DEVELOPMENT AND EVALUATION OF ADVANCED MOLECULAR STRATEGIES FOR QUALITY MONITORING AND SUSTAINABILITY OF RECREATIONAL WATERS

Development and Evaluation of Advanced Molecular Strategies for Quality Monitoring and Sustainability of Recreational Waters

By FAIZAN SALEEM, B.Sc., M.Sc., MPhil.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

Descriptive Note

McMaster University DOCTOR OF PHILOSOPHY (2024) Hamilton, Ontario (Biology)

TITLE: Development and Evaluation of Advanced Molecular Strategies for Quality Monitoring and Sustainability of Recreational Waters

AUTHOR: Faizan Saleem, B.Sc, M.Sc (University of Karachi), MPhil (International Center for Chemical and Biological Sciences, H.E.J Research Institute, University of Karachi).

SUPERVISOR: Professor Dr. Herb E. Schellhorn

NUMBER OF PAGES: xv, Main Text 212, Appendix 175

List of Publications

1. **Saleem, F**., Atrache, R., Jiang, J.L., Tran, K.L., Li, E., Paschos, A., Edge, T.A. and Schellhorn, H.E., 2024. Characterization of Taxonomic and Functional Dynamics Associated with Harmful Algal Bloom Formation in Recreational Water Ecosystems. Toxins, 16(6), p.263.
2. **Saleem, F**., Schellhorn, H. E., Simhon, A., & Edge, T. A. (2023). Same-day Enterococcus qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. Canadian Journal of Public Health. 114: 676–687.
3. **Saleem, F**., Li, E., Tran, K.L., Rudra, B., Edge, T.A., Schellhorn, H.E. and Gupta, R.S., 2024. Utilizing novel Escherichia coli-specific conserved signature proteins for enhanced monitoring of recreational water quality. MicrobiologyOpen, 13(3),
4. **Saleem, F**., Li, E., Edge, T. A., Tran, K. L., & Schellhorn, H. E. (2024). Identification of potential microbial risk factors associated with fecal indicator exceedances at recreational beaches. Environmental Microbiome, 19(1), 4.
5. **Saleem, F**., Edge, T.A. and Schellhorn, H.E. (2023). Application of Same-Day Enterococcus qPCR-Based Analyses for Quality Assessment of Shorelines (Water and Sand) at Recreational Beaches. Water. 15(38): 2338.
6. **Saleem F,** Jiang JL, Atrache R, Paschos A, Edge TA, Schellhorn HE. (2023). Cyanobacterial Algal Bloom Monitoring: Molecular Methods and Technologies for Freshwater Ecosystems. Microorganisms. 11(4): 851.

7. **Saleem, F**., Edge, T.A. and Schellhorn, H.E. (2022). Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring. Journal of Great Lakes Research. 48(3): 707-716.
8. Li, E., **Saleem, F**., Edge, T.A. and Schellhorn, H.E., 2024. Assessment of crAssphage as a human fecal source tracking marker in the lower Great Lakes. Science of The Total Environment, 912, p.168840.
9. de Vries, J., **Saleem, F**., Li, E., Chan, A.W.Y., Naphtali, J., Naphtali, P., Paschos, A. and Schellhorn, H.E. (2023). Comparative Analysis of Metagenomic (Amplicon and Shotgun) DNA Sequencing to Characterize Microbial Communities in Household On-Site Wastewater Treatment Systems. Water. 15(2): 271.
10. Naphtali, J., Chan, A.W., **Saleem, F**., Li, E., Devries, J. and Schellhorn, H.E. (2022). Comparative Metagenomics of Anaerobic Digester Communities Reveals Sulfidogenic and Methanogenic Microbial Subgroups in Conventional and Plug Flow Residential Septic Tank Systems. Processes. 10(3): 436. Co-Author
11. Li, E., **Saleem, F**., Edge, T.A. and Schellhorn, H.E. (2021). Biological indicators for fecal pollution detection and source tracking: A review. Processes. 9(11): 2058.

Lay Abstract

Traditional culturing-based enumeration methods for qualitative monitoring of freshwater beaches provide results 24-48 h after sample collection, and the resulting information obtained is limited to fecal indicator counts. Delays in acquiring fecal indicator counts can lead to error-prone beach postings and unnecessary beach closures/openings. Modern DNA-based molecular methods such as qPCR and DNA sequencing can serve as rapid and more robust alternatives for routine quality assessment and provide additional, comprehensive information that can be useful for fecal source tracking and improvisation of measures for beach preservation. I evaluated the application of DNA-based molecular methods for qualitative assessment of freshwater beaches and developed a novel *E. coli*-specific qPCR-based method for freshwater beach monitoring. Findings from my thesis are published as a series of peer-reviewed publications, and the insights gained from these studies will help health authorities implement molecular methods for error-free water management decision-making.

Abstract

Traditional culturing-based enumeration methods for qualitative monitoring of freshwater beaches provide a delayed assessment of microbial risk factors, and the resulting derived information obtained is limited to fecal indicator counts. Delays in acquiring fecal indicator counts can lead to error-prone beach postings, and the acquired information does not identify additional microbial factors or specific fecal contamination sources. However, the introduction of novel DNA sequencing methodologies and new forms of bioinformatics analyses has revolutionary potential to augment conventional water monitoring technologies. In a series of studies on Southern Ontario recreational waters, including the Great Lakes, we have explored the use of these DNA-based technologies, including rapid qPCR-based assays, DNA sequencing, use of conserved Signature Genes/Proteins and environmental DNA (eDNA) metabarcoding to improve public health responses. This thesis includes evaluating qPCR methods for routine quality assessment of freshwater beaches, demonstrating the capability for rapid, accurate monitoring and timely decision-making. To address conventional fecal contamination monitoring limitations, a novel *E. coli*-specific qPCR strategy based on Conserved Signature Proteins (CSPs) was also developed as an alternative to conventional microbial markers, offering greater specificity and reduced false positives/negatives. The use of DNA sequencing (metagenomic) analysis revealed microbial community changes associated with fecal indicator exceedances, uncovering cyanobacteria, cyanotoxins, and antibiotic-resistance genes not detected by traditional bacterial culturing methods. eDNA metabarcoding was evaluated to identify a broad spectrum of fecal contamination components and characterize region-specific differences. Identifying the root

causes of water quality deterioration using fecal source tracking can enable targeted interventions and a deeper understanding of recreational water quality changes. This thesis underscores the potential of adopting advanced molecular techniques for comprehensive microbial risk assessment and sustainable management of recreational water ecosystems, ultimately improving water quality monitoring and ensuring safer recreational environments and better public health outcomes.

Acknowledgements

I sincerely thank my PhD supervisor, Professor Dr. Herb Schellhorn, for accepting me into his mentorship and guiding me through every single step of my journey. His support extended beyond just research, as his valuable lessons about life and how to always approach things with good faith have made me a better researcher and human being. I appreciate the sincere guidance, support, kindness and freedom I received under Dr. Schellhorn's supervision. I entered Canada without knowing a single soul, but his constant support and encouragement to move forward helped me adjust to the new environment. I feel like I have substantially improved upon my academic writeup and presentation skills because of Dr. Schellhorn's suggestions, and these skills are rarely taught to international graduate students. He was always extremely patient with all the mistakes I made during presentations and writeups and never complained when he had to repeat the same instructions on multiple occasions. Although it is wishful thinking, I hope to have his companionship moving forward, and I truly hope that every international graduate student can find a supervisor/mentor like him. I am also thankful to my supervisory committee, Dr. Thomas Edge, Dr. Radhey Gupta, and Dr. Brian Golding, for their guidance and support. I would especially like to acknowledge Dr. Thomas Edge for his time and support in designing study plans, arranging samples, and reviewing manuscript drafts. I would also like to acknowledge all the past and present members of Schellhorn lab, including Enze Li, Matthew Cleland, Kevin Tran, Jennifer Jiang, and Elizabeth Podleszanski. I am incredibly thankful to Dr. Ethan Paschos for all the emotional support, advice and breakfast meetups.

Finally, I am most thankful to my mother, my sisters, and my brothers for their constant emotional support as I transitioned from one difficult phase to another. This dissertation is dedicated to my late father, Muhammad Saleem Khan. He was the one who pushed me to pursue a PhD even when I thought I would not be capable enough to do so. Although he is no longer here to witness this achievement, truly, it was my father's ambition that brought me to this point.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Declaration of Achievement

This sandwich thesis contains an introduction (Chapter 1), seven published works (Chapter 2, Chapter 3, Chapter 4, Chapter 5, Chapter 6, and Appendix Section), and a manuscript under review (Chapter 7). Chapters 1 (Thesis Overview) and 8 (Thesis Discussion) are added to provide a general introduction to the problems addressed in this thesis and a general discussion that considers problems encountered during the course of these studies for the whole thesis.

