Deep-amplicon sequencing for monitoring impacted waterways in rural Manitoba

by

Jocelyn Zambrano Alvarado

A Thesis submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

ABSTRACT

Aquatic environments across Canada constantly receive anthropogenic discharge. In this context, water is considered the main vehicle for the dissemination of pathogens into the environment. The current assessment of microbiological water quality indicators focuses on fecal bacteria such as *E. coli*, overlooking viruses and protozoa. Monitoring impacted freshwater ecosystems without focusing on single microorganisms offers fundamental data to develop new microbial indicators for aquatic contamination. A longitudinal, 8-month amplicon-based analysis was performed in the Assiniboine River to identify the microbial composition, relative abundance, and environmental structuring factors in water associated with different anthropogenic activities. Bacteriophages, bacteria, and microeukaryotes were studied using the major capsid protein gene g23, 16SrRNA, and 18SrRNA genetic markers with Illumina technology. Our analysis revealed the riverine ecosystems were compounded in the majority by cyanophages (62%) (*Synechococcus* phages, and *Bellamyvirus* genus phages), *Escherichia coli* phages (26%), *Emdodecavirus* genus phages (8%), and *Slopekvirus* genus phages (2%). The most relatively abundant bacterial families across waterways included *Sporichthyaceae (20) %*, *Pseudomonadaceae (19%), Enterobacteriaceae (14%), Burkholderiaceae (8%), Spirosomaceae (7%) , Flavobacteriaceae (3%), Aeromonadaceae (3%), Micrococcaceae (3%) and Nostocaceae (3%)*. The microeukaryotic structure consisted of *Mediophyceae (24%), Cryptomonadales (14%), Choreotrichia (14%), Oligotrichia (3%), Sphaeropleales (7%), Chlamydomonadales (1%), Kappamyceteceae (5%),* and *Thoracosphaeraceae (3%)*. The effects of location and sampling months for bacteriophages were inconclusive. Sampling months affected the community structure of bacteria and microeukaryotes significantly. Forested, urban, and agricultural activities significantly shaped microeukarytic communities. Microbial communities were highly impacted by environmental factors, especially daylight length, water temperature, dissolved oxygen (DO), and precipitation. Our results also revealed complex networks of trophically interacting microorganisms between primary producers, heterotrophic bacteria, and predators (i.e. *Oligotrichia*). These results are critical to better understanding the dynamic interplay and the effect of temporal changes in microbial fractions as well as providing essential information for new potential indicators for aquatic contamination.

ACKNOWLEDGMENTS

I extend my profound gratitude to my supervisor Dr. Miguel Uyaguari for his outstanding mentorship and genuine support throughout these years. His authenticity, perseverance, and continuous guidance aided my academic accomplishments and inspired my professional and personal development. I sincerely thank Dr. Miguel for his patience and for being an exemplary supervisor, serving as a role model of exceptional teaching for me, his mentees, and all his students.

I would also like to thank my advisory committee including Dr. Ivan Oresnik, Dr. Beata Gorczyca, and Dr. Natalie Knox for sharing their invaluable insights and intellectual contributions to this work and for their thoughtful and considerate recommendations to enhance this research.

I would like to acknowledge all the members of the Uyaguari lab, who provide a friendly work environment. I especially thank Kadir Yanaç for imparting his knowledge which was vital for me to complete the viral analysis. Furthermore, I am also thankful for the help of Juan Diaz, Jasmeet Kaur, and Tiffany Penner for their important help when conducting water sample collection. I extend my gratitude and deep appreciation to everyone who supported me throughout my research journey, even before it began. This includes my undergraduate professors, former colleagues in Ecuador, and people who share and interpret complex scientific knowledge on digital platforms.

I acknowledge the research organizations who funded this work including Indigenous CREATE; Research Manitoba New Investigator Operating grant: No 5385; University of Manitoba Indigenous Research Program Project No. 54345. I am especially appreciative of The Visual and Automated Disease Analytics (VADA) graduate training program for funding me and training me in complex data analysis topics.

Finally, I thank my family: My mom Gina, my sister Jessica, and my brother Joel for their love, constant encouragement, and unconditional support.

DEDICATION

"*May all wounds to forests, rivers, deserts, oceans,*
all wounds to Earth be lovingly restored to bountiful health."

-One Earth Sangha

Table of Contents

ABSTRACT ii

ACKNOWLEDGMENTS iii

List of Figures viii

List of Tables ix

List of Abbreviations ix

1: INTRODUCTION 1

1.1 Freshwater in Canada: Overview and key human-induced threats 1

1.2 Microbial loop in freshwater 2

1.3 Surface Water Monitoring in Canada 3

1.4 Microorganisms in the Guidelines for Canadian Drinking Water Quality 4

1.4.1 Other waterborne pathogens 8

1.5 Limitations of Traditional Microbial Water Quality Indicators 9

1.6 Research Premise 8

1.7 Research Objectives 9

2.0 MATERIALS AND METHODS 13

2.1 Water Sampling 13

2.2 Concentration of bacteria, viruses and eukaryotes, from environmental samples 15

2.3 Nucleic Acid Purification 13

2.4 Characterizing the Microbial Distribution..... 15

2.5 Quantitative PCR analysis..... 22

2..5.1 Bacteria and eukaryotes quantitative PCR analysis 22

2.6 Assessing bacteriological counts using standard methods for water monitoring..... 24

2.7 Comparative analysis of the microbial fractions with seasonal and spatiotemporal patterns 24

3.0 RESULTS..... 27

3.1 Diversity..... 27

3.1.1 Alpha diversity in viruses- g23..... 27

3.1.2 Beta diversity in viruses- g23..... 30

3.1.3 Viral taxonomy profiling- g23 32

3.1.4 Alpha diversity in Bacteria -16S rRNA..... 36

3.1.5 Beta diversity and Bacteria taxonomy profiling 37

3.1.6 Alpha diversity in microeukaryotes -18S rRNA..... 45

3.1.7 Beta diversity and microbial eukaryotes taxonomy profiling 46

3.2 Quantitative PCR analysis..... 54

3.2.1 Bacteria and Eukaryotes Quantitative PCR Analysis 54

3.3 Comparative analysis of the microbial fractions with seasonal and spatiotemporal patterns..... 62

4.0 DISCUSSION..... 70

4.1 Evaluation of Viral Diversity..... 70

4.2 Evaluation of Viral Taxonomy and seasonal and spatiotemporal patterns..... 71

4.3 Evaluation of bacterial diversity 74

4.4 Evaluation of bacterial taxonomy and seasonal and spatiotemporal patterns..... 76

4.5 Evaluation of Micro-eukaryotes Diversity 80

4.6 Evaluation of microbial co-occurrence 86

5.0 CONCLUSION 89

Appendix 92

REFERENCES 96

.

List of Figures

Figure 1.1 Graphic representation of the source water protection strategy…….….5

Figure 2.1 Map indicating water sampling points…….…..13

Figure 2.2 Graphic representation of Methodology……..…..26

Figure 3.1.1: Observed diversity - g23 viral samples…….29

Figure 3.1.2: Non-metric multidimensional scaling (NMDS) g23……...…….....31

Figure 3.1.3: Relative abundance of viral composition ……...…….....35

Figure 3.1.4 Bacteria Alpha diversity-Shannon …….....40

Figure 3.1.5 Bacteria Alpha Diversity-Simpson…… …....42

Figure 3.1.6 Relative abundance of bacterial composition…….44

Figure 3.1.7 Eukaryotic Alpha diversity – Shannon ...……...49

Figure 3.1.8 Eukaryotic Alpha diversity- Simpson…….51

Figure 3.1.9 Non-metric multidimensional scaling 18S rRNA……...52

Figure 3.2.0. Relative abundance of microeukaryotic communities……...53

Figure 3.2.1: Quantification of 16S rRNA gene - log10 GCN…….....58

Figure 3.2.2: Violin plot 16SrRNA and 18S rRNA genes……..60

Figure 3.2.3: Quantification of 18S rRNA…….. ...61

Figure 3.2.4 Principal component analysis of viruses, bacteria,and microeukaryotes…..…….....66

Figure 14 Spearman correlogram microbial fraction and environmental variables……...….... 67

Figure 15 Co-occurrence network…….…..69

Figure A.1: Krona representation (pie chart) of g23 …….…….93

List of Tables

Table 1.3- Microbiological parameters used to asses source and drinking water quality……..……..6

Table 2.1 – Coordinates and description of the sample locations used in this study performed in rural areas of Manitoba,
Canada…….....14

Table 2.2 – Primers and probes used for sequencing and quantification assays …….....18

Table 3.1.1 – Average of Shannon Alpha diversity values of bacteria……..... 41

Table 3.1.2 – Average of Simpson Alpha diversity values of bacteria 43

Table 3.1.3 Average of Shannon Alpha diversity values of microbial eukaryotes…….....50

Table 3.1.4 Average of Simpson (1-D) Alpha diversity values of microbial eukaryotes…….…..50

Table 3.2.1 – GCN average values of the bacterial communities found in water.....57

Table 3.2.2 – GCN average values of microbial eukaryotes 59

Table 3.3.1 – Bacteria and microbial eukaryotes correlated with total coliforms and *E.coli*.…..…..…..…....68

Table A.1: Samples considered for viral analysis ……..…92

Table A.2: Samples included for microeukaryotes analysis …….....94

List of Abbreviations

ART ANOVA	Aligned ranks transformation analysis of variance
ASVs	Amplicon sequence variants
BLAST	Basic Local Alignment Search Tool
DADA2	The Divisive Amplicon Denoising Algorithm improved
DO	Dissolved oxygen
DOC	Dissolved organic carbon
dsDNA	Double-stranded DNA
GCN	Gene copy number
g23	Gene 23
IMG/VR	Integrated Microbial Genome/Virus
IMR	Integrated Microbiome Resource
NMDS	Non-metric multidimensional scaling
PCA	Principal component analysis
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
rRNA	Ribosomal ribonucleic acid
SD	Standard deviation
UViGS	Uncultivated virus genomes

1: INTRODUCTION

1.1 Freshwater in Canada: Overview and key human-induced threats

Canada contains 7% of the world's renewable freshwater, over 2 million lakes, and more than 8500 named rivers. The country contains more inland water than any other nation (Environment and Natural Resources, 2022; Desforges et al., 2022). This proclaims Canada as the fourth largest freshwater reserve in the world, only surpassed by Brazil, Russia, and the United States of America (Zambrano-Alvarado & Uyaguari-Diaz, 2024; Environment and Natural Resources, 2022). Fourteen out of the 25 largest rivers found in North America flow completely or partly within Canada (Monk & Baird, 2010). Although Canadian rivers flow into 5 major ocean drainage basins (Arctic, Pacific, and Atlantic oceans), the majority of the rivers flow north into the Arctic Ocean or Hudson Bay (St-Laurent, Straneo, Dumais, & Barber, 2011). Runoff is one of the main factors producing pronounced seasonal variation in most Canadian rivers. Generally, high flows are driven mainly by spring snowmelt, followed by seasonal rainfall and glacial meltwater influencing water flow in mountainous areas (Monk & Baird, 2010). For most unmodified (streams without dams or diversions) low flows are apparent after reduced precipitation or high-water evaporation occurring in late summer or in late winter, when precipitation is collected as ice or snow (Environment and Climate Change Canada, 2024).

Canadian waterways are essential for critical environmental functions that sustain life and are a key driver for the economic and social development of the country. An increasing number of anthropogenic activities threaten the river's water quality in Canada. With population growth and increased water demand, dam constructions in combination with industrial and agricultural development river water quality suffers significant alterations (Monk & Baird, 2010).

Agricultural threats: For example, agricultural cropland runoff including organic contaminants such as hormones, pesticides, and fungicides derived from agricultural industries can also influence water quality and alter the physiology of endemic aquatic microorganisms and other species (NIWA, 2024). Moreover, enteric protozoa from infected calves such as *Cryptosporidium spp*, contained in poorly contained waste are an important source of

contamination. Inorganic contaminants such as Nitrogen or Phosphorus arising from fertilizers used in Agriculture represent a source of nutrients for microorganisms which can lead to harmful algal blooms, that can directly endanger the life of aquatic life including fish and invertebrates, microorganisms, and the environment (Tanabe et al., 2016).

Urban threats: In urban settings, urban runoff, stormwater originating from rain or snow melt can acquire contaminants such as petroleum, oil, grease, microplastics (Xiang et al., 2022), and chemical compounds that end up in water (Ruiz, Vogel, & Taghvaeian, 2017). In addition to this, wastewater treatment plants, especially the ones not designed efficiently, contribute to the discharge of pharmaceuticals, antibiotics, and pathogens into river water. Some pathogens include *Giardia spp, Campylobacter* spp., *Escherichia coli* (*E. coli*) *Shigella* spp., *Hepatitis A* virus, among others (Health Canada, 2022). Moreover, mobile genetic elements (i.e. plasmids, integrons, and so forth), important for transferring antibiotic-resistance genes have also been found in wastewater discharges (Jankowski et al., 2022). Ubiquitous freshwater bacteria and biofilms formed on natural substrates can exchange hereditary material and contribute to the spread of antibiotic resistance genes and pathogens which represents a growing health risk to humans, animals, and the environment (Makk et al., 2024; Wingender & Flemming, 2011).

1.2 Microbial loop in freshwater

Freshwater ecosystems are more functionally diverse than other environments. (Sagova-Mareckova et al., 2021). Microorganisms are estimated to represent 80% of the total marine biomass and are important for global biogeochemical processes (Baron & Milo, 2019; Pomeroy & Hobbie, 2007). Autotrophic protists and photosynthetic bacteria are responsible for the primary production of organic matter in aquatic ecosystems (Pomeroy & Hobbie, 2007). Autotroph protists, for example, convert carbon dioxide (CO_2) to organic carbon via photosynthesis (Worden et al., 2015). These autotrophs then represent a major source of energy for heterotrophic bacterial growth (Worden et al., 2015). These heterotrophic bacteria are later lysed by bacteriophages or grazed by larger predatory microeukaryotes (i.e. microflagellates, ciliates) which play a significant role in transferring energy from low to high trophic levels (Feng et al., 2015). This is how this 'microbial loop' accelerates CO_2 production and recycles nitrogen

and phosphorus compounds (Azam & Malfatti, 2007; Pomeroy & Hobbie, 2007; Rosenberg et al., 2010). Only after 1995, bacteriophages (phages for short) were recognized as key contributors to biogeochemical processes in aquatic ecosystems (Brussaard, 2004; Rosenberg et al., 2010; Hennes & Suttle, 1995). In addition to grazing, executed by larger microeukaryotes, viral lysis is also one of the main ways of microbial mortality in aquatic ecosystems (Sandaa & Larsen, 2006). Bacterial lysis by bacteriophages releases nucleic acids, proteins, and other cell constituents into the water as dissolved organic carbon (DOC). Furthermore, with bacteriophage lysis, DOC is transferred to lower trophic levels, to feed other heterotrophic bacteria (Shiah et al., 2022), whereas with grazing by larger microukaryotes, the carbon becomes available to higher trophic levels (Rosenberg et al., 2010; Cobian Guemes et al.,2016). This strain-specific bacterial lysis performed by bacteriophages controls and shapes prokaryotic populations (Chow & Fuhrman, 2012) ,and lysates play a fundamental role in bacteria-governed carbon transfer and nutrient recycling in aquatic environments (Shiah et al., 2022).

1.3 Surface Water Monitoring in Canada

The surface water (any body of water above ground, including rivers, lakes, and streams) in the country plays a fundamental role in domestic and industrial activities, livestock production and agriculture, recreational activities, and drinking water. Within this context, water microbial pollution represents one of the main threats to human and animal health (Nhantumbo et al., 2023). Surface water protection is essential to avoid drinking water contamination.

It has been recognized that methods for the detection of all the potential pathogens in water are not practical for routine monitoring (Health Canada, 2022), therefore water monitoring efforts focus mostly on preventive approaches. The “source-to-tap” water safety framework, for example, seeks to develop strategies to control potential sources of pollution. The source water assessment aims to i) properly delimitate source water protection areas, ii) identify site-specific contaminants of concern through inventories (i.e. pollutant or land use inventories), and iii) assess the susceptibility to contamination (Canadian Council of Ministers of the Environment, 2004; Health Canada 2020). The identification of site-specific contaminants includes recognizing current and potential fecal contamination sources and/or potential routes by which pathogens can

make their way or increase their concentration in water such as snowmelt, rainfall, and river flow changes (Health Canada, 2019). A graphic representation of the source water protection approach proposed by the Canadian Council of Ministers of the Environment and currently recommended by Health Canada can be found in Figure 1.1.

Depending on the water source, monitoring for enteric protozoa and enteric viruses should also be included. However, water quality monitoring in source and finished drinking water focuses mostly on the identification of indicator microorganisms as a surrogate for the presence of all other potential pathogens. The indicator microorganisms included in the "Guidelines for Canadian Drinking Water Quality" are *E. coli* (fecal contamination indicator) and total coliforms (overall water quality indicator). For the assessment of enteric viruses and protozoa, no indicators are used, instead, water treatment goals are established for their reduction or inactivation based on turbidity levels found in source water. Table 1.3 depicts the microbiological parameters used to assess source and drinking water quality.

1.4 Microorganisms in the Guidelines for Canadian Drinking Water Quality

The "Guidelines for Canadian Drinking Water Quality" on microorganisms include basic parameters and maximum concentrations of specific pathogens to be limited in drinking water (Health Canada, 2020). The microorganisms included in the guidelines are the ones that i) are frequently or expected to be detected or could be found in a large number of drinking water supplies throughout the country; ii) could lead to adverse health effects. Within this framework, the guidelines for Canadian drinking water quality include the following microorganisms:

Bacteria: *E. coli* is the most used fecal indicator organism in drinking water risk management for municipal and residential scales worldwide (WHO, 2024). *E. coli* is a coliform, thermotolerant, bacterium from the family Enterobacteriaceae naturally found in the intestines of warm-blooded mammals including humans, when found in source water it indicates recent fecal contamination (Health Canada, 2020; Wisniewska, Niewolak, Korzeniewska, & Filipkowska, 2007).

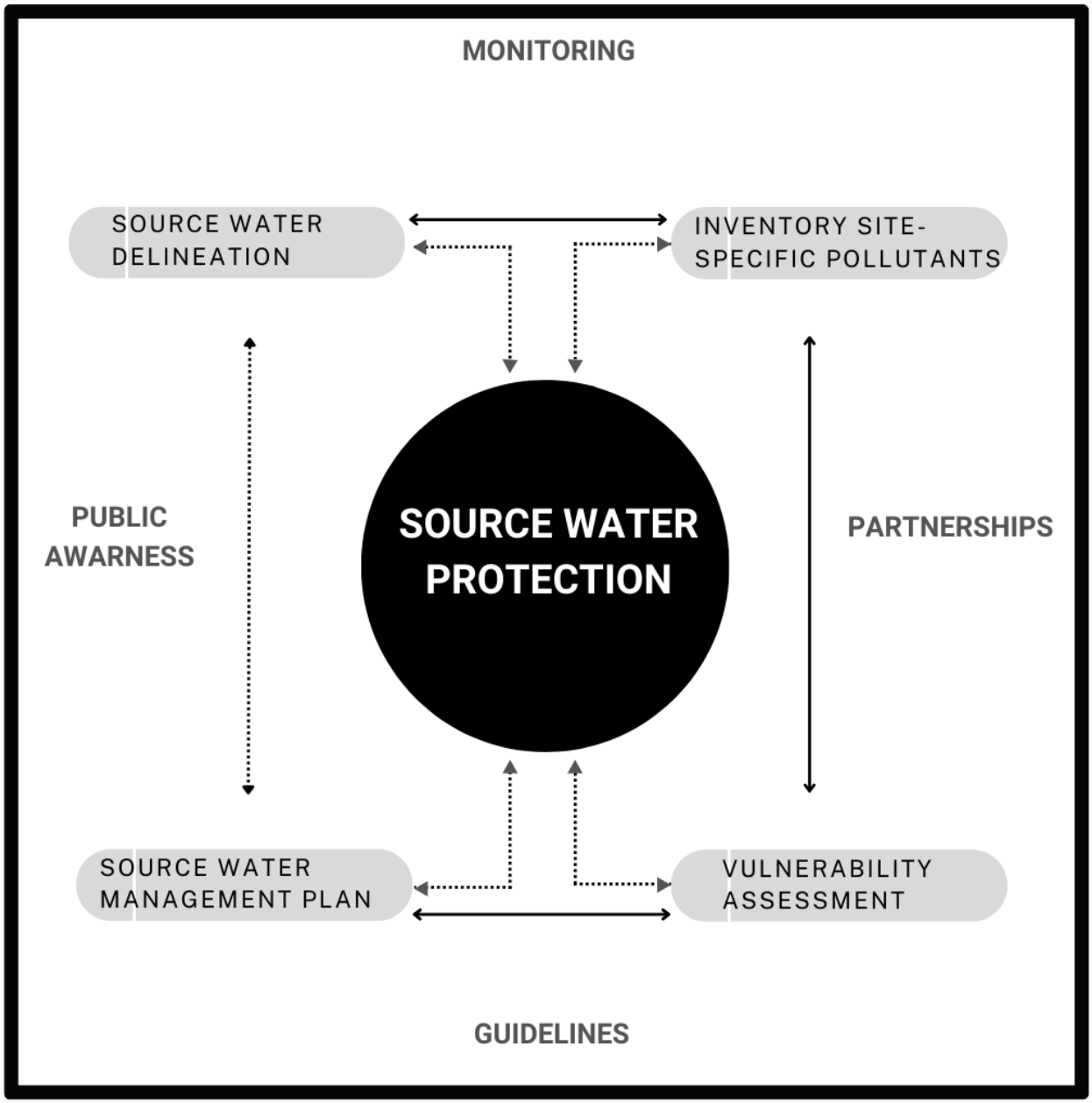

Figure 1.1 Graphic representation of the source water protection strategy. Modified from Canadian Council of Ministers of the Environment, 2004.

Parameter	Guideline	Source water (recommended)	Indication/ Comment
Escherichia coli	0 x 100 mL	<200 CFU x 100 mL	Recent fecal contamination
Total coliforms	0 x 100 mL	<200 CFU x 100 mL	Water vulnerable to contamination
Giardia and *Cryptosporidium*	Minimum 3 log removal or inactivation of cysts/oocysts		Identification in source water determines risk-based approach in DWTP
Enteric Viruses	Minimum 4 log removal or inactivation of viruses		Routine monitoring for viruses no feasible. Disinfection is a critical barrier to reduce risk.
Turbidity	≤0.3 NTU		Water entering the distribution system should have 1.0 NTU or less

Table 1.3 Microbiological parameters used to assess drinking and source water quality; table adapted for surface water using the Guidelines for Canadian Drinking Water Quality . DWTP: Drinking water treatment plant. (Health Canada, 2020; Lévesque & Gauvin, 2007)

Total coliforms: are a group of thermotolerant coliforms from the family *Enterobacteriaceae*. These bacteria are naturally found in fecal and non-fecal origins and the intestines of mammals. Their presence is used to determine vulnerability to contamination by more harmful microorganisms in source water. Members of the total coliforms group include *Enterobacter spp., Citrobacter spp, Klebsiella spp., Serratia spp, Escherichia spp. Budvicia spp, Erwinia spp, Leclercia spp, Pantoea spp*.) (Martin, Trmcic,Hsieh, Boor, & Wiedmann, 2016; Bartram, Cotruvo, Exner, Fricker, & Glasmacher, 2003, Health Canada, 2020)

Protozoa: including the enteric and environmentally resilient i) *Giardia* and ii) *Cryptosporidium* have been recognized as common causes of waterborne disease outbreaks causing infectious enteritis in Canada. In 2015, *Giardia spp*. was the most identified intestinal parasite in the country (Health Canada, 2019). *Giardia* (flagellated protozoan) and *Cryptosporidium spp.* are common parasites found in the intestine of humans and other mammals. These microeukaryotes can produce thick-walled cysts that protect them from external environmental conditions and in the case of *Cryptosporidium* also from water disinfectants, making them survive for long periods under typical environmental conditions (Health Canada, 2019; WHO, 2008)
Giardia cysts are excreted in large numbers in the feces of infected humans and other wild (i.e. beaver, muskrat) and domestic (i.e. horses, dogs) animals, disseminated in the environment, and transmissible via the fecal-oral route (WHO, 2008). On the other hand, human infectious *Cryptosporidium* spp. oocysts are mostly excreted from infected calves and represent an important source of *Cryptosporidium* in surface water. *Giardia* cysts and *Cryptosporidium* oocysts are commonly found in sewage, surface waters, and occasionally in drinking water. The highest concentrations of pathogenic cysts and oocysts in surface water have been associated with spring runoff events (Health Canada, 2019; Daraei et al., 2021; Doménech et al., 2022).

Viruses:5 RNA enteric viruses causing gastrointestinal illness are also recognized including: i) Norovirus, ii) Hepatitis A, iii) Hepatitis E, iv) Rotaviruses, and v) Enteroviruses. In general, these biological entities are spread in the environment through fecal matter. Therefore, surface water can receive virions from wastewater effluents, dumping of sanitary sewage, and leaking sanitary sewers, among others (Powelson, Gerba, & Yahya, 1993).

1.4.1 Other waterborne pathogens

There is a vast number of waterborne pathogens that can be found in source water that differ from those included in the guidelines for Canadian drinking water quality. Some of the pathogens recognized as concerning but not officially counted in the Guideline include:

Bacteria of gastrointestinal origin including the gram-negative Proteobacteria: i) *Campylobacter spp.*; ii) Enteric pathogenic *E. coli* and *Shigella spp*., iii) *Helicobacter pylori*, iv) *Salmonella spp*., and v) *Yersinia spp*. These enteric pathogens can enter surface water from inadequate sewage treatment plant effluents and lagoon discharges, inadequate septic systems, and overland transportation through rainfall or snowmelt (Health Canada, 2013; Chahal et al., 2016).

i) Campylobacter spp., a group of rod-shaped bacteria from the family *Campylobacteraceae*, is among the most frequently reported causes of waterborne outbreaks in developed countries (Moreira and Bondelind, 2017; Health Canada, 2022)) Interestingly, *Campylobacter spp*. and *E. coli* were the pathogens responsible for the most notable Canadian water outbreak where more than 2,300 people became ill in May 2000 in Walkerton, Ontario (Health Canada, 2022; Salvadori et al., 2009). *Campylobacter spp*. can reach surface water from intrusion or runoff containing wild and domestic birds and mammals' feces (Sharafutdinov et al., 2024).

ii) E. coli and Shigella spp.: E. coli and *Shigella spp.* are nearly indistinguishable rod-shaped bacteria that belong to the family *Enterobacteriaceae* (Ud-Din & Wahid, 2014). *E. coli* is naturally found in the intestines of warm-blooded mammals including humans. On the other hand, *Shigella spp.* only infects humans and has no natural animal reservoirs. *Shigella dysenteriae* produces shigatoxin, a potent enterotoxin capable of killing host cells (Wexler, 2024). In the drinking water industry, Shiga toxin-producing *E. coli* and verotoxin-producing *E. coli* (VTEC) are among the most concerning pathogens (Health Canada, 2022). *E. coli* and Shigella can access surface water through sewage and surface runoff similar to *Campylobacter spp*

iii) Helicobacter pylori (*H. pylori*) are highly motile, helical, species from the family *Helicobacteraceae*, which have adapted to the 'inhospitable' conditions of the gastric mucosa of

mammals (Kusters, van Vliet, & Kuipers, 2006). Although *H. pylori* has not been recognized to cause waterborne outbreaks, these bacteria are included in the Canadian guidance on waterborne pathogens (Health Canada, 2022) .

iv) Salmonella spp. are predominantly motile enterobacteria from the family *Enterobacteriaceae,* known for causing gastrointestinal infections in animals and humans. In Canada, *Salmonella* spp. has been recognized as the second-leading cause of bacterial gastrointestinal illness. Wild animals, birds (including chicken and turkey) pigs, and cattle have been recognized as *Salmonella spp* reservoirs (Baron, 1996). In general, (non-typhoidal) *Salmonella* is rarely linked to drinking water outbreaks, however, this pathogen is of concern because of its resistance to antibiotics such as ciprofloxacin, and their potential ability to spread resistance to other microbes through horizontal gene transfer (Tao, Chen, Li, Wang, & Liang, 2022).

v) *Yersenia* spp. is a rod-shaped bacterium known to cause gastrointestinal illness belonging to the recently reclassified *Yersiniaceae* family (Wyllie, Hyams, & Kay, 2020). In Canada, *Yersinia*-associated diseases are generally associated with *Y. enterocolitica.* For this strain, pigs, cattle, goats, and dogs have been recognized as reservoirs (Health Canada, 2022) . Although *Yersinia*-associated disease was the third most reported enteric illness in 2022 in the European Union (Fredriksson-Ahomaa et al., 2024) waterborne outbreaks caused by *Yersinia* spp. have very rarely been reported in Canada *(*Health Canada, 2022).
In addition, there are naturally (environmental) occurring bacteria including i)*Aeromonas spp*. (associated with Gastroenteritis); ii) *Legionella* spp. (source of respiratory illness and pneumonia); iii) *Mycobacterium spp*. (pulmonary diseases in immunocompromised individuals); iv) *Pseudomonas* spp (skin, bloodstream, urinary, and lung infections) ; v) *Acanthamoeba spp*. (eye infection mainly) and vi) *Naegleria fowleri (a*mebic meningoencephalitis) (Health Canada, 2022; Marciano‐Cabral, Puffenbarger & Cabral, 2000).

1.5 Limitations of Traditional Microbial Water Quality Indicators

The assessment of enteric pathogens relies on the identification of *E. coli* (fecal contamination indicator), total coliforms (overall water quality indicator), and water turbidity in most cases (Health Canada, 2022).

For enteric virus monitoring, when direct virus assessment is unavailable, proposed indicators include *E. coli*, total coliforms, Enterococci, *Clostridium perfringens* spores, and coliphages. While coliphages have been demonstrated to reflect the risks of exposure to human enteric viruses, the use of fecal indicator bacteria to monitor surface water is of concern.

With aging sewer infrastructure, the deterioration of water and the environment, and increased contamination from extreme weather events and floodings carrying contaminated runoff to source water, relying on a single-class microorganism as a surrogate for all potential pathogens is insufficient (Holcomb &Stewart, 2020).

Several studies have demonstrated that single indicator organisms in surface water such as *E. coli* failed to correlate with other pathogens including *Giardia spp.*, *Clostridium perfringens* (Dorner et al., 2007) , *Salmonella enterica*, *Listeria monocytogenes* . In addition to this, co-occurrence with many other pathogens has been inconsistent (Baker, Almeida, Lee, & Gibson, 2021). While it is somewhat publicly recognized that the use of *E. coli* as an indicator for protozoan and viral contamination in surface water is not accurate, this enterobacterium remains the standard due to its rapid measurement capability, rather than its effectiveness in predicting contamination. Current methods for water monitoring emphasize the need to develop more suitable indicator microorganisms to protect public health.

1.6 Research premise

Water is a key driver of economic and social development. The modern pressures of the industry, population growth, environmental degradation, and changing land uses are a constant threat to freshwater. Current water quality tests focus on coliform bacterial species as indicators of microbial pollution. However, these techniques are insufficient to monitor potential enteric pathogens and the health of aquatic ecosystems. Effective biomarkers of microbial contamination have yet to be developed. Research done on water tends to be conducted in the same season ignoring seasonal sensitive factors such as water flow, surface runoff, snowmelt, and human

activities surrounding the waterways. Horizontal comparison of time factors between water bodies can help elucidate seasonal stability, differences in abundance, and the overall distribution of microorganisms among waterways.

In this research, culture-independent approaches such as deep amplicon sequencing of targeted genes were used to expand the understanding of the overall microbial biodiversity residing in waterways. Such an approach can reveal taxonomic profiles of bacteria, viruses, and microeukaryotes coexisting in freshwater environments. The value of culture-independent sequencing approaches using ribosomal RNA (rRNA) sequences to catalog several hundred microbial species to reveal a community's microbiome has been illustrated by several studies examining soil, water, groundwater, and feces (Liles et al., 2003; Yadav et al., 2021). Ribosomal RNA (rRNA) genes present in prokaryotes and eukaryotes have conserved and hypervariable regions where more conserved fragments are useful for determining the higher-ranking taxa. In contrast, more hypervariable ones can help identify genus or species.(Kim, Morrison & Yu, 2011; Mincheol, 2014; Bukin et al., 2019). The structure of ribosomes and translation factors have changed remarkably little over the years, that is the main reason rRNA genes are useful in phylogeny. The 16S rRNA gene serves as a molecular target and is by far one of the most common approaches targeting housekeeping genes to study bacterial phylogeny and genus/species classification (Bukin et al., 2019). Similarly, despite the chromosomal complexity and bigger genome of eukaryotes the 18S rRNA gene is the most commonly used molecular marker to explore the microeukaryote community structure due to its high specificity and sequence conservation (Gong & Marchetti, 2019; Massana & López-Escardó, 2022). On the other hand, viruses lack rRNA genes or "universal" gene markers, therefore, we can only target specific viral groups to characterize bacteriophages. In this research, the identification of the major capsid protein gene (g23) of viral groups such as T4-like viruses which affect marine cyanobacteria and members of the family Enterobacteriaceae were studied. Viruses (particularly, bacteriophages) are the most abundant entities in aquatic ecosystems (Potapov et al., 2018). It is well documented that bacteriophages have an imperative effect on the ecosystem of bacterial populations since they mediate horizontal gene transfer and gene flow, microbial community size and variety, modifications of host metabolism and lysis (Brum et al., 2015; Forterre & Prangishvili, 2013; Roux et al., 2015; Suttle, 2007). In fact, the availability of elements such as carbon, phosphorus, and nitrogen, critical for the trophic chain, is directed by the lysis of

biomass performed by bacteriophages in the ocean (Zheng et al., 2013). Characterizing ecological relationships between viruses, bacteria, and microeukaryotes in the waterways is critical to understanding ecosystem function, however, few studies have addressed these interactions.

