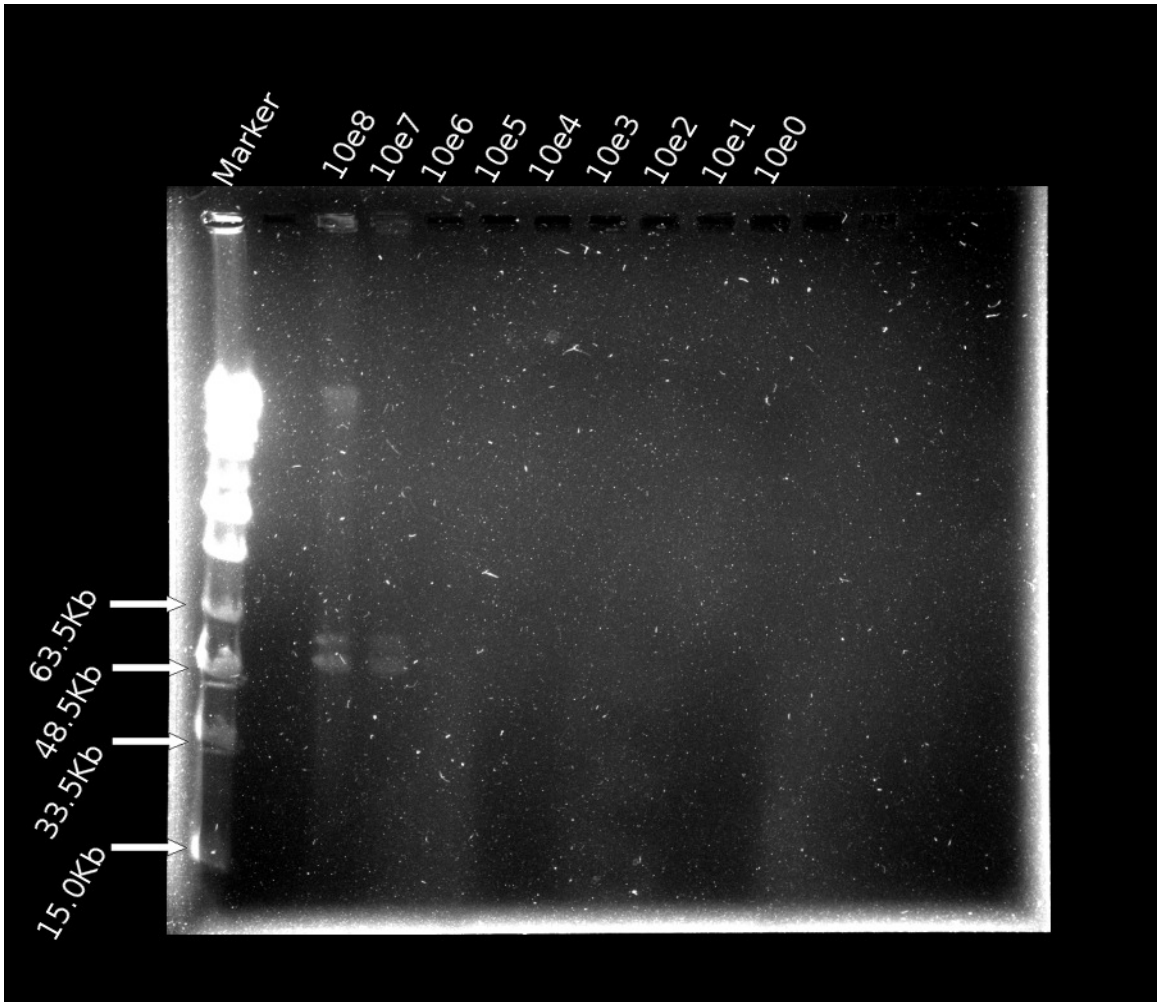

Figure 6 - CrimD Dilution Series: Dilution series of CrimD. Lower titer phage, showing only two bands at 10^8 and 10^7. NEB BioLabs MidRange Marker used for size standard.

In an attempt to reproduce these results, plugs containing T4 were digested at 55°C overnight and then run on a gel (Figure 4). Band location did not change from Figure 3 (room temperature digestion), though band intensity appeared to increase relative to room temperature digestions. The second lane of the gel was not loaded with sample plugs, but shows a strange staining pattern. This may be an artifact that resulted from extended time in the PFGE cell before staining. λ was loaded as an additional size reference for T4.