Chapter 1. Thesis Introduction:

1.1. Thesis Overview

Establishing robust environmental monitoring involves systematically acquiring and analyzing information on ecological parameters, including atmospheric and aquatic quality, pedological health, and the presence of xenobiotic compounds (Kaufman et al., 2012). The deployment of effective monitoring strategies augments traditional sampling methodologies with avant-garde technologies such as remote sensing, molecular diagnostics, and real-time information strategies (Wahab et al., 2024; Bhatia et al., 2024). For example, molecular diagnostics, exemplified by environmental DNA (eDNA) analysis, augments the capacity for comprehensive monitoring and management of biotic and abiotic pollution (Mierzejewska et al., 2020). Quality monitoring programs, such as the National Water Quality Monitoring Council (NWQMC) (Kleiner et al., 2009) and the European Water Framework Directive (WFD) (Voulvoulis et al., 2017), exemplify structured approaches to environmental assessment, providing standardized protocols and frameworks for information collection, analysis, and dissemination. These quality monitoring programs facilitate the generation of longitudinal information essential for tracking environmental trends and compliance with regulatory benchmarks. Integrating advanced monitoring paradigms with traditional methodologies, such as chemical assays and biological assessments, ensures a holistic understanding of environmental fluxes and pollution sources. The integrative approaches enable the prompt identification of ecological perturbations, fortify regulatory enforcement, and inform the cautious management of land and resources. Effective environmental monitoring is crucial for

discerning and mitigating the anthropogenic impacts on natural systems, ensuring ecological integrity, and safeguarding public health. As environmental challenges become increasingly intricate and pervasive, the imperative to invest in and advance these monitoring strategies intensifies, underscoring the necessity for a robust, scientifically informed framework to navigate the challenges of a sustainable future.

With water being an indispensable life resource, water quality monitoring is at the forefront of basic provisions required for the safeguarding of both human health and economic/ecological balance (WHO, 2016). Effective water quality ensures that water sources remain free from contaminants, including pathogens and pollutants, thus reducing the risk of waterborne illnesses and the long-term sustainability of water ecosystems (Ajami et al., 2018). As part of the Sustainable Development Goals (SDGs) of the United Nations (UN), the provision of clean water for consumption/recreational purposes and sustainable management of water ecosystems/reservoirs are of paramount importance (Allen et al., 2018). To ensure water quality standards are met, rapid and robust identification of water-associated risk factors and sources, including industrial/agricultural runoffs, atmospheric runoffs, algal bloom development, and fecal contamination, is necessary (Bradshaw et al., 2018; Holcomb and Stewart, 2020). Among the sources of water contamination, fecal contamination presents a substantial risk due to its link to numerous waterborne diseases (Murei et al., 2024). Therefore, performing a robust assessment of risk factors associated with fecal contamination can be helpful for policymakers and public health authorities to develop comprehensive water quality monitoring strategies.

Engaging in recreational activities, including swimming and boating, constitutes an integral part of lifestyle, contributing to tourism and local economies generating revenue through hospitality, transportation, and related services (DeFlorio-Barker et al., 2018). However, these activities +may result in an increased risk of illness caused by water-borne pathogens (DeFlorio-Barker et al., 2017; DeFlorio-Barker et al., 2018). Improvements to environment monitoring standards in the United States (U.S.) and the European Union (EU) have primarily focused on the evaluation, validation and development of new analytical methods for effective monitoring of health risk implications that can be used by regulators and public health authorities (Martínez et al., 2010; Colford et al., 2012). Nevertheless, effective detection of pathogens in environmental samples, is subject to challenging constraints, including the cost of testing multiple sites, sample collection/delivery to the laboratory, lack of broad-spectrum validated methods for testing, delay in acquisition of results for timely decision making, and technical challenges in detecting the typically low concentrations of pathogens present in the environment (Whitman et al., 1999; Byappanahalli et al., 2003; Boehm., 2007). Therefore, validating and developing newer technologies for the transition from decades-old conventional methodology is imperative for designing sustainable and effective environmental diagnostic programs.

With the recent advances in the applications of molecular diagnostics and the generation of ever-increasing DNA information, it is imperative to re-evaluate traditional/conventional testing along with newer technologies and develop more target-specific diagnostic methods to establish comprehensive and robust diagnostic

applications. This thesis is designed to understand, develop and present the applications of newer molecular diagnostic methods with an emphasis on how these technologies can be useful for public health authorities to transition away from information-limited conventional environmental monitoring for the long-term sustainability of recreational water ecosystems and rapid decision-making in favor of public health. The major questions explored in this thesis are: 1) Can rapid DNA-based assays be a potential alternative to conventional environmental quality monitoring strategies? 2) How will shifting to rapid water testing strategies impact the decision-making process in environmental postings? 3) Are there any environmental microbial risk factors that may not be accounted for when relying upon conventional environmental quality monitoring strategies? 4) Would conserved molecular markers be used to develop a more robust water diagnostic measure? 5) How can next-generation DNA sequencing be applied to environmental diagnostics?

1.1. The Laurentian Great Lakes

The Laurentian Great Lakes, commonly known as the Great Lakes, are a critical resource, holding 84% of North America's freshwater and 21% of the world's available freshwater (Environment and Climate Change Canada, 2019). This group of five lakes covers an area exceeding 200,000 km² and straddles the borders of the United States and Canada (Great Lakes Commission, 2017). Approximately 30% of the economies in both countries are connected to the Great Lakes through sectors such as fisheries, industry, tourism, and recreation (Great Lakes Commission, 2017). The Great Lakes region supports around 30 million people (US EPA, 2019) and is home to 4,000 species of

wildlife (Environment and Climate Change Canada, 2019). The Great Lakes feature freshwater beaches that offer recreational opportunities to millions of people. These freshwater beaches reflect the overall health of the Great Lakes and play a vital role in the economies of many local communities. Besides recreational activities, the Great Lakes provide drinking water for shoreline communities and are a crucial resource for commercial shipping and transportation (Nevers et al., 2014). As anthropogenic activities have grown, the ecological pressures on the lakes and concerns about protecting recreationists from harmful chemical and microbiological agents have also increased. The rising recreational use of the Great Lakes beaches has created a demand for reliable and timely microbial water quality assessments to ensure public health protection. This thesis is undertaken to tackle these challenges by validating, evaluating and developing rapid and comprehensive molecular methods for long-term sustainability strategies for the freshwater beaches in the Great Lakes region.

1.2. Environment Quality Monitoring Challenges

Routine environmental quality monitoring is essential to sustainability measures for recreational waters, including beaches (USEPA, 2012). However, commonly used quality assessment strategies, including bacterial culturing, are decades old and suffer from inherent limitations such as time delays in the acquisition of results, information limited to a single quality marker, labour intensive, and cross-reactivities with non-target quality markers (Alonso et al., 1998; Chao et al., 2004). Although with the advancements in quality monitoring technologies, more robust and comprehensive molecular methods are available now, these methods have rarely been validated for use in different regions

despite the fact that regio-specific changes in environmental factors can affect the functionality of molecular methods (Mukherjee et al., 2023; Bao et al., 2023). Environmental factors, including heavy rain events and agricultural/urban runoffs, result in sudden water quality changes due to a higher waste influx, which may lead to complications due to the time delay associated with conventional quality management strategies (Sivaganesan et al., 2019; Haugland et al., 2021). Additionally, most universal quality markers or single target testing strategies may not provide information regarding unrelated pathogenic factors, which hinders the ability of public health authorities to make robust and definitive decisions. For example, potential pathogenic factors, including antibiotic-resistance genes (Saleem et al., 2024), pathogenic viruses (Lin and Ganesh, 2013), protists (Dorevitch et al., 2011), and cyanobacterial blooms (Saleem et al., 2024), may be undetected if only conventional fecal markers are relied upon. Therefore, exploring, validating and developing newer technologies as an alternative to conventional decades-old strategies is imperative for effective environmental management strategies.

1.2. Fecal Indicator Bacteria

Fecal indicator bacteria (FIB), including *E. coli* and *enterococci,* are commonly used as proxy indicators of pathogenic bacteria (Sadowsky and Whitman, 2014; USEPA, 2012). FIBs are part of the mammalian gut microbiome and are considered good indicators of fecal contamination due to their high concentrations in fecal material (Devane et al., 2020). However, the efficacy of fecal indicators for fecal contamination can become questionable depending on the source of fecal contamination. For example, point sources

of fecal contamination result in much higher concentrations of fecal indicators than non-point sources, but fecal contamination in the latter case can persist for a more extended period and require more advanced genetic/analytical methods to differentiate between a recent or aged fecal contamination event (Devane et al., 2018; Hagedorn et al., 2011; Teixeira et al., 2020; Tran et al., 2015). The relationship between fecal indicators and bacterial pathogens can also change due to factors, including dilution, survival rate/time, and water flow dynamics (Chu et al., 2011; Liang et al., 2013). Additionally, a comprehensive study on the relationship between fecal indicators and waterborne pathogens concluded that no single indicator could predict all pathogens, including bacteria, viruses, and protozoa (Korajkic et al., 2018), which highlights the importance of developing measures that can go beyond just the concept of fecal indicators.

1.3. Conventional and Molecular Technologies for Water Quality Monitoring

1.3.1. Culturing-Based Fecal Indicator Enumeration

The primary methods used for the detection of FIBs are based on their ability to be cultured by inoculating a known volume of sample on a specific growth medium, followed by an incubation period of 18-24h (Rochelle-Newall et al., 2015), which is easy to use and well-validated over the years. Culture-based methods are more widely used because of the feasibility factors, including availability in portable formats, cost-effectiveness, and availability of standard protocols from public health agencies (Oon et al., 2023). However, culture-dependent strategies rely on the cultivability of the bacteria present, a factor that can vary considerably depending on the bacterial species investigated, environmental parameters, and the physiological state of the bacteria

(Colwell, 2000). Culture-based methods are primarily based on the indicator β-galactosidase, which is only considered specific to *E. coli* (Mathew and Alocilja, 2007). However, β-galactosidase activity can be absent in some environmental *E. coli* isolates under specified incubation temperatures, thus leading to false negatives, or it can also be detected in non-*E. coli* species (Alonso et al., 1998; Chao et al., 2004), which can result in false positives. Additionally, culture-based methods require 18-24 h of incubation for the results to be available, which limits decision-makers' ability to generate a timely response (Ho and Tam, 1997). Alternatively, newer molecular diagnostic tools, including PCR amplification, conserved molecular markers and DNA Sequencing, can be used for more robust fecal contamination assessment, and their potential to augment conventional environmental monitoring strategies must be evaluated.

