Seasonal Abundance of Different *Chlorella* Viruses in Two Contrasting Freshwater Environments in Ontario, Canada

by

Oksana Rusanova

A thesis submitted in conformity with the requirements for the degree of Master of Science

Ecology and Evolutionary Biology

University of Toronto

© Copyright by Oksana Rusanova 2010
Seasonal Abundance of Different *Chlorella* Viruses in Two Contrasting Freshwater Environments in Ontario, Canada

Oksana Rusanova

Master of Science

Ecology and Evolutionary Biology
University of Toronto

2010

Abstract

The aims of this study were to identify Chloroviruses in two different Ontario freshwaters and to determine if the seasonal abundance patterns of Chloroviruses in different environments are similar. Gene fragments nearly identical to cultivated Chloroviruses were obtained from Lake Ontario and a nearby pond at the University of Toronto Mississauga (UTM) and novel *Chlorovirus* gene fragments were obtained from Lake Ontario. Quantification of these two *Chlorovirus* genes over several seasons revealed the presence of persistent viruses with different seasonal dynamics suggesting that different Chloroviruses replicate by infecting different hosts. Additionally, patterns of seasonal abundance and timings of peak abundances for individual viruses differed between Lake Ontario and the UTM pond, demonstrating the critical role of the environment in *Chlorovirus* dynamics. The observation of different Chloroviruses with different seasonal dynamics allows speculation that these viruses and their hosts stably coexist in Ontario freshwater environments.
Acknowledgements

The past 16 months have been quite intense but the achievement of this dissertation would simply not have been possible without the help, advice and support of my colleagues, family and friends.

First and foremost, I would like to acknowledge my supervisor, Dr. Steven M. Short, for giving me an extraordinary opportunity to enter the aquatic virology realm. I sincerely thank him for the trust deposited in me, for all the invaluable advice and for numerous discussions he endured in throughout the course of this program. The success of the work presented herein would not have been possible without Dr. Short’s, creation and frame-working of the project, along with his great mentorship and encouragement.

Special thanks should be given to my research committee members, Dr. Sasa Stefanovic and Dr. George Espie. The rigorous comments and knowledge they shared with me, has undoubtedly improved this thesis.

Furthermore, I am very thankful to our lab technician, Cindy M. Short, for her considerable support, mentorship and guidance on quantitative methodology. Especially, I am grateful for her hard work when we faced constant qPCR challenges.

Thanks are also extended to the colleagues of mine, Natasha Musrap and Michael A. Staniewski, for their support and assistance with sample collection, filtration and other procedures discussed in this report. The exciting lunches with them during my time in the lab were of particular importance to me. Mike and Natasha became real friends to me and I am very thankful for their companionship, unforgettable trips to sample collection sites and support with words of wisdom to guide me through.

I am deeply indebted to my family and friends for their trust in me and support through the rough patches of this journey. Very special thanks go out to Mariya Volos, Lexy Fifer, Maryna Smirnova, Helen Chiang and Nataliya Melnyk-Lamont for their long counselling discussions, wise advice and for keeping me running and smiling! Further gratitude is expressed to Mariya Volos for helping me to refine my thesis and to proofread the copy of this manuscript.

Thank you all very much,

Oksana
# Table of Contents

Abstract ........................................................................................................................................... ii
Acknowledgements ............................................................................................................................ iii
Table of Contents ............................................................................................................................... iv
List of Tables ...................................................................................................................................... vi
List of Figures .................................................................................................................................... vii
Abbreviations .................................................................................................................................... viii
Preface ................................................................................................................................................ x

I Introduction ...................................................................................................................................... 1
  1.1 Viruses in the aquatic ecosystems: Why are they important? .................................................... 1
  1.2 Phycodnaviruses: An overview .................................................................................................... 4
  1.3 *Chlorella* viruses and their phytoplankton hosts ...................................................................... 8
  1.4 Thesis objectives ......................................................................................................................... 12

II Materials and Methods ................................................................................................................... 14
  2.1 Sample collection and preparation ............................................................................................ 14
  2.2 Chlorophyll \(a\) filter extractions and fluorescence measurements ........................................... 16
  2.3 Cultivated *Chlorella* virus sequence analysis and PCR primer design .................................... 17
  2.4 Validation of primer specificity .................................................................................................... 19
  2.5 PCR amplification ......................................................................................................................... 20
  2.6 Cloning and sequencing *Chlorella* polB fragments .................................................................. 22
  2.7 Virus sequence analysis ............................................................................................................... 23
  2.8 Nucleotide sequence accession numbers ................................................................................... 24
  2.9 Quantitative PCR ........................................................................................................................ 24
  2.10 Data analysis .............................................................................................................................. 27

III Results ........................................................................................................................................... 29
  3.1 Chlorophyll \(a\) analysis ................................................................................................................ 29
  3.2 PCR primer design analysis ......................................................................................................... 30
  3.3 PCR results and sequence analysis .............................................................................................. 31
  3.4 qPCR results ............................................................................................................................... 36
  3.5 Autocorrelation analysis of Lake Ontario abundance patterns ................................................. 44
List of Tables

Table 1. Summary of Chlorella viruses and the exsymbiotic host they infect .................. 11
Table 2. PCR primers used in this study and their properties ........................................ 31
Table 3. Percent nucleotide identities retrieved from GenBank using BLAST ..................... 36
Table 4. Quantitative PCR primers and probes ............................................................... 37
Table 5. Properties of sequences used to test primer and probe specificity ......................... 37
List of Figures

Figure 1. Viruses are catalysts for biogeochemical cycling ................................................................. 2

Figure 2. Various images of *Chlorella* virus PBCV-1 ........................................................................ 5

Figure 3. Sampling locations .................................................................................................................. 15

Figure 4. Partial sequence alignment of *Chlorella* virus polB genes used in primer design...........18

Figure 5. Generic map of phycodnavirus B-family (α-like) DNA polymerase ...............................19

Figure 6. Chlorophyll *a* abundance in Lake Ontario and the UTM SWM pond .........................30

Figure 7. A neighbour-joining tree .....................................................................................................34

Figure 8. A maximum likelihood tree .................................................................................................35

Figure 9. Quantitative PCR inhibition by the UTM SWM pond samples .......................................38

Figure 10. Quantitative PCR results for unknown *Chlorovirus* target LO.33 ............................40

Figure 11. Quantitative PCR results for the CVM-1-like target LO.41 ........................................42

Figure 12. Correlogram for unknown *Chlorovirus* target LO.33 ................................................44

Figure 13. Correlogram for the CVM-1-like target LO.41 ..............................................................45
Abbreviations

ΔG- Gibbs free energy
°C- degree(s) Celsius
µ- micro (10⁻⁶)
µL- microlitre(s)
µm- micrometres
µM- micromolar
aLRT- approximate Likelihood Ratio Test
ASFV- African swine fever virus
ATCV-1- Acanthocystis turfacea Chlorella virus 1
ATCVS- primer set that amplifies polB fragment of the ATCV-1 virus
BLAST- Basic Local Alignment Search Tool
bp- base pairs
CbV-PW1- Chrysochomulina brevifilum virus PW1
CHLV- Chlorovirus degenerate primer set
cm- centimetres
CO₂- carbon dioxide
Ct- cycle threshold
CVM-1- Paramecium bursaria Chlorella virus 1
Marburg
CVMS- primer set that amplifies polB fragment of the CVM-1 virus
DNA- deoxyribonucleic acid
dNTP- deoxyribonucleotide triphosphate
DOM- dissolved organic matter
EDTA- ethylenediaminetetraacetic acid
EfasV-1- Ectocarpus fasciculatus virus 1
EhV86- Emiliania huxleyi virus 86
EsV-1- Ectocarpus siliculosus virus 1
FAM- 6-carboxyfluorescein
FR 483- Chlorella Pbi virus
FsV-1- Feldmannia simplex virus 1
g- gram(s)
h- hour
HaV01- Heterosigma akashiwo virus 01
HCl- hydrogen chloride
HDPE- high-density polyethylene
hrs- hours
HVCV-1- Hydra viridis Chlorella virus 1
IPTG- isopropyl β-D-1-thiogalactopyranoside
kb- kilobase pairs
kcal- kilocalorie(s)
KCl- potassium chloride
L- litre(s)
m- metres
M- molarity
mg- milligram
MgCl₂- magnesium chloride
min- minute(s)
ML- maximum likelihood
mL- millilitre
mM- milimolar
mm-millimetre(s)
mol- mole(s)
MpV- Micromonas pusilla virus
n- nano(10⁻⁹)
NCLDV- nucleocytoplasmic large DNA virus
ng- nanogram(s)
NJ- neighbour-joining
nM- nanomolar
nm-nanometres
NY-2A- Chlorella NC64A virus
OsV5- Ostreococcus tauri virus V5
OTU- operational taxonomic unit
PBCV-1- Paramecium bursaria Chlorella virus 1
PBCVS- primer set that amplifies polB fragment of the PBCV-1 virus
PCR- polymerase chain reaction
PFU- plaque forming units
PgV-03T- Phaeocystis globosa virus 03T
polB- B-family DNA polymerase
POM- particulate organic matter
vis- visible
xg- acceleration due to gravity
X-gal- bromo-chloro-indolyl-galactopyranoside
YGDTDS- amino acids encoding for catalytical domain of a viral DNA polymerase
λ DNA- lambda phage DNA

qPCR- quantitative real-time polymerase chain reaction
RC- reverse complement
ROX- 6-Carboxyl-X-Rhodamine
rpm- revolutions per minute
rRNA- ribosomal RNA
SAG- culture collection of algae at Goettingen University, Germany
s- second(s)
TAE buffer- Tris base, acetic acid and EDTA
Tm- melting temperature
U- unit(s)
UTM SWM- University of Toronto Mississauga storm water management pond
UV- ultraviolet light
V- volt(s)
VC- viral concentrate
Preface

Results obtained in this study have been included in the following two manuscripts submitted for publication:


I Introduction

1.1 Viruses in the aquatic ecosystems: Why are they important?

The importance of viruses in aquatic environments has become evident with their first discovery in the early 1960s (Safferman and Morris, 1963). Ever since, the growing body of literature on viruses reported that they were active components of marine and freshwater ecosystems (Wilhelm and Suttle, 1999). Recent estimates of ca. $10^{30}$ viruses in the world’s oceans indicate that they are the most abundant biological entities and following bacteria, these viruses form the second largest carbon reservoir on Earth (Suttle, 2005). It is now well established that viruses are significant agents of phytoplankton mortality, global biogeochemical cycles and can also structure algal community dynamics (Thingstad et al., 1993; Bratbak et al., 1994; Wilhelm and Suttle, 1999; Brussaard, 2004; Yamada et al., 2006; Suttle, 2007).

Over half of the total inorganic carbon in the atmosphere is fixed by photoautotrophic bacteria and eukaryotic microalgae (Giovannoni and Stingl, 2005). These microorganisms are responsible for driving the cycle of matter and energy in aquatic ecosystems and have been regarded as key constituents of the biological pump (Brussaard, 2004; Martinez et al., 2007). The biological pump is an amalgamation of biologically mediated processes that lead to the sequestration of carbon into the ocean’s interior as organic matter sinks from the surface of euphotic zones (Chisholm, 2000; Suttle, 2007). It has been proposed that viruses might decrease the biological pump efficiency by short-circuiting the transfer of organic carbon and other essential elements from phytoplankton to higher trophic levels (Suttle, 2005) (Figure 1). It is hypothesized that viral lysis decreases the efficiency of the biological pump by increasing the concentration of dissolved organic matter (DOM; i.e., cytoplasm contents) and particulate organic matter (POM; i.e., cell wall, lipids, etc.) fractions that are inefficiently recycled back into
the main food chain via microbial loop (Fuhrman, 1999). The net result of viral lysis is an increase in respiration of organic carbon that would otherwise sink from the surface into deep waters, where it retains for thousands of years (Chisholm, 2000).

Figure 1. Viruses are catalysts for biogeochemical cycling. Photoautotrophic bacteria and green algae utilize inorganic carbon (such as CO$_2$) and fix it into organic compounds by photosynthesis in order to support new biomass and cellular respiration. Some of this energy will eventually be transferred to higher trophic levels via the main food chain. However, when viruses cause cell lysis, they divert biomass of their hosts (phytoplankton and bacteria) away from higher trophic levels. Instead, the living biomass is effectively converted into particulate and dissolved organic matter (POM and DOM respectively) fractions, which are then utilized by heterotrophic bacteria (as well as Archaea) and other phytoplankton cells. The net outcome is an increase in carbon respiration, which decreases the transfer efficiency of nutrients and energy to higher trophic levels within the aquatic food web. Adapted from Suttle (2005).
On the other hand, another hypothesis has recently suggested that viruses might enhance the efficiency of the biological pump (Suttle, 2007). Viral destruction of bacterial and phytoplankton cells causes them to unequally partition into DOM and POM pools that differ in their chemical content (Suttle, 2007). DOM in comparison to POM contains lower carbon:nitrogen and carbon:phosphorus ratios and is usually retained in the upper water column where it can become a good source of nutrients for phytoplankton growth (Suttle, 2007). POM, on the other hand, is carbon-enriched and heavy, thus it can sink to the bottom of the ocean and in this way, enhance the efficiency of the biological pump (Suttle, 2007). Current evidence for viral lysis and the pump suggests that world’s oceans are net photosynthetic (Suttle, 2007); however, with increasing anthropogenic input of carbon and limiting nutrients into aquatic environments, a shift to net respiration is possible.