T4/λ Mixtures

For mock-phage assemblages, 50%:50%, 90%:10%, and 10%:90% mixtures of T4:λ were created and loaded into plugs. Replicate plugs were loaded (2 per mixture), and the resulting gel is shown in Figure 7. Bands appear clear for all lanes. Distinct bands for both phages were observable. Genome size did not seem to impact band intensity, and the presence of another phage did not appear to impact the location of bands for either genome. Genomes of the same size migrated to the same location regardless of mixture proportions. However, strange patterns of fluorescent intensity were observed. The intensity of T4 bands appeared to decrease with an increasing number of genomes, whereas the intensity of λ bands appeared to increase with an increasing number of genomes.

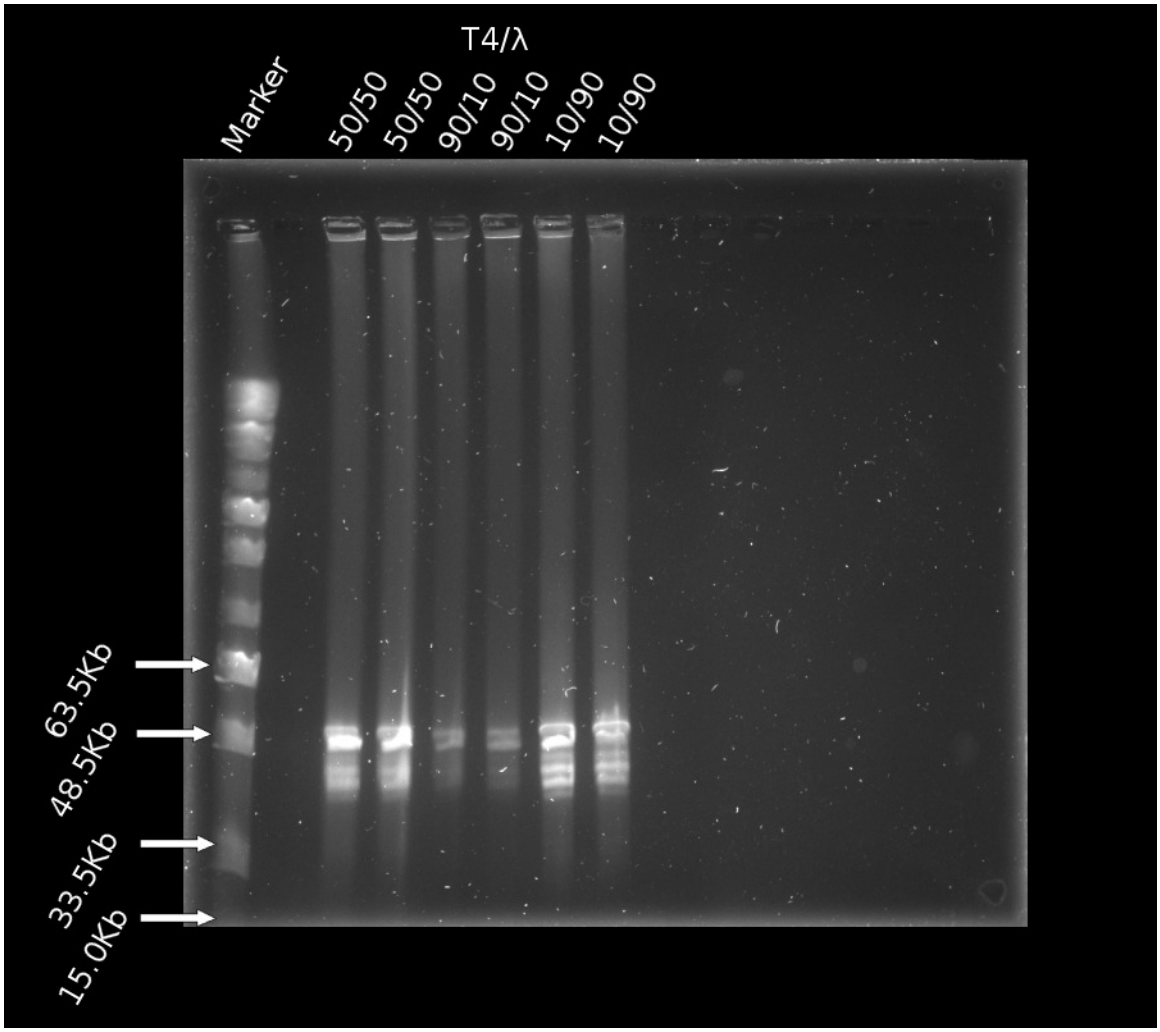

Figure 7 - Phage Mixtures: Phage mixture plugs run in duplicate. 50% T4/50% λ; 90% T4/90% λ; 10% T4/90% λ. λ sizes as the higher band, T4 as the lower band. Brighter banding for 90% λ and slightly dimmer banding for 10% λ were expected, but not observed. NEB BioLabs MidRange Marker used as size standard.

IV. DISCUSSION

Fingerprinting Environmental Viral Assemblages with PFGE

PFGE provides researchers with an idea of the diversity of a viral community and how the community may be changing over time and/or space. Banding patterns produced by a gel run create a proxy measurement from which researchers can draw conclusions about a given community. Changes are observed by the presence or absence of bands from sample to sample.

Threshold of Detection

Previous studies using viral concentrates generated from environmental samples have typically loaded about 10^9 virus particles per lane (Steward, 2001; Filippini & Middelboe, 2007; Tijdens et al., 2008). Within several studies (Wommack et al., 1999; Filippini & Middelboe, 2007), the authors state that the limit of band detection in pulsed-field gels was determined using known phage isolates, and was generally found at 10^6 particles per plug. In other words, if 10^9 particles were loaded into a lane, viruses within a given genome size class that made up only 1% of the population (10^6 particles) should still be detected. Based upon the results obtained in the present study from the single-phage dilution series, the minimum number of phage particles of a given genome size necessary to resolve into a detectable DNA band was 10^7 (Figures 3, 4, 6). This trend was consistent across each isolate tested, independent of phage genome size. Furthermore, this cut-off differs from previous estimates that indicated a minimum of only 10^6 (Steward, 2001), or even 10^4 (Filippini & Middelboe, 2007) viruses were needed per plug