1.7 Research objectives

In the present research, we aim to characterize the ecological correlations among domains to understand the control and role of bacteriophages, bacteria, and microbial eukaryotes in structuring the aquatic microbiomes. Microbial communities can help expose the impacts of anthropogenic activities on the health of aquatic environments. Furthermore, evaluating all microbial fractions may provide fundamental data to set a baseline for new effective biomarkers of microbial aquatic contamination which are essential to better protect public health.

To accomplish this, two objectives were established

1) Identify the microbial communities present in anthropogenically impacted water bodies of rural Manitoba.

2) Perform a comparative analysis of the microbial fractions encountered in these freshwater ecosystems to set a baseline of microbial community structure in this area.

2.0 MATERIALS AND METHODS

2.1 Water Sampling

Two biological replicates consisting of 2L of surface water samples were collected from waterways adjacent to urban and agricultural impacted areas. A case-control study was conducted to compare microbiomes in a less impacted forested area. All the sampling sites were interconnected along the portion (12 Km) of the Assiniboine River located in the city of Portage la Prairie and around different land uses close to First Nation communities. The sampling was performed in the following order (less to more contaminated areas): Location 1: Forested area (3 sub-locations); Location 2: Water samples adjacent to urban areas (3 sub-locations) and Location 3: water samples adjacent to agricultural areas (3 sub-locations). Locations are depicted in Figure 2.1 and coordinates can be found in Table 2.1. Milli Q water was used as a negative control for the sample processing. This totals 10 samples, each with a biological duplicate, with eight sampling events in the year 2022 to evaluate seasonal variability. This accounts for a total of 160 samples x 3 microbial targets. Sampling events occurred once per month from April to November 2022 on the following dates: April 26th, May 27th, June 17th, July 28th, August 23rd, September 12th, October 15th, and November 8th. The samples collected were stored at 4 °C until processing, which occurred within 24 hours after collection.

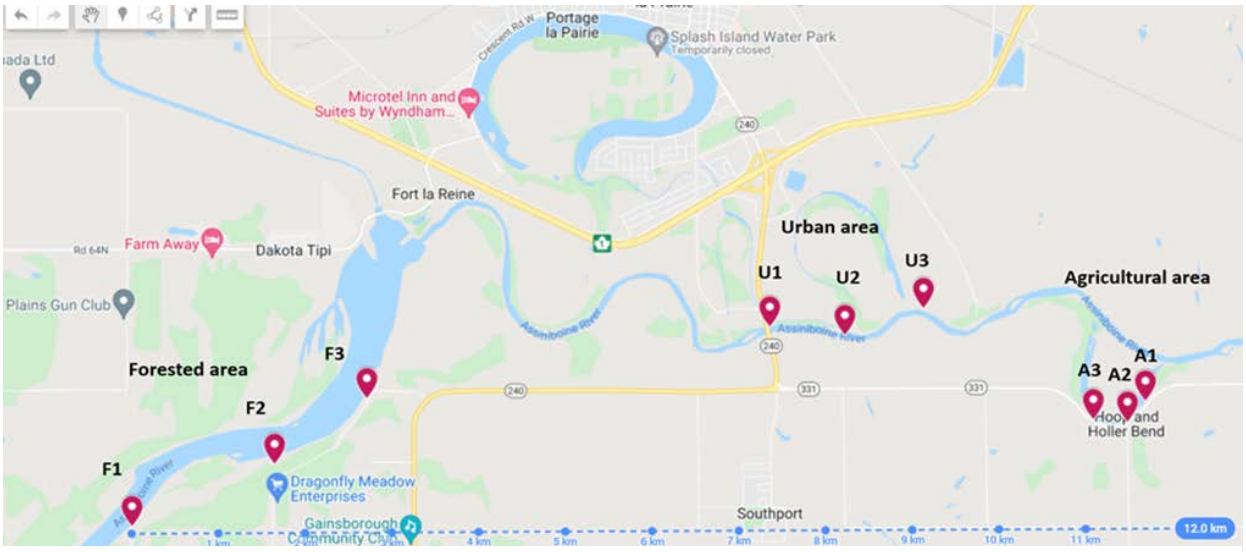

Figure 2.1 Map indicating the sampling points (red pins) along the 12 km of the Assiniboine River in rural Manitoba. Source: Google Maps, 2023

Sampling ID	Coordinate	Description
F1	49°54'59.6"N 98°22'34.3"W	Forested area No. 1; limited anthropogenic activities around the zone.
F2	49°55’22.7”N 98°21’12.3”W	Forested area No. 2; limited anthropogenic activities around the zone.
F3	49°55’48.7”N 98°20’18.9”W	Forested area No. 3; limited anthropogenic activities around the zone.
U1	49°56’14.6”N 98°16’25.7”W	Urban area No. 1; Close to urban activities, 2.2 km (south) from the wastewater treatment plant of Portage la Prairie
U2	49°56’10.0”N 98°15’40.7”W	Urban area No. 2; 4.2 km (south) from the wastewater treatment plant of Portage la Prairie
U3	49°56’21.3”N 98°14’51.2”W	Urban area No. 3; 4.1 km (east) from the wastewater treatment plant of Portage la Prairie
A1	49°55’47.0”N 98°12’47.4”W	Agricultural No. 1; < 5 km from the closest berry farm
A2	49°55’39.0”N 98°12’57.6”W	Agricultural No. 2; <5 km from the closest berry and celery farms
A3	49°55’39.7”N 98°13’17.0”W	Agricultural No. 3; < 5 km from the closest berry and celery farms

Table 2.1: Coordinates and description of the sample locations used in this study performed in rural areas of Manitoba, Canada.

2.2 Concentration of bacteria, viruses, and eukaryotes, from environmental samples

Surface water samples collected from forested, urban, and agricultural waterways were processed to recover the representative bacteria, viruses, and eukaryotes. Raw water that had large macroscopic solids was filtered through a sterile cheesecloth, and the remaining samples were directly treated with the Skimmed milk flocculation procedure described by Calgua et al. (2008) and Gonzales-Gustavson et al. (2017). Pre- flocculated skimmed milk solution [1% (w/v)] was prepared by dissolving 2.5 g of Difco skimmed milk powder in 250 mL of artificial seawater (Ricca chemical, Texas, USA) and carefully adjusting the pH to 3.5 with 1 N HCl. Twenty milliliters of this solution were added to each of the previously acidified 2-liter freshwater samples. The change in the isoelectric point of the microorganisms in an acidic pH improves their precipitation out of the water. The final concentration of skimmed milk was 0.01% (w/v)). The samples and the skimmed milk solution were mixed for 8 hours at room temperature using a magnetic stirrer, and the flocs were allowed to sediment by gravity for another 8 hours. The supernatants were cautiously removed with a peristaltic pump (Thermo Scientific, Massachusetts,USA), and the sediment was collected and transferred to a 50 mL centrifuge tube and centrifuged at 8000 × g for 30 minutes. The obtained pellet tentatively containing viruses, bacteria, and eukaryotes, was weighed and resuspended in 200 μL of 0.2 M phosphate buffer (pH 7.5). The concentrates were kept at -20 °C. The nucleic acid extraction was performed 48 hours after the recovery of the flocs, the weighed values were considered to assess gene copy number per 100 mL of water sample as explained in section 2.5.1.

2.3 Nucleic acid purification

To isolate the genetic material of viruses, bacteria, and microeukaryotes, the commercial kit MagMax microbiome Ultra nucleic acid isolation kit (Applied Biosystems, Massachusetts, USA) was used. Once the flocs were obtained, they were added to the bead-beating tubes (included in the kit) and lysed with 800 μL of lysis buffer. The flocs were vortexed twice, the first time individually, upside down for 10 seconds to mix the beads, the floc, and the lysis

buffer and subsequently, they were vortexed in a group with a vortex adaptor at 2,500 rpm for 20 minutes. Right after the lysis, the samples were centrifuged for 2 minutes at 14,000 x g, as much as possible of the supernatant was transferred to a new microcentrifuge tube containing 40 μL of Proteinase K for digestion of contaminant proteins including nucleases. After that, samples were shaken at 900 rpm for 5 minutes, and then incubated at 65°C for 20 minutes. Afterward, the binding bead mix was prepared using 500 μL of binding solution and 25 μL of total nucleic acid binding beads (per reaction) to add to the samples, and then the tubes were shaken for 5 minutes at 900 rpm. Then, the tubes containing the samples were placed on the magnetic stand for 8 minutes, or until all beads were collected (some samples took up to 15 minutes to fully collect). Next, the supernatant was discarded while keeping the tubes on the magnet, and 1 mL of wash buffer was added directly to the beads. Later, the beads and the wash buffer were shaken at 800 rpm for 30 seconds, the tubes returned to the magnet until all beads were gathered. The mentioned wash process was repeated with wash buffer, then with 1mL of 80% ethanol (2 times). In total 2 washes were performed with wash buffer, and the other 2 with 80% ethanol. After the last ethanol wash, the beads were air-dried by opening the lid of the tubes (for 2 minutes) while being on the magnet, inside a biosafety cabinet to eliminate any leftover ethanol. Finally, 50µL of elution solution was added to each sample. Then the tubes were placed in the incubator at 75°C for 5 minutes and shaken while incubating at 800 rpm. Finally, the tubes were placed on the magnetic stand for the last time until all beads were collected, and the freshly extracted nucleic acids (eluates) were transferred to a new clean microcentrifuge tube. The quantity (ng/µL) of each nucleic acid extracted was assessed using the Qubit dsDNA High Sensitivity Assay Kit (Invitrogen, Massachusetts, USA) by quantifying the total amount of DNA in the samples.

2.4 Characterizing the microbial distribution

2.4.1 Amplicon-based characterization of viruses

1 µl of the nucleic acid extracts was used to perform a quantitative polymerase chain reaction (qPCR). This assay targeted the 'g23' fragment present in the capsid of the T4-like morphotype of phages. Primers used for the first PCR were designed by (Filée et al., 2005) and the respective sequences are described in Table 2.2. A nested approach was suggested with T4 super primers as recommended by Chow et. al (2012) and the Integrated Microbiome Resource (IMR) (IMR, 2024). Initially, the nested approach was performed in the lab with 'T4 super primers', and the amplicons obtained were around 300 bp. After sequencing issues (reported by IMR), only the 1st PCR product with MZIA1bis and MZIA6 was submitted. Each PCR reaction consisted of 0.4 µM of primers, 1.25 U of Hot Start Polymerase (Promega Corporation, Fitchburg, WI), 1.5-mM MgCl2, 0.2 mM of nucleotides, and water in a 50-µl volume. The conditions used were as follows 94 °C × 1.5 min, 35 cycles of 45 s at 94 °C, 60 s at 50 °C, 60 s at 72 °C, and 5 min at 72 °C. PCR amplicons and potential contaminants were purified using the QIAquick Gel Extraction Kit (QIAGEN Group, Maryland, MD, USA) following the manufacturer's instructions. After amplicon clean-up, randomly selected samples were resolved in a 1.5% agarose/1 X TAE gel stained with 1X GelRed (Biotium, Inc., Hayward, CA) to verify the presence of the g23 fragment. A total of 160 amplicons were sent out for sequencing. A volume of 40 µL of g23 amplicons representing each sample (including biological duplicates) was transferred to 0.2 mL PCR tube strips with caps for sequencing (M Comeau, 2022; IMR, 2024). Illumina Miseq technology V3 chemistry, ~400 bp was used. A myovirus propagated in a *Synechococcus* sp. strain was grown in LB broth + kanamycin (50 ug/mL) and used as a positive control. The myovirus DNA was extracted with the MagMax microbiome ultra nucleic acid isolation kit (Applied Biosystems, Massachusetts, USA), amplified through PCR with MZIA1bis and MZIA6 primers, purified with QIAquick Gel Extraction Kit and sent out for sequencing to IMR together with amplicons of the samples and biological duplicates.

Assay	Primer/Probe	Sequence 5'-3'	Amplicon size (bp)	Reference
Myovirus sequencing	MZIA1bis (F)	GAT ATT TGI GGI GTT CAG CCI ATGA	400-500	(File et al., 2005)
	MZIA6 (R)	CGC GGT TGA TTT CCA GCA TGA TTTC		
Myovirus sequencing	T4superF1	[TET] GAY HTI KSI GGI GTI CAR CCI ATG	300-400	(Chow & Fuhrman, 2012)
	T4superR1	[6FAM]GC IYK IAR RTC YTG IGC IAR YTC		
16S rRNA gene (qPCR)	Bac1055YF	ATG GYT GTC GTC AGCT	320 (approx)	(Ritalahti et al., 2006)
	Bac1392R	ACG GGC GGT GTG TAC		
	Bac1115P	(FAM) CAA CGA GCG CAA CCC (/ZEN, IABkFQ)		
18S rRNA gene (qPCR)	18S rDNA (F)	GTA GTC ATA TGC TTG TCT	315->1000	(Wei et al.,2024)
	18S rDNA (R)	ATT CCC CGT TAC CCG TTG		

Table 2.2: Primers and probes used for sequencing and quantification assays to screen for myoviruses, bacteria, and eukaryotes.

2.4.1.1 Viral Taxonomic Profiling

Successful demultiplexed paired-end sequencing reads provided by IMR were uploaded into The Galaxy Europe Server (version 24.1.2) (Galaxy, 2024). Galaxy enables access to several third-party open-source bioinformatic analysis tools to process different types of data including next-generation sequencing (Galaxy, 2024). These analytical tools can be used individually or linked into complex workflows. Merging issues between forward and reverse reads were experienced, therefore, only forward reads were considered for the characterization of viral communities. Generally, Illumina reverse reads have more error reads than forward sequences, producing a lower quality score for reverse sequences (Sunyoung Kwon et al., 2013). When merging paired-end reads, if the forward and reverse sequences do not match, the base with the higher quality score is selected (Sunyoung Kwon et al., 2013). Failure in performing this process due to multiple miscalls was potentially the reason for the merging issues encountered. After building a collection with the forward g23 reads, the primers were cut from the reads using Trimmomatic trimming tool with HEADCROP: 24 operation. Moreover, amplicon lengths above 450 bp were removed using the CROP operation in Trimmomatic (Bolger, Lohse, & Usadel, 2014). The trimmed forward sequences were used for taxonomy assignment using the k-mer-based algorithm Kraken 2 applying the default parameters established in Galaxy Europe. This taxonomic classification system matches short genomic substrings (known as k-mers) to the lowest common taxonomic ancestor (LCA) contained in that given k-mer (Wood, Lu, & Langmead, 2019). The taxonomic outputs generated for each sample were visualized with 'Krona' which is a tool that creates an interactive file allowing hierarchical data to be explored. To depict the relative abundances, the individual species found across location and time were divided by the total number of species population per site and month (in each sample). Visualization of viral relative abundances was created using the ggplot2 package in R 2024.04.1.

2.4.1.2 Viral diversity

A local BLAST nucleotide database was created with the most recent version of the Integrated Microbial Genome/Virus (IMG/VR) database v.4 (Camargo et al., 2023) using the 'makeblastdb' application in (BLAST+ 2.9.0) (Camacho, 2008). More than 15 million virus genomes and genome fragments available in the IMG/VR database were used to obtain the uncultivated virus genomes' or 'UViGs' present in the successful R1 (forward) Illumina fasta files. A nucleotide search was performed against the IMG/VR database customized and only matches within the set threshold (e-value equal to or less than 1e-5) were reported. The resulting files were then filtered based on e-value and best bit scores to obtain only the best UViGs present in the viral reads. The BLAST output files containing the UViGs were transformed into a comma-separated values (CSV) format and compiled using the 'get data' function in Microsoft Excel. A pivot table was created to count different UViGs present in each sample. Once the count was performed, the file was transformed using Pyvis library with Python 3 (Colab, 2024). This step enabled the inclusion of UViGs names and counts and to tracking the UViGs source (sample).

Alpha diversity including Shannon (H) and Simpson (1-D) (also known as Gini-Simpson) was calculated by adding the Shannon and Simpson functions and formula (Magurran, 2003) in R 2024.04.1. Chao-1 richness was calculated using the Chao-1 bias corrected formula to account for data with no doubletons (Gotelli & Colwell, 2010). Kruskal-Wallis test was applied to the diversity and richness indices to establish statistical differences among locations and time. Dunn's test was further conducted for multiple comparisons when the Kruskal-Wallis test was significant. To address the False Discovery Rate (FDR), the p-values were adjusted using the Benjamini-Hochberg statistical method (Mangiafico, 2015).

Beta diversity was calculated to estimate if UViGs obtained from the g23 sequences were different in the forested, urban, and agricultural waterways studied in this research and to analyze if they changed over time. Non-metric multidimensional scaling (NMDS) ordination based on Bray Curtis dissimilarity was used to identify the different patterns of viral composition in the waterways studied. To address the differences in library size between communities a rarefaction threshold was set based on the minimum significant UViG count (Lin & Peddada, 2020). To identify if the differences observed were statistically significant, the multivariate homogeneity of group dispersions, as well as the Permutation test for homogeneity of multivariate dispersions, were

calculated using the betadisper and permutest functions from the vegan (version 2.6-6.1) package in R (R, 2024). A p-value of 0.05 was assumed for all the tests as the minimum level of significance. Beta diversity analysis and graphs were generated in R using the vegan package (metaMDS, adonis2 for PERMANOVA,betadisper, and permutest functions) and visualization was performed with the ggplot2 package.

2.4.2 Amplicon-based characterization of bacteria and microbial eukaryotes

A total of 160 DNA extracts collected over eight sampling events were used for quantification and sequencing. 10 μL of microbial DNA of each sample (including biological duplicates) were transferred to 0.2 mL PCR tube strips with caps properly sealed with parafilm and sent to IMR for sequencing (M Comeau, 2022; IMR, 2024). Illumina Miseq technology V3 chemistry, ~400 bp was used. A mock community of pooled DNA of bacteria and eukaryotes was included to account for sequencing controls for the sequencing of 16S rRNA and 18S rRNA genes respectively. The bacterial DNA used in the mock community included the species: *Pseudomonas aeruginosa*, *Salmonella enterica*, *Campylobacter lari*, *Legionella pneumophila*, *Legionella longbeachae*, *Escherichia coli*, *Staphylococcus aureus*, *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter upsaliensis*. Whereas for eukaryotes, the mock community included *Giardia lamblia*, *Cryptosporidium parvum*, *Claviceps purpurea*, and *Candida albicans*. DNA concentration of each microorganism was assessed before being pooled and the quantities included were adjusted based on their equimolar concentration of 10 ng/μL. As outlined by IMR protocols, the samples were indexed with barcodes through PCR, verified with gel electrophoresis, and products were cleaned up and finally sequenced using Illumina MiSeq technology (M Comeau, 2022).

2.4.2.1 Bacteria and microbial eukaryotes taxonomic profiling and diversity

Demultiplexed paired-end sequencing reads provided by IMR were uploaded into the R software R 2024.04.1. The Divisive Amplicon Denoising Algorithm (DADA2) (Callahan et al., 2016) R software package was used to inspect sequence quality profiles, trim bad quality read ends, chimera identification, merge paired reads, and infer sample sequences. Primers were trimmed using the‘trimLeft’ argument in DADA2 (trimLeft =c(17,21)). Standard filtering parameters were used,

and the maximum expected errors were slightly increased for reverse reads (maxEE=c(2,5) (Callahan, 2024). With DADA2, sample composition can be inferred with the division of amplicon reads into partitions consistent with an error model created, this enables to differentiate between biological variations and Illumina amplicon errors (Callahan et al., 2016). The inferred amplicon sequence variants were used for taxonomy assignment using the naive Bayesian classifier method available in DADA2 and the Silva database version 132 (Quast et al., 2012). Phyloseq R package was used to calculate and visualize alpha diversity including Shannon (H) and Simpson (1-D) (also known as Gini-Simpson). Kruskal-Wallis test was applied to the diversity and richness indices to establish statistical differences among locations and time. Beta diversity and the different patterns in bacterial and eukaryotic composition were calculated using non-metric multidimensional scaling (NMDS) ordination based on Bray Curtis dissimilarity. To identify if the differences observed were statistically significant, the multivariate homogeneity of group dispersions, as well as the Permutation test for homogeneity of multivariate dispersions, were calculated using the betadisper and permutest functions from the vegan (version 2.6-6.1) package in R (R, 2024). A minimum level of significance of 0.05 was assumed for all tests. Beta diversity analysis and graphs were generated in R using the same packages detailed for viruses and visualization was performed with the ggplot2 package.

2.5 Quantitative PCR analysis

2..5.1 Bacteria and eukaryotes quantitative PCR analysis

To better understand the microbial magnitude and composition found in the forested, urban, and agricultural waterways studied, quantitative PCR (qPCR) targeting 16S rRNA and 18S rRNA genes was performed for bacteria and eukaryotes, respectively. For 16S rRNA, each 10 µL Taqman qPCR reaction consisted of 400 nM of each primer, 100 nM of the probe, 5 µL of Taqman environmental master mix (Applied Biosystems, Foster City, CA), and 1 µL of template DNA to screen for the 16S rRNA gene. Regarding the 18S rRNA gene quantification, each 10 µL SYBR green qPCR reaction contained 400 nM of each primer, 5 µL of PowerTrack SYBR Green Master Mix (Applied Biosystems, Foster City, CA), and 1 µL of template DNA. The details about the primers used for quantification are detailed in Table 2.2. Each qPCR reaction was performed in

triplicate using Applied Biosystems QuantStudio 5 real-time PCR (Applied Biosystems, Foster City, CA). The 16SrRNA and 18S rRNA gene thermal conditions were: 50 °C x 2 min (incubation), denaturation, and activation of polymerase at 95 °C x10 min, followed by 40 cycles of 15 s at 95 °C and 60 s at 60 °C (Jankowski et al., 2022). Standard curves were generated from serial dilutions of genomic DNA of *Pseudomonas aeruginosa* 2.58 $X10^5$ copies/uL (ATCC 10145*)* and *Candida albicans* ($5.11x10^5$ copies / uL) for 16S rRNA and 18S rRNA genes, respectively. Each assay included a non-template control of nuclease-free water (Promega Corporation, Fitchburg, United States). All the samples that were amplified above cycle 38 were considered negative, as recommended by the Center for Disease Control and Prevention (CDC) (CDC, 2022). For 18S rRNA gene quantification, samples with no amplification were diluted (1:10) using nuclease-free water. This additional step was conducted to address any reaction failure produced by potential inhibitors. The diluted samples were further adjusted in the gene copy number (GCN) calculation. Raw values obtained from the quantification were evaluated to obtain GCN per sample volume following the formula published by Ritalahti et al. (2006):

$$\frac{(gene\ copies\ per\ reaction)\ x\ (total\ volume\ of\ DNA\ extracted)}{(volume\ of\ DNA\ used\ per\ qPCR\ reaction)x\ (volume\ of\ environmental\ sample)}$$

The equation above considers the gene copies per reaction (based on the standard curves obtained on the cycle threshold number) multiplied by the volume of DNA extracted (50µl) divided by the volume of DNA used per reaction (1 µl) and the volume of sample from which the DNA was extracted (Ritalahti et al., 2006). Bacteria and microbial eukaryote GCN values were normalized per 100 ml of water and nanogram of DNA using equations described by Ritalahti et al. (2006). To address the multicopy nature of the 16S rRNA gene in different bacteria, an average factor of 4.3 was used to normalize GCN per ml of sample (Uyaguari-Diaz et al., 2016; Lee, Bussema & Schmidt, 2009). To date, there is no factor to assess the multicopy nature of the 18S rRNA gene in eukaryotes. To better interpret the data, all values were log10 transformed to reduce the span of a broad range of values obtained. The normality of the data was evaluated with the Shapiro-Wilk test. A linear mixed model was created to test the effects of time and location on gene copy

number. Align rank transform ANOVA (ART ANOVA) was used as the nonparametric method for analysis of variance on the model with months (fixed) and location (random) repeated measurements. A Tukey's Post hoc test for multiple comparisons was performed to identify differences among the variables (Mangiafico, 2016). To understand how much of the variability of GCN could be explained by the predictor variables, a partial eta squared was conducted. Finally, to determine the good fit of the model created, Efron-pseudo-R squared was used. A p–value of 0.05 was assumed as a minimum level of significance for all tests. Analysis and visualization of the data were performed in R software (R 2023.12.1) with lme4, Matrix 1.6.3, ARTool and ggplot2 packages. A schematic representation of the methodology can be found in figure 2.2.

2.6 Assessing bacteriological counts using standard methods for water monitoring

Chromocult Coliform Agar (EMD Millipore, Darmstadt, Germany) was prepared following the manufacturer's instruction for the simultaneous detection of total coliforms and *E. coli* in environmental water samples. Chromocult contains the substrate Salmon-GAL, used for the detection of β-D-galactosidase activity typically found in coliforms. The interaction of total coliforms with this substrate results in pink-red colonies. For *E. coli* the substrates Salmon-GAL and X-glucuronide are used for the recognition of β-D-glucuronidase activity. As *E. coli's* enzymes cleaves both substrates, these cells result in a violet color (Millipore, 2022). Direct plating was used with 100 µl of the freshwater sample to the plate containing Chromocult media (run in duplicate). Colonies were quantified using the standard-plate count and selected isolates were transferred three more times to fresh plates and obtain pure colonies. Bacteria were resuspended in Lysogeny broth and 25% glycerol before storage in 2.0 ml cryogenic storage vials (GeneBio Systems, Toronto,ON, Canada) at -80°C.

2.7 Comparative analysis of the microbial fractions with seasonal and spatiotemporal patterns

To explore the strength of the relationships between microbial communities inferred and seasonal and spatiotemporal patterns collected (metadata), the multivariate data analysis PCA (principal component analysis) was employed after data standardization (scaling) using vegan and factoextra (Reyes, 2020) packages in R (2023.12.1). The non-parametric Spearman's rank correlation was applied to the standardized data to comprehend the pairwise correlations between variables and co-occurrence across taxa using the corrplot library in R (2023.12.1) and networkx and pyvis libraries in Python 3 (Colab, 2024). Correlation coefficients >+0.30 were considered significant for positive correlation, whereas negative coefficient values <-0.30 were considered for negative correlations. Moreover, for the co-occurrence network only linear correlations >+0.30 and significant p- values (<0.05) were considered. In addition to the inferred ASVs and UVIGs of the genes of the microorganisms studied, GCN of 16S rRNA, 18S rRNA per 100 ml of water sample, and standard bacteriological counts, the following physiochemical parameters were included: water temperature (°C), pH, dissolved oxygen (DO) (mg/L) and, atmospheric pressure (mmHg). In addition to this, also environmental, and biological data such as precipitation (cumulative precipitation over three days before each sampling event), (Environment and Natural Resources, 2022), daylight length (mins), (Time and Date, 2022), and total suspended solids (TSS(mg/L)) released from the closest water pollution control facility (City of Portage la Prairie, 2022) were analyzed.

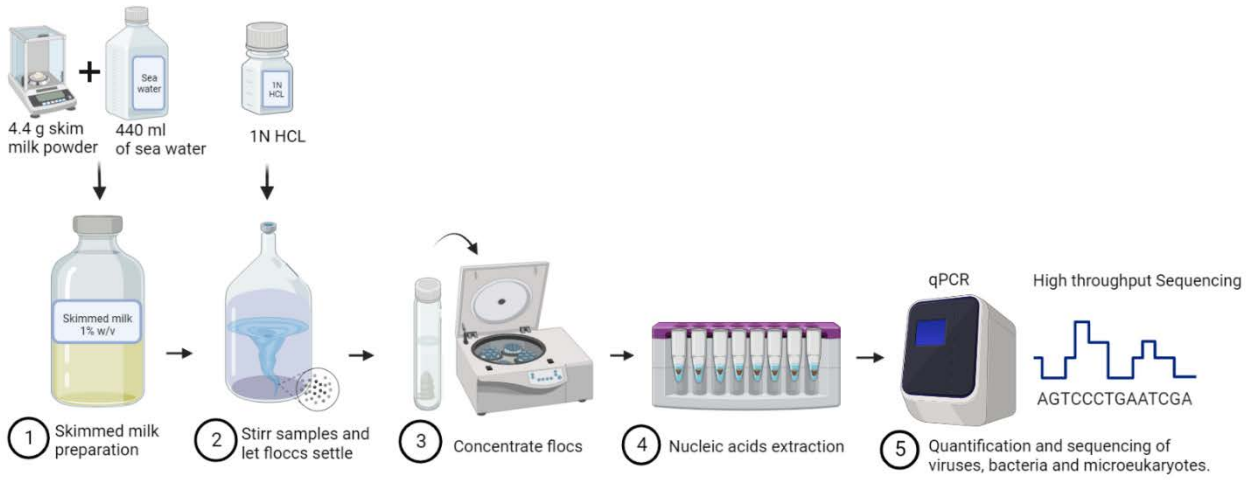

Figure 2.2 Graphic representation of the methodology used for the concentration of bacteriophages, bacteria and microeukaryotes for DNA identification. Created in Biorender

3.0 RESULTS

3.1 Diversity

Alpha and beta diversity indices were measured from the deep amplicon sequencing reads obtained after the sequencing of the nucleic acids of water samples coming from waterways adjacent to urban and agricultural impacted areas in rural Manitoba. The alpha diversity indices assessed were Shannon, and Simpson, or Gini- Simpson (1-D) to better understand the community diversity as well as the seasonal patterns of the microbial communities (Magurran, 2003). Shannon values indicate a proportional distribution of the number of species (evenness) in a sample and exhibit values greater than zero. The Simpson dominance index, which is one of the most robust diversity indexes available was also applied (as 1-D). Simpson values oscillate from 0 to 1 and the measure rises as the diversity increases.

3.1.1 Alpha diversity in viruses- g23

86.8% of the samples failed the deep amplicon sequencing for the g23 viral marker. The specific reasons for such a high failure rate could not be confirmed since the success of the reaction depends on multiple factors. The sequencing of the g23 amplicons was attempted several times with different amplicon cleanup methods. Despite these attempts, the amplicons were poorly sequenced. After an information exchange with representatives of the IMR (M Comeau, 2022; IMR, 2024) sequencing center, we hypothesize the mispriming of the Illumina sequencing primers. This occurs when primers bind to unintended regions of the genome other than the targeted region (Puskas & Bottka, 1995; Qiagen, 2024). This results in nonspecific products, incorrect sequencing, and/ or low-quality reads (Puskas & Bottka, 1995).

Alpha diversity analysis using the Shannon information index and the Simpson dominance index as was originally planned could not be performed in the remaining 13.1% of the samples because of the unevenness in the successful g23 reads regarding replicates and repeated sampling sites. While Chao-1 richness considers the ratio of the singletons and doubletons present in each sample, this nonparametric estimator assumes homogeneity and it would not be accurate to compare the richness of different locations and months with an uneven number of observations

(as explained above). As a comparative analysis of this data with the alpha diversity indexes increases the sources of error, we decided to report only the observed diversity manifested in each sample. The observed diversity was calculated from the uncultivated virus genomes' or 'UViGs' obtained from the successful g23-sequences. The g23 reads that passed sequencing quality control filters include 5 samples from Forested waterways, 12 from urban-located waterways (including 4 biological replicates), and 4 from Agricultural locations (including 1 Biological replicate) for more information about the samples, refer to table A.1. The samples had on average 3339.423 UViGs corresponding to myoviruses, which translates into a Shannon diversity index of 5. Among the successfully sequenced samples (21 out of 160), the sample from forested area No. 2 had the highest observed viral diversity (6504 UViGS) in the month of July. This same site had the lowest observed diversity (504 UViGS) in the month of November (Figure 3.1.1). Although this could suggest seasonal patterns in viral diversity in the sites aforementioned, the sequencing problems with the viral samples contributed to inconsistencies in the data and results to be reliable.

