5-2011

The Temporal Dynamics of the Bacterial Communities in Lake Matoaka, an Eutrophic, Freshwater Lake

Dana Michelle Hardbower
College of William and Mary

The Temporal Dynamics of the Bacterial Communities in Lake Matoaka, an Eutrophic, Freshwater Lake

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

by

Dana Michelle Hardbower

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

_____
Dr. Kurt Williamson, Chair

_____
Dr. Mark Forsyth

_____
Dr. Randy Chambers

_____
Dr. Randy Coleman

Williamsburg, VA
May 3, 2011

Abstract

Little is known about microbial dynamics in freshwater ecosystems, particularly compared to marine systems. The taxonomy of many freshwater bacterial lineages has been determined, but the drivers of community composition dynamics have yet to be clearly determined. Given that community composition determines which important processes occur and the rate at which the occur, an increased understanding of the drivers of community composition change is necessary. The purpose of this study was to and to track changes in the profile of the bacterial species in Lake Matoaka, an eutrophic, freshwater lake, in response to the primary seasonal, environmental and biological drivers. Surface water samples were collected monthly over an 18 month period, along with environmental data (temp, pH, nutrients (P and N), chl-a). Viral and bacterial abundances were determined using epifluorescence microscopy. DNA was extracted from bacterial cells captured on a 0.22 μm filter and bacterial community profiles were generated using terminal restriction fragment length polymorphism (tRFLP). Profiles were compared by converting chromatograms to binary matrices based on the presences or absence of peaks. Pearson correlations and a BioEnv analysis were used to identify relationships between viral and bacterial abundance, viral and bacterial community composition and environmental factors. Bacterial communities were also characterized by sequencing of 16S rRNA clone libraries from two months, June 2009 and December 2009 (summer and winter), within the sample period. Distribution of T-RFs across the 18-month sampling period was analyzed by constructing cluster dendrograms, which suggested that there is no quantifiable seasonal/temporal trend in bacterial community composition. The only significant relationships revealed by correlation analysis were between bacterial and viral abundance, viral abundance and temperature, and

bacterial abundance and temperature. BioEnv analysis indicated that the only environmental factor with any explanatory power in accounting for community composition change was temperature. However, inspection of dendrograms revealed that this relationship is not strictly linear and sample months do not cluster neatly based on temperature fluctuations. Moreover, viral and bacterial community compositions were determined to be very tightly linked by a Mantel test. 16S Clone libraries revealed that several dominant phyla were present within the lake, the foremost of these being *Proteobacteria*, consistent with the findings of previous freshwater studies. The distribution of genera varied between the two seasons. The bacterial assemblage of Lake Matoaka appears to be composed of a core of *Proteobacteria* that are consistently present, with a variable component that changes significantly over the annual cycle. While temperature was identified as having moderate explanatory power, a clear driver for community composition change was not determined. Future studies that include other biological and environmental factors, such as protozoan grazing and dissolved organic carbon. Future studies should also include a survey of the dynamics of community composition over several lakes of highly similar characteristics in order to clearly define drivers of bacterial community composition change in freshwater ecosystems.

Acknowledgements

There are so many people to thank for their generous support that I hardly know where to begin. First and foremost, I would like to thank Dr. Williamson. He is the best of mentors. His guidance and support over the past three years have been critical to my success as a student and as a scientist. I am honored to say that I can not only call him advisor, but friend as well. There is no price that can be placed on my experience in his lab. I will cherish this time for the rest of my life. I will greatly miss Dr. Williamson's lab and all of the wonderful members of it. I would especially like to thank Dan Kiernan, Joe Kendra and Evguenia Orlova for their assistance in sample collection and processing. I would again like to thank Joe Kendra for his work in gathering environmental metadata from the lake, especially chlorophyll A data. I would like to extend much appreciation to Krysten Corzo, Camelia Drissi and Kimy Javier for gathering the abundance data. A special thank you to Jess Dolman and Dr. Williamson for gathering the viral richness data. Much thanks and gratitude goes to Dustin Glasner for his constant support and friendship, both in and out of the lab. My research experience would have been wanting without him. Special thanks to my committee, Dr. Williamson (chair), Dr. Forsyth, Dr. Coleman and Dr. Chambers. I greatly appreciate all the time and energy you put into assisting me with my thesis. Moreover, I would like to thank Mr. Harvey Chappell, Ms. Rhonda Craig, Mr. and Mrs. Stuart Ray, Ms. Margery Daughtrey and Mrs. Susan Vukick for their generous financial support of my research.

I would also like to thank Zac Elmore for offering advice and assistance when needed. I would like to thank Dr. Forsyth, again, for his unwavering and undeserved support throughout my college career. Your advice and support have been invaluable. I would like to thank Brooke Matson for her mutual understanding and constant support. And, lastly, I would like to thank my parents, Dianne and Byron, my brother, Mitchell and my sister, Beth. My life would be absolutely nothing without them. My family is my rock, my heart and my soul. I love you all very much and I would not be here today without you.

Table of Contents

List of Figures and Tables

Figure

Table

1. Introduction

1.1 – Bacteria in Marine and Freshwater Ecosystems

It has been long understood that bacteria represent a significant portion of the biomass in aquatic ecosystems and, to that end, are essential to ecosystem dynamics (Pomeroy, 1974; Fuhrman et al., 1993). Bacteria play crucial roles in most biogeochemical cycles, including, but not limited to, decomposition, mineralization of organic compounds and trophic coupling with higher order eukaryotes (Hiorns et al., 1997; Newton et al., 2011). The role of bacteria has been particularly well characterized in marine aquatic systems, especially the surface waters of the pelagic zones of the world's oceans (Giovannoni et al., 2005). With the advent of the "microbial loop" hypothesis for developing oceanic food webs, oceanic bacteria moved to the forefront of microbial ecology research (Azam et al., 1983; Giovannoni et al., 2005). At its core, the microbial loop is a semi-quantitative means of describing the movement of carbon from lower trophic levels to higher trophic levels (Azam et al., 1983). It recognizes that heterotrophic microbes depend on the release of dissolved organic matter from phytoplankton, the oceans' primary producers (Azam et al., 1983; Fuhrman, 1992). Organisms at higher trophic levels depend, in turn, on the relatively inefficient return of carbon from these heterotrophic microbes (Azam et al., 1983; Fuhrman 1993). Clearly, microbes – particularly bacteria – play a central role in aquatic food chains. With advances in techniques by which to study microbial diversity, bacterial communities have been described in increasing detail and phylogenies have been developed to characterize the bacterial species present in marine systems over the past few decades. The same, however, cannot be said of freshwater systems.

Despite their critical role in aquatic ecosystems, there is a surprising dearth of information regarding the bacterial communities in freshwater (Newton et al., 2011; Clasen et

al., 2008). The phylogentic insights that have been gained for marine systems have not been achieved for freshwater environments to the same degree (Hiorns et al., 1997). However, this trend is changing. Lakes have been recognized to play a major role in the global carbon cycle and as key indicators in global climate change (Cole et al., 2007). It is important to note that lakes can function as both carbon sources and carbon sinks and that lakes can be significant contributors to atmospheric carbon due to microbial respiration (Cole et al., 2001). Given the role of lakes in the global carbon cycle, the breakdown and release of carbon in lake systems becomes critical. Freshwater bacterial lineages are vital to these processes, thus increasing the need to study and understand bacteria in freshwater ecosystems.

Much of the carbon found in lake systems is allochthonous carbon from terrestrial systems that feed into lake systems via rivers, streams, and overland runoff (Jansson et al., 1999). Freshwater bacteria play a critical role in the breakdown of dissolved organic carbon (DOC) entering the lake from terrestrial sources, as well as the autochthonous carbon introduced by phytoplankton (Jansson et al., 1999; Cole et al., 2001). In fact, bacterial respiration of DOC can exceed autochthonous carbon production by primary producers in many lake systems (Cole et al., 2001). This link between terrestrial and freshwater ecosystems and the bacteria therein is critical to the global carbon cycle. Moreover, bacteria play a vital role in the breakdown of dead organic matter, such as algal blooms, and pollutants, such as fertilizers, pesticides and sewage (Akpor et al., 2010 ; Jin et al., 2010) Given the critical functions bacteria fulfill in aquatic systems, it is only logical to study these communities in earnest. The development of several molecular techniques has given microbial ecologists the ability to characterize bacterial communities in detail.

1.2 – Techniques for the Study of Bacterial Communities in Freshwaters

The continuum of techniques for analyzing environmental assemblages of bacteria ranges from culture-dependent methods to culture-independent methods. Culture-dependent methods are inherently limited as about only 1% of the bacterial species estimated to exist on Earth can be cultured in the laboratory. (Newton et al., 2011; Muyzer et al., 1993). Given this severe limitation, culture-independent methods are most often utilized to study bacterial diversity in a broad, complex ecosystem. Techniques commonly used in freshwater studies include denaturing gradient gel electrophoresis (DGGE), automated intergenic ribosomal spacer analysis (ARISA), terminal restriction fragment length polymorphism (tRFLP) and 16S rRNA clone libraries (Shutte et al., 2008). It is very important to note that, in spite of their differences, each of these methods depends on the universally-conserved bacterial 16S rRNA gene segments and polymerase chain reaction (PCR) amplification of the gene.

DGGE is based on the decreased electrophoretic mobility of a partially melted DNA molecule as compared to the mobility of the native DNA molecule when a denaturing agent is applied to the DNA. DNA fragments are classified by melting domains based on the point in the gradient gel at which the double stranded DNA is denatured into two single strands. Once a DNA fragment with the lowest melting point reaches said point, it halts its movement through the gel matrix. It is differences in sequence, especially in G+C content, of each fragment within the sample that determines melting point. These differences are represented by a banding pattern of all melted fragments that can be visualized on a polyacrilamide gel. This banding pattern serves as a proxy measure for bacterial richness and the presence and absence of bands over a sample period provides information about shifts in community composition. Analysis of the

patterns of presence and absence against other environmental data furnishes researchers with information about the drivers of community composition change. Because DGGE effectively separates different gene fragments based on differences in fragment sequence, specifically G+C content, direct analysis of the genomic DNA extracted from the gel can also be performed (Muyzer et al., 1993).

DGGE provides a precise degree of resolution and it allows for analysis of an entire community (Muyzer, 1999). However, DGGE has specific limitations that may interfere with microbial ecology studies. Firstly, DGGE requires a specific apparatus in which to perform the gel electrophoresis. Added to that, a high degree of precision and skill in casting and loading gels is critical. Without well developed DGGE gel loading skills, the technique simply does not work. Moreover, DGGE cannot resolve heteroduplexes and, is not sensitive enough to detect co-migration of fragments, which can arise from minute sequence differences and rare community members (Muyzer, 1999). In an ecology study over an annual cycle, these can represent severe limitations, leading researchers to try other techniques.

ARISA is another culture-independent method commonly used for the study of the bacterial community composition in complex ecosystems. It is based on the PCR amplification of the 16S-23S intergeneic region (Fisher et al., 1999). The ARISA-PCR fragments are then discriminated based on fragment size using an automated electrophoresis system, typically capillary electrophoresis (Fisher et al., 1999). The distribution of fragment sizes serves as a proxy for bacterial community richness, much like the banding patterns in DGGE. The same analysis of the presence and absence of fragments over a sample period can be used to track changes and determine drivers of community composition changes. This technique is rapid, relatively inexpensive and highly reproducible (Fisher et al., 1999). However, ARISA is subject

to PCR bias in that shorter templates are favorably amplified over longer fragments, resulting in an underestimation of diversity (Fisher et al., 1999). Moreover, the heterogeneity and overlapping regions of the intergenic spacer region may also result in underestimations of diversity (Fisher et al., 1999).

1.2.1 – Terminal Restriction Fragment Length Polymorphism

Terminal restriction fragment length polymorphism analysis (tRFLP) is a molecular fingerprinting technique that has proven especially useful in the study of complex bacterial communities (Osborne et al, 2006). A culture-independent technique, tRFLP allows microbial ecologists to track community composition of all operational taxonomic units (OTU) – bacteria with 16S rRNA gene sequences that are ≥ 97% similar to each other (Blackwood et al., 2003). The term "operational taxonomic unit" is used in studies of bacterial diversity because it more accurately identifies true differences between identified bacteria. With the prevalence of genetic mutation and horizontal gene transfer, and the much-outdated and inconsistent nomenclature used in microbiology, the term "species" is simply not specific enough to identify true differences between individual bacteria. Bacteria that can be cultured in a laboratory setting represent only a fraction of the microbial diversity in any environment (Blackwood et al., 2003). It is critical for a complete survey of bacterial diversity in a complex ecosystem – such as a freshwater system – that the number of OTU's represented in the samples populations matches the actual number of OTU's present in the ecosystem. tRFLP, DGGE and ARISA all allow for this to a certain extent.