in order to resolve DNA bands. The transition from visible band to complete absence of band seems to occur very abruptly, but this may be the result of performing only a 10-fold serial dilution; loading plugs at smaller increments ($1x10^6$, $2x10^6$, etc.) would provide a more accurate measurement as to where PFGE loses sensitivity.

Despite only testing 3 simple phage mixtures, it appears that – given at least 10^7 particles in a lane – bands will appear for a single viral species in a mixture; that is, it is reasonable to assume that results would be similar when looking at larger, more complex viral communities. Questions remain, however, as to whether or not viral species of similar genome sizes loaded at 10^7 particles would resolve clearly in a lane.

Accuracy of Viral Genome Size Estimates

In the course of completing the single phage dilutions, a significant – and previously unreported – problem became apparent in using PFGE to estimate viral genome sizes. In spite of its known genome size, determined by sequencing to be 168.9Kb (Miller et al., 2003), the DNA bands for T4 genomes consistently appeared at approximately 40 kb (Figures 3, 4). Despite the purchase of completely new stock of T4 from ATCC, and meticulous attention to detail in preparing phage concentrates, T4 repeatedly appeared at approximately a quarter of its reported genome length on each of the PFGE runs (Figures 3, 4, 7). Following the protocol used by Tetart et. al. (Tetart et al., 2001), digestion was carried out at 55°C overnight in order to loosen any potential secondary structures in the DNA (Figure 4); however, the bands remained at the same location as the standard room temperature digest (Figure 3). Supercoiling of phage DNA could explain why the T4 genome would act as a shorter molecule, migrating further

down the gel than expected (Wang & Lai, 1995). However, this theory is unlikely because T4 possesses a linear genome; for supercoiling to occur in a linear genome, structural proteins are required, and they would be degraded during the Proteinase K digestion. Though superhelicity has been noted by other researchers, it is unlikely that it would account for the drastic change in genome size observed using PFGE (Sinden & Pettijohn, 1982).