1.3.2. DNA Amplification for Environmental Monitoring

Limitations of the culture-based enumeration can be bypassed using rapid FIB-specific gene amplification technologies, including quantitative/qualitative PCR (Rudko et al., 2020). Gene amplification-based methods require a gene target specific to the targeted microbial species and can be selectively amplified using gene-specific primers (Nappier et al., 2020). Signal amplification or fluorescence can be detected using dye-based methods, where the dye selectively binds to double-stranded DNA, resulting in fluorescence (Botes et al., 2013), or TaqMan Probe-based methods, which utilize a third sequence situated between the primer binding sites is cleaved by DNA polymerase during amplification, leading to the generation of a fluorescence signal (Lavender and

Kinzelman, 2009). The PCR cycle number corresponding to the reliable detection of the fluorescence signal is termed the threshold cycle (Ct or Cq) value and is inversely proportional to the number of gene copies in a sample (Tuomi et al., 2010). Considering the applications of DNA amplification technologies, FIB quantification methods, including the *enterococci* (USEPA, 2015) and *E. coli* (Sivaganesan et al., 2019) specific ones from the United States Environmental Protection Agency (USEPA) have been introduced over the years, but are rarely tested on the field samples and are mostly validated in lab or marine settings. For example, freshwater beaches constitute most of the North American region, but the enterococci qPCR approach from USEPA has been mainly validated in marine waters, while *E. coli* qPCR method is still under validation and is not available for implementation by the public health authorities (Sivaganesan et al., 2019; Haugland et al., 2021). Although DNA amplification technologies are a rapid alternative to conventional strategies, limitations, including PCR inhibitors (Gentry-Shields et al., 2013), DNA recovery (Hinlo et al., 2017), and lack of availability of gene copy-based water quality thresholds, can hinder the application in field testing. PCR-based technologies are also restricted to the detection/quantification of a limited number of target genes or organisms in a single reaction, due to which additional microbial risk factors aside from the one being tested may go undetected. Therefore, extensive validation of amplification-based technologies is imperative to understand the logistics involved in replacing conventional environmental quality monitoring programs.

1.3.3. DNA Sequencing-based Methods and Their Applications for Water Monitoring

DNA sequencing has emerged as a powerful tool for environmental monitoring, offering unparalleled insights into the diversity and abundance of microbial communities in various ecosystems (Staley and Sadowsky, 2017). In the context of environmental quality assessment, traditional approaches have limitations in detecting all microbial species present, especially those that are non-culturable or present in low abundance (Epstein, 2013). In contrast, DNA sequencing allows for the comprehensive analysis of microbial communities by directly examining their genetic material. This approach provides a holistic understanding of microbial diversity and dynamics in aquatic environments, enabling researchers to identify potential pathogens (Gwinn et al., 2019), track pollution sources (Caldwell et al., 2011), and understand the microbial dynamics associated with the changes in environmental factors (Wani et al., 2021). Metagenomics is one of the prime examples that has revolutionized water quality testing by allowing the analysis of the entire microbiome (Mohiuddin et al., 2021). The comprehensive taxonomic/functional overview obtained from metagenomic sequencing can help public health authorities detect changes in community structure in response to environmental perturbations or pollution events (Mukherjee et al., 2023; Bao et al., 2023). Although DNA sequencing cannot differentiate between genetic material from live and dead cells (Emerson et al., 2017) and cannot be categorized as a rapid testing technology, it can augment the conventional environment monitoring programs to identify and incorporate additional microbial risk factors that may go undetected by only testing for FIBs.

1.3.4. Conserved Signature Proteins/Genes as Robust Alternatives to Conventional Microbial Markers

Universal taxonomic gene markers, including 16S rRNA (Chern et al., 2011) and LsrRNA (USEPA, 2015), are most commonly used for DNA amplification-based microbial testing. Despite being heavily tested, the universal taxonomic markers are not entirely specific to the targeted organisms and can generate false positives (Papin et al., 2004). With the advancements in DNA databases, comprehensive genomic information of most bacterial species is publicly available, which can be used to identify organism-specific gene/protein markers that are solely part of the targeted organism's genome (Gupta, 2010). The prime example of such taxa-specific markers is Conserved Signature Proteins (CSPs) and Conserved Signature INDELS (CSIs) (Gao and Gupta, 2012). CSPs and CSIs are evolutionary-conserved proteins or insertions/deletions specific to a particular taxonomic group or clade of microorganisms (Gupta, 2010; Gao and Gupta, 2012). In the past decade, CSPs and CSIs have been used for taxonomic reclassification of microbial species into distinct clades (Adeolu and Gupta, 2014; Gupta et al., 2015) and in principle, these proteins or INDELS can be used to design/develop molecular methods for highly specific water monitoring strategies to avoid false positives/negatives. Therefore, developing DNA amplification-based strategies using CSPs and CSIs would allow more robust environment quality assessments and public health authorities to make error-free management decisions.

1.3.5. Conclusion

Environmental monitoring is a critical component of ecological sustainability and public health protection, particularly in the context of water resources, including the Laurentian Great Lakes. The effective monitoring of environmental parameters requires a multifaceted approach that integrates traditional methods with advanced molecular diagnostic technologies, including environmental DNA (eDNA) analysis, quantitative PCR, and DNA sequencing, which offer significant improvements in the detection and management of pollutants and pathogens over conventional methodologies.

Traditional environmental monitoring approaches, while well-established, have inherent limitations, including delays in obtaining results, restricted analytical scope, and challenges in detecting a broad range of pathogens. The incorporation of molecular diagnostics into environmental monitoring frameworks enhances the capability to rapidly and accurately assess water quality and enable the detection of a wider array of microbial contaminants, which is crucial for addressing the increasing complexity of environmental challenges. The development of diagnostic strategies that account for regional environmental variability and specific microbial risk factors is essential for improving the efficacy of environmental monitoring programs. The utilization of advanced molecular markers, such as Conserved Signature Proteins (CSPs) and Conserved Signature INDELS (CSIs), presents promising potential for enhancing the precision and reliability of environmental assessments. In the context of the Great Lakes, molecular diagnostic technologies are particularly valuable for the timely

identification of ecological perturbations, informing public health decisions, and safeguarding recreational water users.

1.3.6. References

Adeolu, M. and Gupta, R.S., 2014. A phylogenomic and molecular marker based proposal for the division of the genus Borrelia into two genera: the emended genus Borrelia containing only the members of the relapsing fever Borrelia, and the genus Borreliella gen. nov. containing the members of the Lyme disease Borrelia (Borrelia burgdorferi sensu lato complex). Antonie Van Leeuwenhoek, 105, pp.1049-1072.

Agenda for Great Lakes restoration and economic revitalization Retrieved from

Ajami, N., Habicht, H., Fewell, B., Lattimer, T. and Ng, T., 2018. Water finance: The imperative for water security and economic growth.

Allen, C., Metternicht, G. and Wiedmann, T., 2018. Initial progress in implementing the Sustainable Development Goals (SDGs): A review of evidence from countries. Sustainability science, 13, pp.1453-1467.

Alonso, J. L., Soriano, A., Amoros, I., & Ferrus, M. A. (1998). Quantitative determination of *E. coli*, and fecal coliforms in water using a chromogenic medium. Journal of Environmental Science & Health Part A, 33(6), 1229–1248.

Bao, Y., Ruan, Y., Wu, J., Wang, W.X., Leung, K.M. and Lee, P.K., 2023. Metagenomics-Based Microbial Ecological Community Threshold and Indicators of

Anthropogenic Disturbances in Estuarine Sediments. Environmental Science & Technology, 58(1), pp.780-794.

Bhatia, D., Paul, S., Acharjee, T. and Ramachairy, S.S., 2024. Biosensors and their widespread impact on human health. Sensors International, 5, p.100257.

Boehm, A.B., 2007. Enterococci concentrations in diverse coastal environments exhibit extreme variability.

Botes, M., de Kwaadsteniet, M. and Cloete, T.E., 2013. Application of quantitative PCR for the detection of microorganisms in water. Analytical and bioanalytical chemistry, 405, pp.91-108.

Bradshaw, J.K., Snyder, B.J., Oladeinde, A., Spidle, D., Berrang, M.E., Meinersmann, R.J., Oakley, B., Sidle, R.C., Sullivan, K. and Molina, M., 2016. Characterizing relationships among fecal indicator bacteria, microbial source tracking markers, and associated waterborne pathogen occurrence in stream water and sediments in a mixed land use watershed. Water research, 101, pp.498-509.

Byappanahalli, M., Fowler, M., Shively, D. and Whitman, R., 2003. Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. Applied and environmental microbiology, 69(8), pp.4549-4555.

Caldwell, J., Payment, P. and Villemur, R., 2011. Mitochondrial DNA as source tracking markers of fecal contamination. Microbial source tracking: methods, applications, and case studies, pp.229-250.

Chao, K. K., Chao, C. C., & Chao, W. L. (2004). Evaluation of Colilert-18 for detection of coliforms and Eschericha coli in subtropical freshwater. Applied and Environmental Microbiology, 70(2), 1242–1244.