It is apparent that viruses are very important players in the global ecosystem and can have a crucial impact on the composition of phytoplankton communities. Viral predation in conjunction with protist grazing is able to maintain microbial numbers at values less than the carrying capacity of the system, thereby promoting the biodiversity of species (Brussaard, 2004; Suttle, 2007; Rohwer et al., 2009). The importance of viruses infecting the phytoplankton is enhanced considerably as they have been implicated in algal bloom termination (Bratbak et al., 1993; Nagasaki et al., 1994; Brussaard et al., 2005). Studies related to this topic demonstrate that viral lysis accounts for a significant percentage of daily mortality of *Synechococcus* spp. (Suttle and Chan, 1994) and lysis of eukaryotic algae (*Emiliania huxleyi*) can account for up to 28% of total daily mortality (Bratbak et al., 1993). As a result, research on viruses and virus-mediated processes is of great importance to our understanding of aquatic biology, global biogeochemistry and sustainability of ecologically important groups of organisms on Earth.
1.2 Phycodnaviruses: An overview

Most phytoplankton viruses that infect marine and freshwater eukaryotic algae are members of Phycodnaviridae, a family of large double-stranded DNA viruses. This family consists of six genera (Chlorovirus, Coccolithovirus, Prymnesiovirus, Phaeovirus, Prasinovirus and Raphidovirus), which were named after the eukaryotic algae that they infect (Van Etten et al., 2002; Yamada et al., 2006; Wilson et al., 2009). Members of the genus Chlorovirus are known to infect freshwater algae, whereas members of the other five genera infect marine algae (Dunigan et al., 2006; Fitzgerald et al., 2007c). All viruses assigned to this family are characterized by an icosahedral morphology with a diameter of 100-200 nm, lack of tails and a genome size of 100-560 kb (Van Etten et al., 2002; Brussaard, 2004; Larsen et al., 2008; Wilson et al., 2009) (Figure 2). Phycodnaviruses are ubiquitous in aquatic environments and have been isolated from many locations worldwide, including the oligotrophic and eutrophic freshwater, and marine ecosystems, as well as sediments (Cottrell and Suttle, 1991; Bratbak et al., 1996; Sandaa et al., 2001; Castberg et al., 2002; Brussaard et al., 2004).

Of particular importance to the research project described here, are the members of the genus Chlorovirus (or Chlorella viruses), that infect freshwater Chlorella-like endosymbiotic green algae (or zoochlorellae) freshly released from their symbiotic hosts (Kang et al., 2005; Fitzgerald et al., 2007c) (Figure 2). Chlorella viruses are without a doubt, the most well characterized viruses from the family of Phycodnaviridae and studies on cultivated Chloroviruses have provided a great deal of information on their physiology, genetics and virus-host interactions (Reisser and Kapaun, 1991; Van Etten and Meints, 1999; Van Etten et al., 2002; Wilson et al., 2009).
Figure 2. A) Three-dimensional representation of *Chlorella* virus PBCV-1 reconstructed from images produced by electron microscopy. Note an icosahedral shape of the virus. (B) The PBCV-1 virus-like particles visible inside the host at 3-4 hrs post infection (C) Release of the PBCV-1 virus from a symbiotic host at 6-8 hrs post infection. Images from Van Etten et al. (2002).

Although numerous efforts have been made to study the biology of algal viruses in aquatic environments, the information on the ecological importance of these “tiny giants” in freshwaters remains limited (Wilhelm and Matteson, 2008; Short and Short, 2009; Wilson et al., 2009). In fact, the majority of information available to date on the role of viruses in natural ecosystems comes from studies of marine viruses and their hosts, most of which are based on bacteria-phage systems (Hennes and Suttle, 1995; Noble and Fuhrman, 1998; Ortmann and Suttle, 2005).

Recently, several efforts have been made to study the community compositions and dynamics of bacteriophages and phycodnaviruses in both marine and freshwater environments (Short and Suttle, 2003; Goddard et al., 2005; Kang et al., 2005; Short and Suttle, 2005; Wilhelm et al., 2006; Lymer et al., 2007; Martinez et al., 2007; Clasen et al., 2008; Short and Short, 2008, 2009). Some of these studies have also attempted to identify potential phytoplankton hosts for uncultivated algal viruses in freshwaters (Clasen and Suttle, 2009). The development of algal
virus-specific (Chen and Suttle, 1995; Chen et al., 1996; Larsen et al., 2008) and cyanophage-specific (Fuller et al., 1998) genetic markers were among the greatest breakthroughs in the field of aquatic viral ecology, because they allowed direct examination of the genetic richness of natural virus communities (Short and Suttle, 2005; Sandaa and Larsen, 2006; Larsen et al., 2008; Short and Short, 2008) and provided the means to investigate the ecological role of viruses in any aquatic environment.

Despite a recent increase of interest in aquatic viruses, knowledge of the mechanisms underlying virus production, even in marine environments, remains sparse. The interaction between viruses and their hosts in nature, particularly what governs virus replication, remains an open question. Numerous studies have suggested that the majority of prokaryotic viruses in the oceans are temperate and lie dormant within their hosts awaiting to be induced by lysogens such as ultraviolet light and other DNA damaging agents (Bratbak et al., 1990; Heldal and Bratbak, 1991; Bratbak et al., 1992; Wilson and Mann, 1997). Conversely, it has been proposed that bacteriophage production in seawater occurs through both processes, lysogen induction and cellular lysis and factors, such as, nutrient availability and multiplicity of infection (the ratio of viruses to host cells) govern lytic replication or lysogeny (Wilson and Mann, 1997). Nevertheless, a study conducted by Wilcox and Fuhrman (1994) demonstrated that when natural bacterial communities were exposed to unattenuated sunlight, lytic infection rather than lysogenic induction was responsible for bacteriophage production in coastal seawaters. Similarly, it has been suggested that most (if not all) phycodnaviruses infecting Chlorophyte and Raphidophyte algae propagate exclusively via lytic pathway (Cottrell and Suttle, 1995; Lawrence et al., 2002; Kang et al., 2005) and latent infections have only been observed for phycodnaviruses infecting Phaeophyte algae (Wilson et al., 2009).
It is still unclear how lytic viruses can survive and propagate when their hosts are not present or are simply below detection limits (Short and Short, 2009). The notion of an “expiration date” for viral capsids in aquatic environments outside of their phytoplankton hosts has been well documented within the literature (Wommack and Colwell, 2000). Studies on bacteriophage particle decay reported rates as high as 1.1 h$^{-1}$ in marine environments and up to 0.6 h$^{-1}$ in freshwater lakes (Heldal and Bratbak, 1991). Similarly, work on decay rates of the phycodnavirus *Micromonas pusilla* virus (MpV), reported losses of infectious particles up to 7.1 day$^{-1}$ (0.296 h$^{-1}$) when viruses were incubated in unattenuated sunlight (Cottrell and Suttle, 1995).

Subsequently, maintenance of a stable viral population requires a specific number of host cells, yet abundances of individual phytoplankton species wax and wane throughout the seasons (Sommer, 1989; Reynolds, 2006). As host densities decline, the time required for virus contact increases and the chance that virus particle will be destroyed before it infects its host also increases (Wommack and Colwell, 2000). Studies on bacterioplankton abundance in the Southern California Bight waters revealed that concentrations between $4.7 \times 10^5$ and $7.2 \times 10^5$ bacteria mL$^{-1}$ were necessary for measurable levels of virioplankton production (Steward et al., 1992). Similarly, Cottrell and Suttle (1995) estimated that concentrations of $1.1 \times 10^3$ to $5.4 \times 10^3$ *Micromonas pusilla* cells mL$^{-1}$ were required to support the MpV production rate in coastal waters in the Gulf of Mexico. Although threshold densities of host cells required for viral propagation vary greatly with location and with individual virus-host systems (Wommack and Colwell, 2000), it is generally accepted that viral abundance and the rates of virus infection are a function of host abundance (Short and Short, 2009). Therefore, minimum densities of hosts are required to maintain a steady-state concentration of infectious viruses (Short and Short, 2009).
1.3 *Chlorella* viruses and their phytoplankton hosts

In order to have a better understanding of the ecological roles of phycodnaviruses in freshwater environments, the development of an easily studied model system is essential. To this day, many excellent studies on phycodnavirus biology utilized cultivation-dependent approaches (Allen and Wilson, 2008). However, such laboratory-based methods offered little insight into virus diversity, virus-host interactions *in situ* and their function within aquatic ecosystems. As a result, cultivation-independent methods have become increasingly more common in the study of virus ecology. Therefore, the impetus for this research was to establish a local virus-host model system that would enable the examination of virus-algae interactions *in situ* in freshwaters via cultivation-independent approaches. Presently, the best genetically and biologically characterized freshwater virus-host system in culture remains *Chlorella* virus PBCV-1 and its exsymbiotic host *Chlorella* NC64A (Wilhelm and Matteson, 2008). Thus, the development of molecular methods that can identify specific host-virus relationships *in situ* is tractable and this well-studied culture system could become a potential candidate for assessment of complex interactions between viruses and host cells in nature.

*Chlorella* viruses, like their hosts, are common in freshwaters and have been isolated from many geographic locations worldwide (Schroeder et al., 2003; Sheath and Wehr, 2003; Brussaard, 2004; Kang et al., 2005; Yamada et al., 2006; Fitzgerald et al., 2007c). These viruses are strain-specific (Van Etten and Meints, 1999) and currently, the only known hosts for *Chlorella* viruses are exsymbiotic zoochlorellae (Kang et al., 2005). These are small ovoid, unicellular, asexually reproducing, non-motile cells found in symbiotic association with many classes of protozoans including *Ciliata*, *Hydrozoa*, and *Rhizopoda* (Prescott, 1951; Kawakami and Kawakami, 1978; Van Etten and Meints, 1999; Bubeck and Pfitzner, 2005; Kang et al.,
Most endosymbiotic algae assigned to the genus *Chlorella* can be readily grown in laboratory cultures free of both their symbiotic hosts and their virus, providing the opportunity to manipulate and study host-virus interactions under controlled environmental conditions (Van Etten et al., 1983; Kang et al., 2005).

Presently, four species have been isolated within the genus of *Chlorovirus* that consists of NC64A, Pbi, SAG 3.83 and Hydra viruses. The NC64A viruses (e.g., PBCV-1) infect *Chlorella* strain NC64A, which was isolated from the *Paramecium bursaria* (*Ciliata*) symbiont. These viruses were originally collected in the North American freshwaters, but were later also detected in freshwaters of Europe, Asia, South America and Australia (Van Etten et al., 2002; Kang et al., 2005). The Pbi viruses (e.g., CVM-1) infect *Chlorella* strain Pbi, which was also isolated from the *P. bursaria*. These viruses were initially collected from freshwaters in Europe, but were later detected in freshwaters of North America (Reisser et al., 1988; Kang et al., 2005). The SAG 3.83 viruses (e.g., ATCV-1) infect exsymbiotic *Chlorella* strain SAG 3.83, which is normally a symbiont in the heliozoon *Acanthocystis turfacea* (*Rhizopoda*). These viruses were collected from freshwaters of Europe and the Americas (Bubeck and Pfitzner, 2005; Fitzgerald et al., 2007c). Lastly, the Hydra viruses (e.g., HVCV-1) infect exsymbiotic *Chlorella*-like algae (several strains) isolated from coelenterate *Hydra viridis* (*Hydrozoa*). These viruses were originally collected from North American freshwaters, but were later detected in freshwaters of Europe (see Table 1 for summary) (Van Etten et al., 1982; Van Etten et al., 1983; Van Etten et al., 1991).

All hosts for *Chlorella* viruses known to date came from laboratory cultures of exsymbiotic green algae originally isolated from their symbiont. When *Chlorella* cells are inside their protozoa, they are resistant to viral infection and can be only infected when they are isolated from their symbiont (Kawakami and Kawakami, 1978; Reisser and Kapaun, 1991).
Presently, it is unknown if zoochlorellae exist free of their symbionts in nature (Kang et al., 2005) and the extent of host diversity for *Chlorella* viruses in the environment is also unknown. Studies of green paramecia indicate that they are rare in freshwaters (Kang et al., 2005) suggesting that *Chlorella* viruses replicate in nature by infecting other host(s) and perhaps have happened to infect the zoochlorellae (Van Etten et al., 1985). In addition, *Chlorella* species form a polyphyletic assemblage (Nakahara et al., 2004) and so far, the taxonomic positions of these coccoid green algae have not been satisfactorily resolved (Krienitz et al., 2004). Further investigations are required to identify other potential hosts for these viruses in nature and to determine whether these viruses exclusively infect algae from the genus of *Chlorella*. It should also be noted that the studies on *Hydra viridis* viruses and their hosts remain very limited, because these viruses were lost from culture many years ago and are no longer available for study (D. Dunigan, pers. comm.).

Complete genome sequences and annotations were previously reported for PBCV-1 and ATCV-1 viruses (Van Etten et al., 1991; Fitzgerald et al., 2007c). Similarly, whole genomes have been sequenced for several *Chlorella* Pbi and NC64A viruses (Fitzgerald et al., 2007a; Fitzgerald et al., 2007b) and some sequence information is available for the CVM-1 virus in GenBank database. Given that our laboratory maintains cultures of CVM-1, PBCV-1 and ATCV-1 viruses, these three became the focus of my research project. Genome sizes for these three plaque-forming, nucleocytoplasmic viruses have been previously reported to range from 288 kb for ATCV-1 to >300 kb for PBCV-1 and CVM-1 (Table 1) (Van Etten and Meints, 1999; Kang et al., 2005; Yamada et al., 2006). Typical virus titres in freshwaters for *Chlorella* viruses are from 1-100 plaque-forming units (PFU) per millilitre with peak abundance during spring months; however, titres as high as 100 000 PFU per millilitre have been reported for some freshwater environments (Dunigan et al., 2006).
Table 1. Summary of *Chlorella* viruses and the exsymbiotic host they infect

<table>
<thead>
<tr>
<th>Virus Type</th>
<th>Known host(s)</th>
<th>Symbiont from which the host was isolated</th>
<th>Genome size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVM-1</td>
<td><em>Chlorella Pbi</em></td>
<td><em>Paramecium bursaria</em></td>
<td>&gt; 300,000</td>
</tr>
<tr>
<td>PBCV-1</td>
<td><em>Chlorella NC64A</em></td>
<td><em>Paramecium bursaria</em></td>
<td>330,792</td>
</tr>
<tr>
<td>ATCV-1</td>
<td><em>Chlorella SAG 3.83</em></td>
<td><em>Acanthocystis turfacea</em></td>
<td>288,047</td>
</tr>
<tr>
<td>HVCV-1</td>
<td><em>Chlorella</em> (several strains)</td>
<td><em>Hydra viridis</em></td>
<td>≥ 250,000</td>
</tr>
</tbody>
</table>

Viruses assigned to the genus *Chlorovirus* encode many genes with highly conserved amino acid sequences (Wilson et al., 2009). Provided with such an advantage, the development of *Chlorella* virus-specific markers that amplify the B-family (α-like) DNA polymerase gene (polB) in viruses may be possible. The gene polymorphisms within polB correlate well with loose taxonomic groupings of phycodnaviruses and prove to be sensitive enough to discriminate between strains of viruses with little genetic difference (Wommack and Colwell, 2000). Therefore, the development of polymerase chain reaction (PCR) methods that can target *Chlorella* virus polB genes in natural environments has great potential for identifying specific host-virus relationships *in situ*.