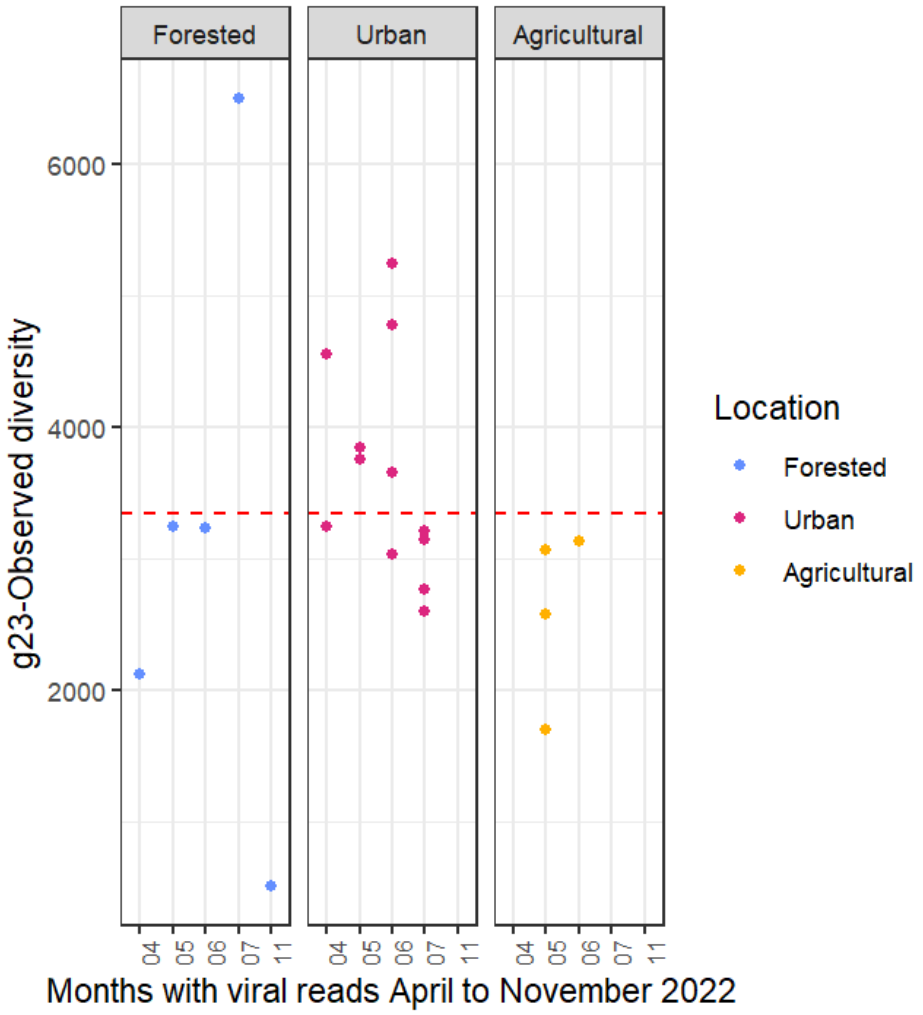

Figure 3.1.1: Observed diversity among the successfully sequenced g23 viral samples from the Forested, Urban, and Agricultural locations in the rural section of the Assiniboine River in Manitoba during the months of April(04), May (05), June (06), July (07), and November (11). Samples had on average 3339.423 UViGs observed (red dashed line). The sample corresponding to the Forested site No.2 had the highest observed diversity (6504 UViGS).

3.1.2 Beta diversity in viruses- g23

To identify the different patterns of viral composition in the forested, urban, and agricultural sampling locations as well as seasonal variations, NMDS ordination based on Bray Curtis dissimilarity was used. In NMDS, the closer the two points are, the more similar the variables (Taguchi & Oono, 2005; Genomics, 2024). From the analysis of the measured viral samples (21 out of 160), the viral composition present in the forested and urban locations appears to be more similar to each other. Within this context, the average distance to the median was 0.448, 0.476, and 0.095 for forested, urban, and agricultural locations, respectively (Figure 3.1.2). The multivariate homogeneity of group dispersions or betadisper (p-value $2.886e^{-07}$), as well as the Permutation test for homogeneity of multivariate dispersions or permutest (p-value 0.001), established that the difference in viral composition between locations was significant.

To assess viral composition variability across months, the betadisper, permutate test, as well as PERMANOVA, was applied. Analysis of the variance of the betadisper output as well as the permutate test indicated that seasonal variability was not statistically significant (p-value $\geq$ 0.434). On the other hand, PERMANOVA result evidenced that the differences in viral composition over time were significant (p-value= 0.001). It is important to mention that this method is susceptible to different within-group variations (dispersions) as compared to different mean values of the groups (Anderson, 2001; R-Documentation, 2024). Therefore, we consider betadisper and permutate results as values that better represent the samples presented in this analysis. When calculating the interaction of time (months) and locations, a p-value of 0.144 shows that the combined interaction of both variables did not contribute significantly to the observed variation While the results here obtained suggest that the viral composition did not change over time, it is important to highlight that all the samples collected and processed could not be included in the analysis because of the failed sequencing therefore, these results are indicative only of the samples included in the viral analysis (see table A.1).

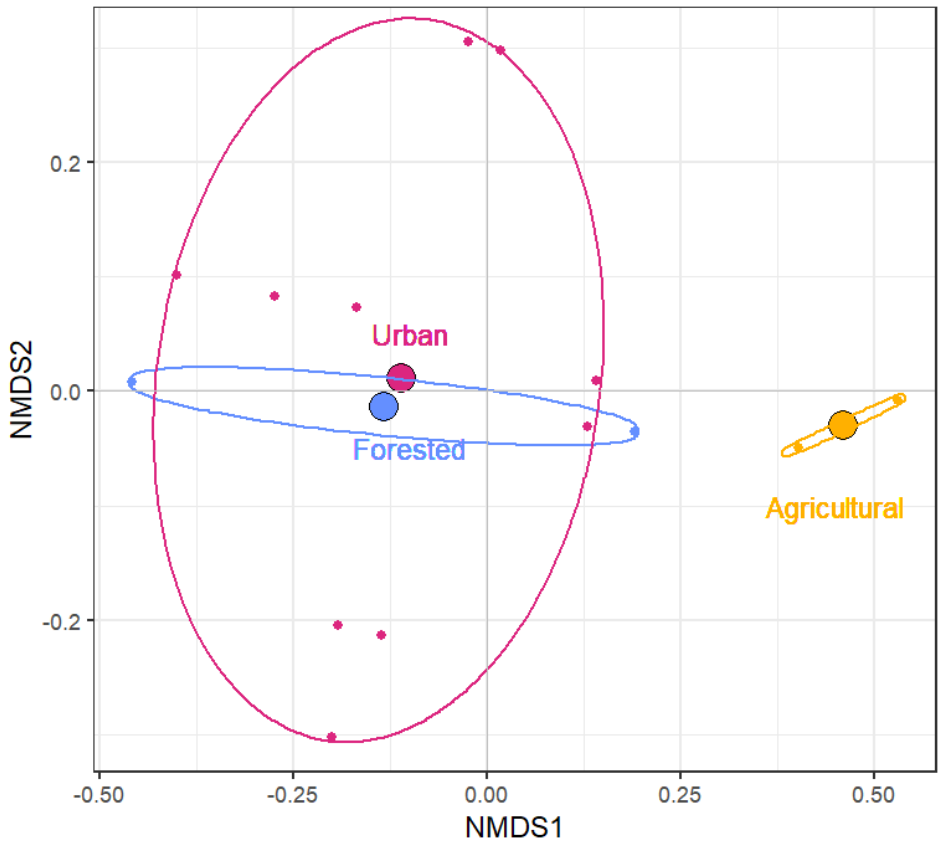

Figure 3.1.2: Non-metric multidimensional scaling (NMDS) ordination plot representing the dissimilarities between uncultivated viral genomes (UViGs) obtained from the successfully sequenced g23-viral reads from Forested, Urban and Agricultural waterways sampled. The viral composition of Forested (light blue dots) and Urban (dark pink dots) waterways, resulted more similar to each other than the UViGs found in Agricultural (yellow dots) waterways (p-value $2.886e^{-07}$). Larger colored circles represent the centroids or average position of each location.

3.1.3 Viral taxonomy profiling- g23

Although head-tailed viruses of the order *Caudovirales* have been documented to be the most numerous and widespread group of bacteriophages characterized to date (Cisneros-Martınez, Eguiarte, & Souza, 2023; Duda, 2008), the rate of taxonomic classification to the viral g23 samples was 1% in this study (Figure A.1). Unfortunately, these results are not uncommon, as there is an immense number of unclassified phage genomes in public databases (Korf et al., 2019; Uyaguari-Diaz et al., 2016). The taxonomic analysis performed in Galaxy Europe with kraken2 and the RefSeq viral database at the species level (version 2022- 06-07) revealed seven major species of viruses. Figure 3.1.3 shows the relative abundance of the viruses encountered in the locations considered in this study.

The viral groups include *Synechococcus* phages (56% adding up the different isolates), *Escherichia* phages (26%), *Emdodecavirus* genus (8%), Cyanophages (4%), *Orgyia leucostigma* nucleopolyhedrovirus (2%), *Slopekvirus* genus (2%), and *Bellamyvirus* genus (2%).

Among the most well-represented groups discovered are *Synechococcus* phages a group of bacteriophages with contractile tails from the formerly called *Myoviridae* family that infect the autotrophic cyanobacteria *Synechococcus* (Zubkov, Sleigh, Burkill, & Leakey, 2000). *Synechococcus* is one of the most abundant and prevalent picophytoplankton in marine and freshwater ecosystems and it is the dominant genus in polar and subpolar freshwater (Chénard, Chan, Vincent, & Suttle, 2015; Coello-Camba et al., 2020; Zhang, He, & Gin, 2021). Although the information regarding phages in freshwater ecosystems is limited (Zhang et al., 2021), *Synechococcus* phages have been found in the marine environment, and several studies have shown evidence for viral propagation in different environments (Dann et al., 2016; Wilhelm et al., 2006).

The second most relatively abundant phage found across all locations and time is the *Escherichia coli* phage (coliphage) G4507. Coliphages have been widely studied as proxies for drinking water quality and the presence of fecal bacteria and enteric viruses. Different *Escherichia coli* phages with contractile tails have been found in similar aquatic environments such as seawater, rivers, and

surface water (Michniewski et al., 2019; Korf et al., 2019).

On the other hand, there is limited information regarding the *Emdodecavirus* genus (8%) which was found in Forested (2.91%) and Urban (2.33%) locations only. One remarkable member of *Emdodecavirus* genus is the jumbo rhizobiophage that infects nodule bacteria from the genus *Sinorhizobium sp.* that are symbionts of leguminous grass and are present in terrestrial environments rather than aquatic environments (Kozlova et al., 2024).

Cyanophage S-RIM32 had a less abundant presence (4%) in the aquatic environments studied in this research. These cyanophages infect another major component of the picophytoplankton which are bacteria from the *Prochlorococcus* genus. *Prochlorococcus* is a sister clade of Synechococcus and even though *Prochlorococcus* has commonly been reported in marine environments, there is evidence of their presence in riverine areas like the ones included in this study (Zhang et al., 2021; Mitbavkar, Rajaneesh, Anil, & Sundar, 2012).

Among the least relatively abundant viruses, *Orgyia leucostigma* nucleopolyhedrovirus (2%) was found only in the urban locations considered in this study (0.19%). This organism is a bacaculovirus known to infect insects in the orders Lepidoptera, Diptera, and Hymenoptera (Thumbi et al., 2011). The aforementioned virus is from a different viral order *(Lefavirales)* than the ones we aimed to target in this research (Caudovirales), but both belong to the four supermodules of the dsDNA virosphere (Krupovic, Cvirkaite-Krupovic, Iranzo, Prangishvili, & Koonin, 2018). The sequencing of this baculovirus, which is not a head-tailed phage may be another consequence of mispriming of the Illumina sequencing primers. When testing the primer pair in the genome of *Orgyia leucostigma* nucleopolyhedrovirus T4super primer binding regions were found when allowing up to 8 mismatches.

Regarding the *Slopekvirus* genus (2%), members of this genus include T4-like phages infecting *Klebsiella* strains, which have been cataloged among the deadliest bacteria problematic to human health (Weber-Dabrowska et al., 2023; Townsend et al., 2021). Phages from the *Slopekvirus* genus have been considered potential candidates for phage therapy because of their lytic activity against *Klebsiella* strains (Townsend et al., 2021). Both environmental and pathogenic *Klebsiella*

have been found in freshwater (Podschun, Pietsch, Holler, & Ullmann, 2001) and rivers (Araújo et al., 2024). In this study, the *Slopekvirus* genus was found only in agricultural waterways (48.58% of the relative abundance in this location).

Finally, the genus *Bellamyvirus* (2%) includes the *Synechococcus* phage Bellamy which is another cyanophage that has been found in freshwater and hypersaline environments (Cisneros-Martinez, Eguiarte, & Souza, 2023). The hosts of this genus are also cyanobacteria (*Synechococcus* phages). These results suggest that viral communities tend to reflect the diversity patterns of their hosts and that some species of bacteria and bacteriophages reported not to be found in cold water, are indeed present. Seasonal patterns could not be demonstrated with the viral sequences obtained.

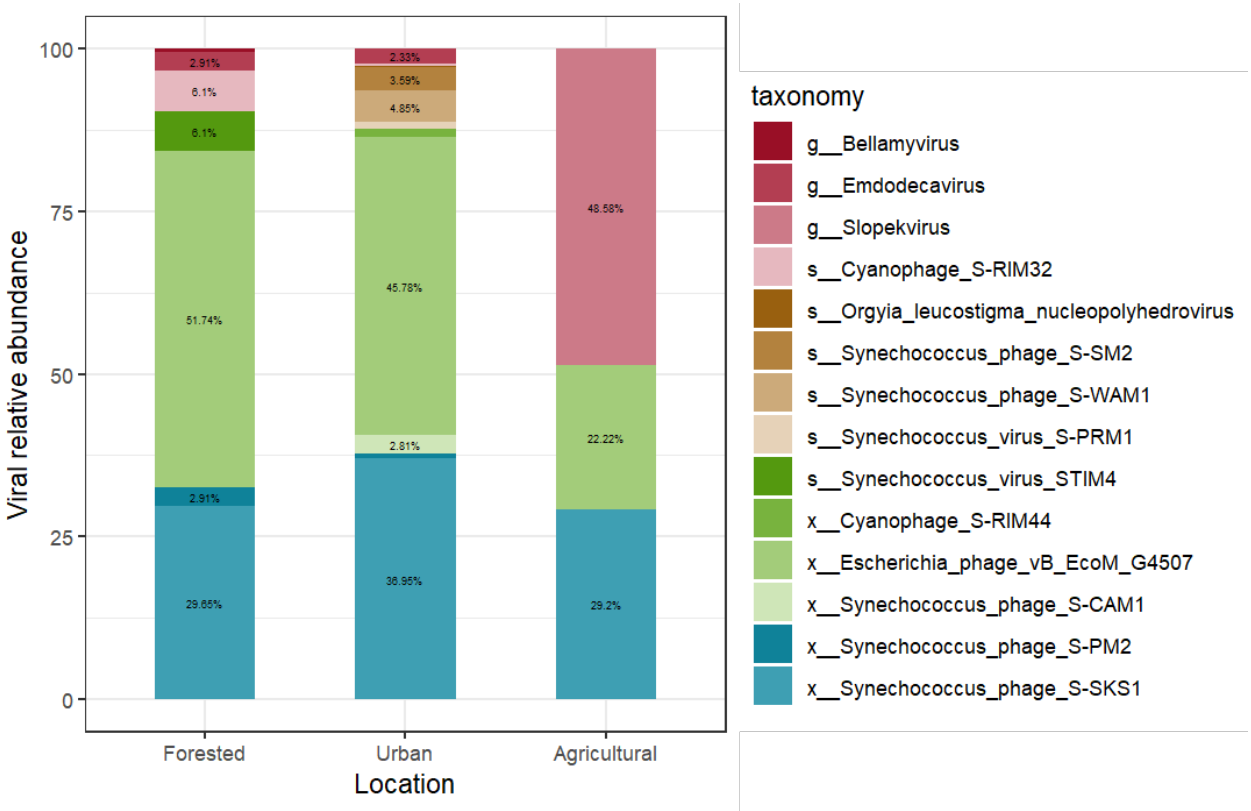

Figure 3.1.3: Relative abundance of viral communities found in waterways surrounding Forested, Urban, and Agricultural locations in rural Manitoba from April to July 2022. The viral composition was obtained using deep amplicon sequencing of the gene 23 (g23) viral marker specific for T4-like phages. The figure depicts the g23 reads that passed quality control including Forested: 5 samples, Urban: 12 samples, and Agricultural 4 samples.

3.1.4 Alpha diversity in Bacteria -16S rRNA

Shannon (information) and Simpson (dominance) diversity indexes were calculated for the inferred bacterial amplicon sequence variants (ASVs) obtained using the naive Bayesian classifier method available in DADA2. Bacterial ASV had on average a Shannon information index of 4.92 across the forested, urban, and agricultural locations and over time (April to November 2022). The Kruskal-Wallis test applied to the data confirmed that the differences in diversity across locations were significant (p-value 0.05). The Forested location had on average the highest Shannon diversity value (5.21±1.04) followed by the Urban location (5.02±1.21) and the Agricultural location (4.53 ± 1.14).

Regarding variations of diversity observed across months, the month of June had on average the highest Shannon diversity index (6.13±0.73) followed by July (5.41±0.89) and September (5.05±5.0.86). The months with the lowest Shannon diversity indexes were April (4.45±1.37), May (3.80±0.85), and August (4.66±0.48) (Figure 3.1.4, table 3.1.1). The Kruskal-Wallis test applied confirmed that the differences in diversity across months were significant (p-value 2.40 e^{-05}). Clear seasonal patterns could not be observed (diversity values increased and decreased randomly). The Dunn test used to compare Shannon diversity indexes across months confirmed that the months with the highest differences in diversity were June (highest Shannon diversity index) and May (lowest diversity index).

Simpson (1-D) values oscillate from 0 to 1 and the measure rises as the diversity increases.

The average Simpson dominance diversity index obtained across all the locations and months was 0.95. Similarly, to the Shannon diversity index, the forested location had the highest Simpson value on average (0.97 ± 0.03), followed by urban (0.944 ± 0.08) and Agricultural location (0.93 ± 0.10). The differences observed across the locations used in this study were statistically significant (p-value = 0.04). In parallel with Shannon values, the months with higher diversity dominance were June, July, and September (0.99±0.00, 0.97±0.03, 0.97±0.01, respectively) (Figure 3.1.5, table 3.1.2). On the other hand, the months with lower dominance were November, October, and May with values of 0.91±0.15, 0.92±0.11, and 0.92±0.05, respectively. The diversity dominance encountered across months is statistically significant (p-value= 6.84e^{-06}). The months with the highest differences according to the Dunn test were June (highest Simpson diversity index) with May, October, and November (lowest Simpson diversity index).

3.1.5 Beta diversity and Bacteria taxonomy profiling

The rank-based ordination NMDS using Bray-Curtis dissimilarity was applied to assign dimensions to the bacterial ASVs in order to identify seasonal and site-specific variations in the bacterial composition of the samples analyzed. The average distance to the median found in forested, urban, and agricultural locations was 0.60, 0.59, and 0.62 respectively. Among the bacterial families found across locations, the betadisper test and the p-value of 0.055 confirmed that there were not significant changes in the bacterial composition across locations (Figure 3.1.5). However, the bacterial composition changed significantly over time (p-value= $1.34e^{-05}$).

The inferred ASVs were used for taxonomy assignment using the open-source software package DADA2 and the Silva database at the family level (version 138.1). The analysis revealed eight major bacterial families including *Sporichthyaceae (20.27%), Pseudomonadaceae (19.93%), Enterobacteriaceae(14.8%), Burkholderiaceae (8.94%), Spirosomaceae (7.90%), Flavobacteriaceae (3.78%), Aeromonadaceae (3.78%), Micrococcaceae (3.43%),* and *Nostocaceae (3.09%)*. The remaining 14.77% could not be assigned. As previously mentioned, the sampling months considered in this research were April to November 2022. April, October, and November are considered "cold months" for the province of Manitoba (average water temperature 3.42°C±2.40 °C), while May and September are warm months (average water temperature 21.00°C± 2.11°C), (Cruaud et al., 2020). Figure 3.1.6 depicts the top 10 bacterial communities found across locations and over time in this study.

Sporichthyaceae, which has been reported to contain commonly freshwater microorganisms (Cruaud et al., 2020), was the most relatively abundant family of bacteria and comprised 25, 14, and 20 % of the bacterial ASV encountered in Forested, Urban, and Agricultural locations, respectively. *Sporichthyaceae* was found in higher abundance during June, July, August, and September (warm months).

Members of the family *Pseudomonadaceae* were found to be the second most relatively abundant across locations. Interestingly, ASV of these organisms appeared more abundant in the

agricultural location (38%) in comparison with forested (10%) and urban (10 %) locations. *Pseudomonadaceae* ASV were found in both warm and cold months. Members of the *Pseudomonadaceae* family include species that are pathogenic for humans and animals (Lobb et al., 2020). Within this context, this taxa has been connected with cosmopolitan or urban-impacted waterways. However, they have also been reported in rural lakes and agricultural watersheds (Numberger et al., 2022; Uyaguari-Diaz et al., 2016).

The family *Enterobacteriaceae* was the third most common taxa encountered, with higher relative abundance in agricultural locations (24%) than in forested (17%) and urban (4%) waterways. ASV from this gram-negative family were found to be more frequent in warm months such as May, July, and September. However, a high relative abundance was also observed during April and November. While this family is primarily known for its enteropathogenic members, it also includes non-pathogenic bacteria commonly found in the environment. In general, freshwater with a higher abundance of *Enterobacteriaceae* is often linked to fecal contamination (Aizenberg-Gershtein, Vaizel-Ohayon, & Halpern, 2012; Wisniewska, Niewolak, Korzeniewska, & Filipkowska, 2007).

The following observed bacterial ASVs belonged to the family *Spirosomaceae*. Members of this family are associated with freshwater health because of their ability to break down complex organic pollutants in water (Baek, Jang, Goh, & Choi, 2024). Interestingly, ASVs from the *Spirosomaceae* were found to be more abundant in the urban location (21%) compared to the other locations (1% in both, forested and agricultural).

Flavobacteriaceae was the following most relatively abundant family found in forested and urban locations with 6% and 5% of ASVs, respectively. Higher ASV counts were found during April and May for both locations. Members of this family have been isolated in freshwater and various Antarctic ecosystems (McCammon et al., 1998). Furthermore, studies of *Flavobacteriaceae* suggest that they play an important role in the degradation of organic matter in aquatic environments (Kacagan, Inan, Belduz, & Canakci, 2013).

Finally, other dominant bacterial families included *Micrococcaceae* in 2 locations (6% in forested, 4% in urban), and the cyanobacteria family *Nostocaceae* in one location (9% in urban). While information about *Micrococcaceae* in freshwater is limited, "Unclassified *Micrococcaceae*" has been reported in anthropogenically impacted waterways (Shao et al., 2021) and freshwater bodies with salinity intrusion (Li et al., 2022; Ayayee, Custer, Tronstad, & van Diepen, 2024). Regarding *Nostocaceae*, these microorganisms have been reported to be among the most antique prokaryotic groups found in marine and freshwater environments. They are known for their capacity to form blooms and for being key players in biogeochemical cycles (Duval et al., 2018).

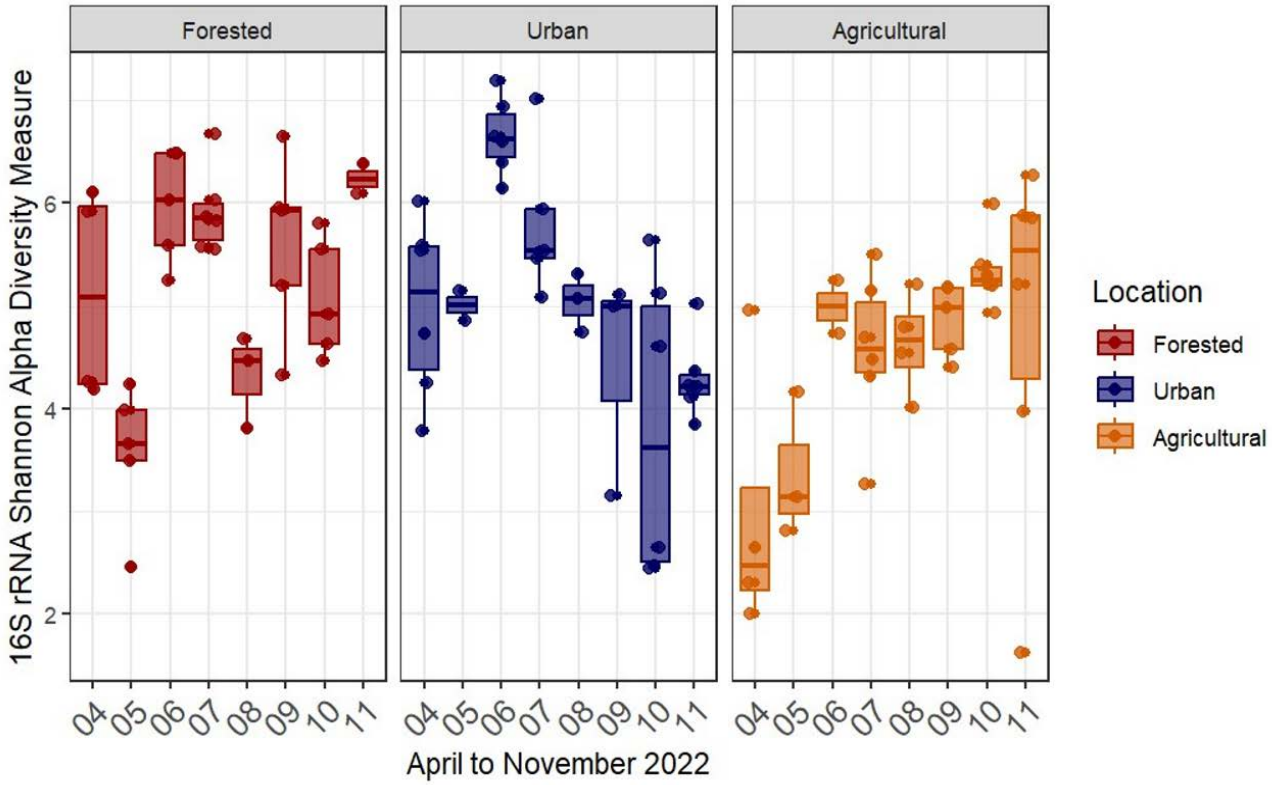

Figure 3.1.4 Bacteria Alpha diversity measure based on Shannon information index obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. June (06) had on average the highest Shannon diversity index (6.13±0.73), while May (05) had the lowest (3.80±0.85). The differences in diversity were significant across months (p-value 2.40 e-05) and locations sampled (p-value 0.050). Waterways near Forested locations depicted the highest diversity value (5.21±1.04) for this measure.

Month	$\bar{x}$ Shannon	SD
April	4.45	1.37
May	3.80	0.85
June	6.13	0.73
July	5.41	0.89
August	4.66	0.48
September	5.05	0.86
October	4.72	1.13
November	4.79	1.30

Table 3.1.1 Average of Shannon Alpha diversity values of bacteria communities obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. The averages are presented per sampling month.

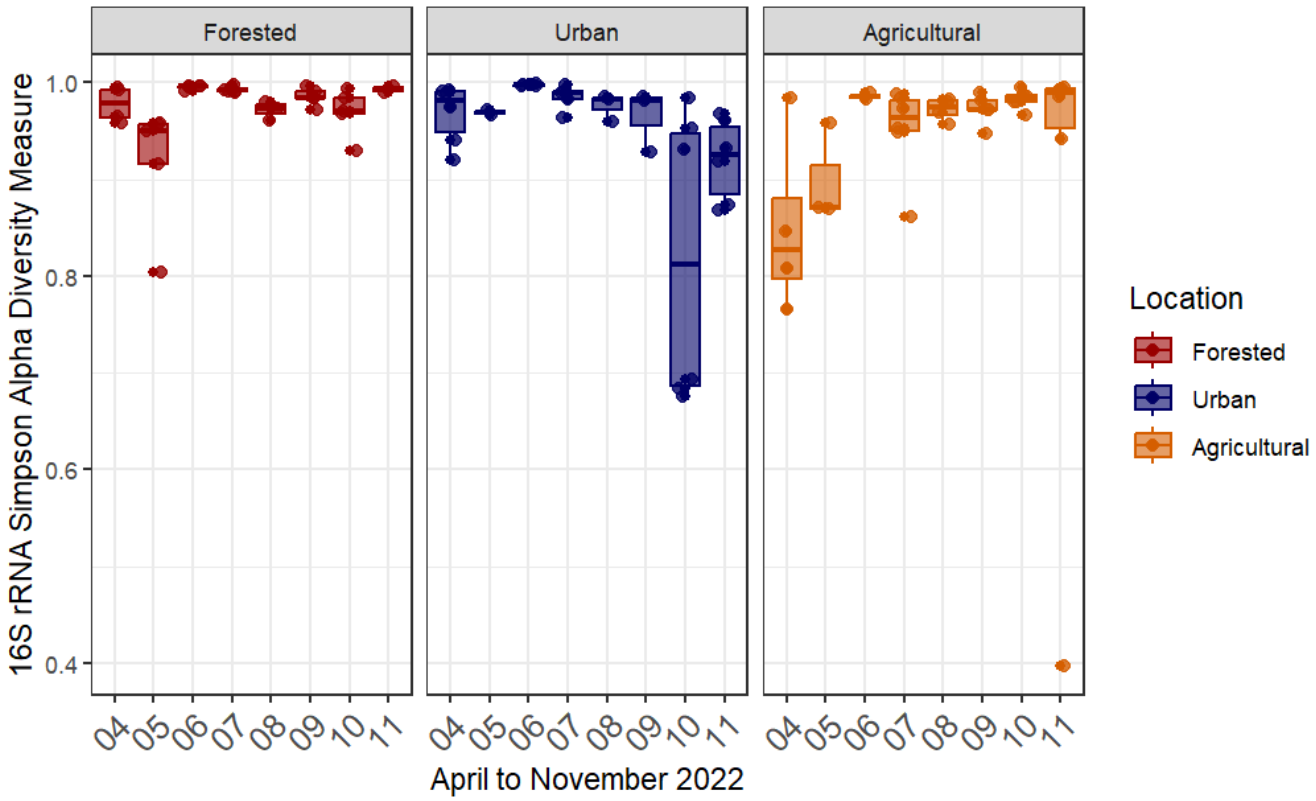

Figure 3.1.5 Bacteria Alpha diversity measure based on Simpson (1-D) dominance index obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. June (06) had on average the highest Simpson index (0.99±0.00), while November (11) had the lowest (0.91±0.15). The differences in diversity were significant across months (p-value $6.84e^{-06}$) and locations sampled (p-value= 0.04). Waterways near Forested locations depicted the highest Simpson diversity value (0.97 ±0.03).

Month	$\overline{x}$ Simspon (1-D)	SD
April	0.937	0.07
May	0.922	0.05
June	0.995	0.00
July	0.976	0.03
August	0.973	0.01
September	0.976	0.01
October	0.921	0.11
November	0.915	0.15

Table 3.1.2 Average of Simpson (1-D) Alpha diversity values of bacteria communities obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. The averages are presented per sampling month. To differentiate the diversity obtained between June, July, August, and September 3 decimals were considered.

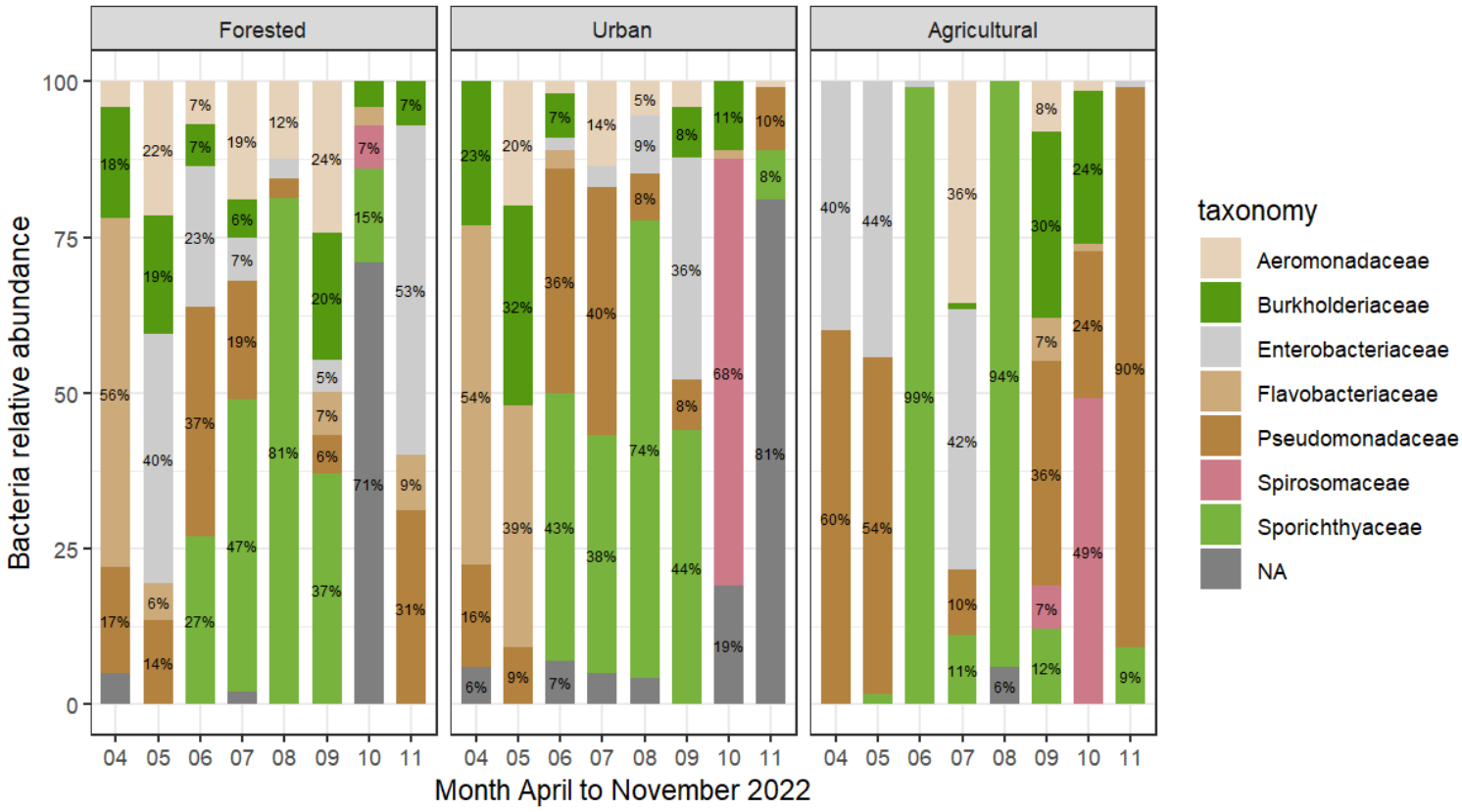

Figure 3.1.6 Bacterial community composition found in Forested, Urban, and Agricultural waterways in rural Manitoba from April to November 2022. The observational trends depict the bacterial families found per sampling month in each location.