When applied to bacterial communities, tRFLP is based on PCR amplification of the 16S rRNA gene representing different bacteria from environmental nucleic acid extracts (Leuders et

al., 2003). The presence of universally conserved sequences within bacterial 16S rRNA genes is critical for the success of this technique. A fluorescently tagged primer is used in each PCR reaction to ensure each amplicon is terminally labeled for downstream applications. The amplicons are then subjected to a restriction digest, resulting in many gene fragments of which only the terminal fragment is labeled. Capillary electrophoresis allows for easy detection of these terminally labeled fragments and high-resolution sizing, down to 1bp differences between terminal restriction fragments (T-RFs). This level of resolution provides a distinct advantage over DGGE. The chromatogram output provides information about each T-RF, specifically, base-pair length of terminal fragments in the sample and relative abundance of each fragment. It is important to note, however, that tRFLP itself does not provide any information as to the identity of the bacterial OTU(s) that comprise each peak.

That having been said, tRFLP does offer several advantages as a molecular fingerprinting technique in complex bacterial environments. tRFLP is a very efficient, high-throughput and highly reproducible method of relative mechanistic simplicity – much like ARISA (Shutte et al., 2008). It provides enough information to track changes in community composition over time or space or both (Osborne et al., 2006). While it is subject to potential homology of terminal fragment sequence and length, tRFLP provides a degree of resolution that accounts for even the rarest OTU's (provided there is adequate PCR amplification of rarest templates) in an ecosystem. Given the availability of universal 16S rRNA primers, tRFLP allows ecologists to account for the vast majority of bacterial diversity (Osborne et al., 2006). Moreover, the availability of equipment and fragment analysis expertise at the College of William and Mary make tRFLP the logical choice for use as a molecular fingerprinting technique for this particular study.

While treatment of chromatogram data has not been completely standardized across individual research groups, particularly in terms of statistical analyses, tRFLP is a powerful and widely used technique for the exploration of changes in community composition in complex bacterial communities (Blackwood et al., 2003; Osborne et al., 2006; Shutte et al., 2008). It is especially useful for assessing community level change due to spatial and temporal factors, or along gradients such as trophic status.

1.2.2 – 16S rRNA Clone Libraries

Like tRFLP, the construction of clone libraries is not dependent on the ability to culture all bacterial OTUs from a complex community in a laboratory setting. However, unlike tRFLP, 16S rRNA clone libraries enable researchers to identify specific microbial taxa in a sample, often down to the species level. Because of the universality of the 16S rRNA gene, the availability of 16S rRNA primers and the advances in sequencing technology, individual OTUs from a mixed environmental sample can be identified, classified and characterized (Singleton et al., 2001). Thus, while tRFLP allows for comparative analyses of changes in community composition across time or space, analysis of 16S gene clone libraries allows researchers to understand exactly how two communities differ in terms of the representation and distribution of OTUs in each library. In short, clone libraries provide the most specific and detailed information about bacterial OTUs from environmental samples.

Furthermore, recognizing the importance of the 16S rRNA gene and the need for bioinformatic tools for analyzing the growing pool of sequence data originating from 16S clone libraries, the GreenGenes database (DeSantis et al., 2006) was developed for the specific identification of bacteria based on 16S sequence data. GreenGenes selects the most likely

taxonomic identity of each clone based on nucleotide sequence homology with sequences found in the BLAST database (DeSantis et al., 2006). This is especially useful in environmental studies given that the vast majority of environmental bacteria cannot be cultured in a laboratory setting.

However, 16S rRNA clone libraries have certain limitations. Clone libraries are highly labor intensive and costly mechanization is required to make the technique high throughput (Fisher et al., 1999). Thousands of clones are needed in each library to capture a full picture of bacterial diversity. This requirement is difficult to meet for several reasons. The time, computational ability and resources available on the part of the researcher may not be sufficient for many libraries to be constructed and analyzed. And the creation of many libraries dictates that a high throughput method of analysis be used. This prevents the researcher from being able to manually audit the resulting sequence data. Either case holds the potential for an incorrect estimation of bacterial diversity to be made. Moreover, the highly detailed output of the sequence data – this technique's greatest strength – can also be its greatest weakness. Given the fact that most aquatic and soil bacterial OTUs are uncharacterized, BLAST results often do not allow for taxonomic identification of clones beyond the genus level, as many clones share the highest homology with other uncultured (and therefore, unidentified) bacteria in the BLAST database. When this is the case, sequence identification provides little information for researchers to use to characterize the bacterial community in an ecosystem. However, when taxonomic information is available, it greatly illuminates the microbial composition of a freshwater ecosystem. 16S rRNA clone libraries, when used in combination with other techniques, are a very effective tool for the exploration of bacterial community dynamics.

1.2.3 – In Silico Digestions

While each of the aforementioned techniques can easily stand on their own in addressing questions of bacterial diversity, they do so in very different ways. It would be most helpful to microbial ecologists to be able to combine the quantifiable changes in community composition from tRFLP and the specific OTU identities from clone libraries. This would create a more holistic picture of the actual community diversity in an ecosystem. Fortunately, this feat can be accomplished through *in silico* digestions of sequence data. A specially created program – TRiFLe – allows ecologists to input sequence data and subject that sequence to a restriction digest based on the known restriction sites of the chosen enzymes (Junier et al., 2008). The output is a chromatogram with a single theoretical T-RF based on the predicted terminal fragment size (Junier et al., 2008). This program can potentially be used to identify specific T-RFs in tRFLP patterns by comparing theoretical and experimental results (Junier et al., 2008). Given this information, microbial ecologists could potentially identify dominant OTUs in an ecosystem based only on tRFLP data, and identify specifically how these OTUs vary in given conditions – time, space, temperature, etc.

1.3 – Temporal Dynamics of Freshwater Bacterial Communities

Since freshwater lakes and their resident bacterial populations were recognized for their importance, dozens of studies have been conducted to better understand the microbial communities therein. Taxonomic characterization of freshwater bacterial communities has been the historical focus of the field, and continues to be a primary focus today. Many studies have been conducted using 16S rRNA clone libraries and various fingerprinting techniques to identify

freshwater lake OTUs (Hiorns et al., 1997; Lindstrom et al., 1998; Zwart et al., 1998; Lindstrom, 2000; Zwart et al., 2002; Bel'kova et al., 2003; Comte et al., 2006; Newton et al., 2011). Because of its ability to provide genomic DNA for direct sequencing, DGGE was the fingerprinting method of choice for many of these studies (Lindstrom et al., 1998; Lindstrom, 2000; Allgaier et al., 2006). Still, many taxa remain unidentified and the effort continues to characterize the diversity of freshwater bacteria. It was not until recent years that studies began focusing on the drivers of community composition change. DGGE and tRFLP fingerprinting techniques were utilized to track composition changes over time, space or both (Lindstrom et al., 2005; Boucher et al., 2006; Lymer et al., 2008). tRFLP became especially popular in tracking community composition change because of its easily produced results and the fact that taxonomic identifications from genomic DNA were not necessarily needed to track community changes (Lymer et al., 2008; Boucher et al., 2006). Despite the increase of studies of this nature, the drivers of bacterial community composition change in freshwater lakes have yet to be identified.

The purpose of my project is to gather more information about freshwater bacterial community dynamics using Lake Matoaka as a model system. Monthly samples were collected over the span of 18 months for the purpose of determining the abundance and diversity (richness) of the bacterial community within Lake Matoaka. Moreover, environmental metadata were collected monthly in an attempt to identify relationships between environmental factors and community composition change. These environmental data include viral and bacterial abundance, viral richness, water temperature, pH, nutrient content (nitrogen and phosphorus) and chlorophyll *a* content. These data were be used to identify the most important temporal drivers of bacterial community composition changes.

2. Material and Methods

2.1 – Sample Site

The site of study is Lake Matoaka, an eutrophic freshwater lake located on the campus of the College of William and Mary in Williamsburg, VA. All samples were collected from the Keck Pier, located just below the Keck Laboratory at a site approximately half way between the lake's main inlet and outlet.

2.2 – Environmental Sample Collection and Processing

Surface water samples (6 L) were collected from the sample site once a month from May 2009 through November 2010, excluding January 2010 and May 2010. The water was pre-filtered through a 10 μm capsule filter (GeoTech Environmental) to remove debris, followed by a series of 142 mm diameter nylon filters [5μm, 1 μm (Pall Corporation) and 0.22 μm (Millipore)] to remove protozoa, zooplankton, phytoplankton and bacterioplankton. Bacterial cells were captured on the 0.22 μm filter. Filters were stored at -80 °C until processed for DNA extraction.

2.3 – Abundance Data Collection

To determine monthly viral and bacterial abundance, epifluorescence microscopy was used. Abundance data was obtained from unfiltered water samples. Aliquots (100-200 μl) of raw water were loaded onto a 25 mm diameter, 0.02 μm pore-size Anodisc (Whatman) and drawn through the filter under vacuum. The Anodiscs were stained with 100 μl of 2.5x SYBR Gold (Invitrogen) in the dark for 15 minutes. After drawing off the stain, the Anodiscs were air-dried and mounted onto a glass 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).

Bacterial cells and 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 raw water sample using a Hamamatsu CCD (charge-coupled device) camera. Bacterial cells and virus particles were enumerated using the Metamorph Basic software package (Molecular Devices). Bacterial and viral abundance was determined based on the average number of particles in the 10 images, the total volume of sample loaded onto the Anodisc and the area of the image observed.

2.4 – Water Chemistry Data Collection

2.4.1 – Temperature

Temperature was measured using a standard mercury thermometer at the time samples were taken.

2.4.2 – pH

pH was measured from a separate 50 ml raw water sample drawn at the time of sampling. pH was measured using an *Ultra*Basic pH meter (Denver Instruments) with 4.0 and 7.0 pH standards.

2.4.3 – Nutrients (P and N)

A Separate 500 ml sample was collected in an opaque bottle at the time of sampling for nutrient analysis. The levels of dissolved inorganic phosphorus and nitrogen (NO_2 + NO_3 and NH_4) in the water were determined by the Keck Laboratory on the campus of the College of William and Mary.

2.4.4 – Chlorophyll A and Chlorophyll B

A separate 500 ml sample was collected in an opaque bottle at the time of sampling for determination of chlorophyll A and chlorophyll B levels in the water. The water was filtered through an APFF 47 mm glass fiber filter (GFF, Millipore) using a Millipore filter tower under low vacuum. The GFF was placed in a 15 ml conical tube with 15 ml of 90% acetone and covered in aluminum foil. The tube was vortexed at maximum speed for one minute. The GFF was incubated in the acetone overnight at 4 °C in the dark to extract phytopigments. The tube was then centrifuged at maximum speed for 10 minutes at room temperature. A sample of the supernatant was decanted into a spectrophotometer cuvette and absorbance was measured at 750 nm, 664 nm, 647 nm and 630 nm.

2.4.5 – Rainfall

Monthly rainfall data was collected from the Williamsburg Government website: www.williamsburgva.gov/Index.aspx?page=242.

2.5 – Bacterial DNA Extraction

To extract the bacterial genomic DNA from the bacterial cells captured on the 0.22 μm filter, a protocol combining physical and chemical extraction methodologies was used (Rusch et al., 2007). The filters were cut into quarters. Two of the four quarters were cut into small pieces, divided among two 50 ml conical tubes and covered with 10 ml of DNA extraction buffer (50 mM EDTA, pH 8.0; 50 mM EGTA, pH 8.0). Lysozyme (MP Biomedicals) was added to a final concentration of 2.5 mg ml^{-1} and the samples were placed in a shaker incubator for 30 minutes at 250 rpm and 37 °C. The samples were then transferred to a static incubator for 30 minutes at 37

°C. Proteinase K (Ameresco) was then added to a concentration of greater than 200 μg ml^{-1}. The samples were then frozen in N_2 (*l*) for five minutes and thawed completely in a 55 °C water bath. The freeze/thaw procedure was repeated once. More Proteinase K was added to a concentration greater than 200 μg ml^{-1} and SDS was added to a final concentration of 1% (0.1 g SDS). The samples were incubated in a 55 °C water bath for two hours.

After incubation, the liquid was removed from the sample tubes and placed into new 50 ml conical tubes. Transferring any of the small filter pieces to the new tubes was carefully avoided. A 1/10 volume (1 ml) of 3 M sodium acetate and an equal volume (10 ml) of Buffer Saturated Phenol (Invitrogen) were added to the samples. The samples were placed in a swing bucket centrifuge for 15 minutes at 4750 rpm and 22 °C. The supernatant was removed and placed in a 50 ml Oak Ridge tube. An equal volume of isopropanol (10 ml) and 5 μl of GlycoBlue™ (Ambion) were added to the samples. The samples were incubated at 4 °C for one hour. The samples were then placed in an SS-34 fixed angle rotor centrifuge for 30 minutes at 7500 rpm and 14 °C. The isopropanol was decanted and discarded. The pellet was resuspended in 500 μl of 1x TE Buffer (10 mM Tris, pH 8.0; 1 mM EDTA, pH 8.0; 988 ml diH_2O) and transferred to a 1.5 ml Eppendorf tube.