Previous study by Short and Short (2009) demonstrated that a quantitative real-time PCR (qPCR) assay based on the TaqMan® chemistry, can be a very sensitive, highly specific and reliable method to monitor abundance of individual phycodnavirus genes in natural environments. Given that viral infection and propagation is a function of host abundance (Wommack and Colwell, 2000; Brussaard, 2004; Short and Short, 2009), the development of *Chlorella* virus-specific markers to monitor the abundance of individual virus genes in the freshwater environment may lead to a better understanding of how *Chlorella* viruses propagate...
in nature. It may be possible to relate the changes in the virus gene abundance to changes in the gene abundance of individual *Chlorella* species. By coupling dynamics of *Chlorella* viruses to their candidate hosts, the identification of host-virus pairs in natural systems should be feasible. Therefore, in aid of efforts to establish a local virus-host model system and to examine interactions of closely related algal viruses in nature, my Master’s project focused on assessing seasonal dynamics of different *Chlorella* viruses in two contrasting freshwater environments.

### 1.4 Thesis objectives

Gene fragments nearly identical to DNA polymerase sequences from viruses infecting *Chlorella* Pbi (e.g., CVM-1 and FR 483) were previously obtained from Lake Ontario, Canada (S. M. Short, unpublished data). Subsequent quantitative study on these Chloroviruses (e.g., CVM-1) at a single location in Lake Ontario over a period of 14 months (S. M. Short, unpublished data) revealed the presence of persistent viruses with peak abundances emerging during late spring and early summer. Following these findings, the intention was to extend the intra-annual dynamics of the CVM-1-like viruses in Lake Ontario and to resolve important research questions addressed hereafter. More specifically, the objectives of this study were to determine if:

1) there are different types of Chloroviruses in Lake Ontario and the nearby storm water management pond on the University of Toronto Mississauga’s campus (from here on out referred to as ‘UTM SWM’).

2) different *Chlorella* viruses share similar patterns of seasonal abundance within individual environments and if their seasonal abundances are comparable to the patterns observed in previously published reports of phycodnavirus seasonality (Short and Short, 2009).
3) the seasonal abundance patterns of individual viruses in different environments differ.

To this end, new *Chlorovirus*-specific (targeting PBCV-1 and ATCV-1 viruses) and degenerate (targeting CVM-1, PBCV-1 and ATCV-1) PCR primers will be designed to amplify genes encoding for a highly conserved amino acid region of a viral DNA polymerase gene. Consequently, clone libraries of *Chlorella* virus polB gene fragments will be constructed from Lake Ontario and the UTM SWM pond samples for the purpose of developing quantitative assays (qPCR) to estimate the abundance of individual viruses. Thereafter, a data analysis, which includes an autocorrelation analysis of seasonal abundance patterns produced by two *Chlorella* virus targets in Lake Ontario, will be conducted.
II Materials and Methods

2.1 Sample collection and preparation

Surface water samples were collected weekly (with the exception of occasional weeks when collection of water sample was not possible due to inclement weather and ice cover) from two locations in Ontario, Canada (Figure 3) using a bucket and 1L acid rinsed HDPE bottles. Water samples were collected from Lake Ontario at a jetty in Mississauga, Ontario, Canada (43° 32.614’N 79° 34.995’W, elevation 74 m) from January 2008 to June 2010. Likewise, surface water samples were collected from a small, newly established (ca. 2008) storm water management (SWM) pond located on the University of Toronto, Mississauga’s (UTM) campus (43° 32’35.74’N 79° 39’35.53’ W, elevation 105 m) between May 2009 and June 2010. Immediately following the collection, all samples were transported to the laboratory for further processing. The virus size fraction was separated by gravity filtration through 115 mm diameter 35 µm monofilament nylon screen (Nitex mesh, Tetko Inc., Briarcliff Manor, USA) and then vacuum-filtered (at 100 mm Hg) through a 47 mm diameter 0.45 µm pore-size Durapore® membrane filter (Millipore, Billerica, USA).

From each sample, 72 mL (2 × 36 mL) of filtrate was centrifuged in a SW 32Ti rotor (Beckman Coulter, Fullerton, USA) at 20°C for 3.5 hrs at 164,139 x g. Following centrifugation supernatants were decanted by rapidly inverting the centrifuge tubes and allowing them to drain completely. 300 mL of 10 mM Tris-HCl (pH 8.0) was added to each centrifuge tube and the pelleted material was left to soak overnight at 4 ºC.
After soaking overnight, the pelleted material was vortexed for 30 s, resuspended into solution using a pipettor and transferred (ca. 600 µL) into pre-weighed 2 mL screw-cap micro centrifuge tubes. The tubes containing concentrates were weighed again to determine the volumes of resuspended samples and virus concentration factors. The concentration factors were estimated by dividing the total volume spun (i.e., 72 000 µL) by the determined volumes of resuspended samples.

As a final step in the preparation of viral concentrates (VCs) for PCR reactions, all samples were subjected to hot and cold treatment. This procedure was conducted to increase PCR amplification yields through the release of viral DNA from capsid particles upon the treatment (Chen et al., 1995). The hot/cold procedure consisted of alternating four heating cycles at 95°C for 2 min with three freezing cycles at -20°C for 15 min and a final treatment on ice at 0°C for 2 min. All VCs and hot/cold preparations were placed at -20°C for long-term storage.
2.2 Chlorophyll a filter extractions and fluorescence measurements

During weekly sample collections, 25-100 mL of water from each sample bottle was vacuum filtered (100 mm Hg) onto GF75 glass-fiber filters (nominal pore size 0.3 µm) (Advantec MFS, Dublin, USA), which were wrapped in aluminum foil and stored at -20°C. Chlorophyll a extraction and fluorometric analysis were based on a previously described method (Welschmeyer, 1994). Prior to extractions, all lights in the work area were dimmed and the equipment was rinsed with distilled water and 90% acetone. The following protocol was performed in batches of five samples so as to minimize the time the samples spent outside of refrigeration. Briefly, glass-fiber filters containing filtrate were thawed on ice packs in a closed Styrofoam container. Each filter was placed inside a clean glass grinding tube containing 4-5 mL of 90% acetone and the grinding was performed by a grooved pestle. The pestle was attached to a bench drill press and minced the contents with its revolving speed of 900 rpm. The grinding procedure continued until the filter was completely homogenized, after which all contents were transferred into a pre-labelled 15 mL falcon tube and the total solvent volume was brought up to 10 mL. The falcon tube was shaken vigorously, vortexed briefly and placed inside the Styrofoam container with ice packs until the entire batch (of five samples) was finished.

Once the entire batch of samples had been extracted, the tubes (covered in aluminum foil) were incubated for 4 hrs at 4°C. Following incubation, the samples were vortexed briefly and centrifuged at 1000 x g for 5 minutes in an Eppendorf Centrifuge 5810R (Eppendorf, Hauppauge, USA). Chlorophyll a measurements were conducted using the Turner Designs laboratory fluorometer (Turner Designs, Sunnyvale, USA) under a Chl-NA module. Before each measurement, the fluorometer calibration was checked with a secondary solid standard (37-39 µg/mL) and the concentrations in each sample were estimated from a saved calibration file.
2.3 Cultivated *Chlorella* virus sequence analysis and PCR primer design

In order to examine if different *Chlorella* viruses were present in Lake Ontario and the UTM SWM pond, the design of *Chlorella* virus-specific and degenerate primer sets was required. Prior to primer design, DNA polymerase sequences available for several cultivated *Chlorella* viruses were retrieved from GenBank database. The GenBank accession numbers for all sequences used in this study are listed below. After initial sequence curation involving deletion of an intron sequence in PBCV-1 virus (as previously described by Short and Short (2008)), poIB nucleotide sequences of CVM-1, PBCV-1 and ATCV-1, along with other cultivated *Chlorella* viruses were aligned using multiple sequence alignment program ClustalW (Thompson et al., 1994) with Mega 4.0 (Tamura et al., 2007) default parameters (Figure 4). Aligned sequences were used to design specific (PBCV-1 and ATCV-1) and degenerate (amplify CVM-1, PBCV-1, ATCV-1, etc.) *Chlorella* virus PCR primers that targeted genes encoding for a highly conserved amino acid region of a viral DNA polymerase catalytic domain (Figure 5). The CVMS primers (that amplify CVM-1 poIB gene fragments) were designed by S.M. Short (unpublished data) and were generously provided for this study.

Using OligoAnalyzer© 3.1 (Integrated DNA Technologies, Coralville, USA), oligonucleotide primers were analyzed and optimal oligonucleotides were selected based on the following parameters: 1) the length of PCR primers varied between 20-30 bp; 2) the melting temperature ($T_m$) ranged from 53 - 60.5°C with no more than a 5°C difference between upstream and downstream primers; 3) the GC content of each PCR primer ranged between 40-60%; 4) consecutive G’s or C’s at the 3’ end of the primer were avoided; 5) oligonucleotide secondary structure was avoided. More specifically, for the hairpin formation at the 3’ end of the primer, the maximum $\Delta G$ of -2 kcal/mol was tolerated. Similarly, for the internal hairpin formation, a
ΔG of ≤ -3 kcal/mol was accepted. In addition, ΔG for the self-dimer did not surpass 50% of the maximum ΔG and for the cross-dimer, no more than 25% of the maximum ΔG was tolerated; 6) the degeneracy factor for forward and reverse primers did not surpass 64.

Figure 4. Partial sequence alignment of *Chlorella* virus polB genes used in primer design. *Chlorovirus* polB fragments were aligned using multiple sequence alignment program ClustalW with Mega 4.0 default parameters. Numbers on top of alignments represent nucleotide positions relative to PBCV-1 genome. Sequence alignments are presented in 5‘-3’ direction. A) Alignment of three *Chlorella* viruses for which specific primers were designed. Residues highlighted in colour represent the forward and reverse PCR primers. B) Sequence alignment of six different *Chlorella* viruses used to design degenerate primers that could amplify CVM-1, PBCV-1 and ATCV-1. Blocks shaded in purple indicate the region from which forward and reverse degenerate primers were acquired. Asterisks below shaded blocks indicate conserved nucleotide regions, whereas the variable nucleotides are shaded with different colors (A-red; G-yellow; T-green; and C-blue).
Figure 5. Generic map of phycodnavirus B-family (α-like) DNA polymerase, showing relative positions and segments amplified by upstream and downstream primers (specific or degenerate). The segment flanked by PCR primers ranges from 560 bp to 645 bp due to difference in size of PCR fragments amplified by different primer sets. The approximate location of residues encoding the highly conserved DNA polymerase amino acid motif “YGDTDS” is 475 bp, and is flanked by the PCR primer locations. Adapted from Chen and Suttle (1995).

2.4 Validation of primer specificity

Independent reactions were conducted to examine the specificity of each specific PCR primer set (CVMS, PBCVS and ATCVS) using 0.45 µm filtered, hot/cold prepped viral lysates from the cultivated viruses CVM-1, PBCV-1 and ATCV-1. Primer specificity was tested by comparing amplification of targets to non-targets (cross-amplification) in a single round of PCR. Cross-amplification tests were conducted in duplicate and the specificity of each primer set was examined by manually inspecting agarose gel electrophoresis of PCR products visualized under the UV light. Primer sets were designated as specific if they exclusively amplified the target for
which they were designed. In addition, the degenerate primer set (CHLV) was also tested for target specificity by examining PCR amplification of polB fragments from the aforementioned virus lysates. Each PCR reaction was replicated three times. The degenerate primer set was considered Chlorovirus specific, if a single PCR product of ca. 600 bp was produced for each Chlorella virus target in every replicate. All PCR reactions were conducted under optimized conditions, described in detail below. Cultures of all Chlorella viruses used in this study were provided to S. M. Short from J. L. Van Etten (Department of Plant Pathology, University of Nebraska-Lincoln; 205 Morrison Center, Lincoln, NE 68583-0900).

### 2.5 PCR amplification

*Chlorella* virus polB fragments were amplified from natural samples collected predominately during late spring or summer (since, according to preliminary findings, the peak abundance of CVM-1 occurred in these months) in a single round of PCR using 50 µL reaction volumes. Each PCR reaction tube contained 5 µL of hot/cold prepped VC, 1.5 mM MgCl₂, 200 µM each deoxyribonucleotide triphosphate (dNTP) and 1U/µL of Platinum® Taq DNA Polymerase (Invitrogen Life Technologies, Carlsbad, USA) with 1x manufacturer supplied reaction buffer (200 mM Tris-HCl (pH 8.4), 500 mM KCl). Nuclease-free water (Integrated DNA Technologies, Coralville, USA) was used to make up full reaction volumes. The concentration of primers per single reaction tube varied for each primer set. For CVMS and ATCVS, 200 nM final concentrations of each forward and reverse primer were included; for PBCVS, a 100 nM final of each was included and for the CHLV degenerate set, a 300 nM final concentration of forward and reverse primer was included. Negative controls contained all reagents except viral DNA; instead, 5 µL of nuclease-free water was substituted for the template. For the positive control, 2µL of template DNA from hot/cold prepped viral lysates of CVM-1,
PBCV-1 and ATCV-1 were used. PCR was carried out using C1000™ thermal cycler (Bio-Rad Laboratories, Hercules, USA), where the cycling protocol differed for each specific primer set. Each PCR reaction began with initiation and denaturation incubation at 95°C for 2 min. For PBCVS and ATCVS primer sets, initiation was followed by 38 cycles of 95°C for 30s, annealing at 44° and 48°C, respectively, for 45s and extension at 72°C for 1 min. For the CVMS primer set, the PCR reaction was carried out using the same conditions as described above, except the number of cycles was lowered to 35 and the annealing temperature was set to 50°C. At the end of cycling, all PCR reactions were subjected to a final extension step at 72°C for 30 min.

Optimization of PCR annealing temperature for the degenerate primer set (CHLV) was necessary to improve amplification yields. PCR reactions were carried out under the following cycle parameters: initial incubation at 95°C for 2 min; 40 cycles of denaturation at 95°C for 30s; an annealing gradient from 46° - 56°C for 1 min and extension at 72°C for 1 min with a final 20 min extension. The negative control contained all reagents except template and was placed at the lowest annealing temperature. Once the PCR annealing temperature was optimized, *Chlorella* virus DNA polymerase amplification was carried out under the same reaction conditions as described previously for the specific primer sets, except the annealing step was changed to 52°C and only 36 cycles were performed.