3.1.6 Alpha diversity in microeukaryotes -18S rRNA

Only 26.87% of the samples could be successfully sequenced for the V4 region of the 18S rRNA gene. Although higher sequencing rates for these genes were expected, DNA sequence fragments derived from environmental samples are a well-known issue in deep amplicon sequencing analyses as further explained in section 4.5 (Tanabe et al., 2016, IMR, 2024).
Shannon and Simpson diversity indexes were calculated from the 43 samples that passed the sequencing which include 11 samples from forested, 15 from urban, and 17 from agricultural waterways (Table A.2). Alpha diversity analysis revealed that microbial Eukaryotes ASV had on average a Shannon diversity index of 3.57 across all locations and over the months included in the study (May, June, August, September, October, and November 2022). Similarly to bacterial results, the forested location had on average the highest diversity (3.85±0.30), followed by urban (3.64±0.46) and agricultural waterways (3.34±0.72). Differences in diversity across locations were not significant (p-value 0.18). Regarding diversity variations across months, August had on average the highest Shannon diversity index (4.18±0.26), followed by June (4.03±0.00) and September (3.72±0.48). On the other hand, months with the lowest Shannon diversity indices were May (3.17±0.00), October (3.42±0.75), and November (3.50±0.39) (Figure 3.1.7, Table 3.1.3). Kruskal-Wallis test revealed that the differences in diversity found across months were not significant (p-value 0.14). To assess dominance, Simpson (1-D) index was calculated from 18S rRNA ASVs. Across all months and locations, eukaryotes had on average a relatively high Simpson index (0.90). The waterways with the highest dominance were forested (0.93 ±0.02) followed by urban (0.92±0.04) and agricultural (0.88±0.08) locations. Differences in dominance across locations were not significant (p-value 0.37). Regarding months, August, June, September, and November were on average the time points with higher species dominance (0.96±0.00, 0.95±0.00, 0.91±0.04, 0.91±0.03, respectively). Whereas, October and May had on average the lowest dominance values (0.88±0.08, and 0.89±0.00, respectively) (Figure 3.1.8, table 3.1.4). Dominance differences across time points were not significant (p-value 0.13). These results suggest that few microbial eukaryote families dominate the waterways sampled.

3.1.7 Beta diversity and microbial eukaryotes taxonomy profiling

To identify seasonal and site-specific variation in the composition of eukaryote ASVs from the eukaryotic reads that passed the sequencing, NMDS using Bray-Curtis dissimilarity was employed. The average distance to the median in the forested, urban, and agricultural waterways was 0.60, 0.51, and 0.62 respectively. Based on the betadisper test applied, the composition of microeukaryotes changed significantly across locations (p-value $5e^{-03}$) (Figure 3.1.9) and over time (p-value $4e^{-04}$).

The inferred ASVs exposed 8 major eukaryotic groups containing classes, orders, and families including *Mediophyceae* (24.02%), *Cryptomonadales* (14.98%), *Choreotrichia* (14.15%), *Sphaeropleales* (7.70%), *Kappamycetaceae* (5.21%), *Thoracosphaeraceae* (3.50%), *Oligotrichia* (3.07%) and, *Chlamydomonadales* (1.87%). The remaining 25.45% could not be classified. Figure 3.2.0 depicts the top 10 eukaryotic communities found across locations and over time in this study.

Mediophyceae is a class of microbial eukaryotes that enclose the most diverse type of algae called diatoms (Wang, Chen, Wang, Liu, & Chen, 2021). Diatoms are known to play an important role in the aquatic food web and they have been recognized as bioindicators of environmental changes (Luethje & Snyder, 2021; Stoermer & Smol, 1999). ASVs from this class were found more abundantly in urban (40.59%) than in agricultural (18.7%) and forested (12.6%) waterways. From the 43 samples that were analyzed, *Mediophyceae* was found in higher abundance during warm months (May, September, early October).

Cryptomonadales order was the second most relatively abundant group found across locations, with higher ASVs assignations in forested locations than agricultural and urban waterways (20, 19.80, and 5.04%, respectively). Members of this order have been reported to be bacterivorous and abundant in the ocean and most freshwater habitats (Shalchian-Tabrizi et al., 2008; Cruaud et al., 2019). ASVs assigned to this order were found more during warm August and September compared November.

The subclass *Choreotrichia* was reported as the third most relatively abundant micro-eukaryotic community in the water samples studied. ASVs from these taxa were found more abundantly in forested, agricultural, and urban locations (32%, 6.08%, 4.27%, respectively). This group of ciliated protists has been found in freshwater plankton (Santoferrara, Alder, & McManus, 2017). Some studies have highlighted *Choreotrichia* members play significant trophic roles as bacteria and algae consumers (Santoferrara, Alder, & McManus, 2017; Rakshit, Ganesh, & Sarkar,2016). However, a deeper understanding of these ciliates' ecological importance and diversity information is limited. *Choreotrichia* was found both during warm and cold months, with slightly higher numbers found at the start of winter in Manitoba (November).

The fourth most common eukaryotic taxa found was the *Sphaeropleales* order. These green microalgae belong to the class *Chlorophyceae* (Lemieux, Vincent, Labarre, Otis, & Turmel, 2015) and have been reported previously as a component of the Canadian algal flora in Northwest territories streams and lakes (Sheath & Steinman, 1982; Lemieux, Vincent, Labarre, Otis, & Turmel, 2015).

Sphaeropleales members were found with higher abundance in Forested (11.9%) waterways followed by urban (6.88%) and Agricultural (4.26%) sampling points. ASVs from this order appeared only during warm months (August, September, and early October).

The fifth dominant micro-eukaryotic group was the fungi family *Kappamycetaceae*. ASVs from this family appeared in urban (14.2%) and forested waterways (1.4%) only. Members of this family belong to the order *Rhizophydiales* (Calabon et al., 2023) which has been recognized as a group of freshwater fungi. ASVs assigned to this fungi family were found during warm and cold months (early October and November).

Thoracosphaeraceae family was the 6th most abundant taxa and was present only in Agricultural waterways (10.5%) predominantly during November (even though it was also found in October). Protists from this family have also been found to contribute to the eukaryome of the Pavilion Lake in British Columbia (Bonacolta, Visscher, Del Campo, & White III, 2024;McCarthy et al., 2018).

Finally, the 7th and 8th micro eukaryotic dominant taxa were assigned to the subclass *Oligotrichia* and the family *Chlamydomonadales*. *Oligotrichia* belongs to the order *Choreotrichida* (related to the *Choreotrichia* subclass mentioned earlier) (Agatha & Strder-Kypke, 2007) and they were only described in forested (4.30%) and agricultural (4.89%) waterways in September, October, and November. Protozoa from this subclass have been reported to be predominant in European lakes covered in ice from autumn to spring / early summer (Macek, Callieri, & Vazquez, 2006).

Lastly, ASVs allocated to the order *Chlamydomonadales* were registered in agricultural (4.2%) and urban (1.42%) sampling points only during October and November, respectively. Green algae from this order have been widely studied due to their psychrophilic characteristics (ability to survive and reproduce in temperatures <15°C). Within this context, it is postulated that the orders *Chlamydomonadales* and *Sphaeropleales* (mentioned earlier*)* are sister clades, based on recent chloroplast phylogenomic studies (Bonacolta et al., 2024).

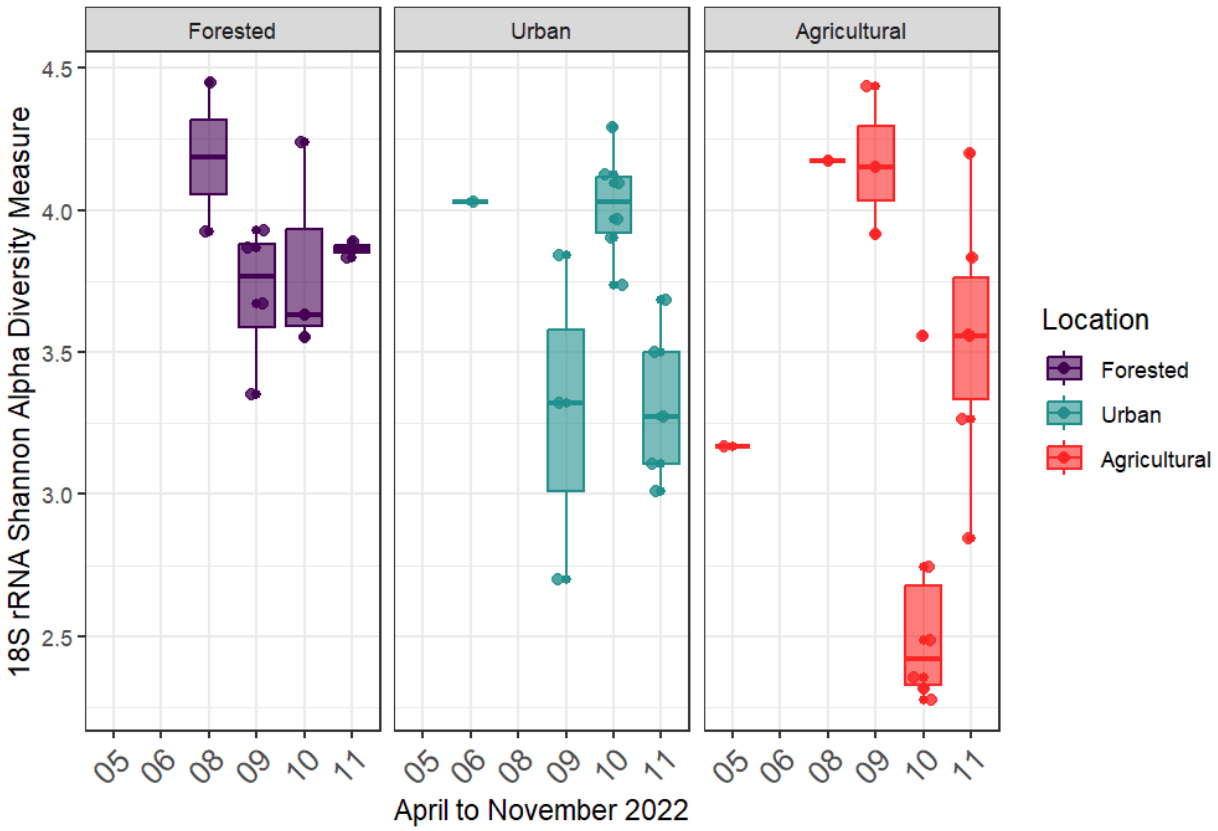

Figure 3.1.7 Eukaryotic Alpha diversity measure based on Shannon information index obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba during May (05), June (06), August (08), September (09), October (10) and November (11). August had on average the highest Shannon diversity index (4.18±0.26), while May had the lowest (3.17±0.00). Forested locations depicted higher diversity than others (3.85±0.30). However, the differences found across months and locations were not significant (p-value 0.14, 0.18, respectively)

Month	$\bar{x}$ Shannon	SD
May	3.17	0.00
June	4.03	0.00
August	4.18	0.26
September	3.72	0.48
October	3.42	0.75
November	3.50	0.39

Table 3.1.3 Average of Shannon Alpha diversity values of eukaryotic communities obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. The averages are presented per sampling month.

Month	$\bar{x}$ Simpson	SD
May	0.89	0.00
June	0.95	0.00
August	0.96	0.00
September	0.91	0.045
October	0.88	0.08
November	0.91	0.03

Table 3.1.4 Average of Simpson (1-D) dominance values of eukaryotic communities obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba. The averages are presented per sampling month.

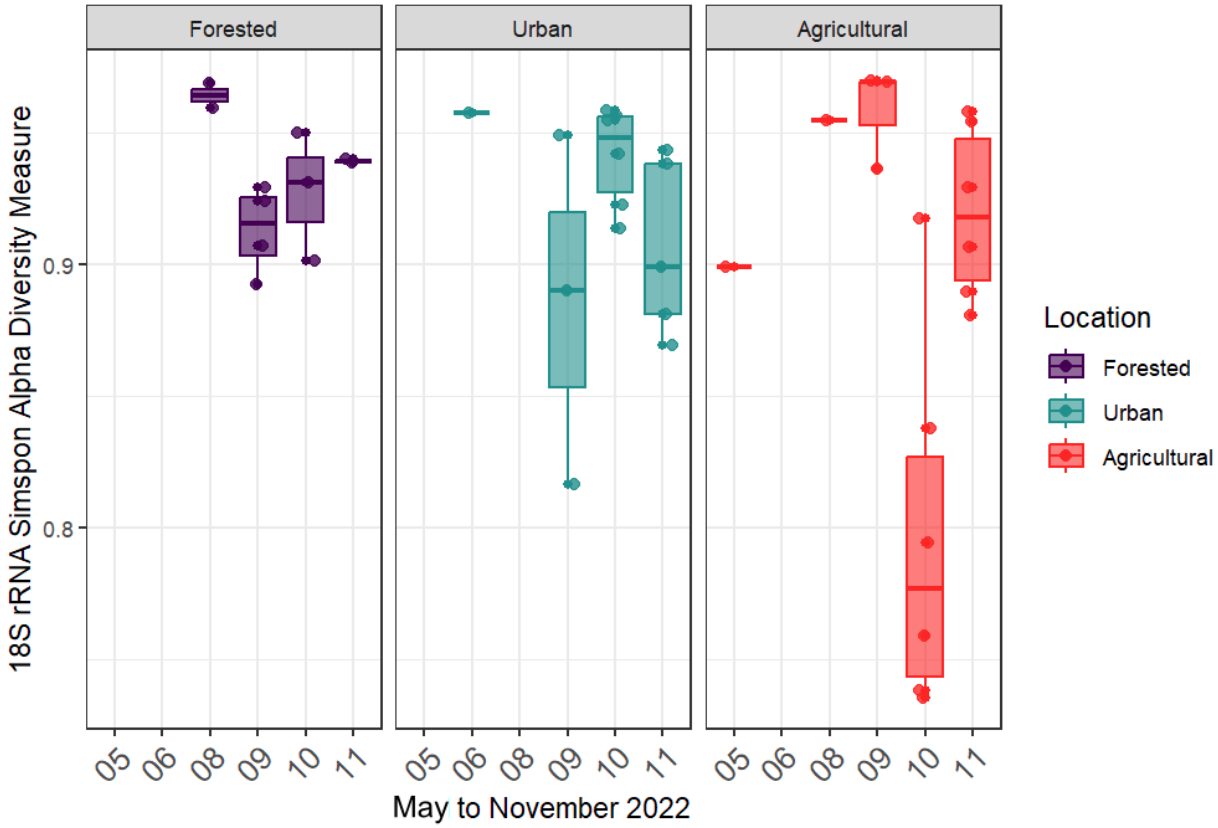

Figure 3.1.8 Eukaryotic Alpha diversity measure based on Simpson dominance index obtained from surface water collected in the rural section of the Assiniboine River in rural Manitoba during May (05), June (06), August (08), September (09), October (10) and November (11). August had the highest dominance index (40.96±0.00), while October had the lowest (0.88±0.08). Forested locations depicted higher dominance than others (0.93±0.02). However, the differences found across months and locations were not significant (p-value 0.13, 0.37, respectively).

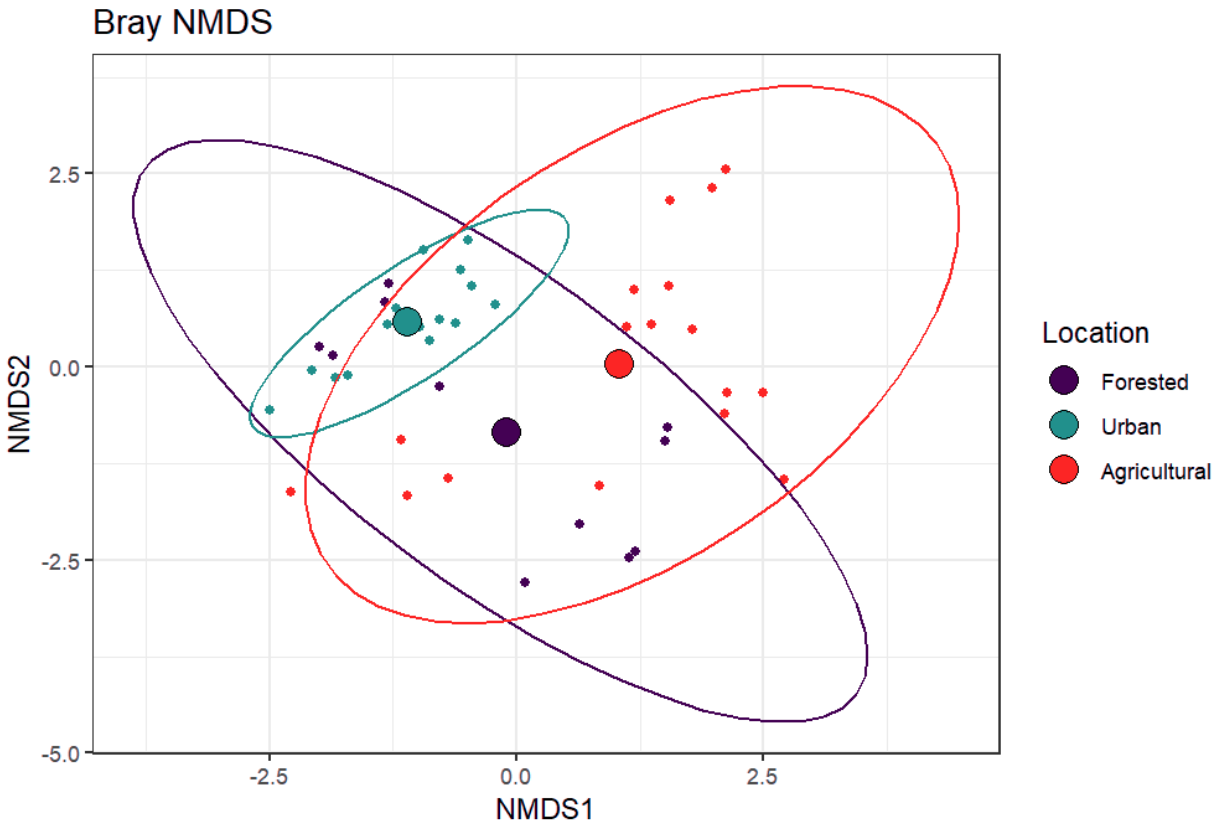

Figure 3.1.9 Non-metric multidimensional scaling (NMDS) ordination plot representing the dissimilarities between Amplicon sequence variants (ASVs) obtained from the successfully sequenced 18S rRNA-V4 eukaryotic reads from Forested, Urban, and Agricultural waterways sampled. The eukaryotic composition changed significantly across locations (p-value 5e-03). Bigger colored circles represent the centroids or average distance of each location.

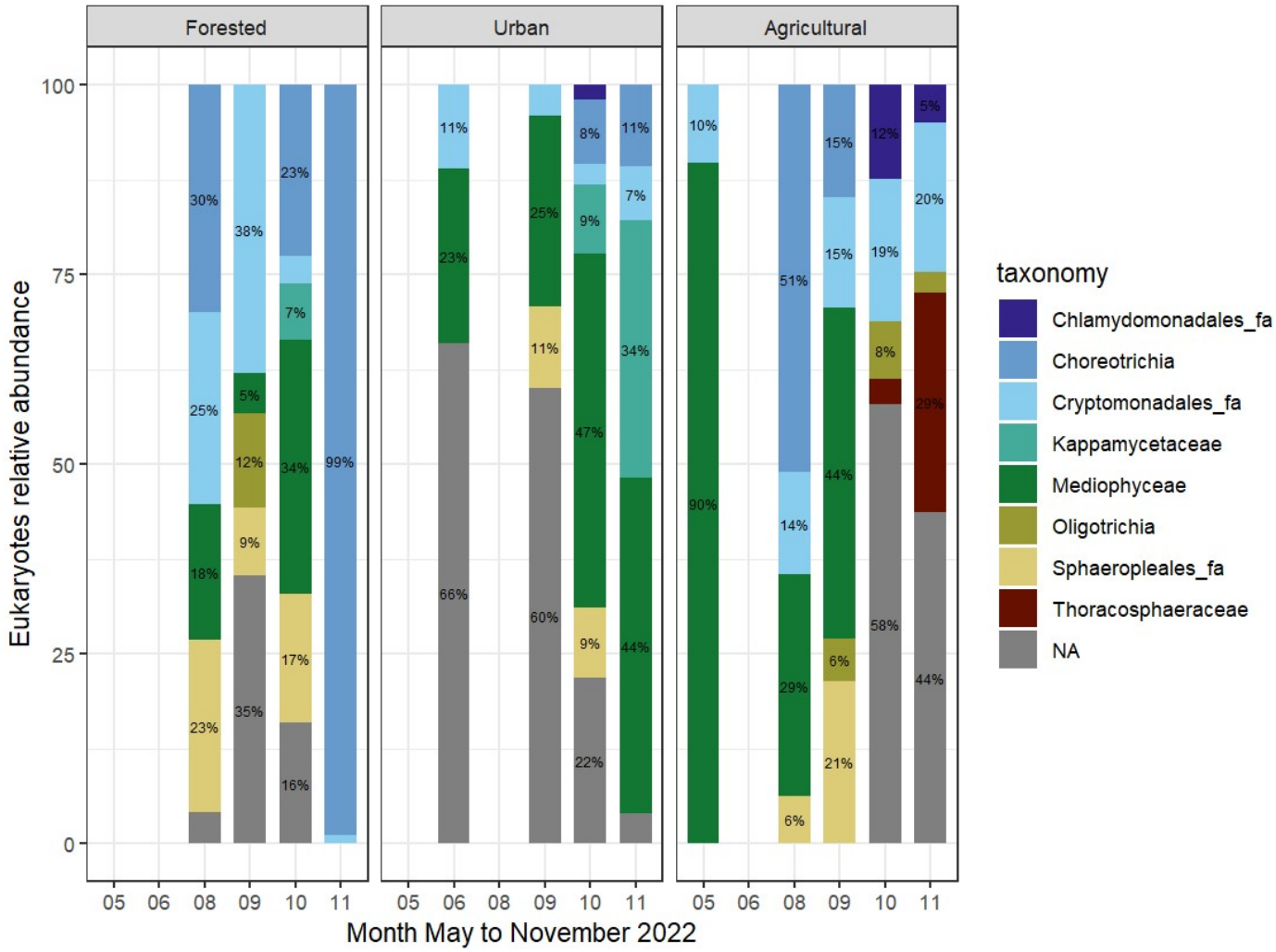

Figure 3.2.0 Microeukaryotic community composition found in Forested, Urban, and Agricultural waterways in rural Manitoba from May to November 2022. The observational trends depict the micro eukaryotic groups found per sampling month in each location. Months without taxonomy assignment (“05”, “06” in Forested; “05”,”08” in Urban; “06” in Agricultural) reveal sampling sequencing failure for those specific locations. NA refers to the sequences that could not be classified.

3.2 Quantitative PCR analysis

3.2.1 Bacteria and Eukaryotes Quantitative PCR Analysis

16S rRNA and 18S rRNA gene markers were used to assess bacterial and eukaryotic cell counts using qPCR reactions in triplicate. Raw values obtained from the Applied Biosystems QuantStudio 5 real-time PCR (Applied Biosystems, Foster City, CA) were normalized per milliliter and ng of DNA to obtain GCN per 100 ml of water, and log10 transformed for further analysis. In addition, bacteria GCN were normalized by the average factor of 4.3 to account for the multicopy nature of the 16S rRNA gene in these microorganisms. 16S rRNA and 18S rRNA qPCR negative controls had amplification before the CT set. Although undesirable, this event is not uncommon (Jankowski et al., 2022; Olomu 2020) and can be associated with cross-contamination during DNA extraction or because of the presence of bacterial and eukaryotic traces in the nucleic acid purification kit and/or other laboratory reagents (Salter et al., 2014). Bacterial samples had on average 3.40 $X10^7$ copies of the 16S rRNA gene, equaling a log10 value of 6.12. The urban location had the highest number of bacterial GCN ($4.92X10^7$ copies; SD $\pm 8.26X10^7$) followed by the allegedly less contaminated forested waterways ($3.70X10^7$ copies; SD$\pm 9.99X10^7$), and the agricultural waterways ($2.64X10^7$copies; SD $\pm 4.38X10^7$) which had the lowest GCN bacterial values. The linear mixed model applied to the normalized and log10 transformed bacterial GCN per 100ml sample and the analysis of variance of aligned rank transform data (ART ANOVA) demonstrated that there are statistical differences among locations. Waterways close to different anthropogenic activities considered in this research have a significant effect on GCN of bacteria (p-value < 0.05). However, based on the partial eta squared value obtained, location can explain only 0.01 (1%) of the variances exhibited, suggesting that other variables could contribute with more significance to the GCN variation. Likewise, the effect of the different months considered in this study (from April to November 2022) in bacteria GCN was analyzed. The months of April and May had on average the highest gene copy number counts of all the sampling months ($1.21X10^8$; SD $\pm 1.64X10^8$ and $5.25X10^7$; SD$\pm 6.09X10^7$ copies of bacteria respectively). The months with the lowest average bacteria GCN counts were November ($6.26X10^6$; SD$\pm$ $6.96X10^6$), June ($7.38X10^6$; SD$\pm$ $1.30X10^7$), and July ($1.15X10^7$; SD$\pm 1.43X10^7$). Figure 3.2.1 visualizes the log10 GCN of bacteria in each

location and month analyzed in this study. Table 3.2.1 contains the average bacteria GCN obtained monthly in this research. While these results are inconsistent to be able to demonstrate seasonal patterns, the analysis of variance of aligned rank transformed data applied to the linear mixed model to assess the effects of time in bacterial GCN, established that months have a significant influence on the bacterial GCN obtained (p-value $\leq 2.22e^{-16}$). The Tukey post-hoc test confirmed that the months with the highest differences in GCN were April, May (highest GCN counts), and November, June, and July (lowest GCN counts). The part.eta.sq value demonstrated that sampling months (April to November 2022) account for 0.26 (26%) of the variance in bacterial GCN. The Efron R squared applied to test the good fit of the models created confirmed that time (measured as months) has more influence in bacterial GCN than the locations sampled.

Quantitative analyses for the 18SrRNA gene demonstrated that eukaryotes had on average $9.47X10^3$ gene copies which equals a log10 value of 3.90 (Figure 3.2.2). Bacterial GCN surpassed eukaryotes by 4 orders of magnitude (Figure 3.2.2). In contrast to bacterial cells, the GCN of eukaryotes was higher on average in agricultural waterways ($1.85X10^4$; SD± $6.27X10^4$), followed by forested waterways ($7.00X10^3$; SD±$1.20X10^4$), and lastly in Urban waterways ($5.75X10^3$: SD ±$8.48X10^3$). The linear mixed model and the analysis of variance of aligned rank transform data (ART ANOVA) applied to the eukaryotic GCN per 100ml of sample demonstrated no significant differences among the values of GCN obtained in different locations (p-value = 0.26). Based on the partial eta squared value obtained, location only accounts for 0.006 (0.6%) of the variances exhibited in eukaryotic GCN. On the other hand, when evaluating the effect of the different sampling months on eukaryotic GCN, a p-value of $4.25e^{-05}$ demonstrates that there are statistical differences across months. May and August had on average the highest GCN values calculated ($3.12X10^4$; SD ±$8.76X10^4$, and $9.29x10^3$; SD ±$1.05X10^4$ respectively). Whereas September ($3.08X10^3$; SD ±$4.88X10^3$), November ($4.86X10^3$; SD ±$9.18X10^3$), and April ($5.49X10^3$; SD ±$9.68X10^3$) had the lowest average of eukaryotic GCN found. GCN of eukaryotes per location and month can be visualized in Figure 3.2.3. Table 3.2.2 displays the monthly average values of GCN of eukaryotes obtained in this investigation. Interestingly, May is one of the months with the highest average GCN counts for both bacteria and eukaryotes. The Tukey post-hoc test confirmed that the months with the highest differences observed were depicted during May (highest eukaryotic GCN count), April, July, August and

September. In eukaryotes, the sampling events performed in different months of this research, account only for 0.07 of the GCN variance. While the effect of months over eukaryotic GCN is stronger than the effect of the locations here considered (based on Efron R squared and p-values), these results suggest that other variables have a higher impact on the variance of eukaryotic GCN and that bacteria are more likely to change in quantity over time than eukaryotes.

Month	$\bar{x}$ GCN copies/100 mL	SD	Min	Max
April	1.21E+08	1.65E+08	1.51E+05	6.16E+08
May	5.26E+07	6.09E+07	6.29E+05	2.77E+08
June	7.39E+06	1.31E+07	3.29E+00	4.63E+07
July	1.16E+07	1.43E+07	5.57E+01	4.77E+07
August	2.18E+07	2.88E+07	4.32E+03	1.11E+08
September	3.47E+07	6.23E+07	4.73E+04	2.50E+08
October	1.62E+07	1.86E+07	3.08E+05	7.37E+07
November	6.27E+06	6.97E+06	8.29E+03	2.55E+07

Table 3.2.1: Gene copy number (GCN) average values of the bacterial communities found in waterways near Forested, Urban, and Agricultural locations in rural Manitoba from April to November 2022.

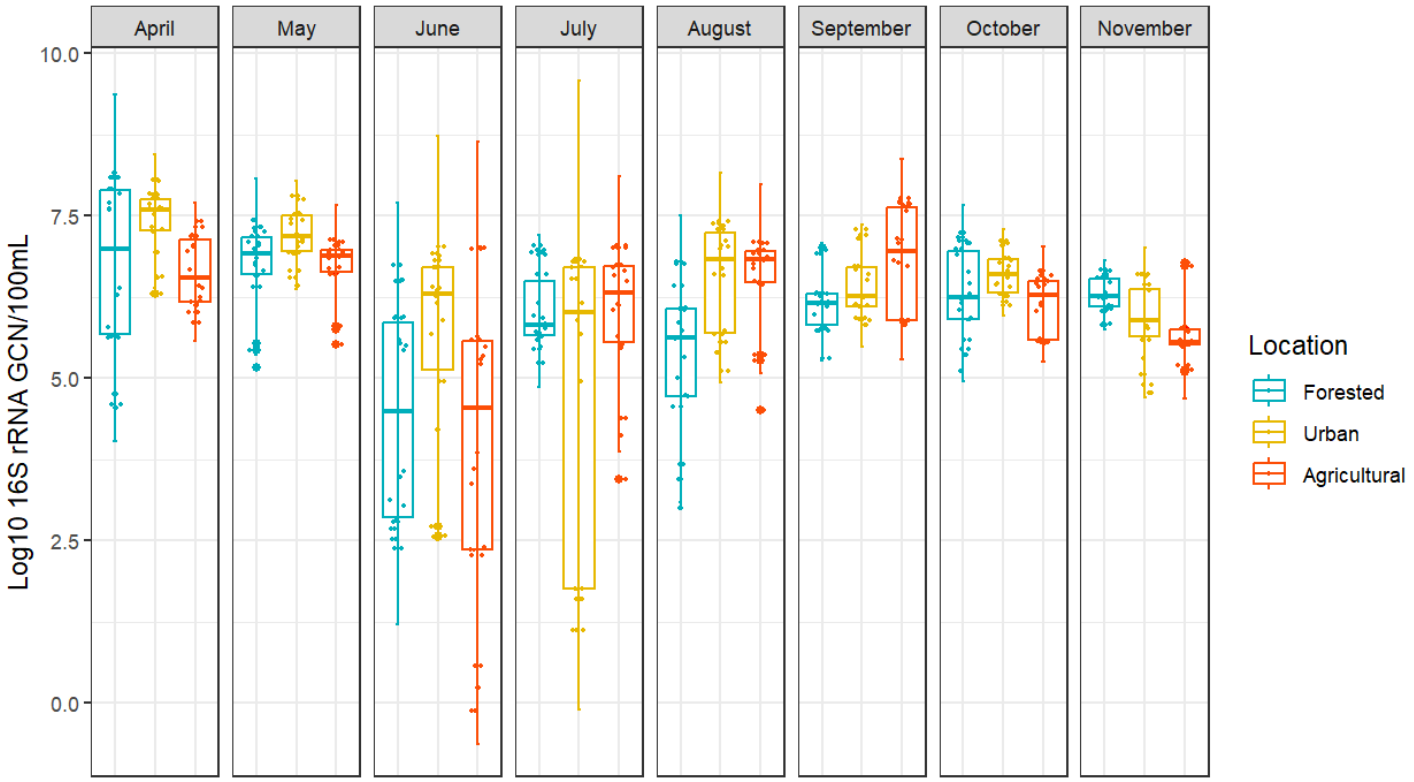

Figure 3.2.1: Quantification of 16S rRNA gene represented as log10 gene copy number (GCN) per volume (100 mL) of surface water collected from the Assiniboine River located in rural Manitoba from April to November 2022. Biological duplicates are included. Statistical differences were found between the GCN values obtained during the different months (p-value < $2.22e-16$). April was on average the month with the highest GCN number (6.91). Significant differences were also found across locations (p-value <0.05). Waterways near Urban locations had on average the highest GCN value found (6.30).