A 1/10 volume (50 μl) of 3 M sodium acetate and an equal volume (500 μl) of Buffer Saturated Phenol (Invitrogen) were added to the samples. The samples were placed into a benchtop centrifuge for 15 minutes at 14,000 rpm and 22 °C. The supernatant was removed and placed in a new 1.5 ml Eppendorf tube. An equal volume (500 μl) of phenol:chloroform:isoamyl alcohol (Fischer) was added to the samples. The samples were placed into a benchtop centrifuge for 15 minutes at 14,000 rpm and 22 °C. The supernatant was removed and placed in a new 1.5 ml Eppendorf tube. An equal volume (500 μl) of isopropanol and 1 μl of GlycoBlue™ were

added to the samples. The samples were placed into a benchtop centrifuge for 30 minutes at 14,000 rpm and 22 °C. The isopropanol was aspirated. The DNA pellets were allowed to air dry for 15 minutes and were then resuspended in 50 µl of 1x TE Buffer. The samples were combined for a final volume of 100 µl. The DNA was quantified using a NanoDrop® 1000 Spectrophotometer (Thermo Scientific). DNA was stored at -20 °C until needed.

2.6 – 16S rRNA PCR and Purification of PCR Products

A partial fragment of the 16S rRNA gene was amplified by polymerase chain reaction (PCR) for each sample. Primers were obtained from Integrated DNA Technologies, and include 519r (5'-GWATTACCGCGGCKGCTG-3'), 27f (5'-AGAGTTTGATCCTGGCTCAG-3') for amplicons to be used in clone libraries (from June 2010 and December 2010) and 27f-HEX (5'-HEX-AGAGTTTGATCCTGGCTCAG-3') for amplicons to be used for tRFLP (from May 2009 to November 2010, excluding January 2010 and May 2010). Template DNA was added from environmental samples at dilutions of 10^{0} to 10^{-3}. The PCR products were then run on a 0.8% agarose gel (0.8 g agarose; 500 ml 1x TAE buffer) in 1x TAE Buffer (40 mM Tris-acetate, 1 mM EDTA) at 90 V for 1 h. Following the run, gels were stained in 400 ml 1x TAE buffer with 25 µl of 10,000x SYBR Gold (Invitrogen) in the dark at 22 °C for one hour. Gels were visualized using a Kodak Gel Logic 100 Imaging System with Kodak Molecular Imaging Software. When the 16S rRNA PCR was confirmed success, the PCR products were column purified using a QIAquick® PCR Purification Kit (Qiagen). The final volume of the column-purified PCR product was 50 µl.

2.7 – Removal of Terminal Overhangs

To remove terminal overhangs on HEX-labeled amplicons prior to tRFLP analysis, amplicons were treated with 1 µl of mung bean nuclease (New England BioLabs) buffered in 5 µl of the corresponding 10x reaction buffer for 30 minutes at 37 °C. The mung bean nuclease was removed by adding 50 µl of phenol:chloroform:isoamyl alcohol (Fischer) and placing the samples in a benchtop centrifuge at 14,000 rpm and 22 °C for five minutes. The supernatant was placed in a new 1.5 ml Eppendorf tube. A 1/10 volume (5 µl) of 3 M sodium acetate, two volumes (100 µl) of 100% ethanol and 1 µl of GlycoBlue™ were added to the samples. The samples were placed in a benchtop centrifuge at 14,000 rpm and 22 °C for 15 minutes. The ethanol was aspirated and the DNA pellet allowed to air dry for five minutes before resuspending in a final volume of 60 µl of sterile diH_2O.

2.8 – Restriction Enzyme Digestion of PCR Products for tRFLP

The HEX-labeled amplicons to be used in tRFLP were split into subsamples of 20 µl in separate 1.5 ml Eppendorf tubes. To each tube, 2 µl of 10x NEBuffer 4 (New England BioLabs) was added. To one of the two subsamples 25 U of HinfI (5'-GANTC-3', New England BioLabs) was added. 50 U of MspI (5'-CCGG-3', New England BioLabs) was added to the other subsample. The samples were incubated at 37 °C for 18 hours. The restriction enzymes were removed using the phenol:chloroform:isoamyl alcohol extraction protocol as described in section 2.7. The pellet was resuspended in a final volume of 20 µl of sterile diH_2O.

2.9 – Capillary Electrophoresis and tRFLP Analysis

A mixture of 12 µl of HI-DI Formamide and 0.25 µl of Rox500® (GeneScan) size standard was added to the restriction fragments. The fragments were then loaded in a 96-well plate for fragment analysis in an ABI3130 Avant Gene Sequencer (Applied Biosystems) with Pop7 polymer (Applied Biosystems). Each run was performed in a 36cm capillary with Dye set DS-30 (ROX, 6FAM, HEX, NED; Applied Biosystems) as a matrix standard. The resulting chromatograms from the above gene fragment analysis were analyzed using GeneMapper® Software Version 4.0 (Applied Biosystems). Replicate tRFLP runs were aligned by manual inspection of peak shift. Peak thresholds were set using a minimum peak height of 50 FU (fluorescence units). The light smoothing algorithm available in the GeneMapper software was applied to the peaks to remove any background noise or artifacts of peak detection. Peak data from each chromatogram were then converted to presence/absence matrices.

2.10 – Creation of Clone Libraries for Sequencing

Non-HEX labeled, purified PCR amplicons were ligated into pCR®4-TOPO Cloning® vectors and transformed into One Shot® Mach1™-T1^{R} Competent *E. coli* cells according the protocol provided in the TOPO TA Cloning® Kit for Sequencing (Invitrogen). Cells were cultured on LB plates (5.0 g tryptone; 2.5 g yeast extract; 5.0 g NaCl (s); 7.5 g agar; 475 ml diH_2O) with 1.5x kanamycin (15 µg kanamycin per liter of agar, added to the liquid media after autoclaving and before plating) for 18 hours at 37 °C. Colonies indicating successful transformation were picked with a sterile loop into 2 ml of LB broth (5.0 g tryptone; 2.5 g yeast extract; 5.0 g NaCl (s); 475 ml diH_2O) with 1.5x kanamycin. The cells were cultured at 37 °C for 18 hours. Plasmid DNA was extracted from the bacterial cell cultures using a QIAprep® Spin

Miniprep Kit (Qiagen). DNA was quantified using a NanoDrop® 1000 Spectrophotometer (Thermo Scientific).

2.11 – Sequencing

Extracted plasmids were sequenced by Sanger Sequencing using an ABI3130 Avant Gene Sequencer (Applied Biosystems). The M13f (-20) primer (5'-GTAAAACGACGGCCAGT-3') (Invitrogen) was used in each sequencing reaction. Approximately 200 to 500 µl of PCR product was used in each sequencing reaction. Each sequencing reaction was performed in a 36 mm capillary with a BigDye® Terminator Kit (Applied Biosystems).

2.12 – Sequence Analysis

DNA sequences were trimmed of vector using GeneSequencer® Software Version 1.0 (Applied Biosystems™) and converted into FASTA files. Multi-FASTA files were created using DNA Baser® Version 3.0.36 (Heracle BioSoft™). Sequences were identified using the NAST align and compare tools available from the GreenGenes 16S rRNA Gene Database (DeSantis et al. 2006; DeSantis et al. 2006).

2.13 – In Silico Digestion

Using a combination of TRiFLe (Junier et al., 2008) and NebCutter (Vincze et al., 2003), the sequences obtained from the clones libraries were subjected to an *in silico* digestion in order to predict the specific T-RFs that would be produced by each clone.

2.14 – Statistical Analyses

All environmental data were log-transformed to obtain normal distributions. To test for relationships between bacterial richness, bacterial abundance, and environmental data, Pearson's correlation analysis was performed in Prism 5 (GraphPad Software). The presence/absence matrix data for the HinfI and MspI tRFLP profiles were converted into similarity matrices (Dice similarity coefficient), and subsequently, cluster dendrograms using the unweighted pair group method with arithmetic means (UPGMA) in PAST (Hammer et al., 2001). BioEnv tests were performed using the vegan package in R (Version 2.12.1) to identify potential environmental factors that best explain shifts in tRFLP banding patterns. To test for co-variance between bacterial (HinfI and MspI profiles) and viral community profile shifts, the Mantel statistic was computed in R (Version 2.12.1), using Spearman rank correlation and 9,999 random permutations.

3. Results

3.1 – Water Chemistry

The temperature of Lake Matoaka was highest in August 2009 and September 2010 and lowest in December 2009 and February 2010, when a thin layer of ice covered shallower sections of the lake (Table 1). Chlorophyll A and B levels were highest in the summer months, with maxima in August 2009 and July 2010. This corresponded with the abundant algal growth in the summer months in the lake. Levels of dissolved nitrogen compounds also appeared to follow a relatively cyclical trend and were highest in August and September of 2009 and October and November of 2010. Dissolved inorganic phosphate levels varied only slightly over the sample period, averaging 0.73 μmol L^{-1}. pH also remained relatively constant over the 18-month observation period, with an average value of 7.45, just above neutral.

3.2 – Bacterial and Viral Abundance

Bacterial and viral abundance followed a relatively cyclical trend (Figure 1). Viral abundance was consistently higher than bacterial abundance by a factor of approximately 30 to 1. Both viral and bacterial abundance were highest in the summer months and lowest in the winter months. Abundance decreased drastically between the summer of 2009 and the winter of 2009. However, the pattern was not seen in abundance trends between the summer of 2010 and the winter of 2010. It is difficult to account for the differences in abundance between the two years.

Table 1. Environmental data from each monthly sample

Month (year)	Chl A (µg/L)	Chl B (µg/L)	Temp	pH	NOx (µmol/L)	NH4 (µmol/L)	DIP (µmol/L)	Rain (30d)
April 09	0.755895	0	18.5	8	ND	ND	ND	ND
May 09	ND	ND	ND	ND	5.770	3.045	1.477	ND
June 09	9.18645	6.1836	27	7.2	1.103	0.000	0.598	5.31
July 09	15.4416	3.3765	27.5	7.87	0.000	0.250	0.000	3.54
August 09	69.1515	69.28875	29.5	7.88	23.246	0.000	0.846	7.54
September 09	3.69705	2.45565	25	7.83	27.979	0.000	0.801	3.94
October 09	-7.9296	8.57655	21	8.01	ND	ND	ND	6.77
November 09	1.54305	-25.04985	13.5	7.07	4.454	0.294	0.935	3.01
December 09	6.73365	3.1578	8.5	7.19	6.310	2.842	0.534	7.3
February 10	12.1044	9.9513	5	7.43	13.4596	0.279	0.561	4.97
March 10	9.98415	1.2204	11.5	7.56	18.392	0	1.071	1.05
April 10	3.9015	2.17845	18.2	7.43	7.0642	0	0.408	6.14
June 10	13.7718	17.26575	26.5	7.04	4.3472	2.046	0.306	2.9
July 10	20.7984	0.65745	27	7.14	3.135	15.159	0.459	1.84
August 10	15.9915	1.0092	27	6.96	4.8906	9.765	0.357	3.2
September 10	10.7988	1.5999	30	7.59	1.2958	10.323	0.408	2.34
October 10	10.7988	1.5999	22	7.1	29.76	77.568	1.7723	14.19
November 10	0.3204	-0.24645	15	7.36	54.312	6.144	1.1496	4.81

Data points include chlorophyll A (Chl A), chlorophyll B (Chl B), temperature (temp), pH, NO_3^- and NO_2^- (NOx), ammonia (NH4), dissolved inorganic phosphate (DIP) and rainfall (over 30 days prior to sampling). Some data points were not determined (ND).