Chern, E.C., Siefring, S., Paar, J., Doolittle, M. and Haugland, R.A., 2011. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Letters in applied microbiology, 52(3), pp.298-306.

Chigbu, P. and Sobolev, D., 2007. Bacteriological analysis of water. In Handbook of water analysis (pp. 111-148). CRC Press.

Christensen, H., Andersson, A.J., Jørgensen, S.L. and Vogt, J.K., 2018. 16S rRNA amplicon sequencing for metagenomics. Introduction to bioinformatics in microbiology, pp.135-161.

Chu, Y., Salles, C., Tournoud, M.G., Got, P., Troussellier, M., Rodier, C. and Caro, A., 2011. Faecal bacterial loads during flood events in Northwestern Mediterranean coastal rivers. Journal of Hydrology, 405(3-4), pp.501-511.

Colford Jr, J.M., Schiff, K.C., Griffith, J.F., Yau, V., Arnold, B.F., Wright, C.C., Gruber, J.S., Wade, T.J., Burns, S., Hayes, J. and McGee, C., 2012. Using rapid indicators for Enterococcus to assess the risk of illness after exposure to urban runoff contaminated marine water. Water research, 46(7), pp.2176-2186.

Colwell, R.R., 2000. Viable but nonculturable bacteria: a survival strategy. Journal of Infection and Chemotherapy, 6, pp.121-125.

DeFlorio-Barker, S., Wade, T.J., Jones, R.M., Friedman, L.S., Wing, C. and Dorevitch, S., 2017. Estimated costs of sporadic gastrointestinal illness associated with surface water recreation: a combined analysis of data from NEEAR and CHEERS studies. Environmental health perspectives, 125(2), pp.215-222.

DeFlorio-Barker, S., Wing, C., Jones, R.M. and Dorevitch, S., 2018. Estimate of incidence and cost of recreational waterborne illness on United States surface waters. Environmental Health, 17, pp.1-10.

Devane, M.L., Moriarty, E., Weaver, L., Cookson, A. and Gilpin, B., 2020. Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. Water Research, 185, p.116204.

Dorevitch, S., Doi, M., Hsu, F.C., Lin, K.T., Roberts, J.D., Liu, L.C., Gladding, R., Vannoy, E., Li, H., Javor, M. and Scheff, P.A., 2011. A comparison of rapid and conventional measures of indicator bacteria as predictors of waterborne protozoan pathogen presence and density. Journal of Environmental Monitoring, 13(9), pp.2427-2435.

Emerson, J.B., Adams, R.I., Román, C.M.B., Brooks, B., Coil, D.A., Dahlhausen, K., Ganz, H.H., Hartmann, E.M., Hsu, T., Justice, N.B. and Paulino-Lima, I.G., 2017. Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. Microbiome, 5, pp.1-23.

Environment and Climate Change Canada [ECCC], 2019. Great Lakes Protection:

Epstein, S.S., 2013. The phenomenon of microbial uncultivability. Current opinion in microbiology, 16(5), pp.636-642.

Gao, B. and Gupta, R.S., 2012. Microbial systematics in the post-genomics era. Antonie Van Leeuwenhoek, 101, pp.45-54.

Gentry-Shields, J., Wang, A., Cory, R.M. and Stewart, J.R., 2013. Determination of specific types and relative levels of QPCR inhibitors in environmental water samples using excitation–emission matrix spectroscopy and PARAFAC. Water research, 47(10), pp.3467-3476.

Great Lakes Commission, 2017. Great Lakes commission federal priorities: an

Great Lakes Retrieved from https://www.canada.ca/en/environment-climatechange/

Gupta, R.S., 2010. Molecular signatures for the main phyla of photosynthetic bacteria and their subgroups. Photosynthesis research, 104, pp.357-372.

Gupta, R.S., Naushad, S. and Baker, S., 2015. Phylogenomic analyses and molecular signatures for the class Halobacteria and its two major clades: a proposal for division of the class Halobacteria into an emended order Halobacteriales and two new orders, Haloferacales ord. nov. and Natrialbales ord. nov., containing the novel families Haloferacaceae fam. nov. and Natrialbaceae fam. nov. International Journal of Systematic and Evolutionary Microbiology, 65(Pt_3), pp.1050-1069.

Gwinn, M., MacCannell, D. and Armstrong, G.L., 2019. Next-generation sequencing of infectious pathogens. Jama, 321(9), pp.893-894.

Hagedorn, C., Blanch, A.R. and Harwood, V.J. eds., 2011. Microbial source tracking: methods, applications, and case studies. Springer Science & Business Media.

Haugland, R., Oshima, K., Sivaganesan, M., Dufour, A., Varma, M., Siefring, S., Nappier, S., Schnitker, B. and Briggs, S., 2021. Large-scale comparison of *E. coli* levels determined by culture and a qPCR method (EPA Draft Method C) in Michigan towards the implementation of rapid, multi-site beach testing. Journal of microbiological methods, 184, p.106186.

Hinlo, R., Gleeson, D., Lintermans, M. and Furlan, E., 2017. Methods to maximise recovery of environmental DNA from water samples. PloS one, 12(6), p.e0179251.

Ho, B.S.W. and Tam, T.Y., 1997. Enumeration of *E. coli* in environmental waters and wastewater using a chromogenic medium. Water science and technology, 35(11-12), pp.409-413.

Holcomb, D.A. and Stewart, J.R., 2020. Microbial indicators of fecal pollution: recent progress and challenges in assessing water quality. Current environmental health reports, 7, pp.311-324.

https://www.glc.org/wp-content/uploads/2017/03/GLC-LegPri-2017-FINALweb.

Kaufmann, H., Förster, S., Wulf, H., Segl, K., Guanter, L., Bochow, M., Heiden, U., Müller, A., Heldens, W., Schneiderhan, T. and Leitão, P.J., 2012. Science plan of the environmental mapping and analysis program (EnMAP).

Kleiner, Y., Rajani, B. and Sadiq, R., 2009. Drinking water infrastructure assessment: the National Research Council of Canada perspective. In World Environmental and Water Resources Congress 2009: Great Rivers (pp. 1-13).

Korajkic, A., McMinn, B.R. and Harwood, V.J., 2018. Relationships between microbial indicators and pathogens in recreational water settings. International Journal of Environmental Research and Public Health, 15(12), p.2842.

Lavender, J.S. and Kinzelman, J.L., 2009. A cross comparison of QPCR to agar-based or defined substrate test methods for the determination of *Escherichia coli* and enterococci in municipal water quality monitoring programs. Water research, 43(19), pp.4967-4979.

Liang, Z., He, Z., Zhou, X., Powell, C.A., Yang, Y., He, L.M. and Stoffella, P.J., 2013. Impact of mixed land-use practices on the microbial water quality in a subtropical coastal watershed. Science of the Total Environment, 449, pp.426-433.

Lin, J. and Ganesh, A., 2013. Water quality indicators: bacteria, coliphages, enteric viruses. International journal of environmental health research, 23(6), pp.484-506.

Martínez, I.L., Díaz, C.Á., Díaz, J.L.G., Cortezón, J.A.R. and Juanes, J.A., 2010. The European bathing water directive: application and consequences in quality monitoring programs. Journal of Environmental Monitoring, 12(1), pp.369-376.

Mathew, F.P. and Alocilja, E.C., 2004. Enzyme-based detection of *Escherichia coli*. Transactions of the ASAE, 47(1), pp.357-362.

Mierzejewska, E. and Urbaniak, M., 2020. Molecular methods as potential tools in ecohydrological studies on emerging contaminants in freshwater ecosystems. Water, 12(11), p.2962.

Mohiuddin, M.M., Salama, Y., Schellhorn, H.E. and Golding, G.B., 2017. Shotgun metagenomic sequencing reveals freshwater beach sands as reservoir of bacterial pathogens. Water Research, 115, pp.360-369.

Mukherjee, A., D'Ugo, E., Giuseppetti, R., Magurano, F. and Cotter, P.D., 2023. Metagenomic approaches for understanding microbial communities in contaminated environments: Bioinformatic tools, case studies and future outlook. In Metagenomics to bioremediation (pp. 103-156). Academic Press.

Murei, A., Kamika, I. and Momba, M.N.B., 2024. Selection of a diagnostic tool for microbial water quality monitoring and management of faecal contamination of water sources in rural communities. Science of the Total Environment, 906, p.167484.

Nappier, S.P., Ichida, A., Jaglo, K., Haugland, R. and Jones, K.R., 2019. Advancements in mitigating interference in quantitative polymerase chain reaction (qPCR) for microbial water quality monitoring. Science of the Total Environment, 671, pp.732-740.

Nevers, M.B., Byappanahalli, M.N., Edge, T.A. and Whitman, R.L., 2014. Beach science in the Great Lakes. Journal of Great Lakes Research, 40(1), pp.1-14.

Oon, Y.L., Oon, Y.S., Ayaz, M., Deng, M., Li, L. and Song, K., 2023. Waterborne pathogens detection technologies: advances, challenges, and future perspectives. Frontiers in Microbiology, 14, p.1286923.

Oster, R.J., Wijesinghe, R.U., Haack, S.K., Fogarty, L.R., Tucker, T.R. and Riley, S.C., 2014. Bacterial pathogen gene abundance and relation to recreational water quality at seven Great Lakes beaches. Environmental science & technology, 48(24), pp.14148-14157.

Papin, J.F., Vahrson, W. and Dittmer, D.P., 2004. SYBR green-based real-time quantitative PCR assay for detection of West Nile Virus circumvents false-negative results due to strain variability. Journal of Clinical Microbiology, 42(4), pp.1511-1518.