Month	$\bar{x}$ GCN copies/ 100 mL	SD	Min	Max
April	5.50E+03	9.69E+03	8.33E+00	3.50E+04
May	3.13E+04	8.77E+04	2.32E+00	4.33E+05
June	6.67E+03	7.91E+03	5.77E+00	2.67E+04
July	9.27E+03	4.05E+04	7.94E-01	3.11E+05
August	9.30E+03	1.05E+04	6.02E+00	3.15E+04
September	3.09E+03	4.89E+03	1.88E+00	1.67E+04
October	5.70E+03	9.58E+03	1.41E-01	3.90E+04
November	4.86E+03	9.18E+03	7.63E+00	3.53E+04

Table 3.2.2: Gene copy number (GCN) average values of microbial eukaryotes found in waterways near Forested, Urban, and Agricultural locations in rural Manitoba from April to November 2022.

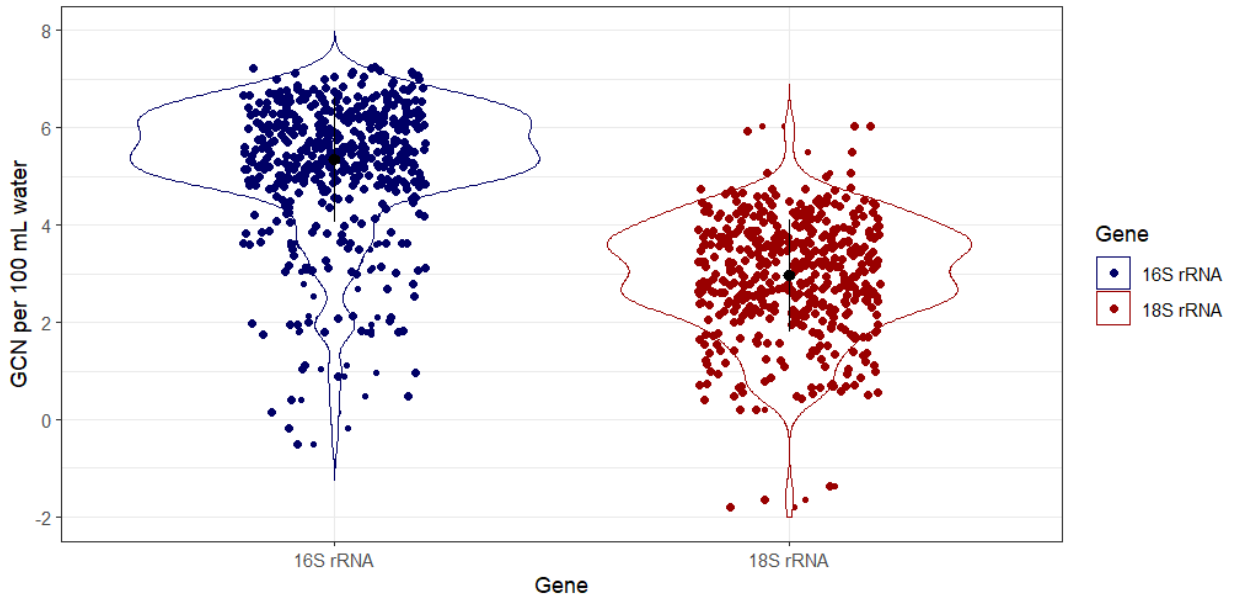

Figure 3.2.2: Violin plot representing the distribution of the gene copy number (GCN) per 100 mL of water sample obtained from the quantification of the 16SrRNA and 18S rRNA genes. The 16S rRNA gene had on average 6.12 log10 copies/100 ml, whereas the 18S rRNA gene had 3.90 log10 copies/100 ml. In this research conducted from April to November 2022, bacteria quantification through qPCR surpassed eukaryotes by 4 orders of magnitude.

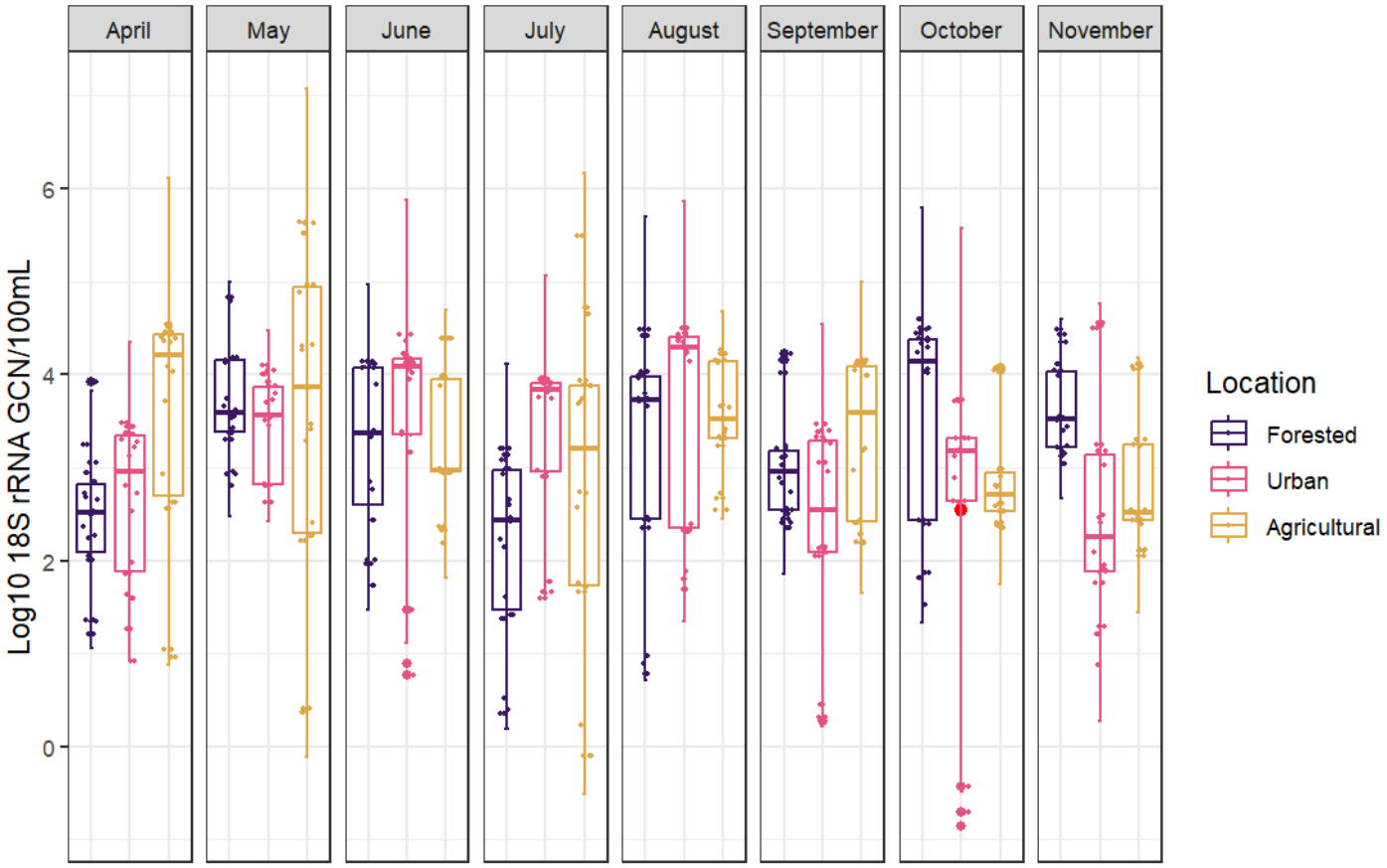

Figure 3.2.3: Quantification of 18S rRNA gene represented as log10 gene copy number (GCN) per volume (100 mL) of surface water collected from the Assiniboine River located in rural Manitoba from April to November 2022. Biological duplicates are included. Statistical differences were found between the GCN values obtained during the different months (p-value < 4.25e − 05). May was on average the month with the highest GCN number (3.55) whereas the lowest GCN average was found in July (2.77). There were not significant differences across locations (p-value 0.266)

3.3 Comparative analysis of the microbial fractions with seasonal and spatiotemporal patterns

To assess the contribution of each of the variables on the variance observed, PCA was applied to the scaled data containing the inferred ASVs of bacteria and microeukaryotes, UVIGs (viruses), GCN of 16S rRNA, 18S rRNA per 100 ml of water sample, standard bacteriological counts of *E. coli,* and total coliforms, water temperature (°C), pH, DO (mg/L) , atmospheric pressure (mmHg), cumulative precipitation, daylight length (mins) and TSS (mg/L) released in the Assiniboine river. Figure 3.2.4 depicts the correlation biplot representing the projection of the observations and the intercorrelation of the quantitative variables.

The cumulative percentage of PC1 and PC2 explained approximately 26% of the variation observed. Overall, samples from cold and warm months were relatively distinct from each other as point clusters occupy different quadrants. This difference was not observed when grouping observations by location. PC1 (or dimension 1), which explained 15% of the observed variance, received its highest contribution from daylight length (15.89%) followed by water temperature (11.60%) and DO (9.90%). As demonstrated by the direction of the principal components in the biplots as well as the Spearman's rank correlation (figure 3.2.5), daylight length and water temperature are positively correlated (rho 0.47,p-value <0.05). On the other hand, daylight and DO, and temperature and DO are negatively correlated (rho -0.49, p-value $2.06e^{-08}$;-0.61, p-value $5.50e^{-14}$ respectively). PC2 (or dimension 2) explained 12.2% of the observed variance and the highest variance contributors were: Precipitation (13.69%), the relative abundance of the bacteria family *Flavobacteriaceae (*9.94%) as well as the relative abundance of the micro eukaryotic family *Cryptomonadales* (9.58%). Precipitation and *Cryptomonadales* were negatively correlated (rho -0.55, p-value $5.48e^{-07}$), whereas the bacteria family *Flavobacteriaceae* could not demonstrate any relation with precipitation or *Cryptomonadales* (rho 0.05, and -0.02). The variables that account for TSS and 18S rRNA GCN measurements are depicted closer to the biplot center which portrays their scarce contribution to the analysis since they are not well represented in this PCA. Interestingly, the bacterial families *Sporichthyaceae* (most relatively abundant) and *Aeromonadaceae* were found in the same quadrant with total coliforms and *E. coli* plate counts in sampling months with warm water temperatures indicating intercorrelation between these variables. Table 3.3.1 summarizes all the microbial fractions that correlated negatively or positively with *E. coli and* total coliform counts.

Bacteriophages

None of the bacteriophages reported in this study demonstrated a significant correlation (>0.3;<-0.3) with GCN of 16S rRNA or 18S rRNA, pH, *E. coli, and* total coliform counts or TSS. *Escherichia* phages and *Emdodecavirus* were the only viruses that correlated positively with atmospheric pressure (rho=0.34, p-value 0.01; rho 0.31, p-value $2.7e^{-3}$, correspondingly). Moreover, *Escherichia* phages also depicted a positive relationship with precipitation (rho=0.34; p-value $1.93e^{-05}$). Finally, only *Synechococcus* phages and *Emdodecavirus* depicted a significant relationship with daylight (rho=0.38,p-value $8.33e^{-13}$; rho=0.31, p-value $7.0e^{-05}$, respectively).

Bacteria

Sporichthyaceae positively correlated with temperature, atmospheric pressure, and *E.coli* counts (rho=0.72,p-value <0.05; 0.38, p-value $4.30e^{-06}$; 0.35; p-value 9.81^{e-08} respectively). On the other hand, this family had a negative correlation with DO and GCN 16S rRNA. (rho=-0.63, p-value $1.22e^{-10}$;-0.32, p-value $9.39e^{-05}$ respectively). The bacterial family *Aeromonadaceae* had a positive correlation with water temperature, daylight, and total coliform counts (rho= 0.37, p-value $5.14e^{-08}$;0.32, p-value $3.07e^{-06}$;0.39, p-value $8.24e^{-04}$ respectively) and a negative correlation with DO (rho= -0.41, p-value $2.34e^{-05}$). Next, as depicted by the concentration ellipses in Figure 3.2.4 *Burkholderiaceae* was observed in both warm and cold months. Intriguingly, this family did not demonstrate a linear correlation with any of the physiochemical, environmental, and biological data studied (rho below the set threshold >0.3, <-0.3). Similarly, *Enterobacteriaceae* did not correlate with any environmental parameter.

On the other hand, *Flavobacteriacee* was the only family that demonstrated a positive intercorrelation with GCN of 16S rRNA (rho=0.35, p-value 1.64^{e-04}). This family also correlated positively with DO while demonstrating a negative correlation with temperature. (rho=0.36, p-value $2.57e^{-06}$; -0.45, p-value 1.72 e^{-06}, respectively). The relative abundance of *Pseudomonadaceae* was positively correlated with precipitation, whereas a negative correlation was demonstrated with pH (rho= 0.37, p-value 0.01; -0.32, p-value $7.78e^{-03}$ respectively). Although the principal component of this family points out to the quadrant where most cold weather observations were discovered, these bacteria were also found in warm months (concentration ellipses in figure 3.2.4, figure 3.1.6). Lastly, *Spirosomaceae* members depicted a negative correlation with daylight. (rho= -0.44, p-value $3.81e^{-07}$).

Microeukaryotes

Most of the members of the Eukarya domain reported in this study pointed towards the same quadrant and remained grouped demonstrating positive intercorrelation among each other. The microeukaryotic group *Mediophyceae* displayed a negative correlation with precipitation, and daylight (rho= -0.47, p-value $8.14e^{-06}$;-0.30, p-value 0.04). Interestingly, microbial eukaryotes did not correlate with GCN of 18S rRNA (Figure 3.2.5).. Regarding pH, only *Choreotrichia, Sphaeropleales,* and the fungi group *Kappamycetaceae,* demonstrated a linear relationship with this parameter (rho= 0.38, p-value $2.55e^{-03}$;0.30, p-value 0.04; 0.40, p-value $1.20e^{-03}$ respectively).

In terms of temperature, *Kappamycetaceae, Thoracosphaeraceae, and Chlamydomonadales* correlated negatively with this parameter (rho -0.37, p-value $1.00e^{-06}$; -0.30, p-value $6.66e^{-05}$; -0.33, p-value $2.34e^{-04}$, correspondingly). In addition, *Kappamycetaceae* had a positive correlation with DO (rho=0.33, p-value $7.70e^{-04}$) Furthermore, Eight out of eight microbial eukaryotes had a negative correlation with daylight. Moreover, *Sphaeropleales* and *Oligotrichia* demonstrated a negative relationship with atmospheric pressure (rho=-0.34, p-value $6.32e^{-03}$; -0.33, p-value 0.05). Most microbial eukaryotes demonstrated a negative correlation with precipitation including *Mediophyceae, Cryptomonadales, Choreotrichia, Sphaeropleales, and Oligotrichia (*rho= -0.47, p-value 8.14e-^{06}; -0.55, p-value $5.48e^{-07}$; -0.36, p-value $9.55e^{-03}$; -0.63, p-value 5.44e-^{08};-0.30, p-value $9.81e^{-04}$, respectively).

In addition, most of the microeukaryotes did not correlate with *E. coli and* total coliform counts. Only *Chlamydomonadales* demonstrated a negative relationship with total coliforms (rho= -0.33, p-value 0.05) (Table 3.3.1). None of the microeukaryotic groups demonstrated a correlation with TSS.

3.3.1 Co-occurrence of bacteriophages, bacteria and Microeukaryotes

To study co-occurrence of microbial fractions across domains a co-occurrence network was created based on significant p-values (<0.05) as well as Spearman's rank correlation coefficients (>0.30) obtained (Figure 3.2.6). The nodes represent either bacteria or microeukaryotes found in the waterways studied across time, and the connection between nodes represents the co-

occurrence. The thickness of the edges is associated with the Spearman coefficient obtained. In this context, thicker edges depict a stronger correlation between domains. As represented in Figure 3.2.6, *Spirosomaceae* demonstrated co-occurrence with *Chlamydomonadales*, *Mediophyceae*, and *Oligotrichia (*rho= 0.55, p-value $4.44e^{-16}$; 0.38, p-value 0.01; 0.38, p-value $8.17e^{-03}$, respectively). The latter (*Oligotrichia*) demonstrated co-occurrence with *Burkholderiaceae* (rho= 0.30, p-value $4.00e^{-06}$*)*. Next, *Pseudomonadaceae* and *Sporichthyaceae* demonstrated co-occurrence with one microeukaryotic group each: *Pseudomonadaceae* with *Thoracosphaeraceae* (rho=0.33, p-value <0.05), and *Sporichthyaceae* with *Sphaeropleales* (rho=0.32, p-value $6.78e^{-03}$*)*. None of the bacteriophages retrieved in this study revealed a significant co-occurrence with the bacterial families included in the network (rho<0.30).

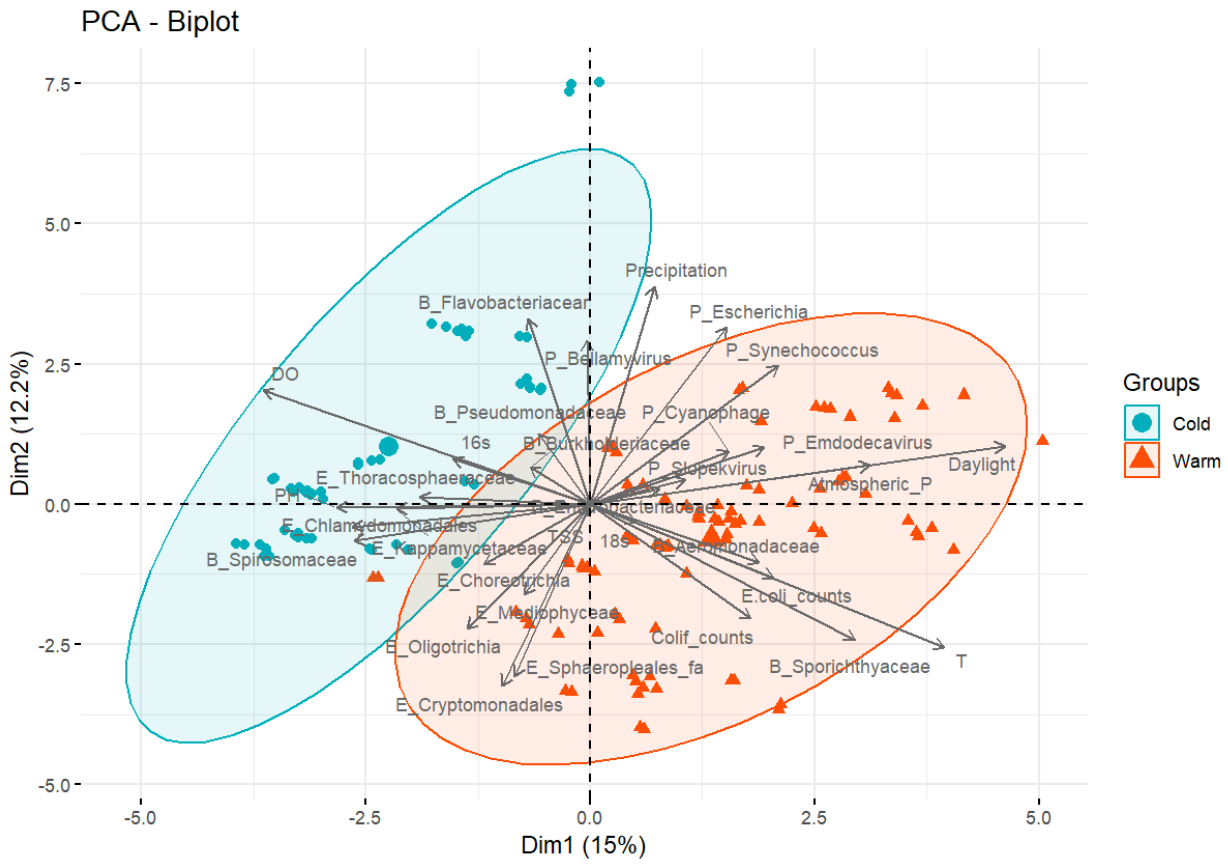

Figure 3.2.4 Principal component analysis (PCA) biplot based on the variance of 32 dimensions considered in the model including relative abundance of bacteria, microeukaryotes, viruses, gene copy number (GCN) of 16S rRNA and 18SrRNA, as well as environmental, biological and physicochemical water parameters. The first two components explained 15.0% and 12.2% of the variances, correspondingly. Longer arrows indicate a higher contribution to the variance, while shorter indicate a lower contribution. Positive correlated variables are grouped while negative correlated ones are positioned in opposite quadrants. Turquoise dots represent observations detected during cold months, while orange triangles indicate those from the warm months. The bacterial families are labeled with "B" at the start of the text, microeukaryotes with "E" and bacteriophages with "P".

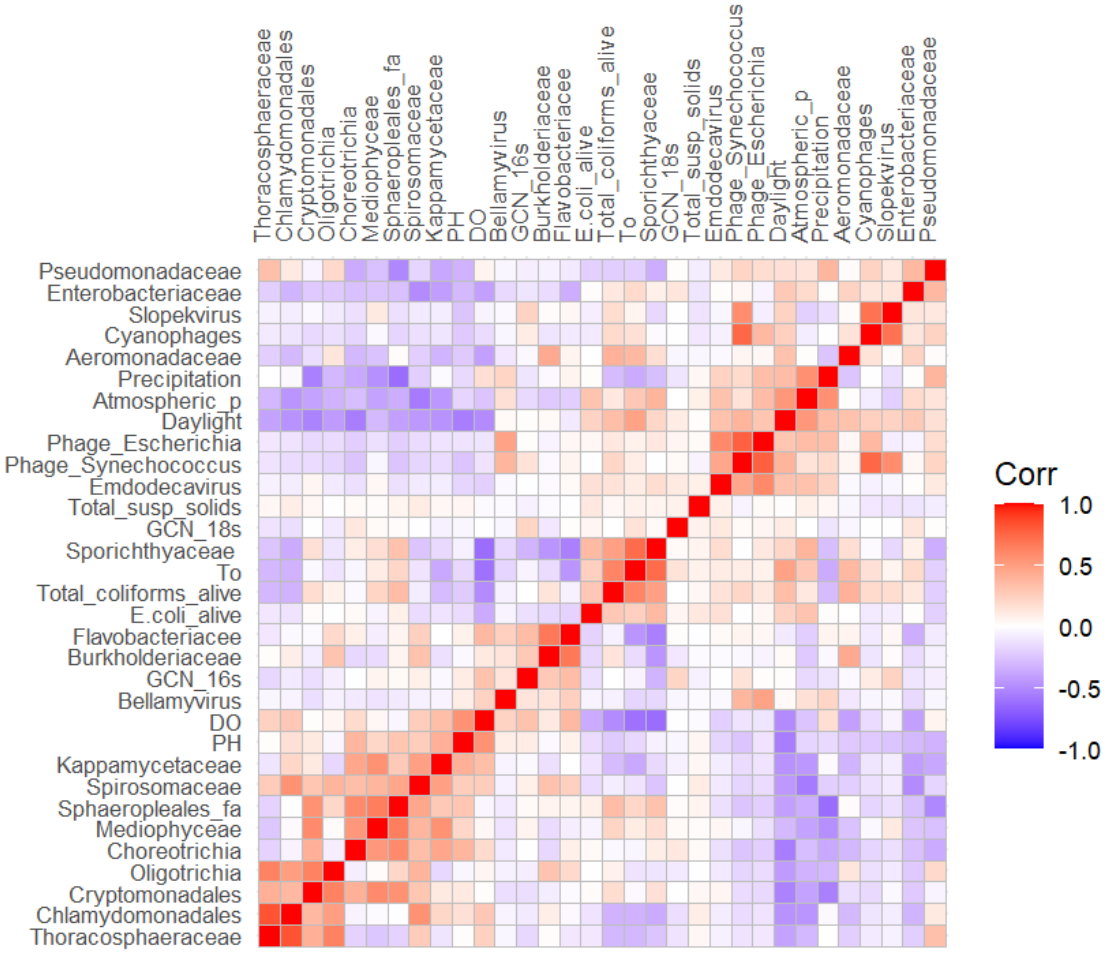

Figure 3.2.5 Correlogram of bacterial, microbial eukaryotes, and viral taxa with environmental, biological, and physicochemical water parameters encountered across time in the sampling locations of the Assiniboine River in rural Manitoba. Spearman correlation coefficients range from 1 to -1. Red indicates a very strong correlation, while blue indicates a very strong negative correlation. Atmospheric_p = atmospheric pressure; To = water temperature ; DO= dissolved oxygen.

Taxa	Correlation	rho	Type	Standard microbial test
Chlamydomonadales	Negative	-0.33	weak	Total coliforms
Sporichthyaceaea	Positive	0.35	weak	*E.coli*
Aeromonadaceae	Positive	0.39	weak	Total coliforms

Table 3.3.1 Bacteria and microbial eukaryotes found in Forested, Urban, and Agricultural waterways in rural Manitoba in 2022 that depicted a positive or negative correlation with *E.coli* and total coliforms, the standard microbial indicators used to asses for contamination in water monitoring.

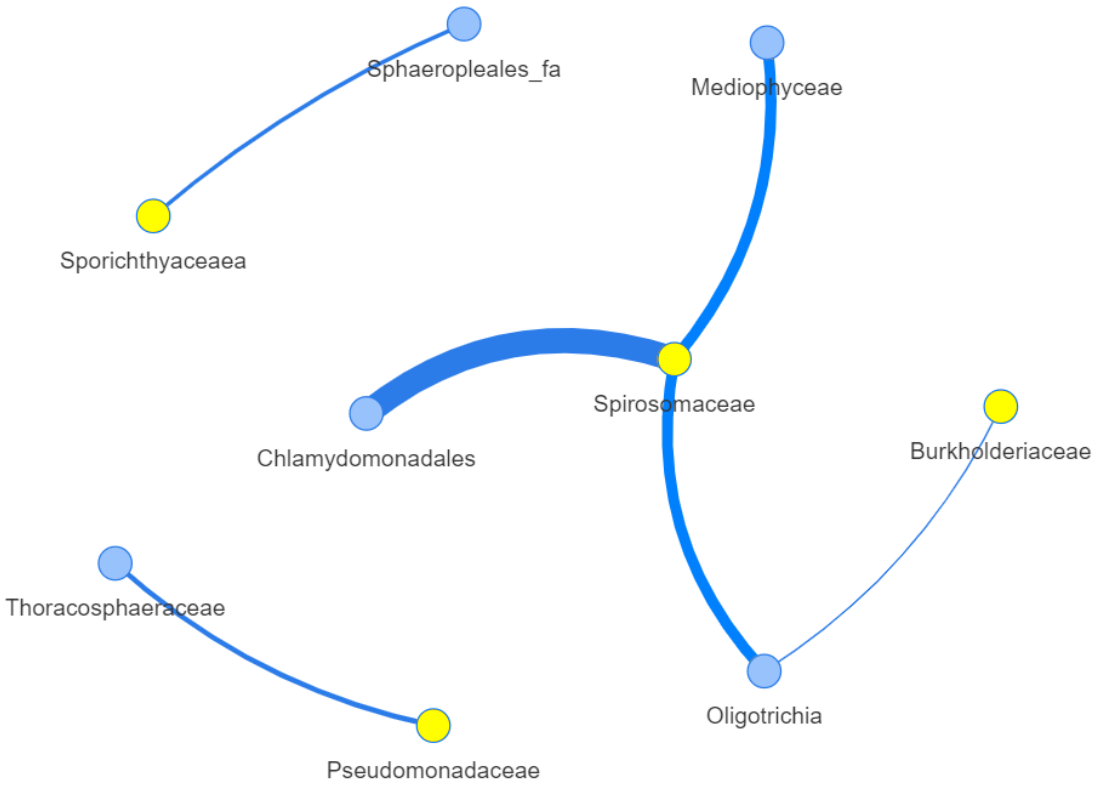

Figure 3.2.6 Co-occurrence network depicting the patterns across the microbial taxa (nodes) found in Forested, Urban, and Agricultural waterways of the Assiniboine River in rural Manitoba from April to November 2022**.** The thickness of the edges is proportional to the Spearman coefficient obtained. Bacteria nodes are yellow while microeukaryotes nodes are blue. Only rho values >0.30 and p-values $\leqslant$0.05 between different domains were included in the network.

4.0 DISCUSSION

4.1 Evaluation of Viral Diversity

The viral diversity and taxonomic composition were explored across locations and over time in anthropogenically impacted waterbodies in rural Manitoba. A total of 139 amplicons containing the major capsid protein g23 gene amplified using MZIA1bis and MZIA6 and T4superF1, T4superR1 degenerate primers (Table 2.2, (Chow & Fuhrman, 2012) failed the deep amplicon sequencing with Illumina Miseq technology V3 chemistry, 300-400 bp. The most probable reason, as mentioned in section 3.1.1, is the mispriming of the Illumina sequencing primers. This occurs when primers bind to unintended genome regions other than the targeted region (Puskas & Bottka, 1995; Qiagen, 2024). The outcome is nonspecific products, incorrect sequencing, and low-quality reads (Puskas & Bottka, 1995). While most studies targeting myoviruses have successfully applied MZIA1bis and MZIA6 primers (Potapov et al., 2018; File, Ttart, Suttle, & Krisch, 2005; Uyaguari-Diaz et al., 2016), only a limited number of literature have used T4superF1, T4superR1 for T4-like phages recovery successfully (Chow & Fuhrman, 2012). T4 super primers provided an alternative to obtain overlapping sequences as MZIA1bis and MZIA6 are around 600 bp and do not overlap, despite this, T4superF1, T4superR1 did not work as expected.

In addition to mispriming which is directly associated with the sequencing process, the quality of sequencing data obtained from environmental samples can also be challenged by low viral abundance (Breitbart et al., 2002;Uyaguari-Diaz et al., 2016), degradation, and fragmentation of viral DNA, competition with other genetic material (Gonzales Gustavson et al., 2017) from bacteria and eukaryotes present in the sample or inhibitory substances present in water (humic acids) (Schrader, Schielke, Ellerbroek, & Johne, 2012). Overall, viral DNA yields have been reported to be less than bacterial and eukaryotic fractions (Munteanu et al., 2023). Furthermore, the method used for microbial concentration holds certain limitations. SMF method and recoveries are susceptible to improvement. Moreover, small flocs containing viruses can be accidentally decanted after centrifugation, contributing to viral loss (Gonzales Gustavson et al., 2017). This approach could have benefited from the filtration of larger particles (bacteria and microeukaryotes) to improve viral recovery and downstream analysis (Yanac, et al, 2023; Kaletta et al.,2020).

Alpha diversity analysis increases the source of error when comparing uneven observations therefore we analyzed only the observed diversity based on the UVIGs recovered from the 21 successfully sequenced samples. The samples had on average 3339.423 UViGs (Shannon diversity index= 5.00). It is important to highlight that the values obtained represent only a small fraction of the entire viral community (targeted approach). The forested area No. 2 had the highest observed viral diversity (6504 UViGS) in July. This same site had the lowest observed diversity (504 UViGS) in November. A month-to-month and location-by-location comparison could not be performed for viral reads in this study, however, fluctuating viral abundances with lows during November without showing any specific seasonality pattern have been previously reported (Chow & Fuhrman, 2012). The viral composition (figure 3.1.3) did not demonstrate a significant change over time (p-value 0.434). However, UVIGs recovered from samples from different locations were significantly different (p-value 2.886e-07). While the "months" variable alone did not demonstrate influence in the compositional viral taxa, other environmental parameters that changed over time revealed a positive correlation with some of the viral groups encountered (as further explained in section 4.2).

4.2 Evaluation of Viral Taxonomy and seasonal and spatiotemporal patterns

The rate of taxonomic classification to the viral g23 samples was 1% in this study. As previously mentioned, this is congruent with public literature since most viruses remain uncultivated, and omics technologies allow for the identification of a massive number of new but unclassified phage genomes (Roux, Emerson, Eloe-Fadrosh, & Sullivan, 2017; Uyaguari-Diaz et al., 2016; Korf et al., 2019). Cyanophages (including *Synechococcus* phages, and members of the *Bellamyvirus* genus) represent 62% of the viral composition encountered. This is consistent with most studies suggesting that cyanobacteria including *Synechococcus* are one of the most abundant and prevalent picophytoplankton in marine and freshwater ecosystems (Chenard, Chan, Vincent, & Suttle, 2015; Coello-Camba et al., 2020; Zhang, He, & Gin, 2021). Therefore, cyanophages as mortality agents, follow their hosts, regulating their population, carbon fixation, and nutrient flow in aquatic environments (K. Wang, Wommack, & Chen, 2011; Sandaa & Larsen, 2006).

Cyanophages and *Synechococcus* phages have been reported to tolerate a wide range of temperatures and to be highly abundant during summer or warm months (Chenard, Chan, Vincent, & Suttle, 2015). This is consistent with our observations since most viral reads analyzed correspond to warm months (Table A1). Although cyanophages could not be positively correlated with temperature, a linear correlation with daylight was confirmed (rho=0.38), which is generally higher during summer and warm months. Moreover, studies have reported that higher infection rates of cyanophages concurred with lowered *Synechococcus* densities, this could explain why the bacterial family *Synechococcaceae* was not among the top 10 most relatively abundant bacteria found in this study. However, the Cyanobacteria *Nostocaceae* was among the top 9 recovered in this study (K. Wang, Wommack, & Chen, 2011). Within this context, high throughput sequencing generally fails in the detection of less abundant microbial groups (Sandaa & Larsen, 2006). This finding suggests that *Synechococcus* and their viruses cyanophages are important microbial components in the Assiniboine River's forested, urban, and agricultural waterways sampled in this research.