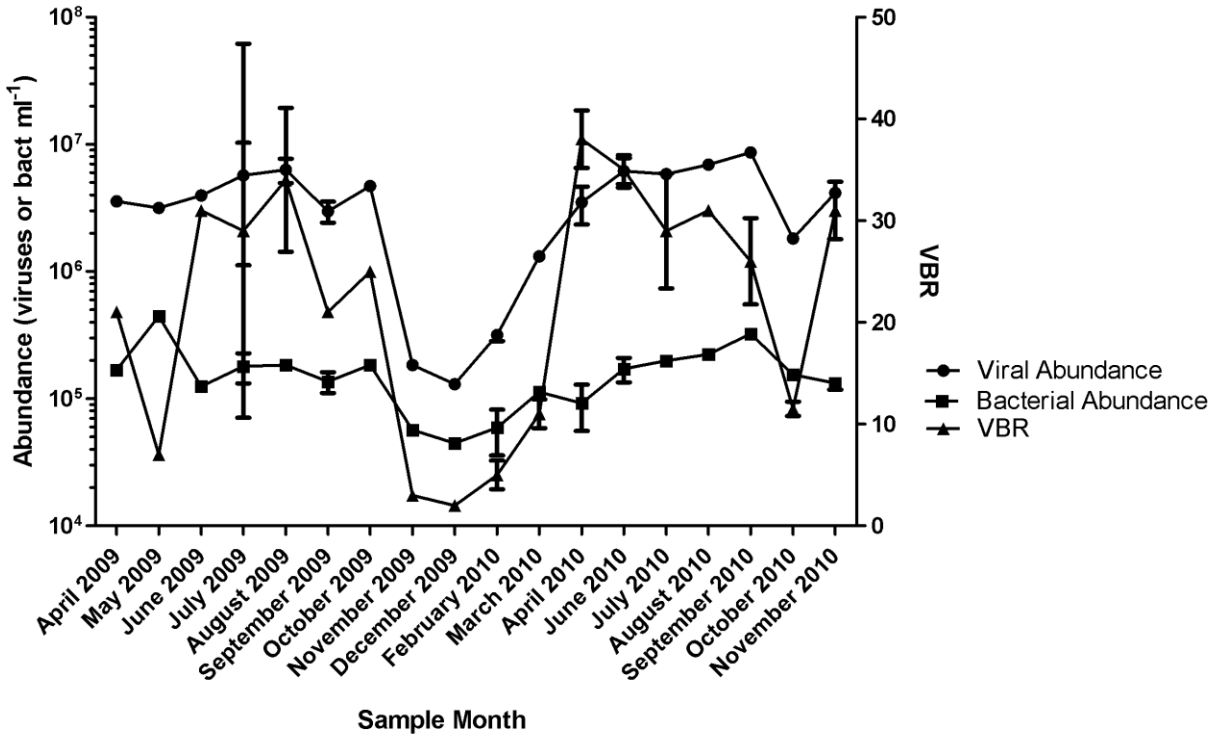

Figure 1. The abundance of viral particles and bacterial cells in Lake Matoaka over an eighteen month cycle. VBR represents the ratio of virus particles to bacteria.

3.3 – Bacterial Community Composition

Bacterial community composition was characterized by two different techniques: 1) tRFLP was used to generate a profile of the bacterial community in the lake and to track changes over time; and 2) 16S rRNA gene clone libraries were used to gather more specific information about the dominant OTUs in Lake Matoaka.

3.3.1 – Terminal Restriction Fragment Length Polymorphism

tRFLP analysis of the bacterial community composition in Lake Matoaka was performed using two different restriction enzymes – HinfI and MspI. tRFLP analysis was carried out for 18 months worth of samples, generating a total of 40 individual chromatograms (Figure 2). The reproducibility of the tRFLP data was measured by running five different samples in duplicate using the HinfI enzyme. Cluster dendrograms showed that tRFLP performed on replicate restriction digests produced T-RF distributions that were ≥ 90% similar (Figure 3). This indicated that the tRFLP patterns of this study were highly reproducible and lends much strength to the validity of trends observed in the tRFLP data.

Overall, both sets of cluster analyses (Figures 3 and 4) displayed no discernible trend for the similarity and dissimilarity of the samples from each month. The most similar samples in the HinfI profile, apart from the replicate runs, were July 2010 and October 2009. The peak distributions for these samples were about 65% similar. October 2010 appeared to be an outlier, with approximately 25% similarity with any other sample. This can perhaps be explained by the elevated rainfall (14.19 in) in that month. In the MspI cluster analysis, the most similar months were December 2009 and February 2010 at about 66% similarity. The next most similar months were November 2010 and September 2010 at about 60% similarity. The cluster analysis for

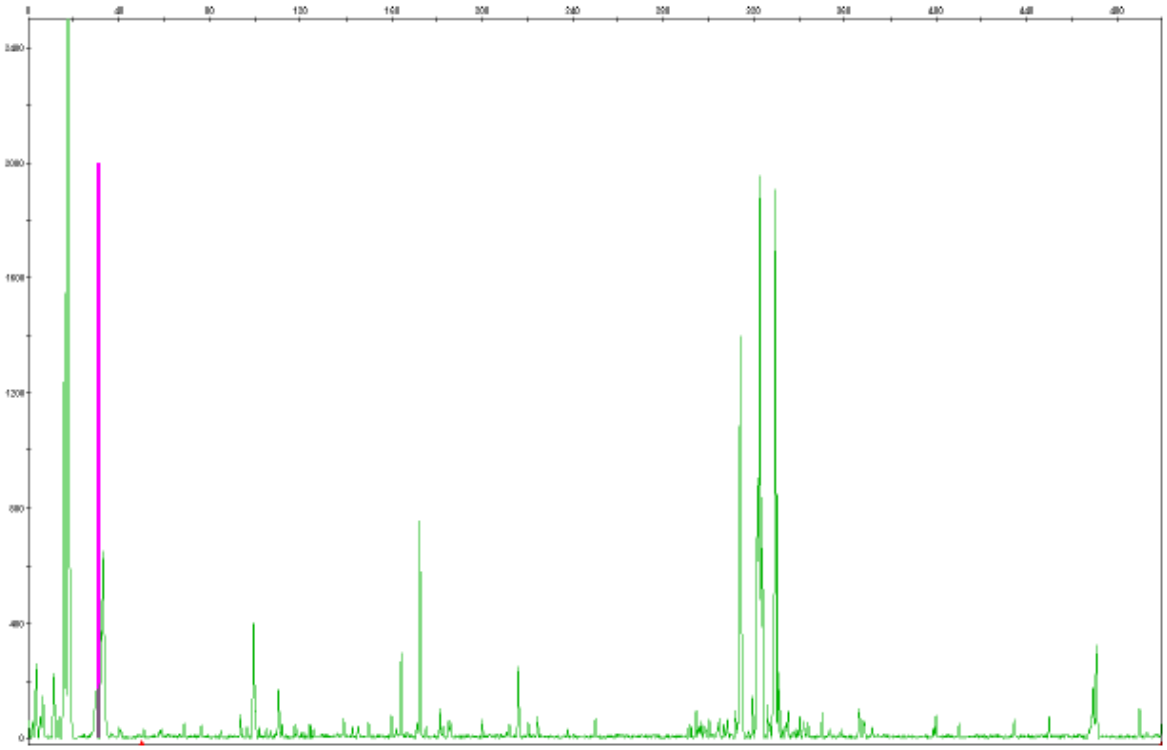

Figure 2. Sample chromatogram after tRFLP. The x-axis is base pair length. The y-axis is arbitrary fluorescence units. The pink bar is the result of primer dimerization. No peaks are considered a part of the data set if they are outside 50 to 500 base pairs, as denoted by the small, red triangles at the bottom of the graph.

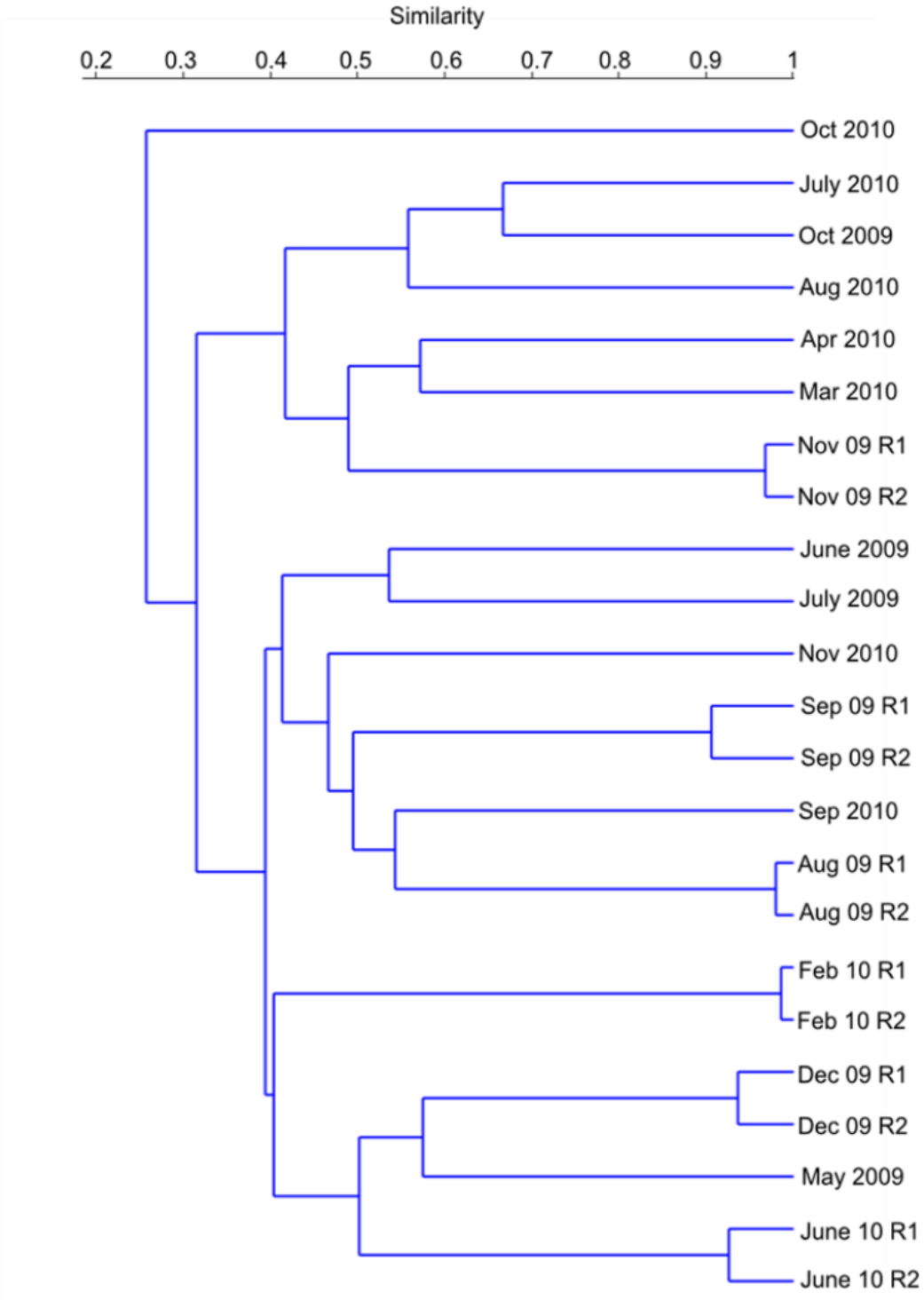

Figure 3. Resulting dendrogram of similarity matrix data obtained by cluster analysis using the Dice method for the HinfI tRFLP data from May 2009 to November 2010, including replicate tRFLP runs.

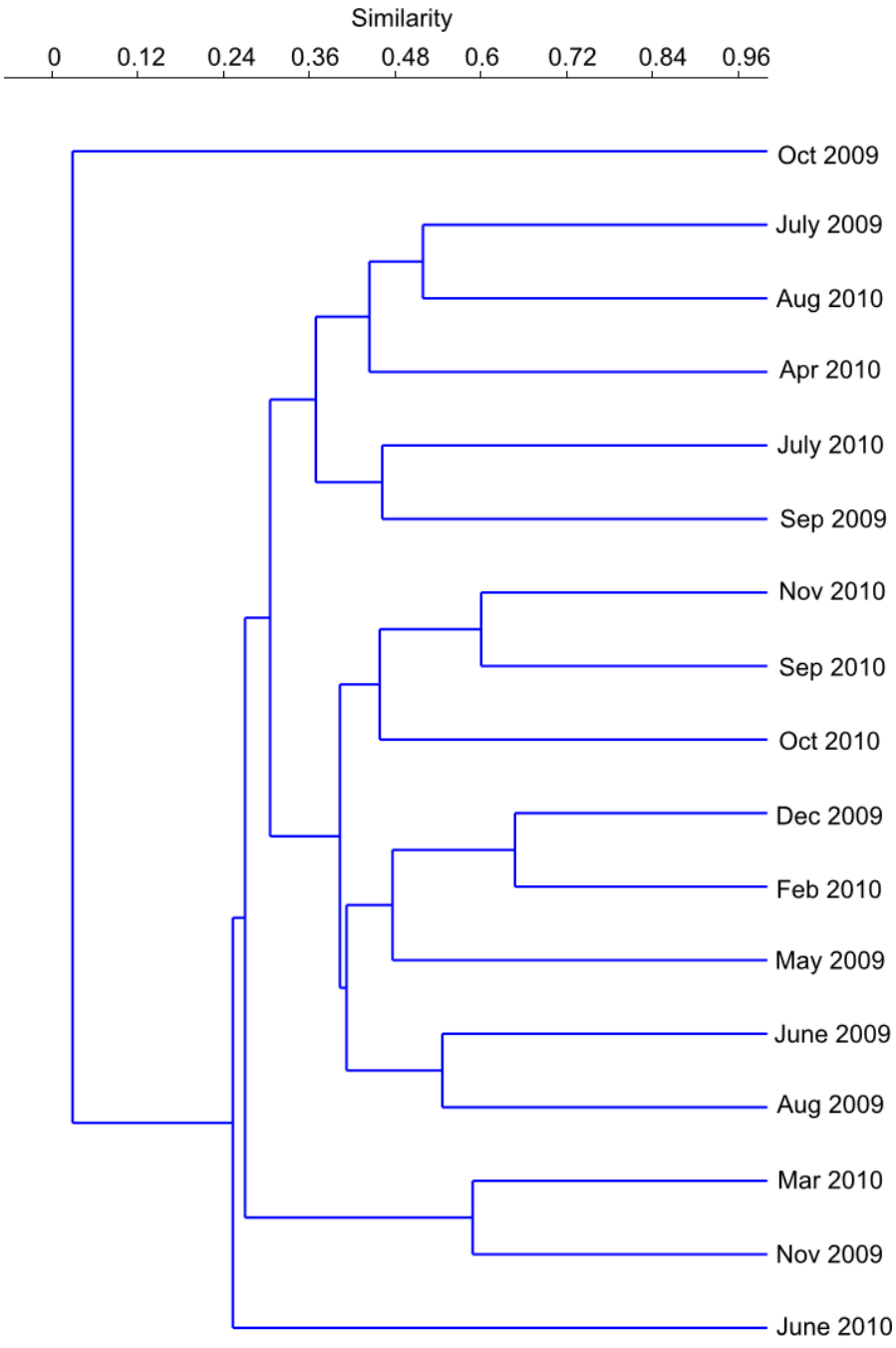

Figure 4. Resulting dendrogram of similarity matrix data obtained by cluster analysis using the Dice method for the MspI tRFLP data from May 2009 to November 2010.