After thermal cycling, the entire volume of all the PCR reactions were loaded into 1.5% agarose gels (Promega, Madison, USA) and were electrophoresed at 120 V/cm for 65 min in 1x TAE buffer (40mM Tris, 20mM Acetic acid, 1 mM EDTA; pH 8.0). GeneRuler® 100 bp ladder (Fermentas, Maryland, USA) was used as a molecular weight marker. The gel was stained with 0.5 µg/mL ethidium bromide for 30 min and visualized using a Molecular Imager ChemiDoc® XRS System (Bio-Rad Laboratories, Hercules, USA) with a 1.44 megapixel CCD camera. In all
cases, it was ensured that the DNA’s exposure to UV light was minimized. Bands of interest were then excised from the agarose gel and the DNA was extracted using a QIAquick Gel Extraction Kit (Qiagen, Valencia, USA) according to manufacturer’s recommendations. Gel extractions were stored at -20°C until further use.

2.6 Cloning and sequencing Chlorella polB fragments

DNA fragments extracted from the gel were ligated into pGEM®-T vector using a pGEM®-T Vector System II (Promega, Madison, USA) kit according to the manufacturer’s recommendations. Plasmids from the ligation mixture were used to transform competent Escherichia coli JM109 cells (Promega, Madison, USA) using a heat shock treatment according to the manufacturer’s protocol. After the transformation, 150 µL of recovered cells were incubated overnight at 37°C on LB plates containing 1.5% agar and 100 µg/mL carbenicillin and spread with 20 µL 50 mg/mL X-Gal and 100 µL 0.1M IPTG. A rapid, PCR-based colony screening was performed to identify which clones had taken up the plasmids containing an insert of interest. Reaction conditions for colony screening were the same as described above, except that bacterial cells were used as templates by transferring directly tooth-picked cells from the edge of a white colony into a PCR reaction mixture. Following the PCR-based screening, colonies of interest were used to set up overnight cultures for the purpose of proceeding to miniprep. Bacterial cultures were suspended in 5 mL of LB broth containing 100 µg/mL carbenicillin and incubated overnight for 15-16 hrs at 37°C with continuous agitation at 300 rpm. After incubation, screened colonies were purified using QIAprep Spin Miniprep Kit (Qiagen, Mississauga, ON, Canada) and quantified via a NanoDrop 1000 UV-vis spectrophotometer (NanoDrop Technologies, Wilmington, USA). The miniprepped plasmids were sent out for automated sequencing at the McGill University and Genome Quebec Innovation Centre in
Montréal, QC, Canada. Only sequences that covered the entire ca. 600-700 bp DNA polymerase gene fragment were used for further analysis. A highly conserved polB amino acid motif, YGDTS, was used to screen for sequences encoding for DNA polymerase (Figure 5) and sequences that consisted of vague base-calls were excluded.

2.7 Virus sequence analysis

DNA polymerase sequences recovered from Lake Ontario and the UTM SWM pond using Chlorovirus specific and degenerate primers were analysed phylogenetically. Prior to phylogenetic analysis, all sequences were edited and translated using BioEdit version 7.0.9.0 (Hall, 1999) and the plasmid sequences were removed from polB fragments obtained. All DNA sequences from cloned PCR fragments were aligned using the multiple sequence alignment program ClustalW with Mega 4.0 default parameters. A neighbour-joining (NJ) tree of pairwise p-distance (proportion of differences observed between sequences) was constructed to determine the unique (i.e., non-redundant) sequences that would be used for more robust phylogenetic analysis. Clade credibility for NJ was assessed using bootstrap values inferred from 500 replicates with random seed 64238. Additionally, using BioEdit, a pairwise DNA identity matrix was constructed from the sequences recovered and only one sequence from any cluster (created by NJ) that shared ≥ 99 % identity on the nucleotide level was included in further analysis.

Sequences amplified from Lake Ontario and the UTM SWM pond were compared with other polB fragments from cultivated nucleocytoplasmic large DNA viruses (NCLDVs) by reconstructing phylogeny via maximum likelihood. An initial curation involving deletion of intron sequences in PBCV-1 and NY-2A viruses and intein sequences in the Mimivirus and HaV-01 was performed as previously described (Short and Short, 2008). An amino acid alignment of polB fragments obtained from environmental clones and cultivated algal viruses
was constructed using MUSCLE (Edgar, 2004) with Phylogeny.fr (Dereeper et al., 2008) default parameters. All positions containing gaps and missing data were eliminated from the dataset using Gblocks (Castresana, 2000) with Phylogeny.fr default parameters. A maximum-likelihood (ML) phylogeny was reconstructed using PhyML (Guindon and Gascuel, 2003) and the tree reliability was assessed using the approximate Likelihood Ratio Test (or aLRT) statistic (Anisimova and Gascuel, 2006) with the minimum of SH-like and Chi2-based values. Tree viewing and drawing was performed using Mega 4.0 and Adobe Illustrator CS 5 (Adobe Systems, San Jose, USA).

2.8 Nucleotide sequence accession numbers

Only unique sequences (or OTUs, which were defined as DNA sequences that were ≤ 97% similar to any other sequence) were deposited into the GenBank database. A representative OTU sequence for Chlorella Pbi viruses (closely related to CVM-1 or FR483) amplified from either Lake Ontario or the UTM SWM pond (i.e., LO.05Jun08.41) was assigned the accession number HM629734 and an OTU closely related to ATCV-1 virus amplified from the UTM SWM pond (UP.21Sep09.7) was assigned the accession number HM776041. Lastly, the two novel Chlorovirus sequences (LO.20May09.33 and LO.24Jun09.14) obtained from Lake Ontario were assigned accession numbers HM629733 and HM776040, respectively.

2.9 Quantitative PCR

Primers and TaqMan® probes for the qPCR targets chosen in this study were designed from full-length Lake Ontario DNA polymerase sequences using Beacon Designer 7.0 (Premier Biosoft International, Palo Alto, USA). The steps were taken to ensure that primers and TaqMan® probes designed were specific for their targets and did not amplify other gene
sequences from environmental samples. Therefore, in search of the optimal primer and probe set
the software was instructed to identify and avoid regions of homology in each template with
other non-redundant Chlorovirus sequences using BLAST. From the list of the candidate sets
generated by the software, the highest scoring primer and probe pair was selected for each target.
All probes were 5’ labelled with FAM (6-carboxyfluorescein) as a fluorescent reporter and 3’
labelled with Iowa Black® FQ as a quencher (Integrated DNA Technologies, Coralville, USA).
From each sample, the two different Chlorella virus polB targets were amplified in separate
reactions with TaqMan® probes via 5’-nuclease assay. Each 25 μL reaction contained 1 x PCR
buffer (20 mM Tris-HCl (pH 8.4) and 50 mM KCl), 5 mM MgCl₂, 200 μM each dNTP, 400 nM
each forward and reverse primer, 200 nM of TaqMan® probe, 0.625 U/μL of Platinum® Taq
DNA polymerase (Invitrogen Life Technologies, Carlsbad, USA), 30 nM of ROX reference dye
and 2 μL of hot/cold treated VC. All reactions were conducted in an Mx3000 qPCR system
(Stratagene, La Jolla, USA) with the following parameters: 95°C for 5 min, then 40 cycles of
95°C for 15s followed by 60°C for 1 min.

Standard curves for each Chlorella virus target were generated from a set of eight ten-
fold serially diluted standards (ranging from ca. 3.7 × 10⁷ to 3.7 × 10⁰ molecules per reaction)
run in duplicate. The diluent for each standard was sterile, nuclease-free water (Integrated DNA
Technologies, Coralville, USA) that contained λ DNA (9 ng/μL final) as a carrier to enhance
target molecule stability during storage. Standards were cloned fragments of the target polB
molecule. The plasmids containing these fragments were linearized by restriction digest, purified
from undigested plasmids by agarose gel electrophoresis, extracted from agarose gels using a
QIAquick gel extraction kit (Qiagen, Mississauga, ON, Canada) and quantified using NanoDrop
1000 UV-vis spectrophotometer (NanoDrop Technologies, Wilmington, USA). The amplification
efficiency for each primer and probe set was calculated from the slopes of the standard curves of
the target molecules using the following formula: \( E = (10^{\frac{1}{-1/slope}} - 1) \times 100 \). The number of gene copies for specific target sequence in each sample was estimated from the linear regression (cycle threshold vs. standard quantity) equation generated by the standard curves with amplification efficiencies ranging from 99% to 102% and \( R^2 \) values greater than 0.998. Each reaction containing environmental DNA was performed in sets of triplicates and each qPCR experiment included one set of standards for control and at least three no-template controls containing 2 µL of nuclease-free water (Integrated DNA Technologies, Coralville, USA).

Independent experiments were conducted to examine primer and probe specificity and inhibition of amplification by natural samples. Primer and probe specificities were examined by comparing cycle threshold (Ct) values from reactions containing \( 10^7 \) target molecules (linearized plasmid containing targeted polB fragments) to Ct values from reactions containing \( 10^7 \) non-target molecules (linearized plasmids containing most closely related non-target polB fragment). Similarly, inhibition of Platinum® Taq DNA Polymerase (Invitrogen Life Technologies, Carlsbad, USA) by the UTM SWM pond samples during PCR amplification was tested using four different samples. The experiment was designed to ensure that the estimates of target polB copies in the UTM SWM pond were generated by the viruses themselves and were not merely inhibited by natural samples. Therefore, one of the VCs chosen had both qPCR targets (LO.20May09.33 and LO.05Jun08.41) with high gene abundance (sample collected on June 18th, 2009). Another VC that was chosen had LO.05Jun08.41 target present and LO.20May09.33 absent (sample collected on December 1st, 2009). A third VC chosen had LO.20May09.33 present and LO.05Jun08.41 at a very low abundance (July 23rd, 2009) and lastly, the forth VC had both targets at the lowest possible abundance (September 21st, 2009). The inhibition tests were conducted using LO.20May09.33 primers and probe by comparing amplification of \( ca. 3.08 \times 10^5 \) target molecules (linearized plasmid) to amplification of targets plus 2 µL of hot/cold
treated pond VCs. Percent amplification was calculated using the formula \(100 \times (1 - \frac{C_{p+VC}}{C_p})\), as previously described in Short and Short (2009). The statistical significance of estimated mean gene copies in a single reaction was tested using a Chi-Square test in Microsoft Excel 2007 (Microsoft, Redmond, USA).

2.10 Data analysis

Quantitative PCR results were analyzed using Excel software (Microsoft, Redmond, USA). The numbers of gene copies for environmental samples, were adjusted using the VC concentration factors and were reported as gene copies mL\(^{-1}\) of lake or pond water. Error terms (i.e., standard deviation of gene copies for each data point) for target abundances in two environments were estimated from gene copies mL\(^{-1}\) that were inferred from triplicate qPCR reactions. Quantitative PCR inhibition tests were also analyzed using Excel. The numbers of gene copies in individual qPCR reactions were estimated from the linear regression equation generated by the standard curve of the LO.20May09.33 target. Standard deviation for individual inhibition reactions was calculated from triplicate gene copy number estimates. Graphs for chlorophyll \(a\) abundances, quantitative PCR results and inhibition tests were all generated using Excel software.

In addition, abundance patterns produced by the two *Chlorella* virus targets (LO.20May09.33 and LO.05Jun08.41) in Lake Ontario over a period of 28 months were subjected to autocorrelation analysis. Since gaps were present in weekly sample collections (when sample collection was not possible) and following the recommendations of Scheiner and Gurevitch (1993) the data points obtained from weekly quantifications of LO.20May09.33 and LO.05Jun08.41 targets in Lake Ontario were binned (averaged) to give a single abundance value for each 28-day period. This was done to ensure that the missing data from the time-series would
not obscure the overall analysis. A total of 31 bins were created for the two qPCR targets. SPSS 17.0 (SPSS Inc., Chicago, USA) statistical software was used to generate correlograms based on binned average abundance of LO.20May09.33 and LO.05Jun08.41 targets in Lake Ontario.
III Results

3.1 Chlorophyll $a$ analysis

The average chlorophyll $a$ concentration measure at a single location in Lake Ontario over a period of 10 months was 1.64 µg/mL, whereas the UTM SWM average chlorophyll $a$ concentration was 20.09 µg/mL (Figure 6). These averages were based on 27 samples collected from Lake Ontario and 24 samples collected from the UTM SWM pond on the same day during September 2009-June 2010 period. Three concentration measures were missing from the UTM SWM pond data set, because sample collection was not possible due to ice cover present during the months of January, February and the beginning of March. When three data points (from the samples where the pond chlorophyll $a$ measures were missing) were excluded from the Lake Ontario average calculation, the mean chlorophyll $a$ concentration was 1.76 µg/mL. Over the entire study period, the Lake Ontario chlorophyll $a$ concentration ranged from 0.16 µg/mL to 15.09 µg/mL, whereas the chlorophyll $a$ concentration in the UTM SWM pond ranged from 3.63 µg/mL to 101.10 µg/mL.
3.2 PCR primer design analysis

Several *Chlorovirus*-specific (target ATCV-1 and PBCV-1) and degenerate (target CVM-1, ATCV-1 and PBCV-1) polB PCR primer sets were designed from aligned *Chlorovirus* polB genes. OligoAnalyzer© 3.1 was used to analyze all designed oligonucleotides and only optimal primer pairs (Table 2) were selected based on the previously mentioned criteria. Before carrying out PCR amplification in environmental samples, independent reactions were conducted to examine the specificity of each primer set using cross-amplification tests. Electrophoresis of PCR products amplified with *Chlorella* virus-specific (CVMS, PBCVS and ATCVS) primer sets
revealed a single amplified DNA fragment of ca. 600 bp and only reactions with target molecules produced the expected PCR product (data not shown). Subsequently, when the CHLV degenerate primer set was tested against CVM-1, ATCV-1 and PBCV-1 virus lysates, all reactions produced the expected ca. 600 bp fragments. Altogether, cross-amplification tests demonstrated that the primer sets designed in this study were specific for their targets.