Escherichia coli phage (coliphage) G4507 was the second most relatively abundant bacteriophage retrieved in this study (26%). Coliphages with contractile tails have been found in similar aquatic environments such as seawater, rivers, and surface water (Michniewski et al., 2019; Korf et al., 2019). Coliphages have been suggested as fecal indicators because of their relationship with *E. coli* (V. Reyes & Jiang, 2010). Furthermore, these phages have been demonstrated to provide additional information to the one granted by bacterial surrogates (V. Reyes & Jiang, 2010). In addition, they frequently co-occur with other viral pathogens and have been proven resistant to wastewater treatments and persistent in the environment (Duran et al., 2002; McMinn, Rhodes, Huff, & Korajkic, 2020). Thus, in agreement with other studies (Skraber, Gassilloud, & Gantzer, 2004), the presence of this phage suggests fecal pollution coming from human and animal waste in addition to viral contamination in the waterways sampled. Although a positive correlation between the standard bacterial indicators to assess water fecal contamination could not be demonstrated in this study (rho below the threshold set =<0.30), *Enterobacteriacea* family was the third most relatively abundant bacterial family encountered in this analysis. These results are in agreement with previous research that has found coliforms to be in lower concentrations than coliphages due to bacterial susceptibility to

environmental conditions and inactivation factors (Duran et al., 2002; Skraber, Gassilloud, & Gantzer, 2004) . In the present study coliphages positively correlated with atmospheric pressure and precipitation (rho=0.34, p-value 0.01; 0.34, p-value $1.93e^{-05}$ respectively). Although studies reporting the relationship between atmospheric pressure and coliphages are limited, there is literature that reinforces the association between coliphages and precipitation, potentially suggesting that the sources of fecal pollution can be carried out by runoff (V. Reyes & Jiang, 2010). These results may confirm the importance of coliphages to assess river water fecal contamination even in the reported absence of standard water microbial indicators such as *E. coli* and total coliforms.

Interestingly, coliphages were positively correlated with cyanophages (rho=0.77, p-value $5.31e^{-07}$) and *Emdodecavirus* (rho=0.59, p-value $2.63e^{-08}$). After coliphages, the *Emdodecavirus* genus (8%) was one of the most commonly found viruses in this study. One important member of this genus is the Rhizobiophage, that appear to be abundant in soil and to play a role in regulating the legume root bacteria *Rhizobia* (Srinivasiah et al., 2008; Cauwenberghe & Simms, 2023). Finding this genus in forested and urban waterways can be explained by the fact that viruses can be dispersed by wind-borne dust in short- and long-term distances (Cauwenberghe & Simms, 2023; Whon et al., 2012). This may be the case since urban and forested geographic locations are surrounded by vegetation a few meters and even centimeters from the sampling sites. As previously stated, bacteriophages can exceed bacterial abundance compared to their co-occurring host (Cauwenberghe & Simms, 2023; V. Reyes & Jiang, 2010). That is why these phages may indicate the presence of nodule bacteria near the sampling sites even if they were not observed in the top 10 bacterial compositions assessed in this study. In a similar way to coliphages mentioned above, *Emdodecavirus* was intercorrelated with daylight (rho=0.31, p-value $2.74e^{-03}$) and atmospheric pressure (rho= 0.31, p-value $8.33e^{-13}$), suggesting susceptibility to environmental conditions that vary with pressure.

Additionally, the dsDNA baculovirus *Orgyia leucostigma nucleopolyhedrovirus* (a virus outside of the targeted group) was recovered from the urban waterways. For instance, the finding of *Orgyia leucostigma nucleopolyhedrovirus* reflects incorrect sequencing due to mispriming. When testing the T4 super primer pair prime in the genome of *Orgyia leucostigma*

nucleopolyhedrovirus, binding regions were found when allowing up to 8 mismatches. This finding cast doubts on the robustness of the primer sequences employed to target T4-like phages. However, due to the scarcity of studies using T4super primers, there were no comparable findings available for non-targeted virus identification.

Finally, the *Slopekvirus* genus has been isolated in sewage (Kondo et al., 2023) and human feces (Smith-Zaitlik et al., 2022). Literature available regarding this genus establishes that *Slopekvirus* shares a 98.8% nucleotide identity with *Klebsiella* phages (Kondo et al., 2023). Finding *Slopekvirus* genus in the agricultural waterways supports the exposure of *Enterobacteriaceae* as the 3rd most relatively abundant family found in the river water examined. Due to the failed sequencing in most of the viral samples, a clear spatio-temporal viral composition could not be elucidated.

Overall, the bacteriophages encountered did not significantly influence the variance observed in this study (section 3.2). More research is needed to fill in the considerable gaps prevailing in viral databases and identify clear viral seasonal patterns in river water. Following this, additional evidence is required to elucidate viral correlation with environmental, physiochemical, and biological parameters as well as their co-occurrence with other microorganisms aside from their hosts in aquatic environments.

4.3 Evaluation of bacterial diversity

Bacteria had on average a Shannon diversity index of 4.92, slightly lower than the average obtained by viruses (5.00) but higher than the diversity of microeukaryotes (3.57). Remarkably, these results were congruent with the GCN of 16S rRNA and 18S rRNA gene quantification through qPCR, where bacteria exceeded eukaryotes by 4 orders of magnitude (section 3.2.1). Furthermore, a higher abundance of viruses compared to the hosts is expected in aquatic environments (Bergh, Brsheim, Bratbak, & Heldal, 1989; Cobian Guemes et al., 2016). Unfortunately, quantification through qPCR of T4-like phages could not be performed in this study to confirm the higher abundance of viruses over bacteria or microeukaryotic fractions.

Forested waterways had the highest prokaryotic diversity in this study, followed by urban and lastly agricultural waterways. Similarly, 1-D Simpson demonstrated high dominance across locations, suggesting that a few species dominate the bacterial community, especially in the forested location (section 3.1.4, figure 3.1.5) which is congruent with studies performed in aquatic environments (Reza et al., 2018; Zwart, Crump, Kamst-van Agterveld, Hagen, & Han, 2002). Contrasting the differences in diversity obtained across locations is challenging since most metagenomic studies do not perform "case-control" studies or comparative waterway assessments. Instead, most research studies are conducted around wastewater treatment plants because of the microorganisms and mobile genetic elements they release into aquatic environments (Jankowski et al., 2022; Reza et al.,2018; Chu et al., 2018); or in urban rivers or lakes without anthropogenic activities comparisons. In one study conducted by Uyaguari-Diaz et al., higher bacterial counts were found in Agricultural waterways in comparison to "Protected" and "Urban" locations. However, these analyses were performed with flow cytometry and study sites were not interconnected with other waterways (Uyaguari-Diaz et al., 2016). In the present study, waterways were interconnected along the section (12 Km) of the Assiniboine River around the city of Portage la Prairie and rural areas. In this context, the "location" variable did not demonstrate an impact on bacterial composition or bacterial relative abundances (sections 3.1.5; 3.2) . On the other hand, regarding 16S rRNA GCN/100 ml of water abundance, locations demonstrated to significantly influence bacteria GCN abundance. Nevertheless, the impact of the influence is minimal (1%) (section 3.2.1).

On the other hand, diversity and bacterial composition changed significantly over time (section 33.1.5). April was the month with the lowest diversity (cold, snow melting point). In contrast, June depicted the highest Shannon value. Interestingly, June is the month with the highest precipitation in the city of Winnipeg (91.4 km away from sampling sites) (Climate Data, 2022). Despite these findings, diversity values increased and decreased randomly. Therefore, clear seasonal patterns could not be observed. A potential explanation for this is the shift in environmental conditions throughout the year. For example, in a study conducted by Reza et al. fluctuations in prokaryotic abundance were influenced especially by temperature (Reza et al., 2018). In the present study, the relative abundance of bacterial groups was mostly influenced by daylight length, temperature, and precipitation (see section 3.3). This is congruent with previous

studies stating that bacteria respond to environmental stressors due to their specific physiological characteristics (Sagova-Mareckova et al., 2021; Laplante & Derome, 2011).

4.4 Evaluation of bacterial taxonomy and seasonal and spatiotemporal patterns

The top 10 bacterial families discovered across forested, urban, and agricultural waterways sampled from April to November 202 from the rural section of the Assiniboine River are explained below.

Sporichthyaceae belonging to the Actinobacteria class was the most relatively abundant (20.27%) family encountered across time and seasons in this study. As mentioned in section 3.1.5, members of this group are commonly freshwater microorganisms, typical of oligotrophic water lineages (Cruaud et al., 2020). In a study conducted by Cruaud et al., (2020), *Sporichthyaceae* was one of the most abundant microorganisms present in Saint-Charles River, the main source of drinking water in Quebec City. Moreover, in the aforementioned study, these taxa were encountered in higher proportions during the summer compared to winter. These results align with our findings since this family positively correlated with temperature and atmospheric pressure (rho=0.72, p-value <0.05 and 0.38, p-value $4.30e^{-06}$, respectively). Additionally, *Sporichthyaceae* was the only family to demonstrate a linear relationship with *E. coli*, one of the standard microorganisms used to assess fecal water contamination (rho= 0.49, p-value $9.81e^{-08}$). These findings may suggest that *Sporichthyaceae* could be associated with recent fecal contamination in water. However, more studies are needed to support this statement.

Pseudomonadaceae (19.93%) from the class Gammaproteobacteria, was the second most abundant family isolated across locations. These taxa appeared indistinctively throughout warm and cold months (section 3.1.5, figure 3.1.6). *Pseudomonadaceae* includes pathogenic species for humans and animals (Lobb et al., 2020). Generally, these prokaryotes have been associated with cosmopolitan or urban-impacted waterways and polluted rivers (Numberger et al., 2022; Uyaguari-Diaz et al., 2016; Cai et al., 2016). Previous studies have shown an inverse correlation of *Pseudomonadaceae* with environmental factors such as DO (Cai et al., 2016), which aligns with our findings (section 3.3). On the other hand, *Pseudomonaceae* positively correlated with

precipitation (rho= 0.37, p-value 0.01). These results may confirm the association between rain and the enrichment of the aquatic microbial communities, including the dissemination of human pathogens into water (Rose et al., 2000; Cruaud et al., 2020).

Enterobacteriaceae (14.8%), was the third most observed bacterial taxa. While this family is primarily known for its enteropathogenic members, it also includes non-pathogenic bacteria commonly found in the environment (Aizenberg-Gershtein, Vaizel-Ohayon, & Halpern, 2012; Wisniewska, Niewolak, Korzeniewska, & Filipkowska, 2007; Ferguson & Signoretto, 2011). In the present study, *Enterobacteriaceae* did not depict any correlation with environmental variables (section 3.3). Interestingly, this gram-negative family did not correlate with either standard heterotrophic counts of *E. coli* or total coliform counts. Some studies have highlighted the poor synchronization between culture-independent and culture-dependent methods (Merino-Mascorro, Hernandez-Rangel, Heredia, & Garcia, 2018). Moreover, there is the possibility that ASVs assigned to *Enterobacteriaceae* are mostly environmental strains (Ferguson & Signoretto, 2011). Important to note that the absence of enterobacteria does not preclude water from being contaminated with other pathogens. Nevertheless, a culture-independent approach at the species level would be needed to discriminate between autochthonous *Enterobacteriaceae* from the microbial community of the Assiniboine River and introduced pathogenic bacteria from different pollution sources such as wastewater treatment plants, or wild animals. In general, freshwater with a higher abundance of *Enterobacteriaceae* is often linked to fecal contamination from humans or animals (Aizenberg-Gershtein, Vaizel-Ohayon, & Halpern, 2012; Wisniewska, Niewolak, Korzeniewska, & Filipkowska, 2007).

Burkholderiaceae (8.943%), a family that compromises opportunistic pathogens with resistance to antibiotics and disinfectants (Xu et al., 2024), was the fourth most common bacterial family encountered across waterways. Although *Burkholderiaceae* belongs to the Proteobacteria class which has been recognized to be abundant in river water samples, these bacteria have been isolated in highly contaminated waterways including urban rivers and livestock breeding water bodies (Xu et al., 2024; L. Zhang, Fang, Li, Lu, & Li, 2020). In this study, *Burkholderiaceae* was present during warm and cold months without a specific pattern (section 3.3). Moreover, *Burkholderiaceae* did not demonstrate a correlation with any of the physiochemical,

environmental, and biological data here examined (figure 3.2.5, section 3.3.). Seasonal patterns of this proteobacteria in freshwater are limited, only in one study performed by Fiedler et al., (2018), *Burkholderiaceae* was found to be the most abundant in groundwater samples during May. In addition to this, in a study performed by Zhang et. al., (2020) in the urban river Qingliu in China, the occurrence of this group was associated with the presence of nitrates and nitric acid coming from wastewater treatment plants, since these prokaryotes play an important role in nitrification.

Spirosomaceae (7.90%) was the fifth most relatively abundant prokaryotic group encountered across river water in this study. This family belongs to the Bacteroidetes phylum along with *Flavobacteriaceae*, also abundant in these waterways and further analyzed in this section. *Spirosomaceae* did not correlate with water temperature (section 3.3). Moreover, this family depicted a negative correlation with daylight (rho=-0.44, p-value $3.81e^{-07}$). Although *Spirosomaceae* has been associated with freshwater health because of their ability to break down complex organic pollutants in water (Baek, Jang, Goh, & Choi, 2024; Bai et al., 2023), one study conducted by Makk et al., (2024) found these taxa in the moderately polluted Danube River in Budapest, Hungary.

Flavobacteriaceae (3.78%) also from the Bacteroidetes phylum is another of the groups found roughly in accordance with studies performed in surface water and other marine systems (Gomez-Pereira et al., 2010; Fiedler et al., 2018). This family is considered to be ubiquitous in different environments and to thrive at different temperatures (Gomez-Pereira et al., 2010) with high abundance reported in cold waters which is consistent with our findings (Abell & Bowman, 2005). In our research, *Flavobacteriacee* correlated positively with DO while demonstrating a negative correlation with temperature (section 3.3). Furthermore, *Flavobacteriacee* was the only bacterial family that contributed highly to the variances observed in this study when performing PCA (9.94%). Interestingly, these prokaryotes have been reported to strongly depend on the availability of organic matter (Fiedler et al., 2018). In addition, members of this family such as *Flavobacterium* have been evidenced as opportunistic fish and marine organism pathogens (Reza et al., 2018; Xu et al., 2024). *Flavobacteriacee* members were encountered in livestock-rearing sites and the presence of this bacteria was associated with pathogenic pollution of the water

environment (Xu et al., 2024). Interestingly, the order *Bacteroidales* (Bateroidetes phylum) has been proposed as an alternative marker to identify fecal contamination for demonstrating more consistency in appearance than fecal indicator bacteria when analyzing produce contaminated with animal and human feces (Merino-Mascorro, Hernandez-Rangel, Heredia, & Garcia, 2018).

Aeromonadaceae (3.78%), which also belongs to the class Proteobacteria, was the 7th most relatively abundant prokaryotic family encountered across forested, urban, and agricultural waterways considered in this study. Moreover, these facultative anaerobic bacteria had a positive correlation with water temperature, daylight, and total coliform counts and a negative correlation with DO (section 3.3, table 3.3.1). According to previous studies, *Aeromonadaceae* has been suggested as sewer indicator bacteria because of its proliferation within sewer systems (McLellan, Huse, Mueller-Spitz, Andreishcheva, & Sogin, 2010; Yang, Li, Gao, Chen, & Zhan, 2019). However, it has been reported that *Aeromonadaceae* have low specificity as they can be found in freshwater, seawater, polluted, and unpolluted water (Vaz-Moreira, Nunes, & Manaia, 2014). Interestingly, most members of *Aeromonadaceae* are susceptible to chlorine and other disinfectants, unlike *Flavobacteriaceae* (mentioned above) which is resistant to chlorine, chlorhexidine, and other antiseptics (Becker, Neves, Angelis, & Navarrete, 2017).

Finally, *Micrococcaceae (3.43)* from the Actinobacteria class and *Nostocaceae* (3.09) (Cyanobacteria) depicted fewer occurrences than the other bacterial families in this study. *Micrococcaceae* was found only in forested and urban locations, while *Nostocaceae* only in urban areas. Because of their low relative abundance, these taxa could not be associated with any of the physiochemical, environmental, and biological data analyzed. Both Actinobacteria and Cyanobacteria are typical groups in freshwater ecosystems according to previous studies (Vaz-Moreira, Nunes, & Manaia, 2014; Fiedler et al., 2018). Previous studies have highlighted the importance of Actinobacteria in maintaining freshwater quality due to its crucial role in decomposing organic carbon, remediating contaminants, and competing with intruding bacteria brought by different contamination sources (Ung et al., 2019; Fiedler et al., 2018). Likewise, Cyanobacteria is fundamental for primary production and trophic interactions in aquatic environments. However, Cyanobacteria accompanied by high temperature and high

nutrient loads can cause cyanotoxic algal blooms that are pathogenic for humans and animals (Walter et al., 2018; Fiedler et al., 2018).

In summary, riverine bacterial composition in the rural section of the Assiniboine River was dominated by Actinobacteria, Proteobacteria, Bacteroidota, and Cyanobacteria, which are typical river water groups as reported by previous authors studies (Vaz-Moreira, Nunes, & Manaia, 2014; Fiedler et al., 2018). From the families encountered *Pseudomonaceae*, *Enterobacteriaceae*, and *Burkholderiaceae* are of major concern because of their relatively high abundance, generally being isolated in highly disturbed waters, their association with pathogenicity in addition to being considered potential antibiotic-resistance genes carriers. The major environmental contributors to bacterial dynamics in this study were daylight length, temperature, DO, and precipitation, respectively. The response of the aquatic environments to environmental disturbances and pollution is reflected in the changes in microbial structure and composition. Identifying bacterial profiles is essential for monitoring the impact of exogenous inputs on water and potential health risks to humans and animals.

4.5 Evaluation of Micro-eukaryotes Diversity

Despite microeukaryotes being widespread and playing fundamental biogeochemical processes in water, they remain less studied than prokaryotes in freshwater environments (Cruaud et al., 2020). In the present study, we aimed to uncover the diversity and temporal dynamics of microbial eukaryotes by studying forested, urban, and agricultural waterways located in the Assiniboine River in rural Manitoba from April to November 2022. Forty-three samples possessed quality scores indicatives of high base call accuracy, therefore only those were included in the analysis (Table A2), (Illumina, 2024). Although higher sequencing rates for these genes were expected, DNA sequence fragments derived from environmental samples are a well-known issue in deep amplicon sequencing analyses. Water samples contain many PCR inhibitors including polysaccharides coming from plants and fecal matter from vertebrates which can mimic the structure of nucleic acids and disturb the enzymatic reaction (Schrader et al., 2012). Moreover, fulvic and humic acids, by-products of the decomposition of animals and plants by microorganisms, have been reported to interact directly with nucleic acids and polymerase

during PCR preventing enzymatic reactions (Schrader et al., 2012; Zhu et al., 2023; Sutlovic, Gojanovic, Anelinovic, Gugic, & Primorac, 2005)

In addition to sample inhibitors, other limiting factors can also contribute to poor sequencing. For example, the high genome size is one of the main challenges when studying microbial eukaryotes since this leads to fragmented and incomplete genomes. (Seeleuthner et al.,2018). Besides, it is still debated among researchers which region of the genome provides the most utility. While V4 (the one used in this study) -5 regions have the highest number of sequences deposited in international nucleotide sequencing databases, V1-3 showed higher identification power (Tanabe et al., 2016). Therefore, V3 has been speculated to be more suitable for environmental samples. However, the length of this region is not suitable for Illumina sequencing (Tanabe et al., 2016).

Microeukaryotes exhibited a Shannon diversity index of 3.57, the lowest in comparison to other microbial fractions as mentioned in section 4.3 and reported in previous studies (Cobian Guemes et al., 2016). Likewise, GCN of 18S rRNA through qPCR gene quantification depicted on average $9.47X10^3$ copies while bacteria showed $3.40\ X10^7$. These differences are consistent with reported literature, where bacteria surpassed eukaryotes by 3 or more orders of magnitude when analyzing GCN abundance (Wei, Xia, Huang, & Wei, 2024). Similarly to bacteria, the location with the highest microeukaryotic diversity was the forested waterway, followed by the urban and lastly the agricultural setting (section 3.1.6). The dominance (1-D) index calculated suggests that few micro-eukaryotic groups dominate the waterways sampled. Furthermore, the abundance of the GCN 18SrRNA gene and the differences in diversity found across locations were not statistically significant, potentially due to the interconnectedness of the sampling sites along the rural section of the Assiniboine River as depicted in Figure (2.1). Microeukaryotic alpha diversity and abundance (measured through GCN 18SrRNA) appeared influenced by sampling months and the environmental changes that accompany the seasonal variation of atmospheric pressure, daylight length, precipitation, pH, water temperature, and DO (section 3.2). Microbial eukaryotes demonstrated the highest diversity on average during August 2022 when the water temperature was the warmest (24.74 °C) unlike bacteria that depicted their highest abundance during June 2022 (highest precipitation observed) (section 3.1.4). Microeukaryotic alpha diversity was partially congruent with 18S GCNs since the highest copies on average were recovered during May ($3.12X10^4$) and August 2022 ($9.29X10^3$). Despite this, the 18S rRNA

GCNs did not correlate with the relative abundance of any of the microbial eukaryotes recovered (Section 3.3). Just like with bacteria, clear seasonal patterns could not be identified. This highlights the role of environmental conditions in shaping microeukaryotes communities as stated by previous research (Bai et al., 2023; Capo et al., 2017; Cruaud et al., 2020). Although the alpha diversity of microbial eukaryotes remained fairly homogenous across waterways, the microeukaryotic groups were different across locations (p-value $5e^{-03}$) (Figure 3.1.9) and over time (p-value $4e^{-04}$) (Figure 3.2.0).

4.5 Evaluation of microeukaryotic taxonomy and seasonal and spatiotemporal patterns

Among the eukaryotic groups encountered, the diatoms class *Mediophyceae* was the most abundant (24.02%) over time and across sampling sites. Diatoms are the most diverse association of unicellular, autotrophic, algae that play a key role in aquatic food webs (Y. Wang et al., 2021). Large percentages of *Mediophyceae* were reported in urban waterways (40.59%) than in any other location (section 3.1.7). Although a correlation with water temperature could not be confirmed, most diatoms from this group were encountered in May, September, and early October. Interestingly, *Mediophyceae* correlated negatively with precipitation, and daylight (rho= -0.47, p-value $8.14e^{06}$;-0.30, p-value 0.04, respectively)
Diatoms have been widely used as bioindicators to assess river quality (Rimet & Bouchez, 2012). However, its assessment relies mostly on morphological changes associated with climate change and temperature shifts (Luethje & Snyder, 2021). Some members of this class have been reported to cause harmful algal blooms in estuarine coastal waters (Tanabe et al., 2016). Although the level of taxonomic resolution needed for biomonitoring and identifying pollution conditions using diatoms is still debated, reports have shown that microeukaryotes at the species and genus level have more restricted geographical ranges than orders or families (Bowman & Bailey, 1997; Rimet & Bouchez, 2012).

Cryptomonadales order was the second most relatively abundant (14.98%) group found across locations, with higher assignations in Forested waterways. In the present study, *Cryptomonadales* were the only unicellular algae that contributed significantly to the variances encountered (9.58%). This order was ubiquitous across months, and more abundant during warm

days (August and September 2022), (Figure 3.2.0). These findings align with previous research where the class *Cryptophyceae* was detected throughout the year in Saint-Charles River in Quebec, Canada with higher relative proportions observed during warm months (Cruaud et al., 2019). Interestingly, *Cryptomonadales* demonstrated a theoretical negative correlation with precipitation (as most microeukaryotes in this study), highlighting the effect of environmental factors in shaping microeukaryotic communities (Kang et al., 2024). Finding these protists in river water is not uncommon as they have been reported to have a cosmopolitan distribution and are well established in oceans and most types of freshwater habitats (Shalchian-Tabrizi et al., 2008).

Choreotrichia (14.15%) and its sister subclass *Oligotrichia* (3.07%) were the 3rd and 7th most relatively ciliated protozoa of the *Spirotrichea* class found in the Assiniboine River in Rural Manitoba, respectively. Of the locations sampled, *Choreotrichia* was more abundant in forested waterways, while *Oligotrichia* appeared more abundant in agricultural waterways (section 3.1.7). *Choreotrichia* and *Oligotrichia* were found throughout the year, without a defined seasonal pattern. Interestingly, from these two sister subclasses only *Choreotrichia* was demonstrated to be significantly influenced by water pH *(*positive correlation), while *Oligotrichia* had a significant response to atmospheric pressure (negative correlation). On the other hand, both microeukaryotes were negatively correlated with precipitation and daylight *(*section 3.3*).* Some members of these ciliated protozoa have been suggested as bioindicators for assessing water eutrophication and quality since they have been shown to tolerate extreme environmental conditions and more sensitivity to anthropogenic nutrient loading than abiotic factors (i.e. temperature, and salinity) (Sivasankar et al., 2018).

Sphaeropleales (7.70%) order from the class *Chlorophyceae* (Chlorophyta division) was the 4th most abundant protist group encountered across waterways during August, September, and early October. A higher abundance of this order was observed in the forested location compared to the other waterways (figure 3.2). These green microalgae were also positively correlated with pH and negatively correlated with precipitation, daylight, and atmospheric pressure. This is congruent with previous findings where *Chlorophyceae* populations declined during rainy months (Hujare, 2008). In addition to this, phytoplankton has been reported to bloom during

summer (Hujare, 2008), which aligns with the months where they were encountered in this study. *Sphaeropleales* harbor some of the most common species of freshwater green algae and embers of this group have been recommended for ecotoxicological assessment of nanomaterials coming from anthropogenic disturbances in freshwater (Suzuki, Yamaguchi, Nakajima, & Kawachi, 2018; OECD , 2013). Despite its recognized importance, *Sphaeropleales* has not been as deeply studied as other chlorophytes (Suzuki, Yamaguchi, Nakajima, & Kawachi, 2018).

Kappamycetaceae (5.21%) family from the order *Rhizophydiales* was the 5th and only freshwater fungi encountered exclusively in Urban and Forested sampling sites. These fungi were found indistinctively during warm and cold months. Just as in the case of *Choreotrichia* and *Sphaeropleales*, members of *Kappamycetaceae* also depicted a linear relationship with pH and DO (section 3.3). Conversely, these fungi correlated negatively with temperature and daylight. These findings go in accordance with previous literature which has reported that fungal species tend to increase if temperature (and salt) concentrations are low (Calabon et al., 2023). Moreover, *Rhizophydiales* were present throughout the year in Lake Stechlin (moderate level of nutrients or mesotrophic lake) in Germany with higher abundance during winter and spring months (Van den Wyngaert et al., 2022). In addition to this, *Kappamyceteceae* and other fungi that reproduce with flagellated spores (zoosporic) are associated with an increment in the dissolved organic matter in water (Calabon et al., 2023).

Thoracosphaeraceae (3.50%) was the 6th most relatively abundant protist isolated, and it was present only in Agricultural locations during October and November (figure 3.2). *Thoracosphaeraceae* depicted a negative correlation between temperature and daylight. On the other hand, these Dinoflagellates were positively correlated to DO (section 3.2). These protists along with diatoms are one of the most important and abundant groups of the marine microphytoplankton (Ribeiro & Amorim, 2008). They have been found to contribute to the eukaryome of the Pavilion Lake in British Columbia (Bonacolta, Visscher, Del Campo, & White III, 2024; McCarthy et al., 2018). Members of this family are acknowledged because of their ability to produce calcium carbonate cysts as part of their life cycle, which is useful to interpret fossil records which are critical in micropalaeontology studies (Ribeiro & Amorim, 2008). However, a higher abundance of these microorganisms may be of concern since some of these

resting cysts have been identified to play an important role in the ecology of harmful algal blooms in marine ecosystems (Tang & Gobler, 2012).

Finally, *Chlamydomonadales* a sister clade of *Sphaeropleales* from the Chlorophyta division was the 8th most relatively abundant protist encountered in this study. These green algae were present in agricultural and urban waterways during October and November 2022 mostly (figure 3.2). Just like *Sphaeropleales* , *Chlamydomonadales* was demonstrated to be negatively correlated with daylight. While *Sphaeropleales* did not show any correlation with temperature, *Chlamydomonadales* exhibited a theoretical negative correlation with temperature and a positive correlation with DO. This is congruent with previous studies that have highlighted the psychrophilic properties of *Chlamydomonadales*, since this group is a model organism for studying cellular adaptation to extreme environments (Cvetkovska, Huner, & Smith, 2017). Perhaps this is one of the reasons *Chlamydomonadales* correlated negatively with the presence of total coliforms in water (rho=-0.33, p-value 0.05). In one study performed by Barbosa et al., (2022) *Chlamydomonadales* demonstrated to thrive in low-turbidity waters. However, other reports indicate that *Chlamydomonadales* have adapted to intense light conditions, especially those capable of producing snow algal blooms as noted by vanHees et al., (2023).

In summary, these findings provide insights into the temporal dynamics and the microbial composition of eukaryotic communities in riverine ecosystems. In this study, daylight and precipitation were observed to be the major environmental drivers for microeukaryotes. Other studies have also reported the influence of salinity which was not assessed in this research (Bonacolta et al., 2024). The most abundant protist, *Mediophyceae* (diatoms) are primary producers in aquatic environments, forming the base of the food web for many marine and freshwater organisms (Y. Wang et al., 2021). Ciliated protozoa such as *Oligotrichia* and *Choreotrichia* represent a link between primary producers (i.e. small algae) and larger microeukaryotes by consuming and recycling nutrients in aquatic ecosystems. Protists can also modulate the prokaryotic community composition through grazing bacteria (i.e. Actinobacteria, such as *Sporichthyaceae*) which are frequently reported as prey for eukaryotes (Cruaud et al., 2020). Microeukaryotes such as diatoms, and green algae (*Sphaeropleales* and *Chlamydomonadales*) could help elucidate river nutrient fortification in addition to assessing the

impact of pollutants on the environment before the effects are noticeable (Sivasankar et al., 2018). Further research is necessary to fully understand the effects of changing environmental conditions on microeukaryotes. Considering microbial eukaryotes in sediments or river biofilm formation could help elucidate the direct impact of microeukaryotes nutrient availability and river biogeochemical cycles.

4.6 Evaluation of microbial co-occurrence

As previously outlined, a co-occurrence network was created based on the ASVs obtained from the anthropogenically impacted waterways sampled in the rural section of the Assiniboine River in rural Manitoba. The inclusion criteria in the network analysis included significant Spearman's rank correlation coefficients obtained (>0.30) as well as p-values <0.05) to further assess the interaction between microbial fractions.

None of the bacteriophages encountered in this study met the inclusion criteria (section 3.3). While there are some studies demonstrating the association and importance between bacteria and green macroalgae (Weiss, Costa, & Wichard, 2017; Ghaderiardakani, Coates, & Wichard, 2017) reports studying the families or lower taxa associations between bacteria and microeukaryotic interactions in riverine ecosystems are limited up to date.
Members of *Spirosomaceae* (Bacteroidetes) were observed to co-occur with *Chlamydomonadales* (green algae), *Mediophyceae (*diatoms*),* and *Oligotrichia* (ciliates), (rho= 0.55, p-value $4.44e^{-16}$; 0.38, p-value 0.01; 0.38, p-value $8.17e^{-03}$, respectively). Likewise, Oligotrichia correlated positively with *Burkholderiaceae* (Proteobacteria), (rho= 0.30, p-value 4.00e-^{06}). Furthermore, *Pseudomonadaceae* (Proteobacteria) demonstrated significant co-occurrence with *Thoracosphaeraceae* (Dinoflagellata) (rho=0.33, p-value <0.05) *and* lastly, *Sporichthyaceae* (Actinobacteria) showed a correlation with *Sphaeropleales* (green algae) (rho=0.32, p-value 6.78e-03).

These findings are congruent with the data available in diverse aquatic ecosystems, as it has been reported that bacteria can rely on photosynthetic phytoplankton to obtain organic matter to promote their growth (X. Zhang et al., 2023). Likewise, Phytoplankton rely on bacteria to

convert organic matter into inorganic nutrients. This would explain the interactions between *Spirosomaceae* with *Chlamydomonadales* and *Mediophyceae*; and *Sporichthyaceae* with *Sphaeropleales* as these microeukaryotes are photosynthetic (X. Zhang et al., 2023; Roth Rosenberg et al., 2021). Another study in agreement with these findings is research conducted by Bai et. al. (2023) in the Sanya River, China. Bai et al. (2023) reported Bacteroidetes, specifically *Flavobacteriaceae* appeared in the same sampling site as *Chlamydomonadales* and *Sphaeropleales.* In general, the co-occurrence of these microorganisms can be associated with their functional characteristics and their roles as decomposers and consumers in microbial food webs (Cruaud et al., 2020). *Chlamydomonadales* and *Mediophyceae* as primary producers (autotrophs) represent a source of organic matter for *Spirosomaceae* and *Oligotrichia* (heterotrophs), which break down organic compounds (Baek, Jang, Goh, & Choi, 2024). *Oligotrichia* has been documented to play a crucial role in transferring energy from low to high trophic levels (Feng et al., 2015). In the present study, *Oligotrichia* likely acts as a predator of *Spirosomaceae* and *Burkholderiaceae*. However, many factors could influence these ciliates prey preferences including nutrient availability, temperature, or absence of some microbial lineages (Cruaud et al., 2020; Terrado et al., 2017). Finally, specific co-occurrences of *Pseudomonadaceae* and the dinoflagellates *Thoracosphaeraceae* have not been documented. Because of the mixotrophic capability of some members of the *Thoracosphaeraceae* family, these microeukaryotes could be primary producers, prey, or predators in aquatic ecosystems (You Ji Hyun, 2023). In this context, *Pseudomonadaceae* may support *Thoracosphaeraceae* by contributing to nutrient availability.