the MspI data displays some grouping based on season. The winter months of December 2009 and February 2010 showed about a 64% similarity. The fall months of September 2010 and November 2010 showed about a 60% similarity. And, the summer months of June 2009 and August 2009 showed about a 54% similarity. However, there is no quantifiable seasonal pattern to community composition. October 2009 was an outlier in this data set, displaying approximately 5% similarity with any other sample. It is difficult to determine the cause of this as there were no corresponding changes in rainfall, temperature, pH, bacterial richness or other measured environmental factors during that month.

3.3.2 – 16S rRNA Clone Libraries

The clone libraries for June 2009 and December 2009 consist of 48 clones each. The partial fragment of the 16S rRNA gene of each of the 96 clones was sequenced. Community composition proved to vary between the two months (Figure 5). Overall, the most common phylum identified in the lake was that of *Proteobacteria*, with 40% and 43% of the total phyla composition in June and December, respectively, identified as *Proteobacteria*. In June 2009, the next most common phylum was *Actinobacteria*, making up 35% of the total identified clones. In December 2009, the next highest phylum was *Bacteroidetes*, making up 21% of the total identified clones. While the June library was dominated by two different phyla, the December library demonstrated a greater variety of phyla with a more even distribution. This trend continues even at the genus level.

There were 20 different genera identified in the December library, as compared with 17 for the June library (Figures 7 and 6, respectively). The dominant genus in the June library was *Tetrasphera*, comprising 31% of the genera identified. *Tetrasphera* displayed clear dominance at

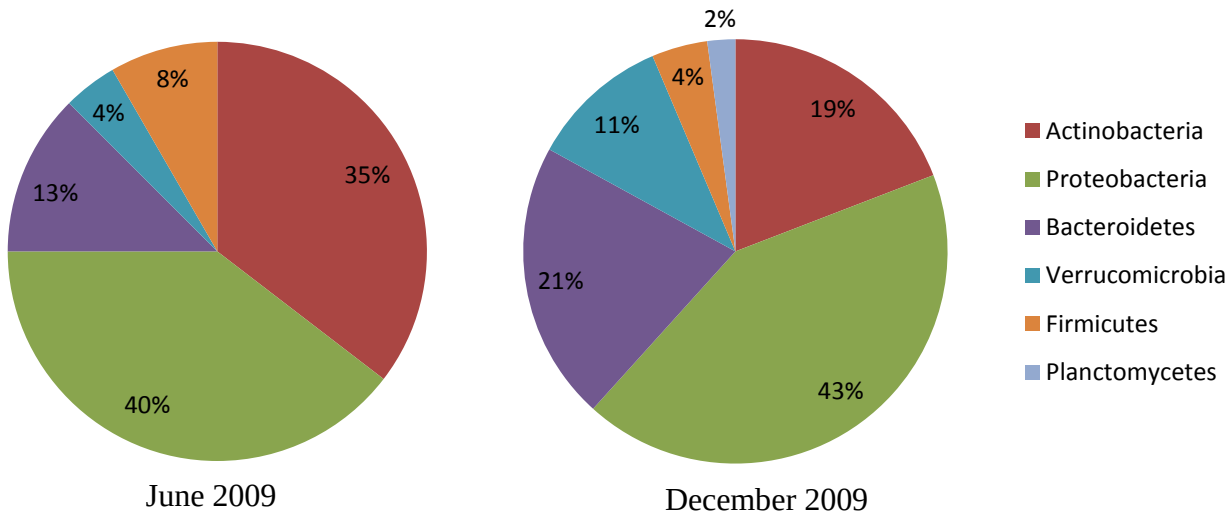

Figure 5. The distribution of phyla as identified by sequence data gathered in clone libraries for the months of June 2009 and December 2009.

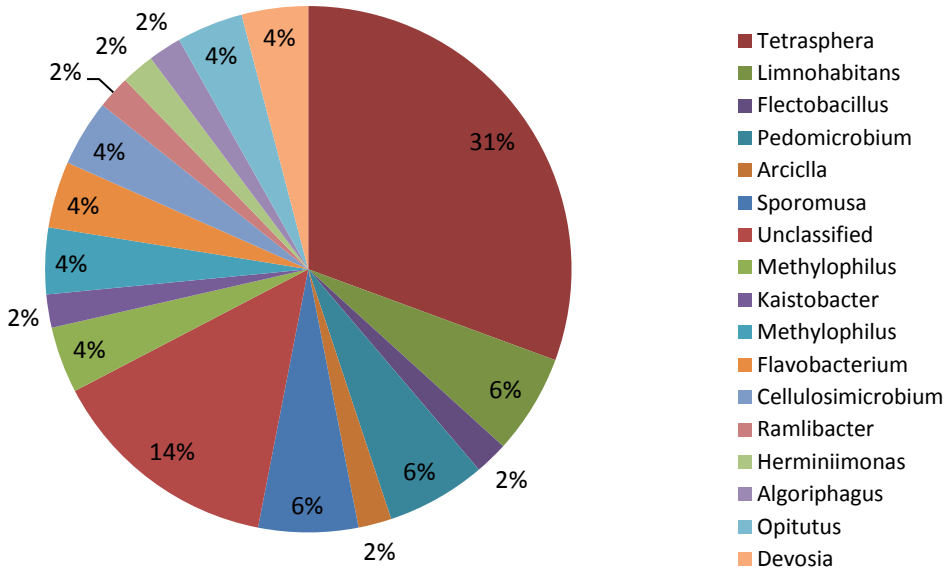

Figure 6. The distribution of genera as identified by sequence data gathered in clone library for the month of June 2009.

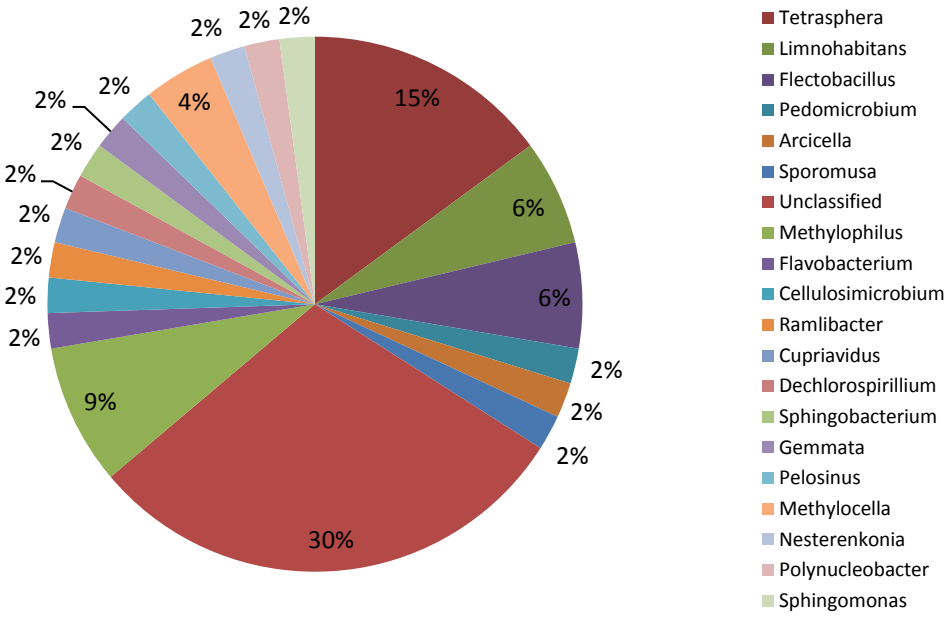

Figure 7. The distribution of genera as identified by sequence data gathered in clone library for the month of December 2009.

the genus level. No other identified genus composed more than 6% of the sequenced clones. The dominant identified genus in the month of December was also *Tetrasphera*. However, *Tetrasphera* only made up 15% of the December library. It was followed closely by *Limnohabitans*, comprising 9% of the December library. No single remaining identified genus comprised more than 6% of the clones in the December library. 14 of the 20 identified genera comprised no more than 2% of the identified clones.

While the sequencing of 16S rRNA genes generally allows for a high-resolution view of OTU richness, accurate classification of individual clones beyond the Phylum level was often difficult. In December, 30% of the clones sequenced could not be identified at the genus level. In June, 14% of the clones sequenced could not be identified at the genus level.

3.3.3 – In Silico Digestions

In silico digestions were performed on each of the sequences gathered from the clone libraries using TRiFLe and NEBCutter (Hammer et al., 2001; Vincze et al., 2003). The *in silico* digestions generated a theoretical chromatogram for each of the sequences, which displayed the most likely position the terminal fragments would be located (Figure 8). Each chromatogram was accompanied by a table that provided the exact terminal fragment length of the fragments. It was noted early that approximately one third of the terminal fragments from a digestion with MspI fell below the 50 base pair minimum for detection by the ABI3130 sequencer. Despite the fact that previous studies have used MspI as one of the restriction enzymes for tRFLP (Lymer et al., 2008), perhaps a different restriction enzyme should have been selected to create the most

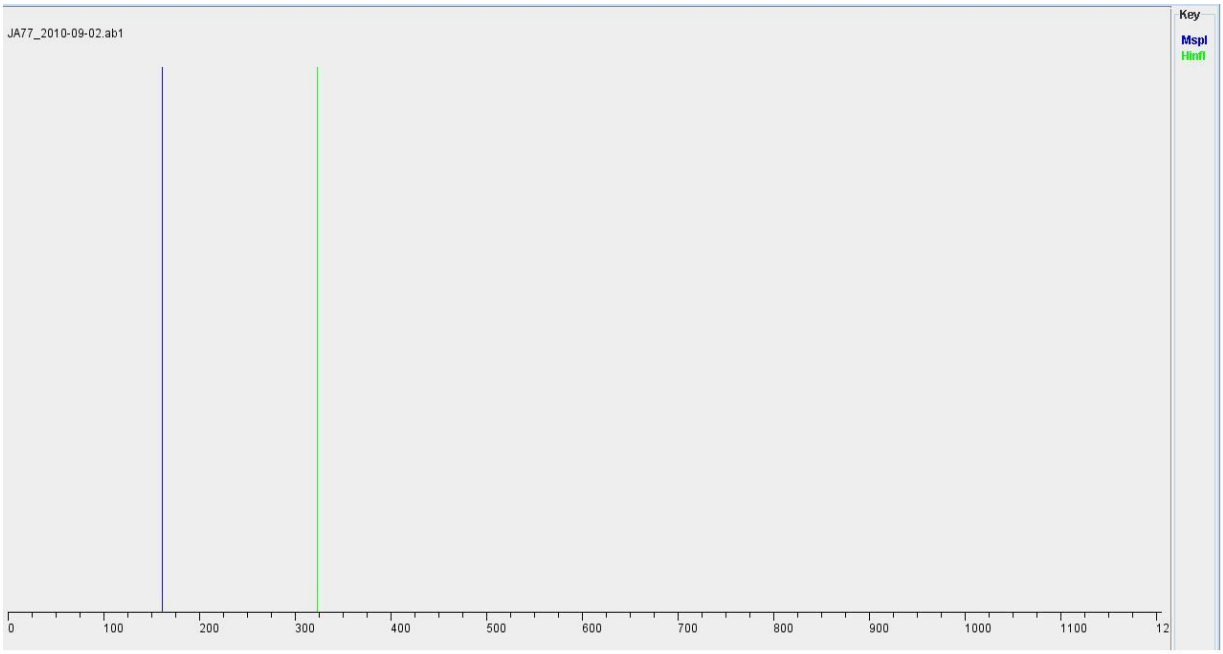

Figure 8. A sample simulated chromatogram after *in silico* digestion by TRiFLe using the HinfI and MspI restriction enzymes. The x-axis is base pair length. The right most bar is the location of the MspI terminal fragment. The left most bar is the location of the HinfI terminal fragment.

complete profile of the bacterial community composition in Lake Matoaka. Two restriction enzymes were used in this study for that exact reason.