In order to determine why T4 sizes so much lower during a PFGE run than its accepted genome size, several experiments can be performed. If some type of supercoiling of the DNA is occurring, treatment of plugs with a nicking agent or restriction enzyme may relax the genome, allowing tension to be relieved and proper migration to occur. Since all of the T4 isolates run for this research originated from ATCC, testing strains of T4 from other sources may produce accurate sizing with PFGE; however, if the results are identical, credence would be added to the argument for some kind of DNA secondary structure. Finally, mutant strains of T4 could be run using PFGE, providing insight into the nature of other T4 genomes.

This finding raises an important question: how often does this type of inaccurate genome sizing (i.e., observed band size in PFGE does not match actual size of viral genome in base pairs) occur in the processing of environmental viral samples for PFGE, and how would this affect genome distributions observed using PFGE? Though Auguet observed a strong correlation between genome size distribution in PFGE and capsid size determined by TEM (i.e. similar numbers of large genomes on PFGE and large capsids with TEM) (Auguet et al., 2006), it is possible that supercoiling or other DNA-DNA interactions play a role in the genome topology of environmental viral strains, and in turn,

in PFGE usage with environmental communities. Finding a solution to this issue with T4 may provide us with insight into how we may alter our protocol for plug preparation and digestion for use with environmental samples.

Artificial Phage Assemblages

The gel containing the mock phage communities had the sharpest bands of any in this study. Solid bands were observed for each phage in each lane. This is likely due to the fact that the loading rates for phages in these mixtures were essentially engineered to produce resolvable bands. The minimum contribution of any single phage to the entire mock assemblage was 10% of 10^8 particles – or 10^7 total particles –the number previously determined as the threshold of detection for PFGE in this study. For the 90% T4:10% λ mixture, however, a stronger band for T4 (the lower of the two bands, despite the larger genome size) and a weaker band for λ was expected because of the loading proportions. There should be significantly more T4 than λ ($9.0x10^7$ T4 particles vs. $1.0x10^7$ λ particles), but T4 presents as the weaker band. It is possible that DNA-DNA interactions within the lane caused some of the T4 genomes to be caught higher on the gel with λ, resulting in a stronger band at λ's location. For the 10% T4:90% λ, the band for λ was expected to be considerably brighter than T4, again as a result of loading proportions. Though this was not the case, it is possible that the larger genome size of T4 resulted in a brighter band because the larger genome provided more DNA for SYBR Gold to bind.

For T4, the highest fluorescent signal came from the band associated with 10% of the sample, followed by the 50% fraction and the 90% fraction. For λ, the pattern was

reversed. This pattern in fluorescence intensity may be an artifact of potential DNA-DNA interactions occurring in a lane, but it also may be the result of loading conditions. It is possible that, when working with multiple genomes in a sample, larger genome sizes are optimally loaded at a lower concentration than smaller genome sizes. This hypothesis could be tested by creating more mixtures of viruses using species of varying genome size and loading proportions.

Limitations of PFGE

Though PFGE can be a powerful profiling tool, it is important to note its limitations. It appears that most viruses in the environment are PFGE-detectable, with double-stranded DNA genomes (Wommack & Colwell, 2000). However, viruses with RNA genomes or single-stranded DNA genomes are known to exist as well; these viruses would not be detected by PFGE. Furthermore, PFGE has a relatively high limit of detection, requiring 10^7 genomes of a given genome size class in order to resolve bands in a gel. Assuming 10^9 viruses loaded per plug, this means that rare classes of viruses comprising less than 1% of the sample population will not be included in the analysis. Thus, while PFGE may capture larger-scale differences in viral community composition, finer-scale details and variations may be lost.

If the results obtained in the laboratory for phage T4 are any indication, then it is possible that phage genomes within environmental viral concentrates can engage in DNA-DNA interactions, creating bands at incorrect genome sizes; this would lead to an inaccurate profile of the community. Virus genomes with cohesive ends, and therefore the ability to concatamerize, may also cause problems with the generation multiple bands

from a single genome; however, my observations of λ genomes under PFGE in the laboratory indicate that this may be less of an issue. Phages with similar genome sizes may not be resolved properly on the gel, resulting in the appearance of a single band, a heavy band, or a smear, despite the presence of two or more viruses. Without prior knowledge of the genome size distribution, it would be impossible to differentiate this on a gel, and further laboratory-controlled experiments are necessary to determine the extent of this phenomenon. It is also interesting to note that overloading of virus particles results in smears, bleeding of bands, and heavy banding to the point of loss of clarity. As a result of the convention of loading 10^9 total particles $lane^{-1}$ in an effort to generate the most complete profile, more dominant viral species may be overloaded, decreasing the clarity of bands representing the most common genome classes. This presents yet another variable to consider when loading gels, and the possibility of overloading common classes to visualize rarer classes becomes an important factor.