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J. and Segata, N., 2017. Shotgun metagenomics, from sampling to analysis. Nature biotechnology, 35(9), pp.833-844.

Rochelle-Newall, E., Nguyen, T.M.H., Le, T.P.Q., Sengtaheuanghoung, O. and Ribolzi, O., 2015. A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions. Frontiers in microbiology, 6, p.308.

Rudko, S.P., Reimink, R.L., Peter, B., White, J. and Hanington, P.C., 2020. Democratizing water monitoring: Implementation of a community-based qPCR monitoring program for recreational water hazards. PLoS One, 15(5), p.e0229701.

Sadowsky, M.J. and Whitman, R.L. eds., 2010. The fecal bacteria. American Society for Microbiology Press.

Saleem, F., Li, E., Edge, T.A., Tran, K.L. and Schellhorn, H.E., 2024. Identification of potential microbial risk factors associated with fecal indicator exceedances at recreational beaches. Environmental Microbiome, 19(1), p.4.

Sivaganesan, M., Aw, T.G., Briggs, S., Dreelin, E., Aslan, A., Dorevitch, S., Shrestha, A., Isaacs, N., Kinzelman, J., Kleinheinz, G. and Noble, R., 2019. Standardized data quality acceptance criteria for a rapid *Escherichia coli* qPCR method (Draft Method C) for water quality monitoring at recreational beaches. Water research, 156, pp.456-464.

Staley, C. and Sadowsky, M.J., 2016. Application of metagenomics to assess microbial communities in water and other environmental matrices. Journal of the Marine Biological Association of the United Kingdom, 96(1), pp.121-129.

Teixeira, P., Salvador, D., Brandão, J., Ahmed, W., Sadowsky, M.J. and Valério, E., 2020. Environmental and adaptive changes necessitate a paradigm shift for indicators of fecal contamination. Microbiology spectrum, 8(2), pp.10-1128.

Tiwari, A., Oliver, D.M., Bivins, A., Sherchan, S.P. and Pitkänen, T., 2021. Bathing water quality monitoring practices in Europe and the United States. International Journal of Environmental Research and Public Health, 18(11), p.5513.

Tran, N.H., Gin, K.Y.H. and Ngo, H.H., 2015. Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. Science of the Total Environment, 538, pp.38-57.

Tuomi, J.M., Voorbraak, F., Jones, D.L. and Ruijter, J.M., 2010. Bias in the Cq value observed with hydrolysis probe based quantitative PCR can be corrected with the estimated PCR efficiency value. Methods, 50(4), pp.313-322.

United States Environmental Protection Agency [US EPA]. The Great Lakes. (2019). Retrieved from https://www.epa.gov/greatlakes

USEPA, 2015. Method 1609.1: Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay. US EPA.

Voulvoulis, N., Arpon, K.D. and Giakoumis, T., 2017. The EU Water Framework Directive: From great expectations to problems with implementation. Science of the Total Environment, 575, pp.358-366.

Wahab, A., Muhammad, M., Ullah, S., Abdi, G., Shah, G.M., Zaman, W. and Ayaz, A., 2024. Agriculture and environmental management through nanotechnology: Eco-friendly nanomaterial synthesis for soil-plant systems, food safety, and sustainability. Science of the Total Environment, p.171862.

Wani, G.A., Khan, M.A., Dar, M.A., Shah, M.A. and Reshi, Z.A., 2021. Next generation high throughput sequencing to assess microbial communities: An application based on water quality. Bulletin of Environmental Contamination and Toxicology, 106, pp.727-733.

Whitman, R.L., Nevers, M.B. and Gerovac, P.J., 1999. Interaction of ambient conditions and fecal coliform bacteria in southern Lake Michigan beach waters: monitoring program implications. Natural Areas Journal, 19(2), pp.166-171.

World Health Organization, 2016. Quantitative microbial risk assessment: application for water safety management.

Chapter 2. Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring

This thesis delves into innovative methodologies for ensuring the safety and sustainability of recreational water environments. Central to this exploration is the integration of molecular techniques, particularly quantitative Polymerase Chain Reaction (qPCR), which offer rapid and reliable alternatives to traditional culture-based methods for water quality assessment. Chapter 2 is pivotal to the theme of this thesis as it validates the use of t qPCR-based method for routine quality monitoring recreational beaches and showcases the practical application of qPCR for real-time water quality monitoring, which aligns with the objective of developing advanced molecular strategies for the sustainability of recreational waters.

*The text I present **here is a peer-reviewed manuscript published** in the Journal of Great Lakes Research. The formatting has been changed from the journal version to be consistent throughout the thesis.* ***Saleem, F.,*** *Edge, T.A. and Schellhorn, H.E., 2022. Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring. Journal of Great Lakes Research, 48(3), pp.707-716.*

Author Contributions:

FS performed the experiments, analyzed the data, communicated the findings with Public Health Agencies, Wrote the original manuscript, and revised it as required by the reviewers. HES and TAE collected the water samples and reviewed the manuscript before submission to the journal.

Abstract

Rapid quantitative PCR-based methods for enterococci monitoring can allow public health authorities to make more timely beach posting decisions. However, qPCR methods must be assessed for proposed sites as locale-specific factors may affect DNA recovery or qPCR inhibition. We assessed the feasibility of the USEPA 1609.1 qPCR-based (*Enterococcus*) method at two urban Toronto beaches and three recreational areas at nearby river mouths in parallel with culture-based methods on the same water samples. A strong positive correlation was observed between the enterococci qPCR method and culturing-based quantification methods for *E. coli* and enterococci at both beaches and two river mouth areas. One river, known to be highly sewage-impacted did not yield DNA suitable for qPCR analyses. qPCR results from biological replicates were strongly correlated and showed coefficients of variation as low as or lower than culture-based methods. With respect to Beach Action Value exceedances, the USEPA 1609.1 qPCR method provided an 80-90% level of agreement with *E. coli* enumeration results and >90% with enterococci enumeration. Results indicated that if recreational water locations and sampling conditions met the requirements of the USEPA 1609.1 qPCR method, the method can meet or exceed all quality control requirements and provide water quality results within 3.5h for diverse recreational water settings around the City of Toronto.

Keywords: Method Validation, Freshwater Beaches, qPCR, *E. coli*, enterococci

Introduction

Culturing-based methods for fecal indicator bacteria are cost-effective and widely used for assessing recreational water quality but can only provide results after 18–96 h of sample collection (Ferguson et al., 2013). However, the fecal indicator profile of recreational waters can change markedly within 24 h potentiating erroneous public health postings (Dorevitch et al., 2017; Shrestha and Dorevitch, 2019).

Advances in water quality assessment methods such as polymerase chain reaction-based assays can serve as rapid, robust alternatives to slower culture-based methods for fecal indicator bacteria (Haugland et al., 2005). The National Environmental Epidemiological Assessment of Recreational water (NEEAR) study in the United States found that a qPCR-based enterococci measure was not only faster, but also a potentially better predictor of gastrointestinal illnesses among beachgoers (Dorevitch et al., 2017; USEPA, 2009; Wade et al., 2008; Wade et al., 2010). Based on the NEEAR study, the United States Environmental Protection Agency (USEPA) has proposed a Beach Action Value (BAV) of 1000 calibra- tor cell equivalents (CCE)/100 mL for enterococci qPCR (USEPA, 2012). In 2015, the U.S. EPA published method 1609.1 as a detailed qPCR-based assay for enterococci enumeration in recreational water ecosystems. The USEPA 1609.1 method can provide results within 4 h of sample processing and can be more reliable than other qPCR assays because of different quality controls for each step of water sample processing.

While method 1609.1 has been tested successfully at a number of beaches in the United States (Dorevitch et al., 2017; Seruge et al., 2019; Sheth et al., 2016), it has not been tested at many beaches in Canada, nor across a wide range of recreational water settings and water

sampling regimes. Water matrix composition can change between different recreational water settings and result in water matrix interferences, leading to either poor DNA recovery or qPCR inhibition (Converse et al., 2012; Nappier et al., 2019). Therefore, it is recommended to test the US EPA 1609.1 qPCR-based method's compatibility with each new water source (USEPA, 2015). As method 1609.1 has only been slowly adopted by beach monitoring programs across the U.S. to date (Shrestha and Dorevitch, 2020), additional studies testing and documenting the advantages of the qPCR across different recreational water settings could assist with wider adoption of the method in beach monitoring programs.

Toronto is a hub for recreational water activities, with diverse opportunities for paddle boarding, kayaking, windsurfing, boating, and freshwater beaches that provide both recreational and economic benefits (Edge et al., 2021; Kidd, 2015). Fecal contamination is monitored each day at Toronto's eleven beaches by culture-based Escherichia. coli enumeration (toronto.ca/beach-water-quality). However, results for beach postings are only available after 24 h. Delayed beach postings can pose a health risk health as well as lead to economic losses due to unnecessary beach postings (Heasley et al., 2021). Therefore, same-day beach water quality monitoring methods based on rapid qPCR-based methodology provide an opportunity for more timely and reliable beach postings. The current study focuses on assessing the applicability of qPCR-based enterococci quantification in concert with culturing-based methods across different recreational water settings and water sampling regimes in Toronto, Ontario. We studied two Toronto beaches (Marie Curtis and Sunnyside Beaches) and recreational waters at the mouths of three Toronto rivers (Etobicoke Creek, Humber River, and Don River). Water samples from the Don River mouth represented a

water sampling regime challenge for the qPCR method as samples were collected and frozen at -20 C until lab processing. Objectives for this study are as follows: i) validate compatibility of USEPA 1609.1 qPCR-based method for enterococci monitoring in Toronto beaches, ii) test for correlation between results of qPCR- based and culture-based quantification methods, and iii) assess limitations associated with the qPCR-based method across diverse recreational water settings and water sampling regimes.