The theoretical T-RF's identified by TRiFLe were matched with actual T-RF's of the same length in June 2009 and December 2009 tRFLP chromatograms. Each chromatograms from the remaining sample months were examined for the presence of the matched T-RF's from June and December. Each identified peak was then given a taxonomic classification based on information gathered from the GreenGenes database (DeSantis et al., 2006) (Table 2). The majority of the peaks that were matched with theoretical peaks from the *in silico* digestions belonged to the phylum *Proteobacteria*. Moreover, peaks belonging to these phyla were identified in both the June and December chromatograms. Peaks belonging to the *Proteobacteria* phylum were identified in chromatograms for 14 out of 16 sample months. This corresponds with the trends observed in the clone library compositions, in which *Proteobacteria* made up about 40% of the community composition in both the June and December libraries. Identification of peaks that are composed of, at least in part, *Proteobacteria* lends strength to the hypothesis that *Proteobacteria* are the dominant bacterial phylum in Lake Matoaka year-round.

It is interesting to note, however, that the majority of *Proteobacteria* identified in June belonged to the *Betaproteobacteria* class and the majority identified in December belonged to the *Alphaproteobacteria* class. Moreover, the same peaks for *Alphaproteobacteria* identified in other months were lower temperature months like November and February. Furthermore, several peaks were identified that only occurred in June 2009. Three of the peaks are *Betaproteobacteria* and two of the three are unclassified at the genus level. The remaining peak is classified in the genus *Limnohabitans*, which was just identified in 2009 and many of the species belonging to the genus have not been identified. It is possible that these peaks belong to unique OTUs that are

Table 2. Results of *In Silico* Digestion of Sequences

Clone Library Sample Month	Enzyme	Identified Peak Fragment Length (BP)	Phylum (Class) and Genus for Identified Peak	Months with Identified Peak
June 2009	HinfI	176	*Actinobacteria, Tetrasphera*	6/09
		189	*Proteobacteria (Beta), Unclassified*	6/09
		196	*Proteobacteria (Beta), Limnohabitans*	6/09
		320	*Proteobacteria (Beta), Unclassified*	6/09
		323	*Actinobacteria, Tetrasphera*	5/09,6/09, 7/09, 8/09, 9/09, 10/09, 2/10, 10/10, 11/10
	MspI	85	*Bacteroidetes, Unclassified*	6/09
		89	*Bacteroidetes, Unclassified*	6/09
		451	*Verrucomicrobia, Unclassified*	6/09, 2/10, 10/10
		490	*Proteobacteria (Beta), Limnohabitans OR Herminiimonas*	5/09, 6/09, 8/09, 9/09, 2/10, 7/10, 9/10, 11/10
December 2009	HinfI	173	*Firmicutes, Unclassified*	12/09, 2/10
		323	*Actinobacteria, Tetrasphera*	5/09, 7/09, 8/09, 9/09, 10/09, 12/09, 2/10, 10/10, 11/10
		324	*Proteobacteria (Beta), Limnohabitans*	5/09, 7/09, 10/09, 11/09, 12/09, 2/10, 3/10, 6/10, 7/10, 9/10, 10/10, 11/10
		331	*Bacteroidetes, Unclassified*	5/09, 7/09, 8/09, 9/09, 10/09, 12/09, 2/10, 10/10, 11/10
	MspI	148	*Proteobacteria (Alpha), Unclassified*	11/09, 12/09, 2/10, 11/10
		150	*Proteobacteria (Alpha), Unclassified*	11/09, 12/09, 11/10
		152	*Verrucomicrobia, Unclassified*	9/09, 4/10

The above table indicates the peaks that were identified after *in silico* digestion of the sequence data. Peaks are classified by the month the restriction site was found in the clone library and by the enzyme used in the digestion. Fragments of specific base pair lengths were identified using the sequence identification. Identified peaks were also located in chromatograms for other months.

only found in June 2009 and are, in reality, rare species. The same could be true for the *Tetrasphera* peak that only appears in June 2009. *Tetrasphera* were identified in many other months at different fragment lengths. It is possible that this particular peak represents a different species than the *Tetrasphera* identified at other fragment lengths.

The *in silico* digestion results also illustrate the fact that tRFLP peaks may consist of more than one OTU. The terminal fragment at 490 base pairs was identified as either belonging to the genus *Limnohabitans* or the genus *Herminiimonas*, as two different clones had restriction sites at the same location. This is a weakness of the technique and it leaves the potential for an underestimation of community richness.

3.4 – Correlations

Pearson's correlations were calculated to test for relationships between abundance, community richness and environmental variables for both bacterial and viral communities in the lake (Table 3). The results indicate that viral and bacterial abundance are very strongly correlated ($r = 0.923$, $p = 0.000000421$; Table 3). Significant positive correlations were also identified between temperature and viral abundance ($r = 0.869$, $p = 0.0000026$; Table 3), and temperature and bacterial abundance ($r = 0.857$, $p = 0.0000045$; Table 3). These correlations suggest that bacterial and viral abundance are driven strongly by each other and by temperature. This would produce the highly similar and cyclical trend observed for bacterial and viral abundance (Fig. 1). It is difficult to determine the cause for this correlation. The viral population may be driving the bacterial population. Equally likely, the bacterial population may be driving the viral population, as cycles of viral and bacterial infection are dependent on the host-pathogen specificity and the availability of said host or pathogen.

Table 3. Pearson Correlations between Abundance, Richness and Environmental Data

Variable	VA	BA	chla	chlb	temp	pH	NOx	NH4	DIP	rain	Vrich	hinRich	mspRich
VA	1												
BA	0.932 (4.21e-7)	1											
chla	0.354	0.349	1										
chlb	0.088	-0.038	0.287	1									
temp	0.868 (2.6e-6)	0.857 (4.5e-6)	0.242	-0.035	1								
pH	0.316	0.263	-0.081	0.406	0.175	1							
NOx	-0.343	-0.414	0.224	0.173	-0.328	-0.158	1						
NH4	0.349	0.503	0.274	-0.389	0.423	-0.417	0.014	1					
DIP	-0.267	-0.196	-0.201	0.025	-0.127	0.327	0.347	-0.127	1				
rain	-0.198	-0.259	-0.095	0.425	-0.100	0.054	0.292	0.090	0.346	1			
Vrich	0.401	0.392	0.144	-0.241	0.665 (0.006)	-0.029	0.074	0.379	0.281	-0.117	1		
hinRich	-0.096	0.058	0.122	0.348	-0.026	0.231	0.017	0.074	0.089	0.274	-0.246	1	
mspRich	-0.104	0.047	0.319	0.432	-0.105	0.010	-0.007	0.153	-0.070	0.360	-0.315	0.791 (4.5e-4)	1

Variables include viral abundance (VA), bacterial abundance (BA), chlorophyll A (chla), chlorophyll B (chlb), temperature (temp), pH, , NO_3^- and NO_2^- (NOx), ammonia (NH4), dissolved inorganic phosphate (DIP), rainfall (rain), viral richness (Vrich), and bacterial richness by each enzyme profile, HinfI and MspI respectively (hinRich, mspRich). Only significant p-values ($p < 0.05$) shown.

Table 4. P-values for Pearson's Correlations between Abundance, Richness and Environmental Data

Variables	VA	BA	chla	chlb	temp	pH	NOx	NH4	DIP	rain	Vrich	hinRich	mspRich
VA	0												
BA	4.22E-07	0											
chla	0.195	0.202	0										
chlb	0.756	0.894	0.300	0									
temp	2.63E-05	4.5E-05	0.384	0.901	0								
pH	0.251	0.344	0.774	0.133	0.533	0							
NOx	0.211	0.125	0.421	0.537	0.233	0.574	0						
NH4	0.202	0.055	0.323	0.152	0.115	0.121	0.960	0					
DIP	0.335	0.484	0.471	0.929	0.653	0.234	0.204	0.651	0				
rain	0.479	0.350	0.735	0.114	0.721	0.849	0.290	0.748	0.206	0			
Vrich	0.138	0.148	0.607	0.386	0.006	0.917	0.792	0.163	0.310	0.678	0		
hinRich	0.733	0.838	0.663	0.204	0.926	0.408	0.951	0.794	0.750	0.323	0.375	0	
mspRich	0.711	0.867	0.246	0.107	0.708	0.971	0.981	0.585	0.803	0.186	0.252	4.51E-04	0

Variables include viral abundance (VA), bacterial abundance (BA), chlorophyll A (chla), chlorophyll B (chlb), temperature (temp), pH, , NO_3^- and NO_2^- (NOx), ammonia (NH4), dissolved inorganic phosphate (DIP), rainfall (rain), viral richness (Vrich), and bacterial richness by each enzyme profile, HinfI and MspI respectively (hinRich, mspRich).

The only other measured variables to show any meaningful degree of correlation were the bacterial richness data obtained from the HinfI and MspI tRFLP profiles, which were positively correlated with each other ($r = 0.791$, $p = 0.00045$; Table 3), and between viral richness and temperature ($r = 0.665$, $p = 0.006$; Table 3). It would appear that bacterial richness (that is, the number of bands or OTU's observed within a given month) has little real correlation with abundance or any environmental factor, but the viral community composition does show some moderate correlation with temperature. This indicates that while bacterial abundance may increase with viral abundance and temperature, bacterial community composition does not. This shows a reasonable parallel with the 16S rRNA clone library data in that richness is only slightly higher in December and there are dominant taxa present at similar concentrations in both clone libraries.

BioEnv tests were performed to determine the relationships between the community composition data (viral and bacterial) and the environmental data (Table 5). The test was based on Dice dissimilarity matrices for the HinfI and MspI tRFLP community data and Euclidean distances of environmental data (Melo, 2009), and attempts to identify environmental factors that best explain the variation observed in the biological data (Clarke et al., 1993). The resulting r-values are a measure of the explanatory power of an environmental variable in accounting for the variation in the biological (community composition) data.

Results from BioEnv analysis indicated that temperature is the single best environmental factor for explaining the changes in community composition with r-values of 0.5602, 0.5676 and 0.5660 for the viral composition, HinfI composition and MspI composition respectively. Moreover, the Pearson's correlation between viral richness and temperature supports the results obtained by the BioEnv analysis. It is unclear why temperature has some effect on viral

Table 5. Results of BioEnv analysis of community composition in relation to environmental variables.

Data Set	Variables	R-Value
Virus Distance Matrix vs. Environmental Data	temp	0.5602
	temp Vrich	0.5249
	BA temp Vrich	0.5002
	BA temp rain Vrich	0.4919
	VA temp rain Vrich hinRich	0.4783
	VA temp DIP rain Vrich hinRich	0.4640
	VA BA temp DIP rain Vrich hinRich	0.4447
	VA BA chla temp DIP rain Vrich hinRich	0.4113
	VA BA chla temp DIP rain Vrich hinRich mspRich	0.3806
	VA BA chla temp pH DIP rain Vrich hinRich mspRich	0.3416
	VA BA chla temp NOx NH4 DIP rain Vrich hinRich mspRich	0.2923
	VA BA chla temp pH NOx NH4 DIP rain Vrich hinRich mspRich	0.2523
HinfI Distance Matrix vs. Environmental Data	temp	0.5676
	temp Vrich	0.5259
	BA temp Vrich	0.5093
	VA BA temp Vrich	0.4987
	VA temp rain Vrich hinRich	0.4813
	VA temp DIP rain Vrich hinRich	0.4655
	VA BA temp DIP rain Vrich hinRich	0.4482
	VA BA chla temp DIP rain Vrich hinRich	0.4149
	VA BA chla temp DIP rain Vrich hinRich mspRich	0.3861
	VA BA chla temp NOx DIP rain Vrich hinRich mspRich	0.3471
	VA BA chla temp NOx NH4 DIP rain Vrich hinRich mspRich	0.2962
	VA BA chla temp pH NOx NH4 DIP rain Vrich hinRich mspRich	0.2543
MspI Distance Matrix vs. Environmental Data	temp	0.5660
	temp Vrich	0.5294
	BA temp Vrich	0.5079
	BA temp rain Vrich	0.4976
	VA temp rain Vrich hinRich	0.4828
	VA temp DIP rain Vrich hinRich	0.4661
	VA BA temp DIP rain Vrich hinRich	0.4477
	VA chla temp DIP rain Vrich hinRich mspRich	0.4210
	VA BA chla temp DIP rain Vrich hinRich mspRich	0.3913
	VA BA chla temp NOx DIP rain Vrich hinRich mspRich	0.3508
	VA BA chla temp NOx NH4 DIP rain Vrich hinRich mspRich	0.3003
	VA BA chla temp pH NOx NH4 DIP rain Vrich hinRich mspRich	0.2583