Table 2. PCR primers used in this study and their properties

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Target DNApolB gene(s)</th>
<th>Forward primer (5’-3’)</th>
<th>Reverse primer (5’-3’)</th>
<th>Annealing Temperature (°C)</th>
<th>Segment length amplified (bp)</th>
<th>Degeneracy factor (forward/reverse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVMS</td>
<td>CVM-1</td>
<td>AAG AAG GGA GCA TAC TTC AGG C</td>
<td>CAA AAT GTA AGG GTA ATA GAT CTT C</td>
<td>50</td>
<td>645</td>
<td>N/A</td>
</tr>
<tr>
<td>PBCVS</td>
<td>PBCV-1</td>
<td>CTT ATC GCA GCT CTC GAT TTT G</td>
<td>GTT CGG TGC TCG GAA ATC CTT C</td>
<td>44</td>
<td>600</td>
<td>N/A</td>
</tr>
<tr>
<td>ATCVS</td>
<td>ATCV-1</td>
<td>AAG AAA GGT GCC TAC TTT GAA C</td>
<td>AGG TCG TTC GGA GCT TTG TAC T</td>
<td>48</td>
<td>610</td>
<td>N/A</td>
</tr>
<tr>
<td>CHLV*</td>
<td>CVM-1, PBCV-1 and ATCV-1 like</td>
<td>CCW ATC GCA CGW CTT CTM GAT TTT G</td>
<td>ATC TCV CCB GCV ARC CAC TT</td>
<td>52</td>
<td>560 +</td>
<td>8/54</td>
</tr>
</tbody>
</table>

*Degenerate primers contain oligonucleotide symbols consistent with IUB coding for mixed bases

3.3 PCR results and sequence analysis

Electrophoresis of PCR products with PBCVS, ATCVS and CHLV primer sets revealed the presence of numerous bands from several VCs (data not shown). For instance, when tested using various pond or lake virus concentrates some PCR reactions (with all 3 primer sets) produced only the expected ca. 600 bp products, whereas other reactions produced products of various sizes ranging from ca. 500 bp to 1000 bp. On the contrary, PCR reactions with the Chlorella virus-specific primer set, CVMS, always produced a single amplified DNA fragment of ca. 600 bp when tested using various virus concentrates. Nevertheless, PCR reactions with four different primer sets in 42 environmental samples were not always able to amplify DNA fragments from either Lake Ontario or the UTM SWM pond VCs and in most cases
electrophoresis revealed the presence of weakly defined bands from only few VCs. Therefore, only fragments that were ca. 600-700 bp in size were excised from agarose gels, cloned, purified and sent out for sequencing.

A total of 68 sequences were obtained from the two environments under study and only three out of the four different primer sets were able to amplify DNA fragments that encoded the conserved DNA polymerase motif ‘YGDTDS’. From the acquired total of 68, the translation of 16 DNA fragments amplified with PBCVS gave rise to sequences that did not encode polB. A total of 37 sequences were recovered using the CHLV degenerate primer set, out of which only 24 encoded the YGDTDS motif. Out of these 24 putative polB gene fragments, 14 fragments were obtained from Lake Ontario and 10 were acquired from the UTM SWM pond. In addition, sequence information recovered from the seven sequences amplified using CVMS primer set, revealed the presence of YGDTDS amino acid motif in all fragments. Out of these seven, two sequences were amplified from Lake Ontario and the other five sequences were obtained from the UTM SWM pond. Translations of the eight sequences amplified with ATCVS forward and reverse primers, revealed the presence of only one good sequence. Therefore, a total of 32 sequences encoding for DNA polymerase amino acids were recovered from two freshwater environments.

Alignment of these 32 polB fragments and subsequent cluster analysis using neighbour-joining algorithm revealed that many sequences could be placed into groups with more than 99% nucleotide identity (Figure 7). Therefore, a pairwise identity matrix of the 32 sequences (data not shown) was used to generate percent identities for all putative polB gene fragments and revealed that 27 formed six clusters of sequences in NJ with >99% similarity to each other. The remaining five sequences were <99% identical to any other sequence in the analysis and as such were considered singletons. One sequence from each cluster as well as the five singleton sequences
were phylogenetically analyzed. However, even though two sequences (LO.20May09.33 CHLV and LO.24Jun09.16 CHLV) from cluster six were >99% identical on the nucleotide level (Figure 7), they were both included in the maximum likelihood analysis.

Phylogenetic inference, based on the maximum likelihood of 12 inferred amino acid sequences from partial polB genes, revealed the presence of robust relationships between the sequences recovered and other known groups of viruses (Figure 8). When compared with various NCLDV viruses, inferred amino acid sequences of Lake Ontario and the UTM SWM pond putative polB fragments were all nested within the phycodnavirus genera Chlorovirus. Furthermore, a 100% likelihood support value at the node separating Chlorella viruses and environmental clones from other phycodnaviruses, indicates that the placement of environmental sequences among Chloroviruses was well supported (Figure 8).

Several polB fragments amplified from the two environments (LO.05Jun08.40 CVMS, LO.05Jun08.41 CVMS, UP.09Jun09.45 CHLV, LO.20May09.31 CHLV, LO.24Jun09.15 CHLV, LO.08May08.28 CHLV, LO.24Jun09.17 CHLV and LO.08May08.26 CHLV [a total of 28 clones]) were closely related to polB sequences encoded by CVM-1, MT325 and FR483 viruses. These sequences were 98-100% or 96-98% identical to Chlorella Pbi viruses (e.g., CVM-1, FR483, etc.) in terms of amino acids or nucleotides, respectively. Only one sequence amplified from the UTM SWM pond (UP.21Sep09.7 ATCVS) was closely related to the DNA polymerase sequence encoded by ATCV-1 virus and was 99% or 89% identical on the amino acid or nucleotide level, respectively. Finally, the three sequences amplified from Lake Ontario formed their own clade (LO.20May09.33 CHLV, LO.24Jun09.14 CHLV and LO.24Jun09.16 CHLV) between PBCV-1 and ATCV-1 viruses with the likelihood support value of 95% at their node.
Figure 7. A neighbour-joining tree of 32 Chlorella-like polB nucleotide sequences amplified from Lake Ontario and the UTM SWM pond. Cluster numbers shown to the right of the tree indicate groups of sequences that were more than 99% identical; only one sequence (shown in black italic type) was chosen from each group for subsequent comparisons to other virus sequences. The names of the sequences obtained in this study indicate the location from which the sequence was obtained (Lake Ontario-LO, UTM SWM pond-UP), the date of sample collection in the format day/month/year, an arbitrary clone number and the PCR primer set used to amplify the sequence. Values at the major nodes indicate percent bootstrap values from 500 replicates and the scale bar indicates proportion of changes per site.
Figure 8. Phylogenetic tree of 12 sequences from Chlorovirus-like polB gene fragments amplified from Lake Ontario and the UTM SWM pond. This cladogram represents their phylogenetic relationship to other NCLDVs based on maximum likelihood of inferred amino acid sequences from partial DNA polymerase genes. Values at the nodes indicate clade credibility using aLRT branch support from minimum SH-like and Chi2-based values. Sequences obtained in this study are indicated in bold type, whereas sequences from cultivated viruses are shown in italics. Gene fragments that were selected as targets for qPCR are shown in color. The African swine fever virus (ASFV) and Mimivirus are used as outgroups for phycodnaviruses. GenBank accession numbers for cultivated viruses are listed within the taxon names following the title of a virus and clusters containing sequences from cultivated viruses are indicated to the right of the tree with the virus genus. The scale bar indicates the proportion of changes per site.
Comparison of the two novel polB fragments (LO.20May09.33 CHLV and LO.24Jun09.14 CHLV) to sequences in GenBank database revealed that these gene fragments were 79%, 76% and 72% identical on an amino acid level to DNA polymerases of the viruses PBCV-1, ATCV-1 and CVM-1, respectively. A summary of nucleotide percentage identities for the two novel polB fragments and cultivated *Chlorella* viruses (CVM-1, PBCV-1 and ATCV-1) can be found in Table 3.

**Table 3.** Percent nucleotide identities retrieved from GenBank using BLAST

<table>
<thead>
<tr>
<th></th>
<th>CVM-1</th>
<th>PBCV-1</th>
<th>ATCV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVM-1</td>
<td>ID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBCV-1</td>
<td>67</td>
<td>ID</td>
<td></td>
</tr>
<tr>
<td>ATCV-1</td>
<td>77</td>
<td>71</td>
<td>ID</td>
</tr>
<tr>
<td>LO.20May09.33 CHLV</td>
<td>69</td>
<td>73</td>
<td>72</td>
</tr>
<tr>
<td>LO.24Jun09.14 CHLV</td>
<td>70</td>
<td>72</td>
<td>72</td>
</tr>
</tbody>
</table>

ID: identical

### 3.4 qPCR results

Two putative *Chlorella* virus targets, LO.20May09.33 CHLV and LO.05Jun08.41CVMS, were chosen for the quantitative study. These two targets were selected because they most likely represent different viruses that infect different hosts. Therefore, based on the previously mentioned criteria, qPCR primers and TaqMan® probes were designed for these *Chlorella* virus targets (Table 4).
Table 4. Quantitative PCR primers and probes

<table>
<thead>
<tr>
<th>Target Sequence/probe name</th>
<th>Forward primer (5’-3’)</th>
<th>Reverse Primer (5’-3’)</th>
<th>TaqMan® probe (5’-3’)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LO.20May09.33 CHLV/LO.33</td>
<td>GATACAGATTCCGTATGCT</td>
<td>CATCTTGAACTTGCCCTC</td>
<td>TGCACAGTTCCGTCCCTC</td>
</tr>
<tr>
<td>LO.05Jun08.41 CVMS/LO.41</td>
<td>GTAAGGGATCTGCCATGTTTC</td>
<td>TCATTGATGCTCTGTTCTCC</td>
<td>AGTCAGGGCAACTGTCGCAACAT</td>
</tr>
</tbody>
</table>

* 5’ labelled with FAM and 3’ labelled with Iowa Black® FQ

Prior to monitoring targets via qPCR in environmental samples, each primer and probe set was tested for specificity. Quantitative PCR detection of $10^7$ non-target molecules ranged from cycle threshold values of 33.14 to no Ct detected. On the contrary, the Ct values for $10^7$ target molecules ranged from 12.30 for LO.41 to 12.34 for LO.33 (Table 5). The limit of detection for both primers and probes was Ct value of 35.13 and was determined from the amplification of ca. $10^9$ target molecules.

Table 5. Properties of sequences used to test primer and probe specificity

<table>
<thead>
<tr>
<th>Target Sequence</th>
<th>Closest match</th>
<th>% nucleotide identity</th>
<th>Forward primer mismatches</th>
<th>Reverse primer mismatches</th>
<th>Probe mismatches</th>
<th>Ct with $10^7$ target molecules</th>
<th>Ct with $10^7$ non-target molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>LO.33</td>
<td>UP.09Jun09.47 CHLV</td>
<td>69</td>
<td>6</td>
<td>7</td>
<td>13</td>
<td>12.34</td>
<td>33.14</td>
</tr>
<tr>
<td>LO.41</td>
<td>UP.21Sep09.7 ATCVS</td>
<td>76</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>12.30</td>
<td>No Ct</td>
</tr>
</tbody>
</table>

Inhibition of Platinum® Taq DNA polymerase resulting from addition of the UTM SWM pond samples was tested using LO.33 primer and probe set and four different VCs. In all cases, the inhibition tests estimated from triplicate reactions demonstrated that the percent amplification of $10^5$ target molecules per reaction was 96.4-108.9% (average 101.10%) upon the addition of the UTM SWM VCs (Figure 9). A Chi-square test ($\chi^2 = 3.32 \times 10^{-36}$) of average gene copies amplified from two reactions with less than 100% percent amplification was found to be lower
than a critical value of 3.84 for 1 degree of freedom, at $p = 0.05$ level of significance. In addition, the deviations in gene copy estimates between ‘expected’ (targets only) and ‘observed’ (target molecules plus VCs) qPCR reactions were less than the standard deviations among the replicates (data not shown).

Figure 9. Quantitative PCR inhibition by the UTM SWM pond samples. Percent amplification of target molecules plus VCs compared with amplification of target molecules alone using Platinum® Taq DNA polymerase enzyme. Error bars were calculated by dividing the standard deviation of gene copies from amplification of target plus VC, by the standard deviation of gene copies from amplification of target alone. A horizontal line at 100% amplification is shown for reference only.
Weekly quantification of the two *Chlorella* virus targets (LO.33 and LO.41) at a single location in Lake Ontario and in the UTM SWM pond, revealed the presence of persistent polB genes with different abundance patterns. Putative *Chlorovirus* genes amplified using LO.33 and LO.41 primers and probes were most abundant when the pond and lake surface water temperatures were relatively high (Figures 10 and 11). A novel gene fragment, LO.33, reached highest abundances in Lake Ontario on July 16\(^\text{th}\), 2008 and June 18\(^\text{th}\), 2009 with *ca.* 1657 ± 233 and 1539 ± 122 gene copies mL\(^{-1}\), respectively (Figure 10). Peak abundances of the same target in the UTM SWM pond occurred on July 8\(^\text{th}\), 2009 and May 13\(^\text{th}\), 2010 with 4998 ± 407 and 6863 ± 550 gene copies mL\(^{-1}\), respectively. The average abundance of LO.33 gene fragments in Lake Ontario over the entire study period was *ca.* 196 ± 35 copies mL\(^{-1}\), whereas in the UTM SWM pond the average abundance was higher (*ca.* 668 ± 87 copies mL\(^{-1}\)). In 2008, the average gene abundance of LO.33 target in Lake Ontario was slightly lower than the average gene abundance in 2009 (216 ± 41 copies mL\(^{-1}\) vs. 246 ± 38 copies mL\(^{-1}\)) with a delay period of *ca.* one month for peak abundance timing between the two years. Moreover, the timing of peak abundance for LO.33 target in Lake Ontario during 2009 was approximately one month earlier than that of the UTM SWM pond (June 18\(^\text{th}\), 2009 vs. July 8\(^\text{th}\), 2009).
Figure 10. Quantitative PCR results for the putative Chlorovirus target LO.33. Top: The average abundance of LO.33 target in Lake Ontario over period of 28 months. Bottom: The average abundance of the same target in the UTM SWM pond over period of 12 months. The target sequences were monitored using the 5’-nuclease assay and were calculated from the triplicate qPCR reactions. Error bars represent standard deviation; in all cases where no error bar is visible, data symbols are larger than the error bars. The water temperature (dotted line) was recorded in situ at the time of sample collection and was included to display the seasonality of both environments.
The LO.33 target was present throughout the year in Lake Ontario with a single peak abundance in spring, summer and fall seasons followed by low abundance throughout the winter months (ca. 23 ± 9 gene copies mL⁻¹; the average value was inferred from 16 samples collected between December 21st and March 21st during 2008-2010). On the contrary, the abundance of the same target in the UTM SWM pond during 2009 was relatively low during spring months, followed by a sharp (single) peak during mid-summer months (from June 24th, 2009 to August 6th, 2009). Subsequently, abundance rapidly decreased and was maintained below or just above the detection limit (for samples collected between August 13th, 2009 and March 18th, 2010 the average abundance was 1 ± 0.81 gene copies mL⁻¹) throughout several months until reappearing again with a sharp (single) peak at the beginning of spring 2010. The limit of detection for LO.33 target in environmental samples was 0.1745 gene copies per reaction (Ct 39.47) or 0.87 gene copies mL⁻¹ and was estimated from the lowest amplification of targets detectable in the standard curve.