Marine vs. Freshwater

Because of the differences between marine environments and freshwater environments (particularly in terms of salinity and water chemistry), it is reasonable to assume that differences would also exist between results produced by PFGE on marine and freshwater samples. Marine environments, especially off-shore, generally have lower concentrations of dissolved organic matter (DOM) than freshwater environments, resulting from a lack of inputs from terrestrial runoff. Furthermore, much of the suspended organic particulate matter carried by freshwater sources will precipitate out of solution in polyhaline seawater. By contrast, freshwaters experience frequent runoff from

the surrounding terrestrial environment, contributing to higher levels of DOM in the water (Battin et al., 2009), and suspended colloids are not generally precipitated in freshwaters. Thus, the increased presence of suspended solids and DOM in freshwater samples may impact PFGE quality, resulting in a lack of clarity and a high background signal for samples obtained from freshwater (Tijdens et al., 2008; Filippini & Middelboe, 2007).

Marine environments have been documented in considerably greater detail than freshwater environments (Wilhelm & Matteson, 2008), and the papers utilizing PFGE to fingerprint marine viral communities have typically shown clearer results. Gels containing marine viral concentrates tend to show less fluorescent smearing, and many distinct individual bands can usually be seen in a given lane (Steward & Azam, 1999; Steward, 2001). As a result of the improved clarity of bands, changes occurring over time and space in marine samples are much easier to distinguish than in freshwater samples. Compared to gels run using marine samples (Wommack, 1999; Steward et al., 2000; Steward, 2001), results from Lake Matoaka are weaker overall and more difficult to interpret.

Several papers documenting viral communities in freshwater have shown similar pulsed field gel results to those obtained from Lake Matoaka (see Figure 1) – smearing in lanes with some areas of stronger fluorescence suggesting a band (Filippini & Middelboe, 2007; Tijdens et al., 2008). The work of Tijdens et al. shows some of the best work with PFGE for freshwater environments. Despite reported problems with PFGE, including smearing and faint banding (Tijdens et al., 20008), their results were both qualitatively and quantitatively stronger than what we observed from Lake Matoaka. The number of

bands per lane observed in Lake Matoaka (2-5 bands) is lower than results observed in other freshwater studies by about 50%; the generally reported range is 6-10 bands per lane (Tijdens et al., 2008). Overall, the fluorescent signal for our gels was weaker and not as distinct; changes in the viral community are clearer in Tijdens' paper.

Based upon the results obtained from Lake Matoaka, it is not unreasonable to hypothesize that genome sizes are continuously distributed across broad ranges (e.g. 40-60 kb), resulting in too few particles of a single genome size to produce clearly resolved bands. This would also suggest that viral diversity in Lake Matoaka is high, and since similar results have been observed in other freshwater studies, it is possible that diversity is generally higher in freshwater environments than marine environments. The hypothesis of higher viral diversity in freshwater is generally supported by the fact that the potential sources of viruses are more varied for freshwater systems (including aquatic, terrestrial, and amphibious host species) than for marine systems, particularly the open ocean.

Despite the lower quality of our results from Lake Matoaka compared to Tijdens' work with freshwater samples, we were still able to view some changes in the viral community of Lake Matoaka over time and space. While we may have been able to analyze viral assemblages more thoroughly with clearer results from PFGE, our results provide us with the knowledge that the viral assemblage in Lake Matoaka is not completely static. Whether clarity of banding patterns like those observed with marine samples can be achieved in freshwater samples has yet to be determined, but because of its ease of use and ability to produce some results, PFGE stands to remain as one of the more popular tools for virology work in aquatic environments.

Future Considerations

Though significant strides have been made towards fully understanding the limitations of PFGE within the field of environmental virology, there is still significant work to perform before we can truly understand how to best apply PFGE to environmental freshwater samples. Several more experiments with laboratory grown phages, along with continued work with environmental samples, would prove beneficial to our understanding of PFGE.