Variables include temperature (temp), viral richness (Vrich), bacterial abundance (BA), viral abundance (VA), rainfall (rain), bacterial richness by each enzyme profile, HinfI and MspI respectively (hinRich, mspRich), dissolved inorganic phosphate (DIP), chlorophyll A (chla), ammonia (NH4), NO_3^- and NO_2^- (NOx).

community composition. Given the tight linkage between viral and bacterial community compositions, temperature should logically affect both communities in the same ways. However, there was no correlation between bacterial community composition and temperature, Thus, it is extremely difficult to account for the difference in temperature's effect on the two communities. The inclusion of additional environmental factors does not serve to increase the r-statistic, and therefore do not appear to contribute strongly to observed changes in community structure (Table 5). This is also supported by the lack of significant Pearson correlations between environmental and biological factors (Table 1)

Mantel Tests based on Spearman's rank correlation were performed to assess the degree of correlation between viral and bacterial composition. Viral and HinfI composition, viral and MspI composition and HinfI and MspI composition were all shown to be significantly positively correlated, with r-values of 0.9975 ($p = 0.0001$), 0.9976 ($p = 0.0001$) and 0.9986 ($p = 0.0001$), respectively. This indicates that the composition, and shifts in composition, of the viral and bacterial communities is strongly linked. Added to the moderate correlation with temperature, it would appear that bacterial community composition is dynamic is driven, in part, by temperature and is likely independent of microbial abundance at this location in Lake Matoaka. This does not preclude abundance as a driver for community changes in other areas of Lake Matoaka

4. Discussion

4.1 – Terminal Restriction Fragment Length Polymorphism and Drivers of Composition Change

As has been noted, the taxonomy of freshwater systems is fairly well characterized, but the divers of changes in community composition have not been well characterized to date. This study seeks to identify environmental or biological factors that have significant explanatory power with regard to community composition change, using Lake Matoaka as a model system. Terminal restriction fragment length polymorphism has become a common technique for the rapid profiling of bacterial composition – especially for the purpose of tracking changes in community composition over time. In spite of numerous studies that have explored freshwater microbial community composition, there has yet to have been one single best factor that explains community composition (Newton et al., 2011; Boucher et al., 2006; Hiorns et al., 1997; Lindstrom, 2000). Previous studies have demonstrated that annual variations in community composition are not generally repeated over annual cycles in eutrophic lakes (Lindstrom et al., 1998; Boucher et al., 2006; Lindstrom, 2000). Repeating seasonal trends have, however, been identified in humic and oligotrophic lakes (Yannarell et al., 2003). In eutrophic lakes, community composition is highly variable over time (Lindstrom, 2000). This is perhaps due, in part, to the dynamic changes in nutrient levels and water quality that are characteristic of eutrophic lakes. This is consistent with the trends – or lack thereof – in annual cycling in bacterial community composition in the current study. Because of lack of repeating annual trends, it is difficult to identify specific environmental factors that drive community change.

To identify specific drivers, researchers will typically perform one of many ordination and/or multiple regression techniques with their data sets. Ordination techniques include canonical correspondence analysis, non-metric multidimensional scaling and principal

component analysis. Multiple regression techniques can be used independently or in conjunction with ordination and include such approaches as BioEnv (Clarke et al., 1993). The purpose of these techniques is to extract global trends from a complex set of multivariate data.

Using these analysis techniques, previous studies have identified several factors – temperature and pH, in particular –as potential drivers of bacterial community composition changes in freshwater lakes (Lindstrom et al., 2005). In one study that examined 15 different lakes in Sweden and Norway ranging from oligomesotrophic to hypereutrophic, humic to non-humic and acidic to alkaline over the course of one summer, temperature was shown to account for 43%-60% of community composition change over time using canonical correspondence analysis (Lymer at al., 2008). While BioEnv analysis was used by this study, it was also revealed that temperature accounts for a moderate degree of the community composition changes over time. Moreover, temperature was also determined to strongly correlate with bacterial and viral abundance in this study. This indicates that there appears to be a seasonal variation in abundance, i.e. as temperature increases in the warmer seasons, bacterial abundance also increases and as temperature decreases in the cooler seasons, bacterial abundance decreases. However, no correlation with community composition and temperature by Pearson's correlation was found. This is consistent with differences in trends for abundance (correlation with season) and community composition (no correlation with season) obtained in other studies (Boucher et al., 2006; Allgaier et al., 2006; Tijdens et al., 2008).

pH has also been identified as a significant driver of community composition change. One study has shown significant clustering of bacterial communities by pH across five different lakes by non-metric multidimensional scaling and a significant Pearson's correlation between pH and bacterial richness of 0.785 (Allgaier et al., 2006). In the present study no meaningful

correlation was observed between community composition and pH (noting that a significant p-value for the Pearson's correlation was not obtained; Table 4). However, it is very important to note that in that same previous study, five lakes – with different properties (i.e. eutrophic, mesotrophic, humic, etc.) were examined and different environmental factors were found to drive community change in each of the five lakes (Allgaier et al., 2006). As the present study only explored one lake and the pH within the lake did not fluctuate considerably (Table 1), it is logical that pH did not have any explanatory power in accounting for bacterial community composition change. Moreover, the previous study showed a range of correlation between bacterial composition and temperature. The five lakes under study showed anywhere from strong negative correlation to no correlation to strong positive correlation between bacterial community composition and temperature (Allgaier et al., 2006). Given that temperature as an explanatory factor for bacterial community composition change varied between lakes of different types, the fact that temperature showed only a moderate correlation with bacterial community composition of Lake Matoaka is not outside of the realm of potential freshwater lake dynamics.

Additionally, the results of the Mantel tests revealed that bacterial and viral communities in Lake Matoaka are tightly coupled, although both sets of community composition appear to change independently from abundance. This is in contrast to the findings in a study of three lakes of mesotrophic or oligotrophic classification, which showed almost no coupling between viral and bacterial community composition in any of the three lakes (Lymer et al., 2008). This difference in outcome can perhaps be explained by the difference in trophic classification of the lakes under study. The present study examined the bacterial and viral communities in an eutrophic lake, with high primary productivity. However, no correlation between viral and bacterial richness with nutrient content was seen in Lake Matoaka. It is unclear if lake trophic

status is a factor in this difference in correlation between viral and bacterial community composition. Another potential explanation of the differences in findings is the specific technique used to obtain the viral community composition data. The present study obtained data by randomly amplified polymorphic DNA polymerase chain reaction (RAPD-PCR), while Lymer et al. used pulsed field gel electrophoresis (PFGE). PFGE is considered the most efficient technique for characterizing viral communities by genome size (Lymer et al., 2008). However, unpublished data obtained in the lab of Dr. Kurt Williamson at the College of William and Mary suggests that PFGE is severely limited in its ability to be applied to freshwater samples and that RAPD-PCR provides much more information about viral community changes (Winget et al, 2008; Winter et al., 2010).

The tight coupling of the viral and bacterial communities in Lake Matoaka also contrasts with a study of the viral and bacterial communities across different depths in the Mediterranean sea, a saltwater, marine system (Winter et al., 2010). The viral and bacterial community compositions were found to be moderately correlated ($r = 0.53$ in the overall water column, $r = 0.63$ in the bathypelagic zone, $p \leq 0.05$) in the Mediterranean Sea (Winter et al., 2010). Perhaps it is the difference in salinity that accounts for the differences in findings. Overall, the differences in findings between this study and other studies imply that it is very difficult to determine the cause for the linkage, or lack thereof, between viral and bacterial communities. It is impossible, without further study, to determine if the viral community is driving bacterial community composition or vice versa.

The lack of clear environmental factors that adequately explain the monthly shifts in bacterial community composition points to the fact that freshwater lakes are extremely complex ecosystems that are affected daily by many different variables (Tijdens et al., 2008). Given that

changes in lake community composition can change very rapidly, it is clearly extremely difficult to parse out the exact drivers (Boucher et al., 2006; Yannarell et al., 2003; Tijdens et al. 2008). Perhaps the drivers would only become clear in a study that included several lakes of almost identical characteristics. Studies to date have been focused on a temporal cycle in one lake or across several very different lakes and on short temporal scales (one month to several months).

4.2 – 16S rRNA Clone Libraries and In Silico Digestions

Several previous studies have used clone libraries to characterize the bacterial community composition of freshwater ecosystems (Bel'kova et al., 2003; Hiorns et al., 1997; Boucher et al., 2006). Each of these studies explored the bacterial community composition of different lakes. The lakes ranged in trophic status from oligotrophic to eutrophic, in latitude from arctic to temperate, and samples were collected over time scales ranging from a single season (three months) to over two years. However, several general trends were identified, despite the vast differences in locale and date. In spite of the extremely broad types of habitat represented in the lakes in these studies, three dominant phyla identified across each lake system were *Proteobacteria, Actinobacteria* and *Bacteroidetes* (Bel'kova et al., 2003; Hiorns et al., 1997; Boucher et al., 2006). The results obtained in the present study also follow this trend, with the same three phyla contributing to the majority of both clone libraries. *Actinobacteria* and *Proteobacteria* were especially dominant in the month of June (Fig. 5), following previous observations from independent studies that patterns of year-round dominance and independence from such environmental factors as temperature and nutrient levels (Boucher et al., 2006).

Commonly identified in lake epilimnia, *Actinobacteria* and *Proteobacteria* have proven to be highly successful aquatic lineages (Newton et al., 2011). *Proteobacteria*, especially

Alphaproteobacteria, are uniquely adapted to be resistant to grazing by organisms of higher trophic levels and to be competitive at low nutrient levels (Newton et al., 2011). This may help to explain the apparent dominance of the phyla in both the warm, nutrient rich month of June and the cold, nutrient starved month of December. The phylum *Actinobacteria* is also highly abundant in freshwater environments. Although less well characterized than *Proteobacteria*, *Actinobacteria* contains several lineages that are specialized to limnetic systems that have been identified globally (Newton et al., 2011). However, what is most interesting about this phylum is its ubiquitous presence in terrestrial soil ecosystems (Boucher et al., 2006). This indicates a potential relationship between the microbial community of the lake and the community present in the surrounding watershed. Given the likely effect of run-off from rainstorms, soil *Actinobacteria* introduced to the lake system – in addition to the lineages specialized for limnetic systems – may help to account for the year-round dominance of the phylum.

The *Verrucomicrobia* are another commonly reported bacterial phylum in freshwater lakes, and were also identified in the microbial community of Lake Matoaka. Little is known about the ecology of this particular phylum, except that it is found at both the lake epilimnion and hypolimnion (Newton et al., 2011). Several other phyla were identified within Lake Matoaka, but each represented only a small percentage of the total. This indicates that there is broad bacterial diversity that is dominated by only a few taxa (Boucher et al., 2006). The overwhelming dominance of only a few taxa in the present study supports this trend. Unfortunately, a large percentage of the clones sequenced in this study could not be identified and were categorized as "unclassified." This trend has been observed in previous studies and is accounted for by the dearth of information available about freshwater bacterial communities (Zwart et al., 2002)

Furthermore, there is a growing body of evidence that suggests the freshwater lineages have a global distribution (Zwart et al., 1998). For example, the recently classified *Limnohabitans curvus*, a freshwater bacterium belonging to the *Proteobacteria* phylum, was isolated in a freshwater lake in Austria (Han et al., 2010). This same new species was identified among the June clones in this study. Despite the physical and chemical differences between Lake Matoaka and Lake Mondesee (Austria), *Limnohabitans curvus* appears to be well adapted enough to be competitive in a wide range of freshwater ecosystems. The same is most likely true of many other freshwater lineages, and as freshwater lake studies increase in frequency, other broadly distributed freshwater lineages may be identified (Zwart et al., 1998).

Due to limited time and resources, only 96 clones were able to be identified by sequencing in the present study. The fact that only 48 clones per library were sequenced means that the potential is there for under-sampling. To lend strength and significance to the data gathered using 16S rRNA clone libraries, upwards of hundreds of clones would be needed (Liu et al., 1997). However, given the labor intensity, and the lack of high-throughput technology for more than that, fewer clones were sequenced. With additional sampling, the observed trends in terms of taxonomic distributions within each library may have changed, but it is difficult to say for sure. Despite this fact, the data gathered begins to shed some light on the bacterial community composition in Lake Matoaka and it follows previously and well established trends.