In comparison to the LO.33 target, the CVM-1-like LO.41 fragment reached its highest abundance in Lake Ontario in 2008 on May 8th, with an abundance of 137 ± 41 gene copies mL⁻¹ and in 2009 it reached its highest abundance on June 18th at 278 ± 20 gene copies mL⁻¹ (Figure 11). In 2009, the same target reached its highest abundance in the UTM SWM pond on June 2nd at 1014 ± 95 gene copies mL⁻¹. The average abundance of LO.41 fragment in Lake Ontario over the entire study period was ca. 59 ± 13 copies mL⁻¹, whereas in the UTM SWM pond the average abundance was ca. 91 ± 16 gene copies mL⁻¹.
Figure 11. Quantitative PCR results for the CVM-1-like target LO.41. Top: The average abundance of LO.41 target in Lake Ontario over period of 28 months. Bottom: The average abundance of the same target in the UTM SWM pond over period of 12 months. The target sequences were monitored using the 5'-nuclease assay and were calculated from the triplicate qPCR reactions. Error bars represent standard deviation; in all cases where no error bar is visible, data symbols are larger than the error bars. The water temperature (dotted line) was recorded in situ at the time of sample collection and was included to display the seasonality of both environments.
In 2008, the average gene abundance of the LO.41 fragment in Lake Ontario was lower than its average abundance in 2009 (23 ± 10 copies mL\(^{-1}\) vs. 111 ± 17 copies mL\(^{-1}\)). The difference in the timing of peak abundance for LO.41 fragment in Lake Ontario between the two years of study was slightly longer (41 days; May 8\(^{th}\), 2008 vs. June 18\(^{th}\), 2009) than the difference for LO.33 target in the same environment. Likewise, the difference in timing of peak abundance for LO.41 target in Lake Ontario during 2009, occurred 16 days later than in the UTM SWM pond (June 18\(^{th}\), 2009 vs. June 2\(^{nd}\), 2009).

The LO.41 target in Lake Ontario was also present throughout the year with much lower, but stable abundance during winter months (ca. 5 ± 4 gene copies mL\(^{-1}\)) followed by numerous oscillations in abundance throughout the spring, summer and fall months. The abundance of the same target in the UTM SWM pond during 2009 was highest in spring, appearing as a single peak followed by a drop to ca. 38 ± 9 gene copies mL\(^{-1}\) during summer months, with few small oscillations. In fall, oscillation with slightly higher peak abundances was evident with average abundance of ca. 132 ± 31 gene copies mL\(^{-1}\). The targets were undetectable from March 11\(^{th}\) to April 2\(^{nd}\), 2010 until reappearing again on April 9\(^{th}\), 2010.

Patterns of abundance and the timing of peak abundance differed between the two putative Chlorovirus targets in either freshwater environment. For instance, the timing of peak abundance for LO.41 target in Lake Ontario appeared much earlier in the year of 2008 (May 8\(^{th}\), 2008), in comparison to LO.33 target (July 16\(^{th}\), 2008). However, in 2009 both targets reached their maximum abundance on the same day (June 18\(^{th}\), 2009). Similarly, the timing of peak abundance for LO.41 target in the UTM SWM pond during 2009-2010 occurred earlier (June 2\(^{nd}\), 2009) than the timing of first peak abundance for LO.33 target (July 8\(^{th}\), 2009). Compared to the LO.41 fragment, the average abundance of LO.33 was ca. three times higher in Lake Ontario and ca. seven times higher in the UTM SWM pond. As an additional comment on abundance
patterns, the average abundance of LO.41 and LO.33 targets in the UTM SWM pond was either 1.5 or 3.5 (respectively) times higher than their average abundance in Lake Ontario.

3.5 Autocorrelation analysis of Lake Ontario abundance patterns

The repeatability of seasonal patterns differed for each polB target monitored in Lake Ontario. An autocorrelation revealed significant positive coefficient for LO.33 target and indicated an annual pattern with a lag of 12 4-week period (r = 0.442) (Figure 12). On the other hand, no significant periodicity was observed for LO.41 target in Lake Ontario and autocorrelation was random at most lags (Figure 13).

![Autocorrelation Function](image)

**Figure 12.** Correlogram for unknown *Chlorovirus* target LO.33 inferred from 31 bins comprised of 4-week average abundance in Lake Ontario.
Figure 13. Correlogram for the CVM-1-like target LO.41 inferred from 31 bins comprised of 4-week average abundance in Lake Ontario.
IV Discussion

Investigations of *Chlorella* virus diversity in two contrasting Ontario freshwater environments were facilitated by the introduction of newly designed primers for PCR-based assay. These primer sets were specifically designed to target DNA polymerase genes of the three different Chloroviruses, CVM-1, PBCV-1 and ATCV-1, and were used to PCR-amplify gene fragments in samples collected from a single location in Lake Ontario and the UTM SWM pond. PCR amplification and sequencing efforts revealed the presence of diverse *Chlorella* virus polB gene fragments in Lake Ontario and the nearby pond. Following these results, qPCR-based assays were developed and applied to monitor the seasonal abundances of the two putative *Chlorella* virus targets (LO.33 and LO.41) in these environments. Our results revealed not only the presence of persistent virus genes with different seasonal dynamics, but also different patterns of seasonal abundances for individual *Chlorella* virus targets in both environments. The following discussion will consider interpretations for the findings observed in this study, as well as address the methodological concerns encountered throughout the data-gathering stage.

4.1 Contrasting freshwater environments

In this study, Lake Ontario and the UTM SWM pond were regarded as two contrasting freshwater environments. Apart from their obvious differences in surface area and volume (e.g., the permanent pool reservoir of the pond is 6136 m$^3$, whereas the volume of Lake Ontario is $1.64 \times 10^{12}$ m$^3$; MGM Consulting Inc., design note, http://www.great-lakes.net/lakes/ref/ontfact.html) as well as other chemical differences (e.g., conductivity and dissolved oxygen in the epilimnion; data not shown) their average chlorophyll $a$ concentrations (based on 24 measurements) were also dissimilar. It is generally accepted that chlorophyll $a$ estimates are a proxy for phytoplankton biomass and can be used as an indicator of trophic status, primary productivity
and water quality assessment (Carlson, 1977; Jones and Lee, 1982; Culley and Welschmeyer, 2002; Hewson and Fuhrman, 2004). Thus, with reference to Carlson’s trophic state index on chlorophyll \( a \) measurements (Carlson and Simpson, 1996), Lake Ontario at a single location can be classified as an oligotrophic environment (average chlorophyll \( a \) concentration 1.76 \( \mu g/mL \)) with some eutrophic phases (e.g., on September 28\(^{th} \), 2009 with a concentration measure of 11.05 \( \mu g/mL \)), whereas the UTM SWM pond can be considered a eutrophic environment (average chlorophyll \( a \) concentration 20.09 \( \mu g/mL \)) with some hypereutrophic phases (e.g., on March 25\(^{th} \), 2010 with a concentration of 101.10 \( \mu g/mL \)). Based on previous research, surface water quality (lake-wide) of Lake Ontario has been regarded as mesotrophic with oligotrophic and eutrophic regions (Chapra and Dobson, 1981; Harris, 1987; Busch and Lary, 1996), demonstrating that conclusions drawn from chlorophyll \( a \) abundance are consistent with reported findings.

The phytoplankton communities supported by Lake Ontario and the UTM SWM pond are very different. Throughout the year, Lake Ontario is dominated (up to 90%) by green algae, diatoms, dinoflagellates and flagellates (Johannsson and Ogorman, 1991). On the contrary, a study of phytoplankton species in the UTM SWM pond indicates that clone libraries of \( psbA \) (that encodes a core photosynthetic protein) sequences are dominated by cyanobacterial genes throughout most of the year (S.M. Short, pers. comm.). Research on mesotrophic environments suggests that lakes with moderate enrichment support seasonally variable phytoplankton biomass comprised of various taxonomic groups (Sommer, 1993), whereas the studies of eutrophic and hypereutrophic environments indicate that they are dominated by very few algal taxa all year round (Paerl, 1988; Jensen et al., 1994). Since Lake Ontario and the UTM SWM pond vary in their trophic statuses, their phytoplankton community compositions are expected to be very different. Thus, given the obvious physical, chemical and biological differences between Lake
Ontario and the UTM SWM pond, these two environments can be regarded as very different aquatic habitats.

4.2 Primer design

Chlorella virus-specific and degenerate PCR primer sets designed for this study worked equally well on positive controls (i.e., CVM-1, PBCV-1 and ATCV-1 hot/cold treated virus lysates), but did not work as well on environmental samples. For instance, PCR reactions with the CVMS primer set always produced a single amplified DNA fragment of ca. 600 bp when tested using various VCs. On the contrary, reactions with PBCVS, ATCVS and CHLV primer sets produced amplicons of various sizes ranging from ca. 500 bp to 1000 bp. It has been previously shown that some Chlorella viruses contain intron sequences in their polB genes (e.g., NY-2A and PBCV-1) (Grabherr et al., 1992), due to which the production of amplicons >700 bp may have been possible. However, in this case such a scenario was very unlikely, as segments flaked by PCR primers were outside of regions interrupted by introns. Therefore, the production of unexpected amplicons (i.e., those that were not ca. 600-700 bp in size) was most likely the result of amplification of extant targets in the environment. It is also possible that chimeras could have formed through a recombination of similar PCR products (Short and Short, 2008) and in the process of amplification produced the unexpected fragments. Nevertheless, three out of the four PCR primer sets presented in this study including the degenerate primer set CHLV and the virus-specific primers CVMS and ATCVS were effective and led to the discovery of Chlorovirus gene fragments in Ontario freshwaters. Future work with all of these primers may lead to the discovery of other novel Chlorella virus genotypes.
4.3 Sequencing efforts

4.3.1 Virus DNA polymerase sequencing

Various fragments closely related to cultivated *Chlorella* virus polB sequences were amplified from both Lake Ontario and the UTM SWM pond. PCR amplifications with a *Chlorella* virus-degenerate primer set resulted in recovery of sequences closely related to viruses infecting *Chlorella* Pbi (e.g., CVM-1, FR483, etc.) and two novel phycodnavirus polB sequences (LO.33 and LO.24Jun09.14 CHLV) that formed their own clade among cultivated Chloroviruses. The initial comparison of the three gene fragments (LO.33, LO.24Jun09.14 CHLV and LO.24Jun09.16 CHLV) recovered from Lake Ontario to sequences in GenBank using BLAST revealed that they were not identical to any phycodnavirus sequence previously deposited in database. As a result, all three sequences were phylogenetically analyzed in ML, even though cluster analysis revealed that LO.33 and LO.24Jun09.16 CHLV were >99% identical on the nucleotide level (Figure 7). PCR reactions with the specific *Chlorella* virus-primer sets CVMS and ATCVS also resulted in recovery of gene fragments closely related to their expected target sequences. Out of the total of 32 sequences recovered, more than 28 sequences were ≥97% identical to others and as a result, only three sequences were submitted to GenBank. As previously argued by Short and Short (2008), a 97% identity cut-off can be used to differentiate sequences that form operational taxonomic units (OTUs) from sequences that are redundant. It is possible that some of the nearly identical sequences could represent strain variations, but the methods used in this study cannot differentiate between authentic differences and the differences acquired merely due to PCR and/or sequencing error. For that reason, the same principal was applied in this study and nucleotide sequences that were less than 97% identical to each other or to other cultivated *Chlorella* virus polB sequences were considered unique OTUs.
A surprising result of the sequencing effort was that no polB sequences closely related to PBCV-1 or other viruses that infect Chlorella NC64A (e.g., NY-2A) were amplified from either Lake Ontario or the UTM SWM pond. Within the literature on phycodnaviruses, the prototype PBCV-1 virus has been regarded as a “northern” virus (Yashchenko et al., 2008) infecting the “American” Paramecium bursaria symbiont NC64A (Reisser et al., 1988); therefore, it was expected that fragments closely related to PBCV-1 polB gene would be found in the freshwater environments we sampled. It is possible that when the samples were collected, virus genes amplifiable with the PBCV-1 specific (PBCVS) primer set were not present or were simply below detection. Future studies involving more thorough PCR efforts (not limited to samples collected during spring and summer months) or attempts to infect and lyse the host in culture with natural virus concentrates might help resolve this issue.

Nevertheless, the results of this research suggest that the Chlorovirus genus is more diverse than previously thought. The novel polB fragments (LO.33 CHLV and LO.24Jun09.14 CHLV) recovered from Lake Ontario samples may represent a new type of Chlorovirus. The notion that CHLV amplicons (LO.33 CHLV and LO.24Jun09.14 CHLV) represent a new group of Chlorella viruses can be supported by the reasoning of Clasen and Suttle (2009). These authors argued that phycodnaviruses infecting the same host species usually form monophyletic assemblages. Moreover, they argued that phycodnaviruses infecting different host species have genetic distances (i.e., distances between monophyletic virus groups) greater than phycodnaviruses infecting the same host species (distances within monophyletic groups). Although exceptions to this principle are known, they are limited to the Phaeovirus genus, or viruses that infect Phaeophyte algae (Wilson et al., 2006). Interestingly, the genetic distances between CHLV amplicons and the polB genes of known Chloroviruses were greater than the distance between ATCV-1 and CVM-1 viruses (Table 3). Furthermore, ML analysis revealed
monophyletic clusters for ATCV-1-like and CVM-1-like viruses, in addition to the novel polB amplicons that formed their own clade (Figure 8). Since it is also known that CVM-1 and ATCV-1 viruses infect different Chlorella spp. (e.g., CVM-1 infects Chlorella Pbi, whereas ATCV-1 infects Chlorella SAG 3.83), the notion that the novel polB amplicons discovered in Ontario freshwaters represent a new type of Chlorovirus is highly plausible.