Materials and methods

Study design

Our study design focused on water quality monitoring at two urban Toronto recreational beaches- Marie Curtis East Beach and Sunnyside Beach. Water quality at these beaches is tested daily by Toronto Public Health. Water samples were also simultaneously collected from comparative reference sites at adjacent river mouths, including Humber River for Sunnyside Beach and Etobicoke Creek for Marie Curtis East Beach. These river mouths serve as recreational water settings for activities such as kayaking and canoeing. Fig. 1 and Electronic Supplementary Material (ESM) Table S1 provide geographical locations and coordinates of sampling sites, respectively. Water sampling was performed in the summer of two consecutive years, 2020 and 2021. For the 2020 sampling season, water samples were collected mid-morning from Etobicoke Creek (E.C.) and at two transects (30 W and 32 W) of Marie Curtis East Beach once a week for six weeks. For the 2021 sampling season, water samples were collected between 5:30 am and 7 am, three days a week for three months, from June 01, 2021, to August 26, 2021. For each sampling day, eight water samples were collected, which included one sample each from Humber River and Etobicoke Creek, and

two samples with a replicate from one of the two beach transects of Marie Curtis East Beach (30 W, 30 W-replicate, and 32 W) and Sunnyside beach (18 W, 21 W and 21 W-replicate). A total of 96 samples were analyzed for the 2020 sampling season, while 309 water samples were collected and processed for the 2021 sampling season. As a challenge test for the *Enterococcus* qPCR method, we also analyzed fifty 500 mL water samples collected by a different water sampling regime from a site at the mouth of the Don River in Toronto. These samples were collected on a biweekly basis over the course of 2020 and kept frozen at -20 C until analysis in 2021. The Don River mouth is in a very urban setting, consistently impacted by high levels of human sewage contamination (Edge et al., 2021).

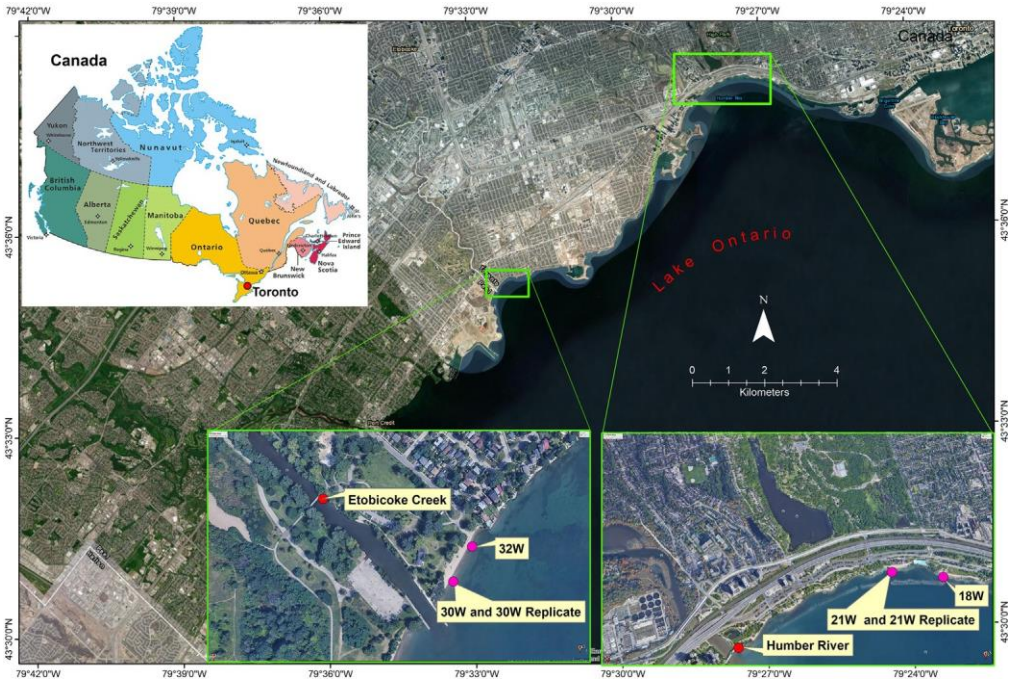


Figure 1 Geographical locations of Marie Curtis Beach, Sunnyside Beach, Etobicoke Creek, and Humber River sampling sites.

Water sample collection and filtration

Grab samples were collected 30–60 cm below the water surface in sterile 1000 mL screw-capped polyethylene terephthalate (PET) bottles. Samples were kept in coolers (4 C) and immediately transported to the lab for processing. All the samples were processed for culture-based enumeration and qPCR within six hours of collection. Bottles were shaken vigorously before passing 100 mL of water sample through a 0.45 µm polycarbonate membrane filter (Millipore Corp., Bedford, MA) to capture bacteria, followed by DNA extraction. In addition, dilutions of water samples were prepared as needed and filtered separately for *E. coli* and enterococci culturing.

***E. coli* and enterococci enumeration by culture**

E. coli enumeration by culture was performed for all the samples from June 01, 2021, to July 27, 2021, while enterococci enumeration by culture was performed for the samples collected from June 17, 2021, to August 26, 2021. For *E. coli* enumeration, differential coliform agar (OxoidTM) was used with cefsulodin, while enterococci were cultured on m-Enterococci Agar (OxoidTM) (APHA, 2012). Membrane filters were placed on 47 mm agar plates and incubated for 24 h at 44.5 °C for *E. coli* and 41.5 °C for enterococci. In addition, 100 mL of sterile phosphate buffer saline was used as a negative control, while cultures of *E. coli* (ATCC 11775) and *Enterococcus faecalis* (ATCC 2921) were used as positive controls.

qPCR analysis

Enterococci qPCR quantification was performed using USEPA 1609.1 method: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) with Internal

Amplification Control (IAC) Assay. Fig. 2 provides a categorical overview of steps followed for the implementation of the qPCR-based quantification method. The section below briefly describes the procedure followed.

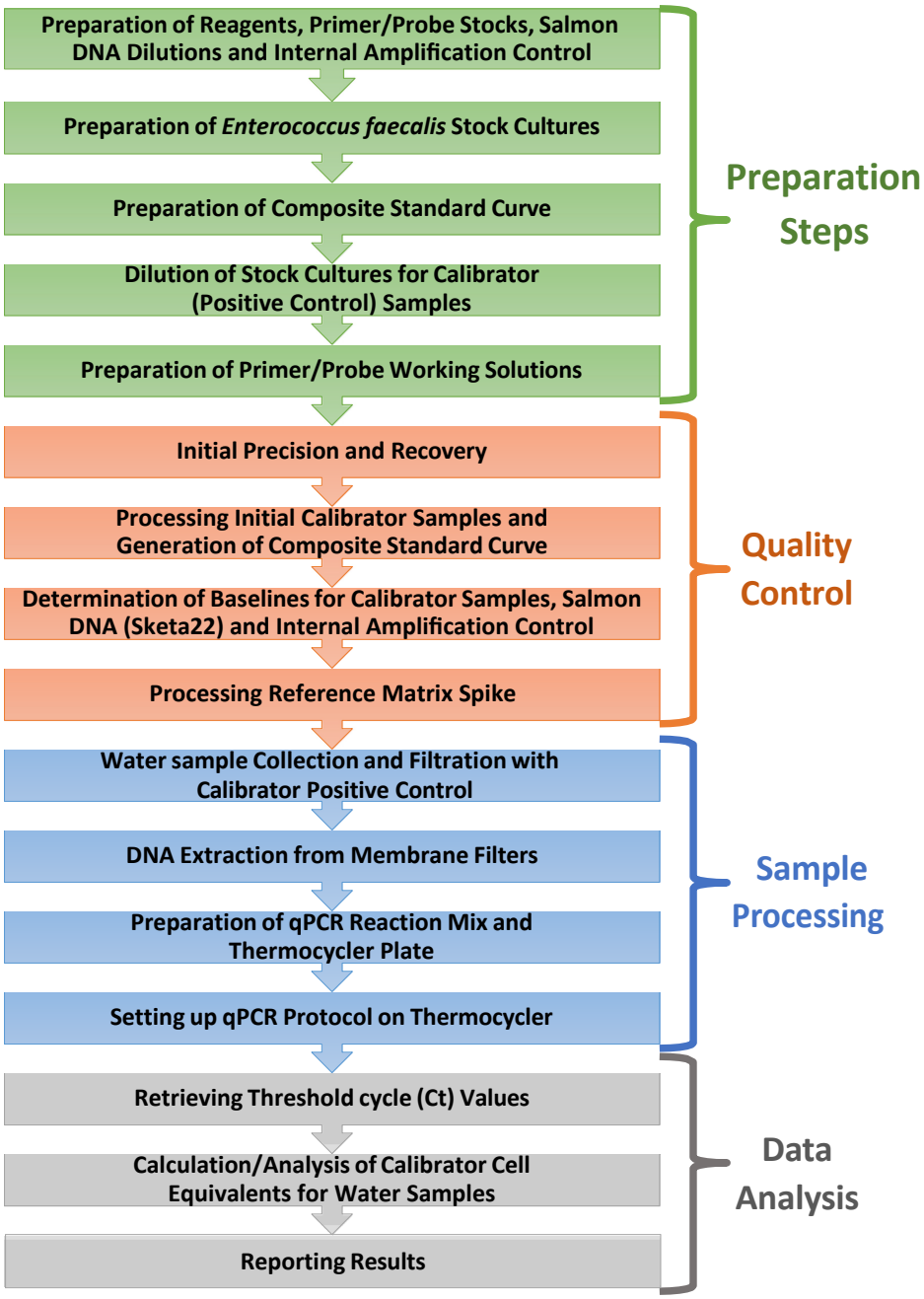


Figure 2. Stepwise flowchart for an implementation of the USEPA 1609.1 qPCR-based method for enterococci quantification in recreational water ecosystems.