Using TRiFLe, theoretical terminal fragments after digestion by HinfI and MspI were obtained and these fragments were converted into theoretical peaks on a chromatogram. These theoretical peaks were easily matched with several peaks found in the actual chromatogram data, as well as being given a taxonomic classification. The classified peaks followed much of the same trends observed in the clone libraries; *Proteobacteria* was the most commonly identified

phylum. While previous studies have used TRiFLe, it has not been used in the same context as it was in this study. Previous studies have used TRiFLe as a means of checking the accuracy of tRFLP data against clone library data when specific phyla were under study (Sisinthy et al., 2011). In this study, TRiFLe was used as another method by which tRFLP and 16S rRNA clone library were compared to better identify trends in community composition change. While more clones would be necessary to make effective use of TRiFLe, its use did serve to augment confidence in identified trends in bacterial community composition change.

4.3 Future Considerations

Because of time and resources, the scope of this study was somewhat limited. Future studies should be conducted to take into account various environmental factors and other variables in lake dynamics that were not included here. Previous studies suggest that dissolved organic carbon plays a significant role in community composition (Jones et al., 2009). Other data suggest that the lifestyle of lake bacteria makes a difference in community composition. For example, differences have been reported between the community composition of free-living and particle-associated bacteria (Allgaier et al., 2006). The 5 μm filters from the sample processing protocol are currently archived at -80 °C. DNA could be extracted from these filters and subject to 16S rRNA PCR and tRFLP to explore the dynamics of particle-associated bacteria in Lake Matoaka. Moreover, the spatial variations across Lake Matoaka should be taken into account. In conjunction with this study, samples were taken from two other locations in Lake Matoaka – the primary inlet and the outlet – which can be used for a similar study as described in this project to test additional hypotheses, such as how bacterial community composition changes over time and

space within the same lake. As tRFLP is such a simple, yet powerful technique, many aspects of bacterial community composition can and should be explored.

Another variable that has played a significant role in community composition in previous studies has been top-down grazing by protozoa and other bacteriovores (Comte et al., 2006; Lindstrom, 2000). Assessments of grazing impacts on bacterial community composition were not included in the present study, and grazing is another factor that should be taken into account for Lake Matoaka. Microcosms containing Lake bacteria, with and without protozoa, could be created to test the effects of protozoan grazing.

4.4 – Closing Thoughts

This study provides a good deal of insight into the dynamics of a freshwater lake ecosystem, despite minor technological limitations. It has clearly demonstrated the close relationship between bacterial and viral abundance and community composition. And while a single best explanation of bacterial community composition change in Lake Matoaka has proven elusive, a few factors including viral abundance and temperature have been identified have having some effect on bacterial community composition. This study is simply one of many in a growing field. Given the importance of freshwater systems for human life and the importance of bacteria in freshwater ecosystems, lake studies are essential. They must continue so as to allow humans to better understand, better maintain and better use Earth's limited freshwater supply.

5. References

Akpor,O.B., and M. Muchie. (2010) Remediation of heavy metals in drinking water and wastewater treatment systems: Processes and applications. *International Journal of the Physical Sciences*, 5(12): 1807-1817.

Allgaier, M. and H-P. Grossart. (2006) Seasonal dynamics and phylogenetic diversity of free living and particle-associated bacterial communities in four lakes in northeastern Germany. *Aquatic Microbial Ecology*, 45: 115-128.

Azam, F., T. Fenchel, J.G. Field, J.S. Gray, L.A. Meyers-Reil, and F. Thingstad. (1983) The Ecological Role of Water-Column Microbes in the *Sea. Mar. Ecol. Prog. Ser.*, 10: 257-263.

Bel'kova, N.L., V.V. Parfenova, T. Ya. Kostornova, L. Ya. Desinova, and E.F. Zaichikov. (2003) Microbial Biodiversity in the Water of Lake Baikal. *Microbiology*, 72(2): 203-212.

Blackwood, T. Marsh, S. Kim and E. Paul. (2003) Terminal Restriction Fragment Length Polymorphism Data Analysis for Quantitative Comparison of Microbial Communities. *Applied and Environmental Microbiology*, 69(2): 926-932.

Boucher, D., L. Jardillier, and D. Debroas. (2006) Succession of bacterial community composition over two consecutive years in two aquatic systems: a natural lake and a lake-reservoir. *FEMS Microbiol Ecol*, 55: 79-97.

Clarke, K.R., and M. Ainsworth. (1993) A method of linking multivariate community structure to environmental variables. *Marine Ecology Progress Series*, 92: 205-219.

Clasen, J.L., S.M. Brigden, J.P. Payet, and C.A. Suttle. (2008) Evidence that viral abundance across oceans and lakes is driven by different biological factors. *Freshwater Biology*, 53: 1090-1100

Cole, J. J., et al. (2007) Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* 10: 171–184

Cole, J.J., N.F. Caraco, G.W. Kling, and T.K. Kratz. (1994) Carbon Dioxide Supersaturation in the Surface Waters of Lakes. *Science* 265(5178): 1568-1570.

Cole, J. J., and N. F. Caraco. (2001) Carbon in catchments: connecting terrestrial carbon losses with aquatic metabolism. *Marine and Freshwater Research*, 52: 101-110.

Comte, J., S. Jacquet, S. Viboud, D. Fontvieille, A. Millery, G. Paolini and I. Domaizon. (2006) Microbial Community Structure and Dynamics in the Largest Natural French Lake (Lake Bourget). *Microbial Ecology*, 52: 72-89.

DeSantis, T. Z., P. Hugenholtz, K. Keller, E. L. Brodie, N. Larsen, Y. M. Piceno, R. Phan, and G. L. Andersen. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res*, 34: W394-9.

DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. (2006) GreenGenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72: 5069-72.

Fisher, M.M., and E.W. Triplett. (1999) Automated Approach for Ribosomal Intergenic Spacer Analysis of Microbial Diversity and Its Application to Freshwater Bacterial Communities. *Applied and Environmental Microbiology*, 65(10): 4630-4636.

Fuhrman, J. (1992) Bacterioplankton roles in cycling of organic matter: the microbial foodweb. In: PG Falkowski and AD Woodhead (eds). Bacterioplankton roles in cycling of organic matter: the microbial food web. New York: Plenum Press, pp. 361-383.

Fuhrman, J., K. McCallum, and A. Davis. (1993) Phylogenetic Diversity of Subsurface Marine Microbial Communities from the Atlantic and Pacific Oceans. *Applied and Environmental Microbiology*, 59(5): 1294-1302.

Giovannoni, S., and U. Sting. (2005) Molecular diversity and ecology of microbial plankton. *Nature*, 437: 343-348.

Hammer, Ø., Harper, D.A.T., and P. D. Ryan, 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* 4(1): 9 <http://palaeo-electronica.org/2001_1/past/issue1_01.htm>.

Han, M.W., V. Kasalick, J. Jezbera, U. Brandt, J. Jezberová and K. Šimek. (2010) *Limnohabitans curvus* gen. nov., sp. nov., a planktonic bacterium isolated from a freshwater lake. *International Journal of Systematic and Evolutionary Microbiology*, 60: 1358-1365.

Hiorns, W., B. Methé, S. Nierzwicki-Bauer, and J. Zehr. (1997) Bacterial Diversity in Adirondack Mountain Lakes as Revealed by 16S rRNA Gene Sequences. *Applied and Environmental Microbiology*, 63(7): 2957-2960.

Jansson, M., A. K. Bergstrom, P. Blomqvist, A. Isaksson, and A. Jonsson. (1999) Impact of allochthonous organic carbon on microbial food web carbon dynamics and structure in Lake Ortrasket. *Archiv Fur Hydrobiologie*, 144: 409-428.

Jones, J.E., R.J. Newton, and K.D. McMahon. (2009) Evidence for structuring of bacterial community composition by organic carbon source in temperate lakes. *Environmental Microbiology*, 11(9): 2463-2472.

Junier, P., T. Junier, and K. Witzel. (2008) TRiFLe: a program for *in silico* T-RFLP analysis with user-defined sequences sets. *Applied and Environmental Microbiology*, 74: 6452-6456.

Lindstrom E.S. (1998) Bacterioplankton community composition in a boreal forest lake. *FEMS Microb Ecol*, 27:163-174

Lindstrom, E.S. (2000) Bacterioplankton Community Composition in Five Lakes Differing in Trophic Status and Humic Content. *Microbial Ecology*, 40(2): 104-113.

Lindstrom, E.S., M.P. Kamst-Van Agterveld, and G. Zwart. (2005) Distribution of Typical Freshwater Bacterial Groups Is Associated with pH, Temperature, and Lake Water Retention Time. *Applied and Environment Microbiology*, 71(12): 8201-8206.

Liu, W., T. Marsh, H. Chang, and L. Forney. (1997) Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *Applied and Environmental Microbiology*, 63(11): 4516-4522.

Lueders, T., and M. Friedrich. (2003) Evaluation of PCR Amplification Bias by Terminal Restriction Fragment Length Polymorphism Analysis of Small-Subunit rRNA and *mcrA* Genes by Using Defined Template Mixtures of Methanogenic Pure Cultures and Soil DNA Extracts. *Applied and Environmental Microbiology*, 69(1): 320-326.

Lymer, D., J. Logue, C. Brussaard, A. Baudoux, K. Vrede, and E. Lindstrom. (2008) Temporal variation in freshwater viral and bacterial community composition. *Freshwater Biology*, 53: 1163-1175.

Melo, A.S. (2009). Explaining dissimilarities in mircoinvertebrate assemblages among stream sites using environmental variables. *Zoologia*, 26(1): 79-84

Muyzer, G. (1999) DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, 2: 317-322.

Muyzer, G., E.C. de Waal, and A.G. Uitterlinden. (1993) Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S rRNA. *Applied and Environmental Microbiology*, 59(3): 695-700.

Newton, R., S. Jones, A. Eiler, K. McMahon. (2011) A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews*, 79(1): 14-41.

Osborne, C., G. Rees, Y. Bernstein, and P. Janssen. (2006) New Threshold and Confidence Estimates for Terminal Restriction Fragment Length Polymorphism Analysis of Complex Bacterial Communities. *Applied and Environmental Microbiology*, 72(2): 1270-1278.

Park, J.H., I. Yoshinaga I, T. Nishikawa T, and I. Imai. (2010) Algicidal bacteria in particle associated form and in free-living form during a diatom bloom in the Seto Inland Sea, Japan. *Aquatic Microbial Ecology*, 60(2): 151-161.

Pomeroy, L. (1974) The Ocean's Food Web, A Changing Paradigm. *BioScience*, 24(9): 499-504.

Rusch, D.B., A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol*, 5: 398-431.

Sambrook, J. (2001). *Molecular Cloning: A Laboratory Manual*. Woodbury, NY: Cold Spring Harbor Laboratory Press.

Schütte, U., Z. Abdo, S. Bent, C. Shyu, C. Williams, J. Pierson, L. Forney. (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Appl Microbiol Biotechnol*, 80: 365-380.

Schindler, D.W., P.J. Curtis, S.E. Bayley, B.R. Parker, K.G. Beaty, and M.P. Stainton. (1997) Climate-induced changes in the dissolved organic carbon budgets of boreal lakes. *Biogeochemistry*, 36: 9-28.

Sisinthy, S., K. Kiran, K. Hara, P. Kumar, R, Sreenivasa, R. Naga, A. Rajesh, and R. Rasik. (2011) Vertical distribution of bacteria in a lake sediment from Antarctica by culture-independent and culture-dependent approaches. *Research in Microbiology*, 162(2): 191-203.

Singleton, D., M. Furlong, S. Rathbun, and W. Whitman. (2001) Quantitative Comparisons of 16S rRNA Gene Sequence Libraries from Environmental Samples. *Applied and Environmental Microbiology*, 67(9): 4374-4376.

Suzuki, M.T., and S.J. Giovannoni. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied Environmental Microbiology*, 62: 625-630.

Tijdens, M., and H. Hoogveld. (2008) Population Dynamics and Diversity of Viruses, Bacteria and Phytoplankton in a Shallow Eutrophic Lake. *Microbial Ecology*, 56: 29-42.

Vincze, T., J. Posfai, and R.J. Roberts. (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.*, 31: 3688-3691.

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

Winter, C., and M.G. Weinbauer. (2010) Randomly Amplified Polymorphic DNA Reveals Tight Links between Viruses and Microbes in the Bathypelagic Zone of the Northwestern Mediterranean Sea. *Applied and Environmental Microbiology*, 76(20): 6724-6732.

Yannarell, A.C., A.D. Kent, G.H. Lauster, T.K. Kratz, and E.W. Triplett. (2003) Temporal Patterns in Bacterial Communities in Three Temperate Lakes of Different Trophic Status. *Microbial Ecology*, 46: 391-405.

Zwart, G., B. C. Crump, M. P. K. V. Agterveld, F. Hagen, and S. K. Han. (2002) Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquatic Microbial Ecology*, 28: 141–155.

Zwart, G., et al. (1998) Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. *Systematic and Applied. Microbiology*, 21: 546–55.