4.3.2 Potential hosts for the viruses encoding LO.33 and LO.41

Evidence from the literature and results obtained in this study strongly suggest that all Chlorella viruses most likely propagate in nature by infecting a free-living Chlorella-like alga susceptible to viral infection. Surveys of Lake Ontario phytoplankton communities in the nearshore regions during 1970s revealed the presence of several Chlorella species, one of which was C. vulgaris (Munawar and Munawar, 1996). Similarly, concurrent work on phytoplankton communities in Lake Ontario and the pond revealed the presence of psbA genes closely related to those from other free-living Chlorella spp. and C. vulgaris (S. M. Short, unpublished data). Interestingly, current evidence on Chlorella systematics suggests that symbiotic Chlorella Pbi and NC64A algae are genetically distinct species belonging to a Chlorella clade that is currently comprised of three free-living species (C. vulgaris, C. lobophora and C. sorokiniana) (Krienitz et al., 2004; Hoshina et al., 2010). Moreover, studies of 18S rRNA-based molecular phylogenies of Chlorella-like symbiotic algae in Paramecium bursaria revealed that symbiotic algae were closely related to a free-living Chlorella vulgaris (Nakahara et al., 2004) and in one case a true C. vulgaris has been found as a symbiont (Hoshina and Imamura, 2008). Altogether, the results of these studies suggest that when Chlorella algae are in symbiotic relationship with protozoa, they receive refuge from viruses and subsequently lower pressure for acquiring resistance, thus making them vulnerable to infection outside of their symbiont. It is very likely that Chlorella
viruses in nature survive by scavenging on sensitive cells within their phytoplankton community (e.g., Waterbury and Valois, 1993). The fact that previous experiments have failed to observe cellular lysis of free-living Chlorella spp. by viruses likely suggests that the choices of hosts used in experiments were simply not appropriate for PBCV-1 or CVM-1 viruses (Reisser et al., 1986). A newly isolated Chlorella virus WH-1, demonstrates that free-living Chlorella spp. can be infected and are infected by viruses in the environment (Wu et al., 2010). Future investigations will attempt to isolate and culture some free-living Chlorella spp. and subsequently try to isolate and identify environmental viruses that infect them.

4.4 Quantitative analysis of LO.33 and LO.41

4.4.1 qPCR technical details

The real-time PCR-based assays developed in this study were used to assess gene abundances of the two Chlorella viruses targeted and all of the following statements about virus abundances in either environment were based on inference from gene copy estimates. For that reason, quantitative assays were designed to fulfill all of the requirements for reliable estimation of the target abundances in environmental samples. Based on an amplification of serially diluted standards, amplification efficiencies of qPCR primer and probe sets were all within the recommended values of 90-110% with an $R^2$ value > 0.990 for an accurate assessment of gene abundances via the 5’-nuclease assay (according to Applied Biosystems application note on real-time PCR: Understanding Ct, 136AP01-01; http://www.appliedbiosystems.com). Independent reactions were also conducted to examine primer and probe specificity and demonstrated that the detection limit for target molecules was six to ca. nine orders of magnitude lower than that of closely related non-targets (as inferred from the difference in Ct from similar numbers of target and non-target molecules). It is very unlikely that close relatives of targeted genes could have
contributed to the estimates of target abundances. This is because differences observed in the detection limit between targets and non-targets were so large that it is extremely unlikely that non-target amplification could contribute to the estimates of LO.33 and LO.41 abundance. For example, it is doubtful that the peak abundance observed for LO.33 target in Lake Ontario on June 18th, 2009 (1539 ± 122 gene copies mL⁻¹) was generated by the amplification of non-target genes. This is for the reason that in order contribute to the estimates of LO.33, the non-target fragments must be present at a minimum abundance of 10⁷ mL⁻¹ (e.g., a Ct value of 33.14 is generated from 10⁷ copies mL⁻¹ of non-target [Table 5], and would lead to an estimated 72.6 copies mL⁻¹ of target), as the amplification of 10⁶ non-target molecules generated no Ct value (data not shown). This scenario, though, is extremely unlikely, as the total virus counts in aquatic ecosystems range from 10⁵ to 10⁸ particles mL⁻¹ (Bergh et al., 1989; Suttle et al., 1990, 1991). Moreover, these non-target fragments (UP.09Jun09.47 CHLV) (Table 5) were amplifiable with LO.41 primers and probe and on June 18th, 2009 reached their abundance at 278 ± 20 gene copies mL⁻¹. It is also unlikely for the UP.21Sep09.7 ATCVS (Table 5) contribute to the estimates of LO.41 fragments in any environmental samples, because this close relative of a target cannot be even amplified with LO.41 primers and probe. As more sequences from natural virus communities become available, it may be possible to re-test the qPCR assay specificities of closely related non-targets. Nevertheless, the gene copy estimates obtained in this study were from amplification of target sequences in environmental samples.

Previous studies on PCR inhibition by Lake Ontario VCs treated with a hot/cold procedure reported that Platinum® Taq DNA polymerase was not sensitive to inhibitory compounds present in the six natural samples (Short and Short, 2009). When the same Taq DNA polymerase was challenged with crude preparations of the UTM SWM pond VCs, inhibition was also not an issue. However, it is unlikely that the sample concentration and extraction of viral
nucleic acids using the hot/cold treatment procedure resulted in a perfect recovery of all particles; therefore, individual estimates of target gene copies in all qPCR reactions most likely underrepresented the actual abundances in the environment. Given that viruses were concentrated *ca.* 100 times by ultracentrifugation, the limit of detection for virus targets in both environments was estimated to be *ca.* five polB copies mL$^{-1}$. This theoretical detection estimate was calculated in the same manner as previously described in Short and Short (2009). However, if only 50% recovery efficiency is presumed, then the theoretical detection estimate will change to 10 polB copies mL$^{-1}$. Therefore, it is very likely that the actual detection limit of polB genes in environmental samples was much higher than the theoretical detection limit inferred from sample concentration factors, as 100% recovery efficiency was always assumed. Since the samples collected in this study did not include estimates of total virus abundances, it was impossible to determine the recovery efficiencies for virus particles by ultracentrifugation and nucleic acid extraction from hot/cold treated VCs.

Despite the technical issues noted in the preceding paragraphs, major abundance trends observed for the two putative *Chlorella* virus targets in this study were indisputable. More specifically, it is very doubtful that the sample-to-sample variations in the estimated abundances of each qPCR target resulted from differences in recovery efficiencies. For instance, between 4$^{th}$ and 16$^{th}$ of July 2008, LO.41 polB abundance in Lake Ontario remained approximately the same (i.e., increased from 18.47 to 18.52 gene copies mL$^{-1}$), while LO.33 abundance increased from 408 to 1657 gene copies mL$^{-1}$. Similarly, between September 28$^{th}$ and October 5$^{th}$, 2009, the LO.41 abundance in the UTM SWM pond increased from 58 to 209 gene copies mL$^{-1}$, while LO.33 abundance remained undetectable. As a result, these observations allow us to safely speculate that the abundance patterns observed for the two qPCR targets were not solely due to variable virus recovery.
4.4.2 Environmental persistence of LO.33 and LO.41

Weekly quantification of the two *Chlorella* virus targets in Lake Ontario and the UTM SWM pond revealed the presence of persistent virus genes; i.e., the polB fragments monitored were detectable throughout the year and in most environmental samples. The fact that persistent virus genes were observed in both freshwater environments suggests that their rate of inactivation and/or destruction was counter-balanced by an equal rate of viral production (or rate of replenishment). These results are very similar to 1) observations reported on persistent *Prasinovirus* polB fragments monitored in Lake Ontario (Short and Short, 2009); 2) persistent phycodnaviruses that have been reported for community fingerprints (Short and Suttle, 2003); and 3) persistent phycodnaviruses observed in bloom termination studies (Baudoux et al., 2006; Martinez et al., 2007). Altogether, these studies demonstrate that some taxa of algal viruses can be temporarily stable and persistent throughout the years with recurrent annual cycles of high (bloom) and low abundances (after bloom-lysis).

It is generally accepted that maintenance of a stable virus population requires a specific number of host cells (Murray and Jackson, 1992) and infectious viruses (i.e., that have not decayed) (Wommack and Colwell, 2000). Intuitively, as the host cell concentration decreases, the time required for a successful host contact and infection increases (Wommack and Colwell, 2000). Therefore, the probability that a virus will be destroyed before it encounters its host also increases (Wommack and Colwell, 2000). In the absence of infection, viruses decay rapidly within the water column (Hill et al., 1971; Kapuscinski and Mitchell, 1980; Fuhrman, 1999). Virus decay is thought to be comprised of two separate processes: virus inactivation (loss of infectivity) and virus destruction (Herrmann et al., 1974). A virus particle undergoing decay can become inactivated long before it decomposes beyond recognition (Fuhrman, 1999). Studies on
virus decay have shown that the most prominent factor responsible for the loss of viral infectivity in surface waters is sunlight-induced damage (Fuhrman, 1999; Wommack and Colwell, 2000). However, chemical (i.e., hydrolytic enzymes, organic aggregates, etc.) and biological (i.e., grazing) mechanisms can also account for the loss of infectious particles from aquatic virus pool (Fuhrman, 1999).

Interestingly, the observation that both LO.41 and LO.33 targets were present at relatively low abundances (ca. 5-23 gene copies mL\(^{-1}\)) in Lake Ontario during winter months raises the question of how these viruses survived in the absence of production. In temperate regions of the Northern Hemisphere, the phytoplankton growth rate during winter is expected to be minimal. Therefore, it is fair to presume that there was no virus production during this period in either environment examined. Numerous studies on phycodnaviruses indicate that viruses continue to persist even when their hosts are present at very low or undetectable abundances (Zingone et al., 1999; Baudoux et al., 2006; Martinez et al., 2007). Thus far, the evidence for phycodnavirus propagation strategies indicates that all of them are lytic, except for the three latent Phaeoviruses (EsV-1, EfasV-1 and FsV-1) that are only known to infect the marine brown algae from the class Phaeophyceae (Wilson et al., 2009; Van Etten et al., 2010). Therefore, assuming that all *Chlorella* viruses are lytic, the rate of their production must exceed the rate of their inactivation and destruction in order for them to be persistent throughout the year.

It is possible that *Chlorella* viruses monitored were able to persist during winter by infecting different *Chlorella* species that form ecotype populations. This scenario, however, is somewhat unlikely, as evidence for *Chlorella* Pbi viruses indicates that they do not attach to the cell wall of *Chlorella* NC64A algae and vice versa (Reisser et al., 1988). Another hypothesis is that they can escape destruction by solar radiation (i.e., UV) through the use of their own pyrimidine dimer-specific glycosylase or host’s photolyase, as has been previously suggested for
some *Chlorella* viruses (Kang et al., 2005; Fitzgerald et al., 2007c). It is also possible that viruses sink into deep water layers or sediments, where they can avoid decay and destruction by biological, chemical and physical processes (Suttle and Chen, 1992). However, the ecological importance of mixing and transfer of viruses from sediment reservoirs to epilimnion are currently subjects of speculation. Recent study on organic and inorganic aggregates in the water column suggests that they can be both a source of virus production and release into environment, or scavengers that form viral traps and promote inactivation of viruses (Weinbauer et al., 2009).

An alternative hypothesis for the *Chlorella* virus polB persistence observed in Lake Ontario during winter is that low virus production could be maintained year round. For example, it is possible for the hosts of LO.41 and LO.33 to be always present above the minimum threshold densities required to maintain a stable virus production. This speculation stems from the research conducted by Muhling et al. (2005). These authors observed that the *Synechococcus* concentrations always exceeded the minimum host cell densities (i.e., $10^2$-$10^3$ cells mL$^{-1}$) required to maintain the persistent population of their respective phages in the Gulf of Aqaba. It is also possible that the minimum threshold densities of the hosts required for these *Chlorella* viruses to propagate in nature are very low. For instance, recent study on *Emiliania huxleyi* viruses (phycodnaviruses) revealed that isolates were able to clear host cultures with multiplicities of infection (the ratio of viruses to host cells) as low as $10^{-5}$ and the host concentrations of only $10^2$ cells mL$^{-1}$ (Vaughn et al., 2010). Regardless of the mechanism, it is evident that Lake Ontario *Chlorella* virus communities are not very abundant or active in the water column during winter. Future studies will examine the rates of virus production throughout the year using individual virus-host systems. Perhaps these studies will help reassess the decay rates of viruses and unravel the mechanisms responsible for freshwater virus persistence, especially during the winter.
The polB abundances inferred in this study do not indicate virus infectivity, but they do suggest that during winter months two virus targets form a standing stock of viruses that ‘wait’ for the concentrations of appropriate hosts to increase the following spring. These speculations are consistent with the “Bank model” proposed by Breitbart and Rohwer (2005), where only the most abundant viruses are active at any given time, leaving the other rare viruses inactive and forming a reservoir for potential recruitment of hosts. This implies that when the appropriate algae hosts reach the critical threshold of abundance, some virus populations from the inactive pool may induce a rapid spread of infection and subsequent collapse of the algal population. Recent studies on virus diversity and temporal variation of viruses in marine and freshwater environments (Marston and Sallee, 2003; Breitbart et al., 2004; Parada et al., 2008; Short and Short, 2009) suggest that this phenomenon might be the case. For instance, the two separate marine mesocosm studies on *Emiliania huxleyi* bloom dynamics in Western Norway revealed the presence of diverse virus genotypes (EhVs) prior to the onset of the bloom (Martinez et al., 2007). During this period, the *E. huxleyi* abundance was $10^3$ cells mL$^{-1}$. For the duration of the bloom event that lasted only seven days, the concentration of *E. huxleyi* increased by one to two orders of magnitude. This was followed by a rapid increase in the number of EhV particles as the bloom declined and the concentration of *E. huxleyi* relapsed to the prebloom level ($10^3$ cells mL$^{-1}$). Such periodic succession of the host and its respective EhVs was nearly identical in both years of the study and was characterized by a single massive peak for both the viruses and the hosts (Martinez et al., 2007). On the whole, these findings suggest that viruses in temperate environments form inactive reservoirs during the winter months and replenish their ‘stocks’ during the spring, summer and autumn months.
4.4.3 Seasonal dynamics of LO.33 and LO.41

The fact that different patterns of abundance were observed for individual Chlorella viruses at both sampling sites demonstrates a critical role of the environment itself in Chlorovirus dynamics. In Lake Ontario two virus targets displayed similar abundance patterns, where multiple peaks were followed by periods of low, but stable abundances through late fall and winter. On the contrary, seasonal abundance trends of the same targets in the UTM SWM pond were characterized by a single massive peak in abundance during late spring and early summer, followed by low or virtually undetectable abundance throughout the rest of the summer, fall and presumably winter. Since Lake Ontario and the UTM SWM pond are very dissimilar, the differences in the seasonal dynamics observed between these habitats for LO.41 and LO.33 viruses may result from complex interactions of their respective hosts with different chemical, biological and physical parameters.