Preparation of calibrator, standard curve, and quality controls

Calibrator-positive controls were prepared by culturing *Enterococcus faecalis* in brain heart infusion broth and then calculating colony-forming units (USEPA, 2012). Stock cultures of

Enterococcus faecalis were prepared and diluted to obtain cell suspensions with 10^9 CFUs for the preparation of genomic DNA standards and 104 CFUs for the preparation of calibrator and positive control samples as described in Method 1609.1. *E. faecalis* genomic DNA was extracted from a 10^9 CFU suspension using a DNA-EZ extraction kit (Generite, NJ, US) and was quantified spectrophotometrically at 260 nm. DNA standards were prepared from serial dilutions of this stock solution and target sequence copy (TSC) numbers in the standards were estimated from their concentration in the stock solution and the corresponding dilution factors, as also described in Method 1609.1. A composite standard curve was created from analyses of the standards in four separate qPCR runs using freshly prepared reagents for each run.

DNA extraction

DNA extraction (Target Sequence Copies) from water samples was performed using the bead beating method described in USEPA 1609.1 protocol (USEPA, 2015). Membrane filters were transferred to polypropylene bead tubes containing 0.3 gm of glass beads and 600 µL of DNA extraction buffer (0.2 µg/mL Salmon sperm DNA in Acetate-EDTA buffer pH 9), followed by bead beating at maximum speed for 60 s. After bead beating, tubes were centrifuged at 12,000 g for 60 s to settle down the beads and filter. The supernatant (400 µL) was collected and transferred to 1.5 mL centrifuge tubes, followed by a second round of centrifugation at 12,000 g for 5 min. Finally, ~350 µL of DNA extract was collected carefully to avoid pelleted debris and transferred to a 1.5 mL centrifuge tube to be used for qPCR. Furthermore, calibrator positive control (1 mL of 10^4 CFU *Enterococcus faecalis* dilution),

phosphate buffer saline matrix spike and reference matrix spike were processed for DNA extraction using the same protocol as the water samples.

qPCR amplification

The qPCR assay for target sequence copy (TSC) a 94 base pair sequence of *Enterococcus* large subunit ribosomal ribonucleic acid (lsrRNA) gene, while the sample processing control targeted a 77 base pair segment the rRNA gene operon from Salmon DNA (USEPA, 2015). The internal amplification control (IAC) assay template included a DNA sequence cloned into the pIDTSMART vector (IDT, Inc., Coralville, IA) that was amplified by the same primer set as the Enterococcus assay but contained a different probe recognition sequence, as specified in Method 1609.1. For each sample, separate reactions were performed for the enterococci/internal amplification control and sample processing control. The primer-probe working solution for the *Enterococcus*/IAC assay remained the same with a single primer set and two different probes for enterococci and internal amplification control qPCR assays. For each sample, separate reactions were performed with the multiplex *Enterococcus*/IAC assay and the simplex SPC assay as described in Method 1609.1. (USEPA, 2015). The qPCR assay mix for the *Enterococcus*/IAC assay was prepared by mixing 12.5 lL of TaqMan Environmental Master mix (TaqManTM), 2.5 lL of bovine serum albumin (2 mg/mL), 2 lL of internal amplification control and 3 lL of primer-probe working solution. For the sample processing control qPCR assay, the composition of the qPCR assay mix remained the same except 2 lL PCR-grade water was used instead of the internal amplification control. The final concentration of primers and probes in the qPCR assay mix was 1.0 lM and 80.0 nm, respectively. The thermocycler program comprised of one cycle of 50 C for 2 min and at 95

C for 10 min followed by 39 cycles at 95 C for 15 s, and 60 °C for 1 min. All reactions were performed in duplicates, and along with the water samples, each qPCR run included positive calibrator control and method blank reactions. All the qPCR reactions were performed in 96-well plates using a Bio-Rad CFX96 Touch Real-Time PCR Detection System thermocycler (Bio-Rad Inc. USA).

Data analysis

Enterococci calibrator cell equivalents were calculated using the comparative DDCt method using enterococci and sample processing control threshold cycle (Ct) values from each water sample (Haugland et al., 2016). Calculations were performed using the Excel datasheet provided by the U.S. EPA (https://www.epa.gov/cwa-methods/other-clean-water-act-test-methods-microbiological#file-183743). Average abundance was calculated by taking the arithmetic mean of enterococci calibrator cell equivalents (for qPCR) and colony-forming units (for culturing) from all the transects for a beach. Beach Action Values used for data interpretation included 235 CFUs/100 mL for *E. coli* enumeration, 70 CFUs/100 mL for enterococci enumeration and 1000 CCE/100 mL for enterococci qPCR (EPA, 2012). Culturing and qPCR data were log-transformed for the correlation analysis, followed by Shapiro-Wilk's statistics at p-value >0.05 to test normality distribution. Bivariate correlation between continuous measures of qPCR and culture results was performed by Pearson correlation with p-value <0.01 and 95% confidence interval. Statistical analyses were performed in R (version 3.6.3) with the stats, summarize, and EnvStats packages, while all the plots were generated with the ggplot2 package.

Results

Sample collection and processing time

Water samples were collected between 5:30–7:00 am for each sampling day and transported within 45 min to the laboratory for culture-based enumeration and qPCR-based testing. Results were typically provided to public health officials between 11:30 am and 12:00 noon. ESM Table S2 provides the time taken for each sample processing step for USEPA 1609.1 qPCR-based method. Laboratory sample processing and results reporting for the qPCR-based method required 3.5 h for each sampling day.

qPCR quality control data

The US EPA 1609.1 asks for six positive control (calibrator) and six negative control (method blank) sample analyses per run. Quality controls allow tracking of DNA recovery during water sample processing and testing for PCR inhibition due to the water matrix. The R^2 value for the composite standard curve was 0.992, while slope and intercept values were -3.25 and 38.66, respectively.

Sample processing and internal amplification controls

The determined mean water sample SPC assay cycle threshold (Ct) value was 22.2 ± 0.3 (Table 1). No sample exhibited a variation of >3 Ct for the sample processing control qPCR compared to positive calibrator controls. Furthermore, the internal amplification control was utilized to assess qPCR inhibition. An offset of >1.5 Ct for a water sample compared to a method blank was taken to indicate qPCR inhibition. The mean threshold cycle (Ct) value

for all the water samples was 31.1 ± 0.6, and no beach water samples gave results that failed the acceptability criterion (Table 1).

Table 1. qPCR data quality controls summary (mean values ± S.D.; n = 304).

qPCR Quality Analytics	Enterococci Threshold Cycle (Ct) Value	Internal Amplification Control (IAC) Threshold Cycle (Ct) Value	Sample Processing Control (SPC) Threshold Cycle (Ct) Value	Expected Results (USEPA 1609.1 Recommended)
Water Samples	34.8 ± 2.3	31.1 ± 0.6	22.2 ± 0.3	IAC Ct value: Offset of no more than 1.5 Ct in comparison to method blank. SPC Ct Value: Offset of no more than 3 Ct in comparison to the calibrator.
Calibrator Positive Control	30.3 ± 0.65	31.5 ± 0.9	21.5 ± 0.9	IAC Ct value: 30–33 SPC Ct value: 18–22
Reference (Phosphate Buffer Saline) Matrix Spike	30.8 ± 0.2	31.2 ± 0.5	21.3 ± 0.5	Enterococci Ct value calibrator positive control IAC Ct value: 30–33 SPC Ct value: 18–22
Method Blank	–	31 ± 0.5	21.2 ± 0.5	Enterococci Ct value = No amplification IAC Ct value: 30–33 SPC Ct value: 18–22

Calibrator positive control and reference (phosphate buffer saline) matrix spike

Altogether, thirty calibrator positive controls were extracted and used for qPCR (one for each qPCR run). The mean enterococci qPCR threshold value for positive calibrator controls

were 30.3 ± 0.65, while sample processing control and internal amplification control amplification means were 21.5 ± 0.9 and 21.5 ± 0.9, respectively (Table 1). All the calibrator positive controls showed high precision (mean coefficient of variation = 0.01%) in Ct value measurements for different sampling days. The phosphate buffer saline (PBS) matrix spike was used to estimate initial and ongoing precision recovery. For the phosphate buffer matrix saline, the mean *Enterococcus* qPCR Ct value was 30.8 ± 0.2, while the sample processing control and internal amplification control amplification means were 21.3 ± 0.5 and 31.2 ± 0.5, respectively (Table 1). For the reference (PBS) matrix spike, amplification Ct values for all three assays were closely associated with positive calibrator controls.

Method blank and non-template control

Method blank (six analyses per run) and non-template controls were used to identify contamination of reagents and buffers. One method blank was extracted for each new batch of phosphate buffer, and a non-template control reaction was performed alongside water samples for each qPCR run. Method blanks provided no amplification for enterococci qPCR, while mean Ct values for internal amplification control and sample processing control were 31 ± 0.5 and 21.2 ± 0.5, respectively (Table 1).