In the lab, I aim to continue work with artificial phage assemblages. If preparation of high titer viral concentrates (10^{11}-10^{12}) can be accomplished, then preparing mixtures of more extreme percentages (99%:1%) would provide us with a better idea of the threshold of detection for samples with multiple viruses at differing titers. Mixtures with three viruses would also be a useful experiment, increasing the similarity to an environmental sample, even if only by a single species. In a very simple experiment, Steward demonstrated that 4 phages in a mixture could produce distinct banding patterns (Steward et al., 2000); one of our goals is to reproduce these results, closely observing potential interactions between different genome size bands. Finally, combining two phages with very similar genome sizes (e.g., λ: 48.5 kb, CrimD: 59.8 kb) would provide us with a way of testing the hypothesis that band smearing arises from tightly distributed viral genome sizes, and provide insights into how two similarly-sized phages affect the gel's final appearance. All of these experiments would provide us with an increased knowledge of what data may be lost while evaluating an environmental sample with PFGE.

Regarding environmental samples, the main experiment remaining is to collect a

very large sample of lake water (~48 L) and process it to obtain a viral concentrate of approximately 2 mL. In theory, this would provide us with six times the viruses in the concentrate compared to viral concentrates prepared from 8 L of lake water. When cast into plugs and run using PFGE, we would hopefully observe stronger and clearer banding patterns because of an increased number of each individual phage species. Another useful experiment would be to characterize the differences in gel quality between marine samples and freshwater samples for ourselves. If equal volumes of starting material were obtained from each environment, filtered and processed in identical fashions, and PFGE was run on the sample plugs, would we observe results similar to what has been reported in the literature?

To fully supplement the limitations of PFGE, other techniques will most likely need to be used to profile viral communities. TEM and RAPD-PCR are the most readily available options for use in lab. By utilizing these methods, we can draw comparisons between results and help build a more complete description of viral assemblages.

Closing Thoughts

Despite the shortcomings observed with PFGE for use with freshwater samples, the technique remains a powerful one. Since it is easy to use and highly reproducible, it is an invaluable tool for environmental virologists. Even though it may not provide a complete picture of a community, it is an excellent starting point – especially when observing how a community changes over time and space. The problems we have observed may not have easy solutions – or solutions at all – but by better knowing the limitations of pulsed field gel electrophoresis with specific application in profiling viral

assemblages, we have the ability to more accurately evaluate data we may obtain. Regardless, by combining PFGE with other profiling methods, we put ourselves in the best position to learn as much as possible about a given viral community.

REFERENCES

Ackermann, H. W., & DuBow, M. S. (1987). Viruses of prokaryotes: general properties of bacteriophages. CRC Press, Inc., Boca Raton, Fla.

Auguet, J. C., Montanie, H., & Lebaron, P. (2006). Structure of virioplankton in the Charente Estuary (France): Transmission electron microscopy versus pulsed field gel electrophoresis. *Microbial Ecology, 51*(2), 197-208.

Battin, T.J., Kaplan, L.A., Findlay, S., Hopkinson, C.S., Marti, E., Packman, A.I., et al. (2008). Biophysical controls on organic carbon fluxes in fluvial networks. *Nature Geoscience, 2*(8), 595-595.

Birren, B., & Lai, E. (1993). *Pulsed field gel electrophoresis: a practical guide*. San Diego, CA: Academic Press, Inc.

Breitbart, M., Felts, B., Kelley, S., Mahaffy, J., Nulton, J., Salamon, P., & Rohwer, F. (2004). Diversity and population structure of a near-shore marine-sediment viral community. *Proc. Biol. Sci., 271*(1539), 565-574.

Brüssow, H., Canchaya, C., & Hardt, W.D. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews, 68*(3), 560-+.

Burke, J., Schneider, D., & Westpheling, J. (2001). Generalized transduction in Streptomyces coelicolor. *Proc. Natl. Acad. Sci. USA, 98,* 6289-6294.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.L., & Brüssow, H. (2004). Phage-host

interaction: an ecological perspective. *Journal of Bacteriology, 186*(12), 3677-3686.

Filippini, M., & Middelboe, M. (2007). Viral abundance and genome size distribution in the sediment and water column of marine and freshwater ecosystems. *Fems Microbiology Ecology, 60*(3), 397-410.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature, 399*(6736), 541-548.