Studies on virus production imply that trophic status might explain the differences observed in the magnitude of peak abundances for individual Chlorella virus targets in both environments (Brown et al., 2006; Parada et al., 2006). For instance, in 2009, the peak abundance estimates for both qPCR targets were ca. 30% higher in the UTM SWM pond than in Lake Ontario. It is possible that in eutrophic environments, such as ponds, nitrogen and phosphorus do not limit primary productivity and can sustain a higher carrying capacity of hosts in comparison to nutrient-limited environments like Lake Ontario (Reynolds, 1998). This is especially true when the lake stratifies during summer and prevents the mixing of nutrient rich deep water with the relatively nutrient poor surface layer (Busch and Lary, 1996). It is also possible that in nutrient-rich habitats viral production is not limited by a shortfall of resources such as nucleotides because these limiting components can be readily replenished through host
biosynthesis, thereby allowing higher virus production (Brown et al., 2006). Further research is required to determine the factors that govern virus production in these environments.

Interestingly, research on bacterioplankton abundance in freshwater environments, suggests that grazer-induced bacterial mortality and not virus-induced bacterial mortality are a major cause of plankton death in lakes, especially with increasing trophic status (Raschke, 1970; Lymer et al., 2008). However, the grazer-induced algal mortality is unlikely to fully explain the seasonal dynamics observed for the individual virus targets in the UTM SWM pond. It is possible that a sudden decline of LO.33 and LO.41 fragments during summer was attributed to algal bloom collapse. This collapse could be an outcome of viral lysis event that in conjunction with massive grazing by zooplankton and/or interspecific competition could significantly reduce the host numbers and subsequent production of viruses. Another partial explanation for the decrease in algal population may be due to accumulation of auto-inhibitory substance such as chlorellin (Raschke, 1970). This compound is produced by Chlorella spp. and when present in sufficient quantities, is capable of inhibiting growth, respiration and photosynthesis of itself and other algae (Pratt and Fong, 1940; Pratt, 1942, 1943; Jorgensen, 1956). It is also possible that massive storm water flushing from the pond upon heavy rainfall can account for the observed virus dynamics. More specifically, it may increase the dilution of both virus and host populations, while removing them and leaving behind low concentrations. Water flushing can have detrimental effects on algal dynamics, because species that undergo low productivity are more affected by additional loses in conjunction to grazing, sedimentation, death or lysis by viruses (Schmitt and Nixdorf, 1999). An insight was gained into this by looking at the rainfall data recorded for Oakville, ON, Canada during the time of the peak abundance observed for LO.33 and LO.41 targets in this environment. When the sample with LO.41 target was collected on June 2nd, 2009, the average abundance estimate was 1014 gene copies mL⁻¹. From June 3rd to
June 18th, 2009, Oakville, ON received a total rainfall precipitation of 26.5 mm and when the next water sample was collected on June 18th, 2009 the average abundance of LO.41 dropped to 556 gene copies mL⁻¹. Similarly, when the sample with LO.33 target was collected on July 8th, 2009, the average abundance estimate was 4998 gene copies mL⁻¹. From July 8th to July 16th, 2009, Oakville, ON, received a total rainfall precipitation of 13.8 mm and when the next sample was collected on July 16th, 2009, the average abundance of LO.33 target was already 2126 gene copies mL⁻¹. Although it is not yet possible to determine which of the aforementioned factors drives seasonal dynamics of Chlorella viruses in the UTM SWM pond, the effect of flushing on polB gene estimates should be considered and warrants further investigation. Even if the evidence from the rainfall data suggests that discharge of water from the pond may affect the intra-annual dynamics of Chlorella viruses, this suggestion for now is only a mere supposition. Hopefully, future studies on virus-host dynamics in these freshwater environments can help unravel factors that govern intra- and inter-annual variability in their abundances and explain the Hutchinson’s “paradox of plankton” on the coexistence of unexpected phytoplankton diversity in aquatic environments (Hutchinson, 1961).

A surprising result of this study was that the average abundance of LO.33 target in both environments was about three and seven times higher than the average abundance of LO.41 target in either Lake Ontario or the UTM SWM pond. These differences in abundances may be explained by the differences in the host abundance or susceptibility of host to viral infection, virus burst size (progeny), or even propagation through multiple host taxa. It is possible for LO.33 target to exhibit a different life strategy than that of the CVM-1-like target by being more virulent, having a higher burst size and exhibiting a more r-selected strategy (rapid replication) relative to the LO.41 target. Such life histories for viruses in nature were previously suggested by Suttle (2007). Even so, when compared to average abundances reported for some
phycodnaviruses (Short and Short, 2009) and high abundance Prasinoviruses in Lake Ontario (ca. $10^4$ gene copies mL$^{-1}$; S. M. Short, unpublished data), the Chlorella viruses targeted here do not seem to be very abundant in nature. A recent psbA sequence analysis of gene fragments amplified from Lake Ontario (S. M. Short, unpublished data) demonstrated that Chlorella spp. did not represent a high proportion of any clone libraries sampled. This suggests that Chlorella-like algae may be present at low abundances in nature and could influence the relatively low abundances of Chlorovirus fragments monitored. These results may explain the inability to detect the PBCV-1 virus in the two environments sampled, thus indicating that perhaps the host for this virus in these freshwaters is not common.

The fact that differences in abundance trends were observed for individual Chlorella virus targets between environments, suggests that water body’s trophic status also plays a critical role in governing the periodic annual succession of viruses. For instance, in nutrient-poor environments, such as Lake Ontario, phytoplankton growth and variability in growth rates between species is expected to be minimal (Tilman, 1982); provided that light is not a limiting factor in determining growth. Therefore, under low nutrient conditions, phytoplankton populations may not be able to achieve high densities required to incur competitive interactions. This would allow many species of phytoplankton to coexist without actual competitive interactions and subsequently maintain high diversity (Liggett, 1999; Miyazaki et al., 2006). In contrast, in nutrient-rich environments, such as the UTM SWM pond, species-specific growth rates are expected to be variable (Kuwata and Miyazaki, 2000; Miyazaki et al., 2006). As a result, interspecific interactions between phytoplankton species are expected to be extremely high and the differences in the growth rate will eliminate the inferior species. Moreover, chemostat experiments indicate that under conditions of continuous flushing and nutrient loading, the timing of peak phytoplankton abundance and the magnitude of the peak biomass in
replicates with identical initial community compositions, can be different (Roelke et al., 2003). These results suggest that under nutrient-rich conditions (e.g., UTM SWM pond), the seasonal abundance trends of phytoplankton in natural ecosystems can be less predictable than in nutrient-poor environments. In contrast, under conditions of pulsed nutrient inflow and continuous flushing, the timing and the magnitude of peak phytoplankton abundances were nearly identical in the replicate experiments (Roelke et al., 2003). These observations indicate that perhaps in nutrient-limited environments (e.g., Lake Ontario) the seasonal dynamics of individual phytoplankton species can be more predictable than in nutrient-rich environments. It is known that Lake Ontario is subject to complex water circulations (Pickett, 1977) as well as episodic upwelling and downwelling events, that can occur even during maximum summer stratification (Rao and Murthy, 2001). These events may supply nutrients during summer and are reflective of the pulsed nutrient inflow and continuous flushing conditions set in the chemostat experiment. Interestingly, nearly identical periodic annual succession was observed for LO.33 target in Lake Ontario. An autocorrelation analysis of these polB fragments indicated a roughly annual pattern with lag of 12 4-week period, thereby suggesting that the seasonal dynamics of the host for this virus in Lake Ontario can be accurately forecasted. In the future, it would be interesting to examine if the periodic inter-annual succession of Chlorella viruses in the UTM SWM pond is less predictable; as the chemostat experiments and autocorrelation analysis suggest this could be the case.

Although significant positive coefficient was not observed for LO.41 target in Lake Ontario, an autocorrelation analyses must be interpreted carefully due to limited size of data set. It is very likely that the patterns of abundance observed for LO.33 gene fragments in Lake Ontario during spring and fall were an outcome of their host being more sensitive to environmental factors than the host for LO.41. These may include water temperature, sunlight,
nutrient availability, competition and grazing by zooplankton. In contrast, the host for the CVM-1-like target, LO.41, might be just better adapted to grow and survive in changing environmental conditions, as some evidence suggests this might be the case (Lalucat et al., 1984; Lee et al., 1996; Martinez et al., 1997; Yoshida et al., 2004). For example, laboratory studies have shown that some Chlorella spp. can grow mixotrophically, especially with decreasing light intensity (Lee et al., 1996; Martinez et al., 1997). Therefore, the possibility that the host for LO.41 target was capable to undergo a mixotrophic growth during fall is very likely. It is also possible for the host of LO.41 to have a better defence against grazers (Yoshida et al., 2004), and dominate the phytoplankton community (in spring) when the grazer-sensitive competitor(s) is not present.

Another hypothesis for the more cryptic periodicity observed for the LO.41 target in Lake Ontario can be due to infection of more than one host strain or taxa. An interesting detail of this study was that the qPCR primer and probe set designed to amplify the CVM-1-like target was also able to detect the polB fragments of other very closely related Chlorella Pbi viruses (e.g., MT325, FR483, etc.). Even thought the DNA polymerase sequences of MT325 and FR483 viruses are 99% or 100% (respectively) identical over the 645-nucleotide region to polB of the CVM-1 virus, all three are considered to be distinct species of Chlorella Pbi viruses (Fitzgerald et al., 2007a). Therefore, it is possible for LO.41 virus target monitored to undergo a similar type of phenomenon observed by Waterbury and Valois (1993) for marine cyanobacteria and their co-occurring phages. These authors discovered that acquisition and maintenance of resistance by some Synechococcus clones to their co-occurring phages played an important role in determining the succession of Synechococcus genotypes and their subsequent phage strains (Waterbury and Valois, 1993). For example, it is possible for the rare Synechococcus strains to be resistant to now dominant phages and outcompete the sensitive conspecifics (Waterbury and Valois, 1993). Subsequently, this would allow the resistant Synechococcus clones to dominate the community.
until a critical threshold of abundance was reached and new cyanophage genotypes could initiate infections on now sensitive *Synechococcus* strains (Waterbury and Valois, 1993). Therefore, if different *Chlorella* Pbi viruses are able to infect different strains of host or taxa of hosts, then one might expect high titres of viruses and multiple peaks of high abundances coupled to subtle patterns in annual succession. These attributes should be particularly recognized at times when *Chlorella*-like algae dominate the phytoplankton communities. The fact that multiple peaks of high abundances were observed for the LO.41 target in Lake Ontario, especially during the 2009 study, with less predictable patterns in abundance over the two consecutive years, suggests that these phenomena might be involved. Moreover, seeing that the LO.41 target was most abundant during late spring and summer months in either year of the study is consistent with prediction, as studies on *Chlorella* spp. indicate that they are most abundant during summer months (Happey-Wood, 1988; Munawar, 2003; Shammas et al., 2009). In the future, as more sequence information becomes available for the CVM-1 virus, it will be interesting to compare the seasonal dynamics of various *Chlorella* Pbi viruses. By designing qPCR primers and probes that target less conservative gene markers it will be possible to examine if different Pbi viruses infect the same strain or different strains (or taxa) of host(s). This hopefully, can provide the evidence to help support or refute the “Killing the Winner” hypothesis (Thingstad, 2000) which argues that viruses become a driving force for community composition and succession, both at the interspecific (Castberg et al., 2001; Brussaard et al., 2005) and intraspecific (Muhling et al., 2005; Martinez-Martinez et al., 2006) levels.

Nevertheless, while the hypothesis about the infection of multiple strains or taxa of host(s) is intriguing, it is also highly speculative. To our knowledge, all *Chlorella* viruses are strain-specific and in culture can infect only one host; thus, it can also be hypothesized that not only the ecology of all *Chlorella* Pbi viruses is very similar but so are their seasonal abundance
trends. The fact that timing, magnitude of peak abundances and predictability of seasonal abundance patterns differed for each *Chlorella* virus target in both environments, suggests that viruses encoding the monitored polB fragments infect different phytoplankton, which themselves exhibit different seasonal dynamics. As previously argued by Short and Short (2009), it is very unlikely that environmental reservoirs, such as sediments could be a reason behind abundances observed for the two qPCR targets in Lake Ontario. This is primarily because the sediment resuspension would most likely generate large intermittent pulses of peak virus abundances, rather than result in constant input of low gene numbers, especially during the winter months (Short and Short, 2009). While it is not yet possible to precisely determine how the dynamics of viruses monitored reflect the dynamics of their hosts, it is reasonable to infer that their abundance patterns were governed by changes in the abundance of hosts that produce them.

4.5 Conclusions

In summary, our results indicate that Lake Ontario and the UTM SWM pond are home to different types of *Chlorella* viruses including previously unknown sequence genotypes. The fact that different Chloroviruses were observed with different seasonal dynamics suggests that these viruses and their hosts, like other well-studied aquatic virus-host systems (e.g., Waterbury and Valois, 1993), can stably coexist in Ontario freshwater environments. Our findings also indicate that viruses belonging to single genera do not perform as a unified ecological entity, because the seasonal dynamics of closely related viruses can be governed by a complex suite of chemical, physical and biological factors. The fact that the same types of viruses behaved differently in different environments, makes an even a stronger case for the complexity we may be observing. Therefore, to better understand the complex interactions between viruses and host cells in lentic ecosystems, individual virus-host pairs must be examined. Future challenges include determining
the mechanisms responsible for virus persistence and identification of culture-independent hosts for *Chlorella* viruses in these two Ontario freshwater environments.
V References