In summary, these findings highlight the complex network of trophically interacting microorganisms in riverine ecosystems. These results support the literature stating that phytoplankton and primary producers are key players in the food webs of riverine ecosystems by using light and inorganic compounds to produce organic matter that is later consumed by heterotrophic bacteria (Zhang, Fang, Li, Lu, & Li, 2020). Bacteria then decompose these complexes, maintaining and being a source of nutrients for heterotrophic protists. In this context, bacterial communities might be controlled by their interaction with microeukaryotic predators (X. Zhang et al., 2023). Further research in microbial co-occurrence networks is needed to better

understand the relationship between domains, the strength of their relationships, and how this contributes to nutrient cycling which is essential for all forms of life in freshwater ecosystems.

5.0 CONCLUSION

Rivers are a receiving body of treated and untreated contaminants and a great reservoir of microorganisms. The microbial signatures present in water can reveal fundamental information about natural biogeochemical processes occurring in water and how the effects of anthropogenic activity could significantly impact the ecosystem's health. Moreover, monitoring freshwater ecosystems without focusing on single classes of microorganisms offers fundamental data to set a baseline for new indicators for aquatic contamination. This approach is crucial, as current methods overlook other pathogens, including viruses and protozoa.
In this study, we established a baseline of the microbial composition found in a rural section of the Assiniboine River in the province of Manitoba using deep amplicon sequencing of the g23 (for T4-like bacteriophages), 16S rRNA (for bacteria) and 18S rRNA for microbial eukaryotes.

The viral community consisted of cyanophages (62%) (including *Synechococcus* phages, and members of the *Bellamyvirus* genus), Coliphages (26%), *Emdodecavirus* genus phages (8%), and *Slopekvirus* genus phages (2%). The bacterial families across waterways included *Sporichthyaceae (20%), Pseudomonadaceae (19%), Enterobacteriaceae (3%), Burkholderiaceae (8%), Spirosomaceae (7%), Flavobacteriaceae (3%), Aeromonadaceae (3%), Micrococcaceae (3%) and Nostocaceae (3%).* The microbial eukaryotic community consisted of *Mediophyceae (24%), Cryptomonadales (14%), Choreotrichia (14%), Oligotrichia (3%), Sphaeropleales (7%), Chlamydomonadales (1%), Kappamyceteceae (5%),* and *Thoracosphaeraceae (3%).*

Although viral observed diversity recorded maximum values in the forested location during June 2022 and the effect of forested, urban, and agricultural activities suggest an impact on viral taxonomic composition in this research, clear spatiotemporal changes cannot be concluded. The majority (86.8%) of samples failed the deep amplicon sequencing for the g23 viral marker, therefore the effects of anthropogenic activities and time on the diversity and composition of t4-like bacteriophages were inconclusive.
The different activities (forested, urban, and agricultural) occurring around the interconnected section of the Assiniboine River sampled in this study had a significant impact on bacterial alpha diversity but not on bacterial taxonomic composition or beta diversity. This can be associated

with waterways being intertwined. Forested waterways had the highest bacterial alpha diversity, followed by urban and lastly agricultural locations. On the other hand, sampling months had a significant impact on both bacterial alpha and beta diversity. June, (the month with the highest precipitation observed) was also the month with the maximum bacterial alpha diversity on average. In the case of microeukaryotes, the different locations and sampling time points did not significantly impact their alpha diversity. However, forested, urban, and agricultural waterways as well as time may have impacted microbial eukaryotic taxonomic profiles significantly. Furthermore, bacteria exceeded eukaryotes in abundance by 4 orders of magnitude (3.40×10^7, 9.47×10^3, respectively).

Overall, the major environmental contributors to microbial dynamics were daylight length, water temperature, DO, and precipitation. *Emdodecavirus* genus and *Aeromonadaceae* were positively impacted by daylight length. On the other hand, *Spirosomaceae,* and all the microeukaryotes retrieved in this study *(Mediophyceae, Cryptomonadales, Choreotrichia, Oligotrichia, Sphaeropleales, Kappamyceteceae, Thoracosphaeraceae, and Chlamydomonadales)* were negatively impacted by daylight length. The taxa positively correlated with temperature in this study include: *Aeromonadaceae* and *Sporichthyaceae.* Conversely, *Flavobacteriaceae, Kappamyceteceae, Thoracosphaeraceae,* and *Chlamydomonadales* were negatively correlated with temperature. *Aeromonadaceae* and *Sporichthyaceae* were negatively correlated with DO whereas *Flavobacteriaceae* and *Kappamyceteceae* were positively correlated with this parameter. Finally, coliphages and *Pseudomonadaceae* were positively impacted by precipitation. On the contrary, the relative abundance of *Mediophyceae*, Cryptomonadales, *Choreotrichia, Oligotrichia,* and *Sphaeropleales* were negatively affected by precipitation. Furthermore, when evaluating microbial co-occurrence, *Spirosomaceae* demonstrated microbial links with *Chlamydomonadales, Mediophyceae,* and *Oligotrichia*. *Oligotrichia* correlated positively with *Burkholderiaceae*. Moreover, *Pseudomonaceae-Thoracosphaeraceae*; and *Sporichthyaceae -Sphaeropleales,* also demonstrated microbial links in the sampled waterways. Finally in this research, the relative abundance of *Sporichthyaceae,* correlated positively with standard colony counts of *E. coli* whereas *Aeromonadaceae* depicted a positive correlation with total coliforms. While correlations between *Aeromonadaceae* members and total coliforms are

not uncommon (Solaiman et al., 2020), further research is required to confirm the association between *Sporichthyaceae* with *E. coli.*

Future directions should include the continuous monitoring of freshwater ecosystems through culture-independent approaches including deep amplicon sequencing and shotgun metagenomics for the accurate detection of microorganisms at lower taxonomic levels (i.e. species). Although this method may be impractical for routine monitoring, it can aid with the identification of complementary microorganisms linked to water pollution, supporting the development of new primers for an effective freshwater surveillance assessment in the future. For such an outcome, longer time series studies (>2 years) may be required.

The fundamental data to be obtained through deep amplicon sequencing or metagenomics could help to properly discriminate between autochthonous freshwater microorganisms and pathogenic or algal bloom-causing microbes. In addition, it is important to find river-specific indicators based on the exclusive ecological and spatiotemporal dynamics of each aquatic ecosystem. This will enable the assessment of the potential generalization and reliability of the indicator taxa encountered. Furthermore, a comparative analysis of the microorganisms found in water with the ones present in epilithic (growing in stonelike material) riverine biofilms or sediment would help complement the consistency of the findings.

Through continuous monitoring and documentation of microbial fingerprints, we could promptly identify the presence of microbial threats to human and aquatic life, support the development of better methods for water monitoring, and help protect freshwater, one of Canada’s (and the world's) most precious natural resources.

Appendix

Sampling ID	Duplicate type	Month
F1	A	04
U1	A	04
U1	B	04
F3	B	05
U3	A	05
U3	B	05
A1	A	05
A3	A	05
A3	B	05
F1	B	06
U1	B	06
U2	A	06
U3	A	06
U3	B	06
A2	B	06
F2	B	07
U1	A	07
U2	A	07
U3	A	07
U3	B	07
F2	B	11

Table A.1: . Samples considered for viral analysis because of successful g23 deep amplicon sequencing. F= Forested waterway, U= Urban waterway, A= Agricultural waterway in rural areas of Manitoba, Canada. Months are represented as numbers 04= April, 05=May, 06=June, 07=July, 11= November. For a full description of locations sample, refer to Table 2.1

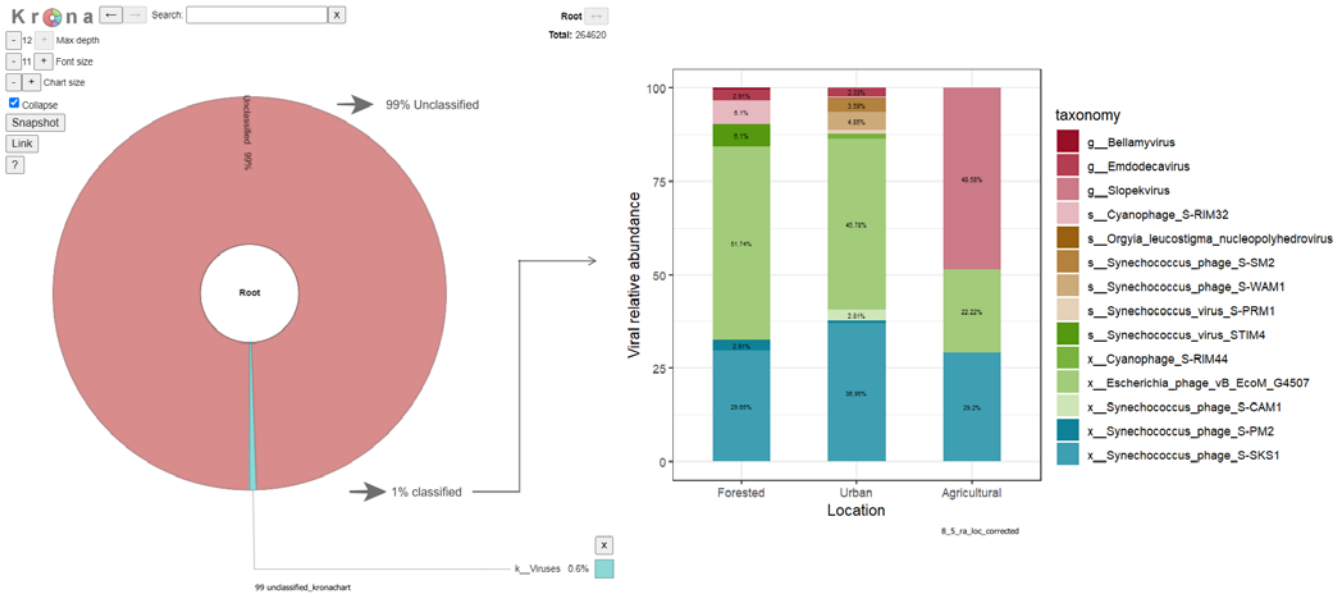

Figure A.1: Krona representation (pie chart) of the taxonomic classification rate (1%) obtained for the forward (R1) g23 sequencing reads through the Galaxy Europe platform. The arrow points out the relative abundance (stacked bar graph) of that 1% of viral reads obtained from samples that were successfully sequenced.

Sampling ID	Duplicate type	Month
F1	B	08
F2	A	09
F2	A	11
F2	B	08
F2	B	09
F2	B	10
F2	B	11
F3	A	09
F3	A	10
F3	B	09
F3	B	10
U1	A	10
U1	A	11
U1	B	10
U1	B	11
U2	A	09
U2	A	10
U2	A	11
U2	B	10
U2	B	11
U3	A	09
U3	A	10
U3	A	11
U3	B	06
U3	B	09
U3	B	10
A1	A	09
A1	A	10
A1	A	11
A1	B	10
A1	B	11
A2	A	09
A2	A	10
A2	A	11
A2	B	05
A2	B	10
A2	B	11
A3	A	08
A3	A	09
A3	A	10
A3	A	11
A3	B	10
A3	B	11

Table A.2: Samples included for microeukaryotes analysis due to successful 18S rRNA (V4) deep amplicon sequencing. F= Forested waterway, U= Urban waterway, A= Agricultural waterway in rural areas of Manitoba, Canada. Months are represented as numbers 05=May, 06=June, 08=August, 09= September, 10=October, 11= November. For a full description of locations sample, refer to Table 2.1

REFERENCES

Abell, G. C., & Bowman, J. P. (2005, 01). Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiology Ecology , 51* (2), 265-277. https://doi.org/10.1016/j.femsec.2004.09.001

Agatha, S., & Strder-Kypke, M. C. (2007). Phylogeny of the order Choreotrichida (ciliophora, spirotricha, oligotrichea) as inferred from morphology, ultrastructure, ontogenesis, and SSrRNA gene sequences. *European Journal of Protistology , 43* (1), 37-63. https://doi.org/10.1016/j.ejop.2006.10.001

Aizenberg-Gershtein, Y., Vaizel-Ohayon, D., & Halpern, M. (2012). Structure of bacterial communities in diverse freshwater habitats. *Canadian Journal of Microbiology , 58* (3), 326-335. Retrieved from https://doi.org/10.1139/w11-138 (PMID: 22339347) doi: 10.1139/w11-138

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology , 26* (1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Arajo, S., Silva, V., de Lurdes Enes Dapkevicius, M., Pereira, J. E., Martins, Igrejas, G., & Poeta, P. (2024). Comprehensive profiling of *Klebsiella* in surface waters from northern Portugal: Understanding patterns in prevalence, antibiotic resistance, and biofilm formation. *Water, 16* (9). Retrieved from https://www.mdpi.com/2073-4441/16/9/1297 doi: 10.3390/w16091297

Ayayee, P. A., Custer, G. F., Tronstad, L. M., & van Diepen, L. T. (2024). Unveiling salinity-driven shifts in microbial community composition across compartments of naturally saline inland streams. *Hydrobiologia, 851* (11), 2627–2639. https://doi.org/10.1007/s10750-024-05479-5

Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology, 5* (10), 782–791.https://doi.org/10.1038/nrmicro1747

Baek, K., Jang, S., Goh, J., & Choi, A. (2024). *Salmonirosea aquatica* gen. nov., sp. nov., a Novel Genus within the Family *Spirosomaceae*, Was Isolated from Brackish Water in the Republic of Korea. Microorganisms, 12 (8), 1671. https://doi.org/10.3390/microorganisms12081671

Bai, S., Zhang, J., Qi, X., Zeng, J., Wu, S., & Peng, X. (2023). Changes of In Situ Prokaryotic and Eukaryotic Communities in the Upper Sanya River to the Sea over a Nine-Hour Period. *Microorganisms*, *11*(2), 536. https://doi.org/10.3390/microorganisms11020536

Barbosa, M., Lefler, F. W., Berthold, D. E., Briggs-Gonzalez, V. S., Mazzotti, F. J., & Laughinghouse, H. D., 4th (2022). Trophic State Drives the Diversity of Protists in a Tropical River (New River, Belize). *Microorganisms*, *10*(12), 2425. https://doi.org/10.3390/microorganisms10122425.

Baron, S. (1996). *Medical Microbiology 4th edition*. University of Texas Medical Branch at Galveston. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK43081/

Bar-On, Y. M., & Milo, R. (2019). The biomass composition of the oceans: A blueprint of our blue planet. *Cell*, *179* (7), 1451-1454. Retrieved from https://www.sciencedirect.com/science/article/pii/S0092867419312747 doi: https://doi.org/10.1016/j.cell.2019.11.018

Becker, A., Neves, F., Angelis, D., & Navarrete, A. (2017). Differential occurrence of heterotrophic bacteria to specific physicochemical characteristics of oil refinery wastewater and adjacent water bodies. *Journal of Applied Biotechnology and Bioengineering* , *2* , 18–23. https://doi.org/10.15406/jabb.2017.02.00020

Bergh, ., Brsheim, K. Y., Bratbak, G., & Heldal, M. (1989, August). High abundance of viruses found in aquatic environments. *Nature* , *340* (6233), 467–468. Retrieved 2024-10-26, from https://www.nature.com/articles/340467a0 doi: 10.1038/340467a0

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30* (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonacolta, A. M., Visscher, P. T., Del Campo, J., & White III, R. A. (2024). The eukaryome of modern microbialites reveals distinct colonization across aquatic ecosystems. *npj Biofilms and Microbiomes* , *10* (1), 78. https://doi.org/10.1038/s41522-024-00547-z

Bowman, M. F., & Bailey, R. C. (1997). Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? *Canadian Journal of Fisheries and Aquatic Sciences* , *54* (8), 1802-1807. Retrieved from https://doi.org/10.1139/f97-085.

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

Brussaard, C. P. (2004). Viral control of phytoplankton populations a review 1. *Journal of Eukaryotic Microbiology* , *51* (2), 125–138. https://doi.org/10.1111/j.1550-7408.2004.tb00537.x

Bukin YuS, Galachyants YuP, Morozov IV, Bukin SV, Zakharenko AS, Zemskaya TI. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. Scientific Data 6:190007. doi: 10.1038/sdata.2019.7.

Cai, W., Li, Y., Wang, P., Niu, L., Zhang, W., & Wang, C. (2016). Effect of the pollution level on the functional bacterial groups aiming at degrading bisphenol A and nonylphenol in natural biofilms of an urban river. *Environmental Science and Pollution Research, 23* , 15727–15738. https://doi.org/10.1007/s11356-016-6757-3

Calabon, M., Hyde, K., Jones, E., Bao, D., Bhunjun, C., Phukhamsakda, C., . . . others (2023). Freshwater fungal biology. *Mycosphere* , *14* (1), 195–413. doi: 10.5943/mycosphere/14/1/4

Calgua, B., Mengewein, A., Grunert, A., Bofill-Mas, S., Clemente-Casares, P., Hundesa, A., Girones, R. (2008, November). Development and application of a one-step low-cost procedure to concentrate viruses from seawater samples. *Journal of Virological Methods* , *153* (2), 79–83. Retrieved 2024-05-24, from

https://linkinghub.elsevier.com/retrieve/pii/S0166093408003078 doi: 10.1016/j.jviromet.2008.08.003

Camacho, C. (2008). *Blast (r) command line applications user manual*. National Center for Biotechnology Information (US) Bethesda, MD.

Camargo, A. P., Nayfach, S., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., . . . others (2023). IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic acids research*, *51* (D1), D733–D743. https://doi.org/10.1093/nar/gkac1037

Canada, H. (2013). Guidelines for Canadian Drinking Water Quality: Guideline Technical Document. Retrieved from https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines canadian-dr.

Canada, H. (2019). *Guidelines for Canadian Drinking water quality. Enteric protozoa (*Tech. Rep.). Ottawa: Government of Canada. Retrieved from https://www.canada.ca/content/dam/hc-sc/documents/services/environmental-workplace-health/reports-

Canada, H. (2020). *Guidelines for Canadian Drinking Water Quality. Es-cherichia coli* (Tech. Rep.). Ottawa: Government of Canada. Retrieved from https://www.canada.ca/content/dam/hc-sc/documents/services/publications/healthy-living/guidelines-

Capo, E., Debroas, D., Arnaud, F., Perga, M.-E., Chardon, C., & Domaizon, I. (2017). Tracking a century of changes in microbial eukaryotic diversity in lakes driven by nutrient enrichment and climate warming. *Environmental microbiology* , *19* (7), 2873–2892. https://doi.org/10.1111/1462-2920.13815

Cauwenberghe, J. V., & Simms, E. L. (2023). How might bacteriophages shape biological invasions? *mBio*, *14* (5), e01886-23. Retrieved from

https://journals.asm.org/doi/abs/10.1128/mbio.01886-23. doi: 10.1128/mbio.01886-23

Chahal C, Van Den Akker B, Young F, Franco C, Blackbeard J, Monis P. (2016). Pathogen and Particle Associations in Wastewater. In: Advances in Applied Microbiology. Elsevier, 63–119. doi: 10.1016/bs.aambs.2016.08.001.

Chénard, C., Chan, A., Vincent, W., & Suttle, C. (2015). Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. *The ISME journal*, *9*(9), 2046–2058. https://doi.org/10.1038/ismej.2015.24

Chow, C. T., & Fuhrman, J. A. (2012). Seasonality and monthly dynamics of marine myovirus communities. *Environmental Microbiology*, *14* (8), 2171–2183. Retrieved 2024-05-24, from https://sfamjournals.onlinelibrary.wiley.com/doi/10.1111/j.1462-2920.2012.02744.x doi: 10.1111/j.1462-2920.2012.02744.x

Chu, B. T., Petrovich, M. L., Chaudhary, A., Wright, D., Murphy, B., Wells, G., & Poretsky, R. (2018). Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. *Applied and environmental microbiology*, *84* (5), e02168–17. https://doi.org/10.1128/AEM.02168-17

Cisneros-Martinez, A. M., Eguiarte, L. E., & Souza, V. (2023). Metagenomic comparisons reveal a highly diverse and unique viral community in a seasonally fluctuating hypersaline microbial mat. *Microbial Genomics, 9* (7), 001063. https://doi.org/10.1099/mgen.0.001063

City of Portage la Prairie. (2022). *Wastewater Treatment Division, Annual Report* (Annual). Manitoba, Canada: Water Pollution Control Facility. Retrieved from https://www.city-plap.com/cityplap/wp-content/uploads/2023/04/2022WPCFAnnualReport.pdf

Climate Data. (2022). *Climate data.* Retrieved from https://en.climate-data.org/north-america/canada/manitoba/win

Cobián Güemes, A. G., Youle, M., Cantú, V. A., Felts, B., Nulton, J., & Rohwer, F. (2016). Viruses as winners in the game of life. *Annual review of virology* , *3* (1), 197–214. https://doi.org/10.1146/annurev-virology-100114-054952

Coello-Camba, A., Diaz-Rua, R., Duarte, C. M., Irigoien, X., Pearman, J. K., Alam, I. S., & Agusti, S. (2020). Picocyanobacteria community and cyanophage infection responses to nutrient enrichment in a mesocosms experiment in oligotrophic waters. *Frontiers in Microbiology*, *11* , 1153. https://doi.org/10.3389/fmicb.2020.01153

Colab, G. (2024, June). *Google Colab.* Retrieved from https://colab.research.google.com/

Cruaud, P., Vigneron, A., Fradette, M.-S., Dorea, C. C., Culley, A. I., Rodriguez, M. J., & Charette, S. J. (2019). Annual protist community dynamics in a freshwater ecosystem undergoing contrasted climatic conditions: The Saint-Charles River (Canada). *Frontiers in Microbiology* , *10*, 2359. https://doi.org/10.3389/fmicb.2019.02359

Cruaud, P., Vigneron, A., Fradette, M.-S., Dorea, C. C., Culley, A. I., Rodriguez, M. J., & Charette, S. J. (2020). Annual bacterial community cycle in a seasonally ice-covered river reflects en- vironmental and climatic conditions. *Limnology and Oceanography* , *65* (S1), S21-S37. Retrieved from https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lno.11130 https://doi.org/10.1002/lno.11130

Cvetkovska, M., Huner, N. P., & Smith, D. R. (2017). Chilling out: the evolution and diversification of psychrophilic algae with a focus on Chlamydomonadales. *Polar Biology* , *40* , 1169–1184. https://doi.org/10.1007/s00300-016-2045-4

Dann, L. M., Rosales, S., McKerral, J., Paterson, J. S., Smith, R. J., Jeffries, T. C., . . . Mitchell, J. G. (2016). Marine and giant viruses as indicators of a

marine microbial community in a riverine system. *MicrobiologyOpen, 5* (6), 1071–1084. https://doi.org/10.1002/mbo3.392

Daraei, H., Oliveri Conti, G., Sahlabadi, F. et al. (2021). Prevalence of *Cryptosporidium spp.* in water: a global systematic review and meta-analysis. *Environ Sci Pollut Res* 28, 9498–9507. https://doi.org/10.1007/s11356-020-11261-6

Desforges, J. E., Clarke, J., Harmsen, E. J., Jardine, A. M., Robichaud, J. A., Serr, S., . . . Cooke, S. J. (2022). The alarming state of freshwater biodiversity in Canada. *Canadian Journal of Fisheries and Aquatic Sciences , 79* (2), 352-365. Retrieved from https://doi.org/10.1139/cjfas-2021-0073

Doménech E, Martorell S, Kombo-Mpindou GOM, Macián-Cervera J, Escuder-Bueno I. (2022). Risk assessment of Cryptosporidium intake in drinking water treatment plant by a combination of predictive models and event-tree and fault-tree techniques. *Science of The Total Environment* 838:156500. doi: 10.1016/j.scitotenv.2022.156500.

Duda, R. L. (2008). Icosahedral Tailed dsDNA Bacterial Viruses. In *Encyclopedia of Virology* (Third ed.).Pittsburgh, PA, USA: Academic Press. Retrieved from https://doi.org/10.1016/B978-012374410-4.00754-8.

Duran, A., Muniesa, M., Mendez, X., Valero, F., Lucena, F., & Jofre, J. (2002). Removal and inactivation of indicator bacteriophages in fresh waters. *Journal of applied microbiology*, *92* (2), 338–347. https://doi.org/10.1046/j.1365-2672.2002.01536.x

Duval, C., Thomazeau, S., Drelin, Y., Yprmian, C., Bouvy, M., Couloux, A., . . . Bernard, C. (2018). Phylogeny and salt-tolerance of freshwater nostocales strains: Contribution to their systematics and evolution. *Harmful Algae, 73* , 58-71. Retrieved from

https://www.sciencedirect.com/science/article/pii/S1568988318300143
https://doi.org/10.1016/j.hal.2018.01.008

Environment and Climate Change Canada. (2024). *Canadian Environmental Sustainability Indicators: Water quantity in Canadian rivers.* (Tech. Rep.). Canada: Government of Canada. Retrieved from https://www.canada.ca/content/dam/eccc/documents/pdf/cesindicators/water-quantity/2024/water-quant

Environment and Natural Resources. (2022). *Weather and climate, Historical data*. Retrieved from https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Feng, M., Zhang, W., Wang, W., Zhang, G., Xiao, T., & Xu, H. (2015). Can tintinnids be used for discriminating water quality status in marine ecosystems? *Marine pollution bulletin*, *101* (2), 549– 555. https://doi.org/10.1016/j.marpolbul.2015.10.059

Ferguson, D., & Signoretto, C. (2011). Environmental persistence and naturalization of fecal indicator organisms. In C. Hagedorn, A. R. Blanch, & V. J. Harwood (Eds.), *Microbial source tracking: Methods, applications, and case studies* (pp. 379–397). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4419-9386-117

Fiedler, C. J., Schönher, C., Proksch, P., Kerschbaumer, D. J., Mayr, E., Zunabovic-Pichler, M., . . . Perfler, R. (2018). Assessment of Microbial Community Dynamics in River Bank Filtrate Using High-Throughput Sequencing and Flow Cytometry. *Frontiers in Microbiology, 9* , 2887.https://doi.org/10.3389/fmicb.2018.02887

File, J., Ttart, F., Suttle, C. A., & Krisch, H. M. (2005, August). Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences , 102* (35), 12471–12476. Retrieved 2024-05-24, from https://pnas.org/doi/full/10.1073/pnas.0503404102

Fredriksson-Ahomaa, M., Grönthal, T., Heljanko, V., Johansson, V., Rantala, M., Heikinheimo, A., & Laukkanen-Ninios, R. (2024). Enteropathogenic *Yersinia* with Public Health Relevance Found in Dogs and Cats in Finland. *Pathogens* (*Basel, Switzerland*), 13(1), 54. https://doi.org/10.3390/pathogens13010054

Galaxy, C. (2024, 05). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research, 52* (W1), W83-W94. Retrieved from doi: 10.1093/nar/gkae410

Genomics, C. (2024, July). *Non-Metric Multidimensional Scaling (NMDS) in Microbial Sequencing Data Analysis: Introduction, Application, and Comparison* (Tech. Rep.). New York: Author. Retrieved from https://www.cd-genomics.com/microbioseq/non-metric-multidimensional-scaling-nmds-in-microbial-sequ

Ghaderiardakani, F., Coates, J. C., & Wichard, T. (2017). Bacteria-induced morphogenesis of *Ulva intestinalis* and *Ulva mutabilis* (Chlorophyta): a contribution to the lottery theory. *FEMS Microbiology Ecology* , *93* (8), fix094. https://doi.org/10.1093/femsec/fix094

Gleason, F. H., Chambouvet, A., Sullivan, B. K., Lilje, O., & Rowley, J. J. (2014). Multiple zoosporic parasites pose a significant threat to amphibian populations. *Fungal Ecology* , *11* , 181-192. doi: https://doi.org/10.1016/j.funeco.2014.04.001

Gonzales-Gustavson, E., Cardenas-Youngs, Y., Calvo, M., Da Silva, M. F. M., Hundesa, A., Amors, I., . . . Girones, R. (2017, March). Characterization of the efficiency and uncertainty of skimmed milk flocculation for the simultaneous concentration and quantification of water-borne viruses, bacteria and protozoa. *Journal of Microbiological Methods* , *134* , 46–53. Retrieved 2024-05-24, doi: 10.1016/j.mimet.2017.01.006

Gotelli, N. J., & Colwell, R. K. (2010). Estimating species richness. *Biological diversity: frontiers in measurement and assessment. A. E. Magurran and B. J. McGill, eds.*, 39–54.

Gomez-Pereira, P. R., Fuchs, B. M., Alonso, C., Oliver, M. J., van Beusekom, J. E. E., & Amann, R. (2010, 01). Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *The ISME Journal* , *4* (4), 472-487. Retrieved from https://doi.org/10.1038/ismej.2009.142

Health Canada. (2022). *Guidance on waterborne pathogens in drinking water* (Tech. Rep.). Ottawa: Government of Canada. Retrieved from https://www.canada.ca/content/dam/hc-sc/documents/services/environmental-workplace-health/reports-

Hennes, K. P., & Suttle, C. A. (1995). Direct counts of viruses in natural waters and laboratory cultures by epifluorescence microscopy. *Limnology and Oceanography, 40* (6), 1050–1055.

Hujare, M. S. (2008). Seasonal variations of phytoplankton in the freshwater tank of Talsande, Maharashtra. *Nature environment and pollution technology* , *7* (2), 253. Retrieved from https://neptjournal.com/upload-images/NL-18-12-12-AB-1132com.pdf

Illumina. (2024). *Measuring sequencing accuracy.* Retrieved from https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-sc

IMR. (2024). *Protocols IMR.* Retrieved 2024-01-05, from https://www.protocols.io/workspaces/integrated-microbiome

Jankowski, P., Gan, J., Le, T., McKennitt, M., Garcia, A., Yana, K., . . . Uyaguari-Diaz, M. (2022, December). Metagenomic community composition and resistome analysis in a full-scale cold climate wastewater treatment plant. *Environmental Microbiomes* , *17* (1). (Publisher: BioMed Central Ltd) doi: 10.1186/s40793-022-00398-1

Kacagan, M., Inan, K., Belduz, A. O., & Canakci, S. (2013). Flavobacterium anatoliense sp. nov., isolated from fresh water, and emended description of Flavobacterium ceti. *International journal of systematic and evolutionary microbiology* , *63* (Pt 6), 2075–2081.https://doi.org/10.1099/ijs.0.040394-0

Kaletta, J., Pickl, C., Griebler, C., Klingl, A., Kurmayer, R., & Deng, L. (2020). A rigorous assessment and comparison of enumeration methods for environmental viruses. *Scientific reports*, *10* (1), 18625. https://doi.org/10.1038/s41598-020-75490-y

Kang, M., Le, V. V., Ko, S.-R., Chun, S.-J., Choi, D.-Y., Shin, Y., . . . Ahn, C.-Y. (2024). Effect of rainfall in shaping microbial community during *Microcystis* bloom in Nakdong River, Korea. *Science of The Total Environment* , *928* , 172482. https://doi.org/10.1016/j.scitotenv.2024.172482

Kim Mincheol, C. J. (2014). Methods in Microbiology (Kim Mincheol & Chun Jongsik, Eds.; Vol. 41). https://www.sciencedirect.com/science/article/abs/pii/S0580951714000130

Kim M, Morrison M, Yu Z. (2011). Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *Journal of Microbiological Methods* 84:81–87. https://doi.org/ 10.1016/j.mimet.2010.10.020.

Kondo, K., Nakano, S., Hisatsune, J., Sugawara, Y., Kataoka, M., Kayama, S., Sugai, M., & Kawano, M. (2023). Characterization of 29 newly isolated bacteriophages as a potential therapeutic agent against IMP-6-producing *Klebsiella pneumoniae* from clinical specimens. *Microbiology spectrum*, *11*(5), e0476122. Advance online publication. https://doi.org/10.1128/spectrum.04761-22

Korf, I. H., Meier-Kolthoff, J. P., Adriaenssens, E. M., Kropinski, A. M., Nimtz, M., Rohde, M., . . . Wittmann, J. (2019). Still something to discover: novel insights into *Escherichia coli* phage diversity and taxonomy. *Viruses* , *11* (5), 454. doi: 10.3390/v11050454.

Kozlova, A. P., Muntyan, V. S., Vladimirova, M. E., Saksaganskaia, A. S., Kabilov, M. R., Gorbunova,M. K., . . . Roumiantseva, M. L. (2024). Soil giant phage: Genome and Biological Characteristics of *Sinorhizobium* Jumbo Phage. *International Journal of Molecular Sciences* , *25* (13), 7388 .https://doi.org/10.3390/ijms25137388.

Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., & Koonin, E. V. (2018). Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research*, *244*, 181-193. https://doi.org/10.1016/j.virusres.2017.11.025

Kusters, J. G., van Vliet, A. H., & Kuipers, E. J. (2006). Pathogenesis of Helicobacter pylori infection. *Clinical microbiology reviews*, *19* (3), 449–490. https://doi.org/10.1128/CMR.00054-05

Laplante, K., & Derome, N. (2011). Parallel changes in the taxonomical structure of bacterial communities exposed to a similar environmental disturbance. *Ecology and evolution*, *1* (4), 489–501. https://doi.org/10.1002/ece3.37

Lee ZM-P, Bussema C, Schmidt TM. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Research* 37:D489–D493. doi: 10.1093/nar/gkn689

Lemieux, C., Vincent, A. T., Labarre, A., Otis, C., & Turmel, M. (2015). Chloroplast phylogenomic analysis of chlorophyte green algae identifies a novel lineage sister to the Sphaeropleales (Chlorophyceae). *BMC Evolutionary Biology*, *15*, 1–13. https://doi.org/10.1186/s12862-015-0544-5 .

Levesque, B., & Gauvin, D. (2007). Microbiological guideline values for recreational bathing in Canada: time for change? *Canadian Journal of Infectious Diseases and Medical Microbiology*, 18 (2), 153–157. https://doi.org/10.1155/2007/180308

Li, Y., Huang, D., Sun, W., Sun, X., Yan, G., Gao, W., & Lin, H. (2022). Characterizing sediment bacterial community and identifying the biological indicators in a seawater-freshwater transition zone during the wet and dry seasons. *Environmental Science and Pollution Research*, *29* (27), 41219–41230. https://doi.org/10.1007/s11356-021-18053-6

Lin, H., & Peddada, S. D. (2020). Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ biofilms and microbiomes*, *6* (1), 60. https://doi.org/10.1038/s41522-020-00160-w

Lobb, B., Hodgson, R., Lynch, M. D. J., Mansfield, M. J., Cheng, J., Charles, T. C., . . . Doxey, A. C. (2020). Time series resolution of the fish necrobiome reveals a decomposer succession involving toxigenic bacterial pathogens. *mSystems* , *5* (2), https://doi.org/10.1128/mSystems.00145-20

Luethje, M., & Snyder, J. (2021). Climate-related morphological changes in *Pantocsekiella* (Mediophyceae) spanning 0-1.2 ma in the Lake El'gygytgyn, northeastern Russia including *Pantocsekiella elgygytgynensis* sp. nov. *Phytotaxa*, *478* (1), 67–91. https://doi.org/10.11646/phytotaxa.478.1.5

Macek, M., Callieri, A.,K. Šimek & AL Vázquez, & L. (2006, 07). Seasonal dynamics, composition, and feeding patterns of ciliate assemblages in oligotrophic lakes covering a wide pH range. Archiv für Hydrobiologie, 166 (2), 261-287. Retrieved from http://dx.doi.org/10.1127/0003?9136/2006/0166?026

Magurran, A. E. (2003). *Measuring biological diversity*. John Wiley & Sons. https://doi.org/10.1016/j.cub.2021.07.049.

Makk, J., Toumi, M., Krett, G., Lange-Enyedi, N. T., Schachner-Groehs, I., Kirschner, A. K., & Tóth, E. (2024). Temporal changes in the morphological and microbial diversity of biofilms on the surface of a submerged stone in the Danube River. *BIOLOGIA FUTURA*, 1–17. .https://doi.org/10.1007/s42977-024-00228-0

Mangiafico, S. S. (2015). An R companion for the handbook of biological statistics. *Available:* rcompanion. org/documents/RCompanionBioStatistics. pdf.(January 2016).

Mangiafico, S. S. (2016). *Summary and Analysis of Extension Program Evaluation in R, version 1.20.07, revised 2024.* New Brunswick, NJ.: rcompanion.org. Retrieved from (Pdf version: rcompanion.org/documents/RHandbookProgramEvaluation.pdf.)

Marciano‐Cabral F, Puffenbarger R, Cabral GA. (2000). The Increasing Importance of *Acanthamoeba* Infections[1]. *Journal of Eukaryotic Microbiology* 47:29–36. doi: 10.1111/j.1550-7408.2000.tb00007.x.

McCammon, S., Innes, B., Bowman, J., Franzmann, P., Dobson, S., Holloway, P., . . . Rankin, L. (1998). Flavobacterium hibernum sp. nov., a lactose-utilizing bacterium from a freshwater Antarctic lake. *International Journal of Systematic and Evolutionary Microbiology* , *48* (4), 1405–1412. https://doi.org/10.1099/00207713-48-4-1405

McCarthy, F. M., Gu, H., Mertens, K. N., Carbonell-Moore, C., Krueger, A. M., Takano, Y., &

Matsuoka, K. (2018). Transferring the freshwater dinoflagellate *Peridinium wisconsinense* (Dinophyceae) to the family Thoracosphaeraceae, with the description of *Fusiperidinium* gen. nov. *Phycological Research*, *66* (2), 137–148. https://doi.org/10.1111/pre.12215

McKindles, K. M., Jorge, A. N., McKay, R. M., Davis, T. W., & Bullerjahn, G. S. (2021). Isolation and characterization of Rhizophydiales (Chytridiomycota), obligate parasites of *Planktothrix agardhii* in a Laurentian Great Lakes embayment. *Applied and Environmental Microbiology* , *87* (4), e02308-20. doi: 10.1128/AEM.02308-20

McLellan, S. L., Huse, S. M., Mueller-Spitz, S. R., Andreishcheva, E. N., & Sogin, M. L. (2010). Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. *Environmental microbiology*, 12(2), 378–392. https://doi.org/10.1111/j.1462-2920.2009.02075.x

McMinn, B. R., Rhodes, E. R., Huff, E. M., & Korajkic, A. (2020). Decay of infectious adenovirus and coliphages in freshwater habitats is differentially affected by ambient sunlight and the presence of indigenous protozoa communities. *Virology journal* , *17* , 1–11. https://doi.org/10.1186/s12985-019-1274-x

M Comeau, A. (2022, November). Preparing multiplexed 16S/18S/ITS amplicons for the Illumina MiSeq v1. Retrieved 2024-07-30, from https://www.protocols.io/view/preparing-multiplexed-16s-18s-its-amplicons-for-th-ci2jugcn doi: 10.17504/protocols.io.4r3l277k3g1y/v1

Merino-Mascorro, J. A., Hernndez-Rangel, L. G., Heredia, N., & Garca, S. (2018). *Bacteroidales* as indicators and source trackers of fecal contamination in tomatoes and strawberries. *Journal of Food Protection*, *81* (9), 1439-1444. Retrieved from https://www.sciencedirect.com/science/article/pii/S0362028X22086914 doi: https://doi.org/10.4315/0362-028X.JFP-18-073

Michniewski, S., Redgwell, T., Grigonyte, A., Rihtman, B., Aguilo-Ferretjans, M., Christie-Oleza, J., . . . Millard, A. D. (2019). Riding the wave of genomics to investigate aquatic coliphage diversity and activity. *Environmental microbiology* , *21* (6), 2112–2128. https://doi.org/10.1111/1462-2920.14590

Millipore, E. (2022). *Chromocult Coliform Agar acc. ISO 9308-1.* Author. Retrieved from https://www.sigmaaldrich.com/CA/en/product/mm/110426#product-documentation

Mitbavkar, S., Rajaneesh, K., Anil, A., & Sundar, D. (2012). Picophytoplankton community in a tropical estuary: Detection of *Prochlorococcus*-like populations. *Estuarine, Coastal and Shelf Science, 107* , 159-164. https://doi.org/10.1016/j.ecss.2012.05.002.

Monk, W., & Baird, D. (2010). Biodiversity in Canadian Lakes and Rivers. *Canadian Biodiversity: Ecosystem Status and Trends* , 2018–2008. doi: 10.13140/RG.2.1.1576.7281

Munteanu, V., Gordeev, V., Saldana, M., Aßmann, E., Su, J. M., Drabcinski, N., . . . others (2023). A rigorous benchmarking of methods for SARS-CoV-2 lineage abundance estimation in wastewater. *arXiv preprint arXiv:2309.16994* https://doi.org/10.48550/arXiv.2309.16994.

Nhantumbo, C., Cangi Vaz, N., Rodrigues, M., Manuel, C., Rapulua, S., Langa, J., . . . Juzo, D. (2023). Assessment of microbial contamination in the Infulene River Basin, Mozambique. *Water* , *15* (2). Retrieved from https://www.mdpi.com/2073-4441/15/2/219 doi: 10.3390/w15020219

NIWA. (2024). *Chemical contamination.* Retrieved from https://niwa.co.nz/freshwater/kaitiaki-tools/what-impacts-interest-you/chemical-contamination

Numberger, D., Zoccarato, L., Woodhouse, J., Ganzert, L., Sauer, S., Marquez, J. R. G., . . . Greenwood, A. D. (2022). Urbanization promotes specific bacteria in freshwater microbiomes including potential pathogens. *Science of The Total Environment, 845* , 157321. Retrieved from https://www.sciencedirect.com/science/article/pii/S0048969722044199 https://doi.org/10.1016/j.scitotenv.2022.157321

OECD . (2013). Environment directorate joint meeting of the chemicals committee and the working party on chemicals, pesticides, and biotechnology.

Organization for Economic Cooperation and Development. .Retrieved from https://one.oecd.org/document/ENV/JM/A(2017)1/REV1/en/pdf

Podschun, R., Pietsch, S., Holler, C., & Ullmann, U. (2001). Incidence of *Klebsiella* species in surface waters and their expression of virulence factors. *Applied and Environmental Microbiology , 67* (7), 3325- 3327. Retrieved from https://journals.asm.org/doi/abs/10.1128/aem.67.7.3325-3327.2001 doi: 10.1128/AEM.67.7.3325-3327.2001.

Pomeroy, P. J. . A. . F., Lawrence R Williams, & Hobbie, J. E. (2007). The microbial loop. *Oceanography ,20* (2), 28–33. Retrieved from https://tos.org/oceanography/assets/docs/20-2_pomeroy.pdf

Potapov, S., Belykh, O., Krasnopeev, A., Gladkikh, A., Kabilov, M., Tupikin, A., & Butina, T. (2018). Assessing the diversity of the g23 gene of T4-like bacteriophages from Lake Baikal with high-throughput sequencing. *FEMS microbiology letters, 365* (3), fnx264. https://doi.org/10.1093/femsle/fnx264

Powelson, D. K., Gerba, C. P., & Yahya, M. T. (1993). Virus transport and removal in wastewater during aquifer recharge. *Water Research, 27* (4), 583-590. https://doi.org/10.1016/0043-1354(93)90167-G

Puskas, L. G., & Bottka, S. (1995). Reduction of mispriming in amplification reactions with restricted PCR. *Genome research, 5* (3), 309–311.https://doi.org/10.1101/gr.5.3.309

Qiagen. (2024). *Identify Mispriming Events.* Retrieved 2024-07-23, from https://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsanalysis/current/index.php

R, D. (2024, June). betadisper: Multivariate homogeneity of groups dispersions (variances). Retrieved 2024-06-18, from https://www.rdocumentation.org/packages/vegan/versions/1.11-0/topics/betadisper

Rakshit, D., Ganesh, P. S., & Sarkar, S. K. (2016). Choreotrich ciliate tintinnid (protozoa: Ciliophora) in a tropical meso-macrotidal estuary, eastern part of India. *Regional Studies in Marine Science, 3,* 89-100. Retrieved from

https://www.sciencedirect.com/science/article/pii/S2352485515000237 doi: https://doi.org/10.1016/j.rsma.2015.06.003

R-Documentation. (2024). Permutational Multivariate Analysis of Variance Using Distance Matrices. Adonis. Retrieved 2024-07-12, from https://search.r-project.org/CRAN/refmans/vegan/html/adonis.html

Reyes, A. (2020). Detailed study of Principal Component Analysis. In *A Machine Learning Compilation.* bookdown. Retrieved from https://f0nzie.github.io/machinelearningcompilation/detailed − study − of − principal − component − analysis.htmlcolor − by − groups – 1

Reyes, V., & Jiang, S. (2010). Ecology of coliphages in southern California coastal waters. *Journal of applied microbiology* , *109* (2), 431–440. https://doi.org/10.1111/j.1365-2672.2010.04676.x

Reza, M. S., Mizusawa, N., Kumano, A., Oikawa, C., Ouchi, D., Kobiyama, A., . . . others (2018). Metagenomic analysis using 16S Ribosomal RNA genes of a bacterial community in an urban stream, the Tama River, Tokyo. *Fisheries science*, *84* , 563–577. doi: 10.1007/s12562-018-1193-6 .

Ribeiro, S., & Amorim, A. (2008). Environmental drivers of temporal succession in recent dinoflagellate cyst assemblages from a coastal site in the North-East Atlantic (Lisbon Bay, Portugal). *Marine Micropaleontology* , *68* (1), 156-178. https://doi.org/10.1016/j.marmicro.2008.01.013

Rimet, F., & Bouchez, A. (2012). Biomonitoring river diatoms: Implications of taxonomic resolution. *Ecological Indicators* , *15* (1), 92-99. Retrieved from https://www.sciencedirect.com/science/article/pii/S1470160X11003013. https://doi.org/10.1016/j.ecolind.2011.09.014

Ritalahti, K. M., Amos, B. K., Sung, Y., Wu, Q., Koenigsberg, S. S., & Lffler, F. E. (2006). Quantitative PCR Targeting 16s rRNA and Reductive Dehalogenase Genes Simultaneously Monitors Multiple *Dehalococcoides* Strains *Applied and Environmental Microbiology*, *72* (4), 2765-2774. Retrieved from

https://journals.asm.org/doi/abs/10.1128/aem.72.4.2765-2774.2006 doi: 10.1128/AEM.72.4.2765-2774.2006

Rose, J. B., Daeschner, S., Easterling, D. R., Curriero, F. C., Lele, S., & Patz, J. A. (2000). Climate and waterborne disease outbreaks. *Journal-American Water Works Association, 92* (9), 77–87. https://doi.org/10.1002/j.1551-8833.2000.tb09006.x

Rosenberg, E., Bittan-Banin, G., Sharon, G., Shon, A., Hershko, G., Levy, I., & Ron, E. Z. (2010). The phage-driven microbial loop in petroleum bioremediation. *Microbial biotechnology , 3* (4), 467–472. https://doi.org/10.1111/j.1751-7915.2010.00182.x

Roth Rosenberg, D., Haber, M., Goldford, J., Lalzar, M., Aharonovich, D., Al-Ashhab, A., . . . Sher, D. (2021). Particle-associated and free-living bacterial communities in an oligotrophic sea are affected by different environmental factors. *Environmental Microbiology , 23* (8), 4295–4308. https://doi.org/10.1111/1462-2920.15611

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A., & Sullivan, M. B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ , 5 ,* e3817.https://doi.org/10.7717/peerj.3817

Ruiz, N., Vogel, J., & Taghvaeian, S. (2017). Minimizing stormwater runoff by disconnecting residential downspouts. *Oklahoma Cooperative Extension Fact Sheets.* Retrieved from https://extension.okstate.edu/fact-sheets/minimizing-stormwater-runoff-by-disconnecting-residential-downspouts.html.

Sagova-Mareckova, M., Boenigk, J., Bouchez, A., Cermakova, K., Chonova, T., Cordier, T., . . . Stoeck, T. (2021). Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. *Water Research, 191 ,* 116767. https://doi.org/10.1016/j.watres.2020.116767

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., . . . Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology , 12 ,* 1–12. https://doi.org/10.1186/s12915-014-0087-z

Salvadori, M. I., Sontrop, J. M., Garg, A. X., Moist, L. M., Suri, R. S., & Clark, W. F. (2009). Factors that led to the Walkerton tragedy. *Kidney International , 75 ,* S33-S34.Retrieved from

https://www.sciencedirect.com/science/article/pii/S0085253815536120 (HUS/TTP Update 2008 Symposium). https://doi.org/10.1038/ki.2008.616

Sandaa, R.-A., & Larsen, A. (2006). Seasonal variations in virus-host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine Synechococcus spp. *Applied and environmental microbiology* , *72* (7), 4610–4618. https://doi.org/10.1128/AEM.00168-06

Santoferrara, L. F., Alder, V. V., & McManus, G. B. (2017). Phylogeny, classification and diversity of Choreotrichia and Oligotrichia (Ciliophora, Spirotrichea). *Molecular Phylogenetics and Evolution*, *112*, 12-22. Retrieved from https://www.sciencedirect.com/science/article/pii/S1055790316303888, https://doi.org/10.1016/j.ympev.2017.03.010

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors–occurrence, properties, and removal. *Journal of applied microbiology* , *113* (5), 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., . . . others (2018). Single- cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nature Communications, 9* (1), 310.https://doi.org/10.1038/s41467-017-02235-3

Shalchian-Tabrizi, K., Br˚ate, J., Logares, R., Klaveness, D., Berney, C., & Jakobsen, K. S. (2008). Diversification of unicellular eukaryotes: cryptomonad colonizations of marine and fresh waters inferred from revised 18s rRNA phylogeny. *Environmental Microbiology*, *10* (10), 2635–2644. https://doi.org/10.1111/j.1462-2920.2008.01685.x

Shao, K., Yao, X., Wu, Z., Jiang, X., Hu, Y., Tang, X., . . . Gao, G. (2021). The bacterial community composition and its environmental drivers in the rivers around eutrophic Chaohu Lake, China. *BMC microbiology* , *21* (1), 179. https://doi.org/10.1186/s12866-021-02252-9

Sharafutdinov, I., Tegtmeyer, N., Rohde, M., Olofsson, A., Rehman, Z. U., Arnqvist, A., & Backert, S. (2024). *Campylobacter jejuni* Surface-Bound Protease HtrA, but Not the Secreted Protease nor Protease in Shed Membrane Vesicles, Disrupts Epithelial Cell-to-Cell Junctions. *Cells*, *13*(3), 224. https://doi.org/10.3390/cells13030224

Sheath, R. G., & Steinman, A. D. (1982). A checklist of freshwater algae of the Northwest Territories, Canada.
Canadian Journal of Botany, *60* (10), 1964–1997. https://doi.org/10.1139/b82-245.

Shiah, F.-K., Lai, C.-C., Chen, T.-Y., Ko, C.-Y., Tai, J.-H., & Chang, C.-W. (2022). Viral shunt in tropical oligotrophic ocean. *Science Advances* , *8* (41), eabo2829. doi:10.1126/sciadv.abo2829

Sivasankar, R., Ezhilarasan, P., Sathish Kumar, P., Naidu, S., Rao, G., Kanuri, V. V., . . . Ramu, K. (2018). Loricate ciliates as an indicator of eutrophication status in the estuarine and coastal waters. *Marine Pollution Bulletin*, *129* (1), 207-211. https://doi.org/10.1016/j.marpolbul.2018.02.027

Skraber, S., Gassilloud, B., & Gantzer, C. (2004). Comparison of coliforms and coliphages as tools for assessment of viral contamination in river water. *Applied and Environmental Microbiology* , *70* (6), 3644- 3649. Retrieved from https://journals.asm.org/doi/abs/10.1128/aem.70.6.3644-3649.2004 doi: 10.1128/AEM.70.6.3644-3649.2004

Smith-Zaitlik, T., Shibu, P., McCartney, A. L., Foster, G., Hoyles, L., & Negus, D. (2022). Extended genomic analyses of the broad-host-range phages vB_KmiM-2Di and vB_KmiM-4Dii reveal slopekviruses have highly conserved genomes. *Microbiology (Reading, England)*, *168*(9), 10.1099/mic.0.001247. https://doi.org/10.1099/mic.0.001247

Solaiman, S., Allard, S. M., Callahan, M. T., Jiang, C., Handy, E., East, C., . . . others (2020). Longitudinal assessment of the dynamics of *Escherichia coli*, total coliforms, *Enterococcus spp*., and *Aeromonas spp*. in alternative irrigation water sources: a CONSERVE study. *Applied and environmental microbiology* , *86* (20), e00342–20. https://doi.org/10.1128/AEM.00342-20

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., & Wommack, K. E. (2008). Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Research in Microbiology* , *159* (5), 349-357. Retrieved from https://www.sciencedirect.com/science/article/pii/S0923250808000612 (Exploring the prokaryotic virosphere) doi: https://doi.org/10.1016/j.resmic.2008.04.010

St-Laurent, P., Straneo, F., Dumais, J.-F., & Barber, D. (2011). What is the fate of the river waters of Hudson Bay? *Journal of Marine Systems* , *88* (3), 352-361. https://doi.org/10.1016/j.jmarsys.2011.02.004

Stoermer, E. F., & Smol, J. P. (Eds.). (1999). *The diatoms: applications for the environmental and earth sciences*. Cambridge, UK ; New York, NY, USA: Cambridge University Press.

Sunyoung Kwon, Seunghyun Park, Byunghan Lee, Sungroh Yoon. (2013). In-depth analysis of interrelation between quality scores and real errors in Illumina reads. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Osaka: IEEE, 635–638. doi: 10.1109/EMBC.2013.6609580.

Sutlovic, D., Gojanovic, D., Andelinovic, v., Gugic, D., & Primorac, D. (2005). Taq polymerase reverses inhibition of quantitative real time polymerase chain reaction by humic acid. *Croatian medical journal.* Retrieved from https://pubmed.ncbi.nlm.nih.gov/16100758/

Suzuki, S., Yamaguchi, H., Nakajima, N., & Kawachi, M. (2018). Raphidocelis subcapitata (=Pseudokirchneriella subcapitata) provides an insight into genome evolution and environmental adaptations in the Sphaeropleales. *Scientific reports*, *8*(1), 8058. https://doi.org/10.1038/s41598-018-26331-6

Taguchi, Y.-H., & Oono, Y. (2005). Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics*, *21* (6), 730–740. https://doi.org/10.1093/bioinformatics/bti067

Tanabe, A. S., Nagai, S., Hida, K., Yasuike, M., Fujiwara, A., Nakamura, Y., . . . Katakura, S. (2016). Comparative study of the validity of three regions of the 18S-rRNA gene for massively parallel sequencing- based monitoring of the planktonic eukaryote community. *Molecular ecology resources* , *16* (2), 402–414. https://doi.org/10.1111/1755-0998.12459

Tang, Y. Z., & Gobler, C. J. (2012). The toxic dinoflagellate Cochlodinium polykrikoides (Dinophyceae) produces resting cysts. *Harmful Algae*, *20* , 71–80. doi 10.1016/j.hal.2012.08.001

Tao, S., Chen, H., Li, N., Wang, T., & Liang, W. (2022). The spread of antibiotic resistance genes in vivo model. *Canadian Journal of Infectious Diseases and Medical Microbiology* , *2022* (1), 3348695. doi: 10.1155/2022/3348695

Terrado, R., Pasulka, A. L., Lie, A. A.-Y., Orphan, V. J., Heidelberg, K. B., & Caron, D. A. (2017, 05). Autotrophic and heterotrophic acquisition of carbon and nitrogen by a mixotrophic chrysophyte established through stable isotope analysis. *The ISME Journal* , *11* (9), 2022-2034. https://doi.org/10.1038/ismej.2017.68

Thumbi, D. K., Eveleigh, R. J., Lucarotti, C. J., Lapointe, R., Graham, R. I., Pavlik, L., . . . Arif, B. M. (2011). Complete sequence, analysis and organization of the Orgyia leucostigma nucleopolyhedrovirus genome. *Viruses* , *3* (11), 2301–2327. https://doi.org/10.3390/v3112301

Time and Date. (2022). *Sunrise, Sunset, and Daylength* [Private]. Retrieved from https://www.timeanddate.com/sun/canada/winnipeg

Townsend, E. M., Kelly, L., Gannon, L., Muscatt, G., Dunstan, R., Michniewski, S., . . . Jameson, E. (2021). Isolation and characterization of *Klebsiella* phages for phage therapy. *PHAGE* , *2* (1), 26-42. https://doi.org/10.1089/phage.2020.0046 (PMID: 33796863)

Ud-Din, A., & Wahid, S. (2014). Relationship among *Shigella spp*. and enteroinvasive Escherichia coli (EIEC) and their differentiation. *Brazilian Journal of Microbiology*, *45* , 1131–1138. https://doi.org/10.1590/s1517-83822014000400002

Ung, P., Peng, C., Yuk, S., Tan, R., Ann, V., Miyanaga, K., & Tanji, Y. (2019). Dynamics of bacterial community in Tonle Sap Lake, a large tropical flood-pulse system in Southeast Asia. *Science of the Total Environment*, *664*, 414–423. https://doi.org/10.1016/j.scitotenv.2019.01.351

Uyaguari-Diaz, M. I., Chan, M., Chaban, B. L., Croxen, M. A., Finke, J. F., Hill, J. E., . . . Tang, P. (2016, December). A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples. *Microbiome*, *4* (1), 20. doi: 10.1186/s40168-016-0166-1

Van den Wyngaert, S., Ganzert, L., Seto, K., Rojas-Jimenez, K., Agha, R., Berger, S. A., . . . others (2022). Seasonality of parasitic and saprotrophic zoosporic fungi: linking sequence data to ecological traits. *The ISME Journal* , *16* (9), 2242–2254. https://doi.org/10.1038/s41396-022-01267-y

vanHees, D., Hanneman, C., Paradis, S., Camara, A. G., Matsumoto, M., Hamilton, T., . . . Kodner, R. B. (2023, 09). Patchy and Pink: Dynamics of a Chlainomonas sp. (Chlamydomonadales, chlorophyta) algal bloom on Bagley Lake, North Cascades, WA. *FEMS Microbiology Ecology* , *99* (11), fiad106. https://doi.org/10.1093/femsec/fiad106

Vaz-Moreira, I., Nunes, O. C., & Manaia, C. M. (2014, 07). Bacterial diversity and antibiotic resistance in water habitats: searching the links with the human microbiome. *FEMS Microbiology Reviews* , *38* (4), 761-778. https://doi.org/10.1111/1574-6976.12062

Walter, J. M., Lopes, F. A., Lopes-Ferreira, M., Vidal, L. M., Leomil, L., Melo, F., . . . others (2018). Occurrence of Harmful Cyanobacteria in Drinking Water from a Severely Drought-Impacted Semi-arid Region. *Frontiers in Microbiology* , *9* , 176.doi: 10.3389/fmicb.2018.00176. PMID: 29541063; PMCID: PMC5835534.

Wang, K., Wommack, K. E., & Chen, F. (2011). Abundance and Distribution of *Synechococcus* spp. and Cyanophages in the Chesapeake Bay. *Applied and environmental microbiology* , *77* (21), 7459–7468. doi: 10.1128/AEM.00267-11. Epub 2011 Aug 5. PMID: 21821760; PMCID: PMC3209163.

Wang, Y., Chen, Y., Wang, J., Liu, F., & Chen, N. (2021). Mitochondrial genome of the harmful algal bloom species *Odontella regia* (Mediophyceae, Bacillariophyta). *Journal of Applied Phycology* , *33* , 855–868. https://doi.org/10.1007/s10811-020-02364-1

Weber-Dabrowska, B., Zaczek, M., obocka, M., usiak Szelachowska, M., Owczarek, B., Orwat, F.,Gorski, A. (2023). Characteristics of environmental *Klebsiella pneumoniae* and *Klebsiella oxytoca* bacteriophages and their therapeutic applications. *Pharmaceutics* . Retrieved from https://www.mdpi.com/1999-4923/15/2/434 doi: 10.3390/pharmaceutics15020434

Wei, F., Xia, H., Huang, K., & Wei, C. (2024). Exogenous mobile genetic elements and their

associated integrons drive the enrichment of antibiotic-resistant genes in the river of a valley basin city (Lanzhou, China). *Environmental science and pollution research international*, *31*(2), 3195–3206. https://doi.org/10.1007/s11356-023-31269-y

Weiss, A., Costa, R., & Wichard, T. (2017). Morphogenesis of *Ulva mutabilis* (Chlorophyta) induced by *Maribacter* species (Bacteroidetes, Flavobacteriaceae). *Botanica Marina*, *60* (2), 197–206. https://doi.org/10.1515/bot-2016-0083.

Wexler, P. (Ed.). (2024). *Encyclopedia of toxicology* (Fourth edition ed.). Amsterdam, Netherlands: Elsevier. WHO. (2008). *Guidelines for drinking-water quality: second addendum. vol. 1, recommendations*. World Health Organization.

Whon, T. W., Kim, M.-S., Roh, S. W., Shin, N.-R., Lee, H.-W., & Bae, J.-W. (2012). Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. *Journal of Virology* , *86* (15), 8221-8231. https://journals.asm.org/doi/abs/10.1128/jvi.00293-12.

Wilhelm, S. W., Carberry, M. J., Eldridge, M. L., Poorvin, L., Saxton, M. A., & Doblin, M. A. (2006). Marine and Freshwater Cyanophages in a Laurentian Great Lake: Evidence from Infectivity Assays and Molecular Analyses of g20 Genes. *Applied and environmental microbiology* , *72* (7), 4957–4963. doi: 10.1128/AEM.00349-06. PMID: 16820493; PMCID: PMC1489316.

Wingender, J., & Flemming, H. C. (2011). Biofilms in drinking water and their role as reservoir for pathogens. *International Journal of Hygiene and Environmental Health*, *214* (6), 417–423. http://dx.doi.org/10.1016/j.ijheh.2011.05.009.

Wisniewska, H., Niewolak, S., Korzeniewska, E., & Filipkowska, Z. (2007, June). *Enterobacteriaceae* Family Bacteria In A Mesotrophic Lake (Lake Dugie Wigierskie) in the Presence of Black Cormorants (*Phalocrocorax Carbo*). *Polish Journal of Natural Science* , *22* (3), 486–499. Retrieved 2024-09-25, from http://versita.metapress.com/openurl.asp?genre=articleid=doi:10.2478/v10020-007-0043-2 doi: 10.2478/v10020-007-0043-2.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome biology* , *20* , 1–13. https://doi.org/10.1186/s13059-019-1891-0.

Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., & Keeling, P. J. (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science , 347* (6223), 1257594. doi:10.1126/science.1257594.

Wyllie, R., Hyams, J. S., & Kay, M. (2020). *Pediatric gastrointestinal and liver disease e-book*. Elsevier Health Sciences.

Xiang, Y., Jiang, L., Zhou, Y., Luo, Z., Zhi, D., Yang, J., & Lam, S. S. (2022). Microplastics and environmental pollutants: Key interaction and toxicology in aquatic and soil environments. *Journal of Hazardous Materials, 422* , 126843. https://doi.org/10.1016/j.jhazmat.2021.126843

Xu, C., Lu, S., Cidan, Y., Wang, H., Sun, G., Saleem, M. U., . . . Li, K. (2024). Microbiome analysis reveals alteration in water microbial communities due to livestock activities. *Environmental Science and Pollution Research, 31* (34), 47298–47314. https://doi.org/10.1007/s11356-024-34334-2.

Yanaç, K., Francis, J., Zambrano-Alvarado, J., Yuan, Q., & Uyaguari-Díaz, M. (2023). Concentration of virus particles from environmental water and wastewater samples using skimmed milk flocculation and ultrafiltration. *JoVE (Journal of Visualized Experiments)*(193), e65058.https://doi.org/10.3791/65058

Yang, Y., Li, S., Gao, Y., Chen, Y., & Zhan, A. (2019). Environment-driven geographical distribution of bacterial communities and identification of indicator taxa in Songhua River. *Ecological Indicators, 101* , 62-70. https://doi.org/10.1016/j.ecolind.2018.12.047.

You Ji Hyun, K. H. C. P. S. A. E. S. H. J. H. J., Ok Jin Hee. (2023). Five phototrophic *Scrippsiella* species lacking mixotrophic ability and the extended prey spectrum of *Scrippsiella* acuminata (Thoracosphaerales, Dinophyceae). *Algae, 38* (2), 111-126. Retrieved from http://www.e-algae.org/journal/view.php?number=2996 doi: 10.4490/algae.2023.38.6.6

Zambrano-Alvarado, J. I., & Uyaguari-Diaz, M. I. (2024). Insights into water insecurity in Indigenous communities in Canada: assessing microbial risks and innovative solutions, a multifaceted review. *PeerJ , 12* , e18277. https://doi.org/10.7717/peerj.18277

Zhang, D., He, Y., & Gin, K. Y.-H. (2021). Novel Freshwater Cyanophages Provide New Insights into Evolutionary Relationships between Freshwater and Marine Cyanophages. *Microbiology Spectrum, 9* (2), e00593–21.https://doi.org/10.1128/Spectrum.00593-21.

Zhang, L., Fang, W., Li, X., Lu, W., & Li, J. (2020). Strong linkages between dissolved organic matter and the aquatic bacterial community in an urban river. *Water Research, 184* , 116089. https://doi.org/10.1016/j.watres.2020.116089.

Zhang, X., Cui, L., Liu, S., Li, J., Wu, Y., Ren, Y., & Huang, X. (2023). Seasonal dynamics of bacterial community and co-occurrence with eukaryotic phytoplankton in the Pearl River Estuary. *Marine Environmental Research, 192* , 106193. https://doi.org/10.1016/j.marenvres.2023.106193

Zhu, X., Liu, J., Li, L., Zhen, G., Lu, X., Zhang, J., . . . Zhang, X. (2023). Prospects for humic acids treatment and recovery in wastewater: A review. *Chemosphere* , *312* , 137193. https://doi.org/10.1016/j.chemosphere.2022.137193

Zubkov, M. V., Sleigh, M. A., Burkill, P. H., & Leakey, R. J. (2000). Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. *Progress in oceanography , 45* (3-4), 369–386. https://doi.org/10.1016/S0079-6611(00)00008-2.

Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., & Han, S.-K. (2002). Typical freshwater bacteria: an analysis of available 16s rRNA gene sequences from plankton of lakes and rivers. *Aquatic microbial ecology , 28* (2), 141–155. https://doi.org/10.3354/ame028141