Hammer, Ø., Harper, D.A.T., & Ryan, P.D. (2001). PAST: paleontological statistics software package for education and data analysis. *Paleontologica Electronica 4*(1), 1-9.

Klieve, A.V., & Swain, R.A. (1993). Estimation of ruminal bacteriophage numbers by pulsed-field gel-electrophoresis and laser densitometry. *Applied and Environmental Microbiology, 59*(7), 2299-2303.

Larsen, A., Castberg, T., Sandaa, R.A., Brussaard, C.P.D., Egge, J., Heldal, M., et al. (2001). Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. *Marine Ecology-Progress Series, 221*, 47-57.

Leff, A.A., Leff, L.G., Lemke, M.J., Heath, R.T., & Gao, X.Q. (1999). Abundance of planktonic virus-like particles in Lake Erie subsurface waters. *Ohio Journal of Science, 99*(2), 16-18.

Maranger, R., & Bird, D.F. (1995). Viral abundance in aquatic systems - a comparison between marine and fresh-waters. *Marine Ecology-Progress Series, 121*(1-3), 217-226.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003). Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews, 67*(1), 86-+.

Move: methods for the isolation of viruses from environmental samples. (2005). Retrieved from http://www.virusecology.org/MOVE/Methods.html

Sambrook, J. (2001). *Molecular cloning: a laboratory manual*. Woodbury, NY: Cold Spring Harbor Laboratory Press.

Sinden, R.R., & Pettijohn, D.E. (1982). Torsional tension in intracellular bacteriophage T4 DNA: Evidence that a linear DNA duplex can be supercoiled in vivo. *Journal of Molecular Biology, 162*(3), 659-677.

Steward, G.F., & Azam, F. (2000). In "Microbial Biosystems: New Frontiers. Proceedings of the 8th International Symposium on Microbial Ecology" (ed Bell, C.R., Brylinski, M. and Johnson-Green, P.), Analysis of marine viral assemblages, Atlantic Canada Society for Microbial Ecology, Halifax

Steward, G. F. (2001). Fingerprinting viral assemblages by pulsed field gel electrophoresis (PFGE). *Methods in Microbiology, 30* (Vol. 30, pp. 85-103). San Diego: Academic Press Inc.

Steward, G. F., Montiel, J.L., & Azam, F. (2000). Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography, 45*(8), 1697-1706.

Suttle, C.A., Chan, A.M., & Cottrell, M.T. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Applied*

Environmental Microbiololgy, 57, 721-726

Suttle, C.A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology, 28*, 237-243

Tetart, F., Desplats, C., Kutateladze, M., Monod, C., Ackermann, H.W., & Krisch, H.M. (2001). Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *Journal of Bacteriology, 183*(1), 358-366.

Tijdens, M., Hoogveld, H.L., Kamst-van Agterveld, M.P., Simis, S.G.H., Baudoux, A.C., Laanbroek, H.J., et al. (2008). Population dynamics and diversity of viruses, bacteria and phytoplankton in a shallow eutrophic lake. *Microbial Ecology, 56*(1), 29-42.

Wang, M., & Lai, E. (1995). Pulsed field separation of large supercoiled and open-circular dnas and its application to bacterial artificial chromosome cloning. *Electrophoresis*, *16*(1), 1-7.

Wilhelm, S.W., & Matteson, A.R. (2008). Freshwater and marine virioplankton: a brief overview of commonalities and differences. *Freshwater Biology, 53*(6), 1076-1089.

Winget, D.M., & Wommack, K.E.. (2008). Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. *Applied and Environmental Microbiology, 74*(9), 2612-2618.

Wommack, K.E., & Colwell, R.R. (2000). Virioplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews, 64*(1), 69-+.

Wommack, K.E., Ravel, J., Hill, R.T., Chun, J.S., & Colwell, R.R. (1999). Population

dynamics of Chesapeake bay virioplankton: Total-community analysis by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology, 65*(1), 231-240.

Yan-Ming, L., Xiu-Ping, Y., Qi-Ya, Z. (2006). Spatial distribution and morphologic diversity of virioplankton in Lake Donghu, China. *Acta Oecologica, 29*(3), 328-334.