Association between qPCR-based and culturing-based quantification methods

E. coli colony forming units demonstrated a strong positive correlation with enterococci calibrator cell equivalents (Pearson correlation coefficient: 0.71–0.81, p-value <0.01) for all the sampling sites (Fig. 3). Similarly, enterococci culturing results also showed a strong

positive correlation (R: 0.65–0.86, p-value <0.01) with enterococci qPCR-based quantification for transects of both beaches and river reference sites (Fig. 4).

In terms of Beach Action Value (BAV) exceedance, *E. coli* colony forming units for Marie Curtis East beach exceeded the USEPA BAV (2:235 CFUs/100 mL) on eleven sampling days, while enterococci calibrator cell equivalents showed exceedances (2:1000 CCE/100 mL) for nine of the same days (Fig. 5). For Sunnyside beach, *E. coli* BAV exceedance was observed for three sampling days, while qPCR-based quantification demonstrated exceedance in two of the same three days (Fig. 6). In addition, on one sampling day (August 21, 2021), qPCR results exceeded the Beach Action Value, but *E. coli* enumeration results remained below the Beach Action Value. In terms of Beach Action Value exceedance, the qPCR-based method showed an 80–90% level of agreement with *E. coli* enumeration and >90% level of agreement with enterococci enumeration. For four sampling days, beach transects closer to river reference sites showed exceedance for both culturing-based and qPCR-based quantification methods, while transects at more distance did not exceed Beach Action Values.

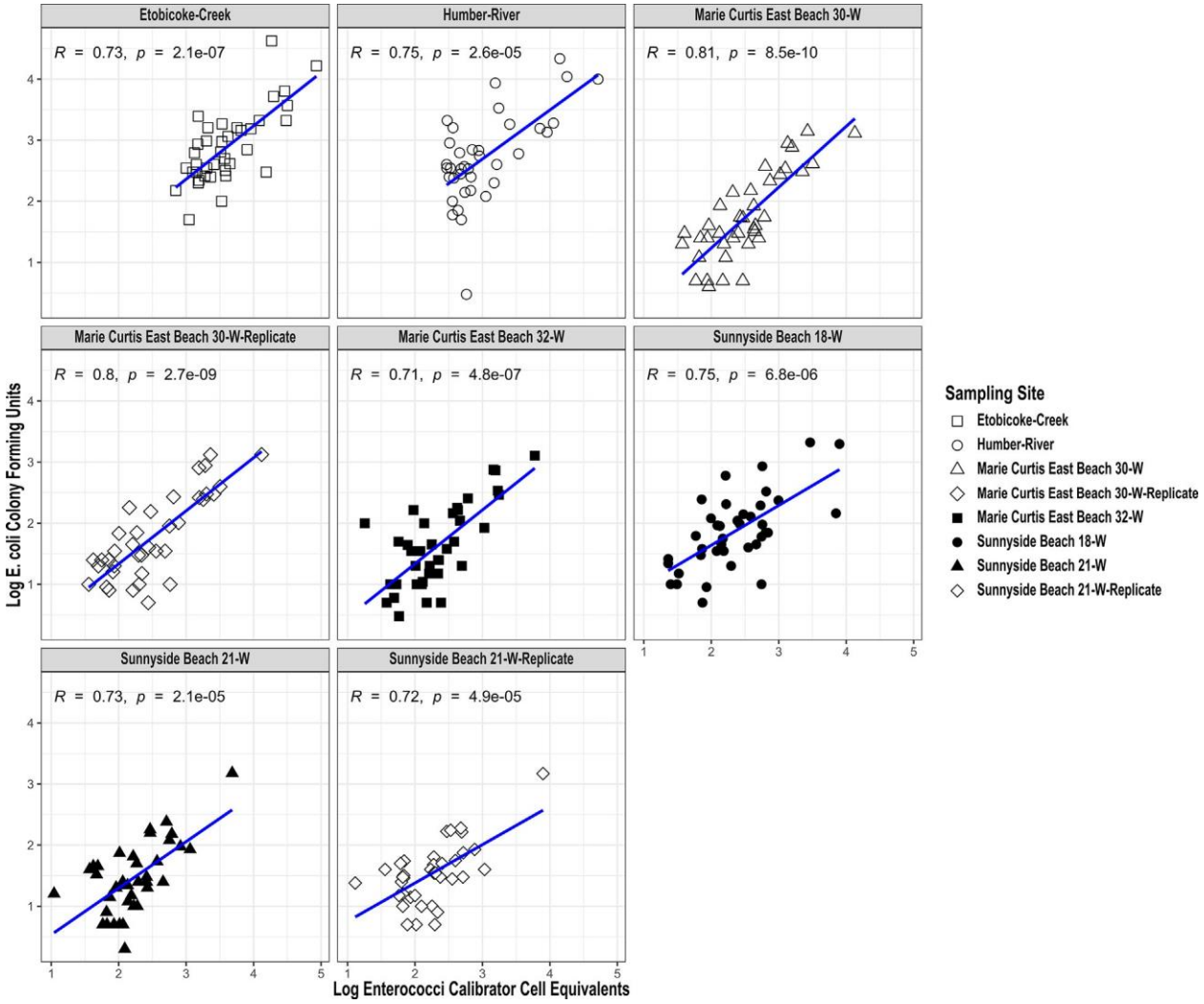


Figure 3. Comparison between log-transformed E. coli enumeration and qPCR-based enterococci quantification results for all the sampling sites using Pearson correlation at 95% confidence interval and p-value <0.01.

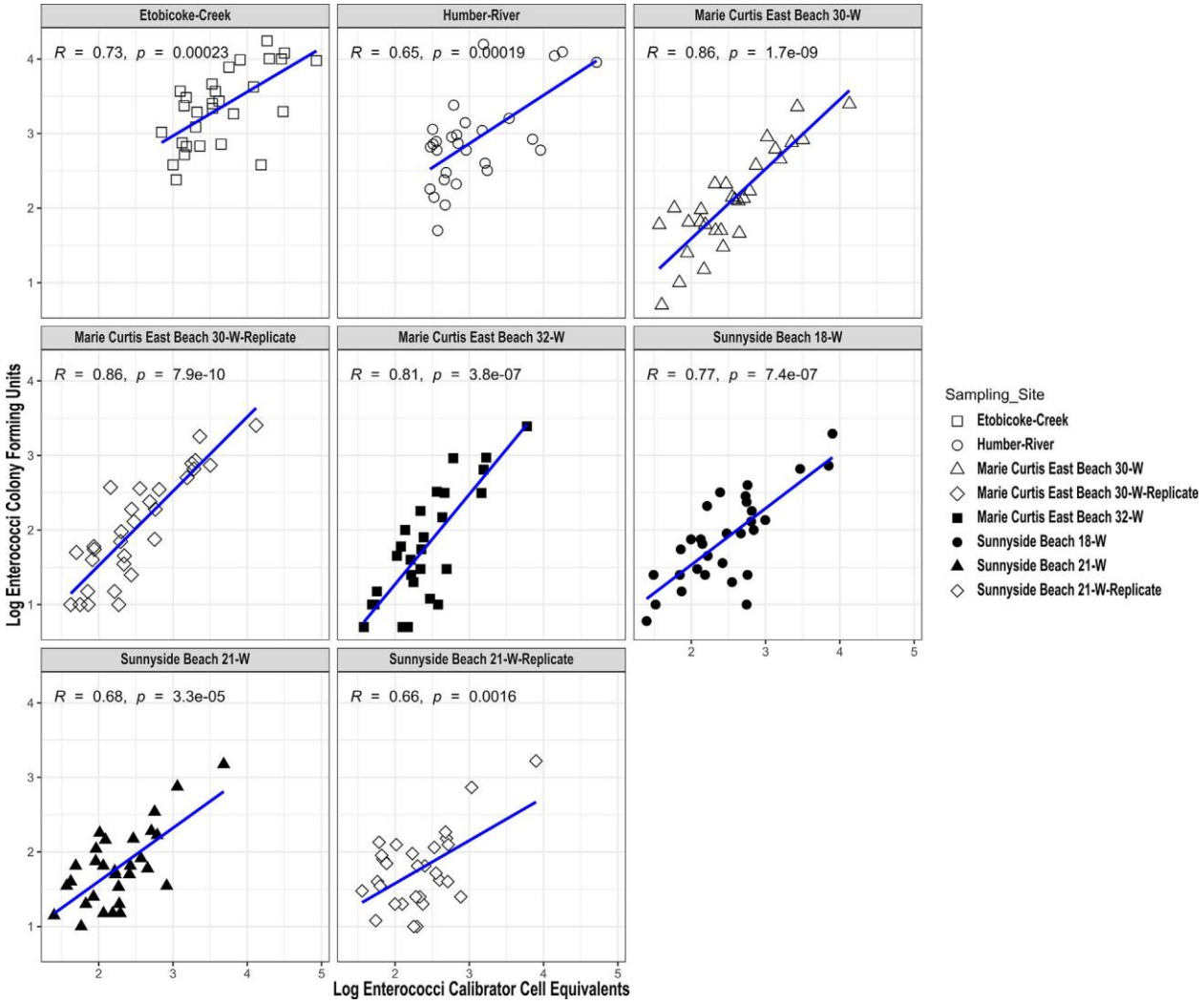


Figure 4. Comparison between log-transformed enterococci enumeration and qPCR-based Enterococcus quantification results for all the sampling sites using Pearson correlation at 95% confidence interval and p-value <0.01.

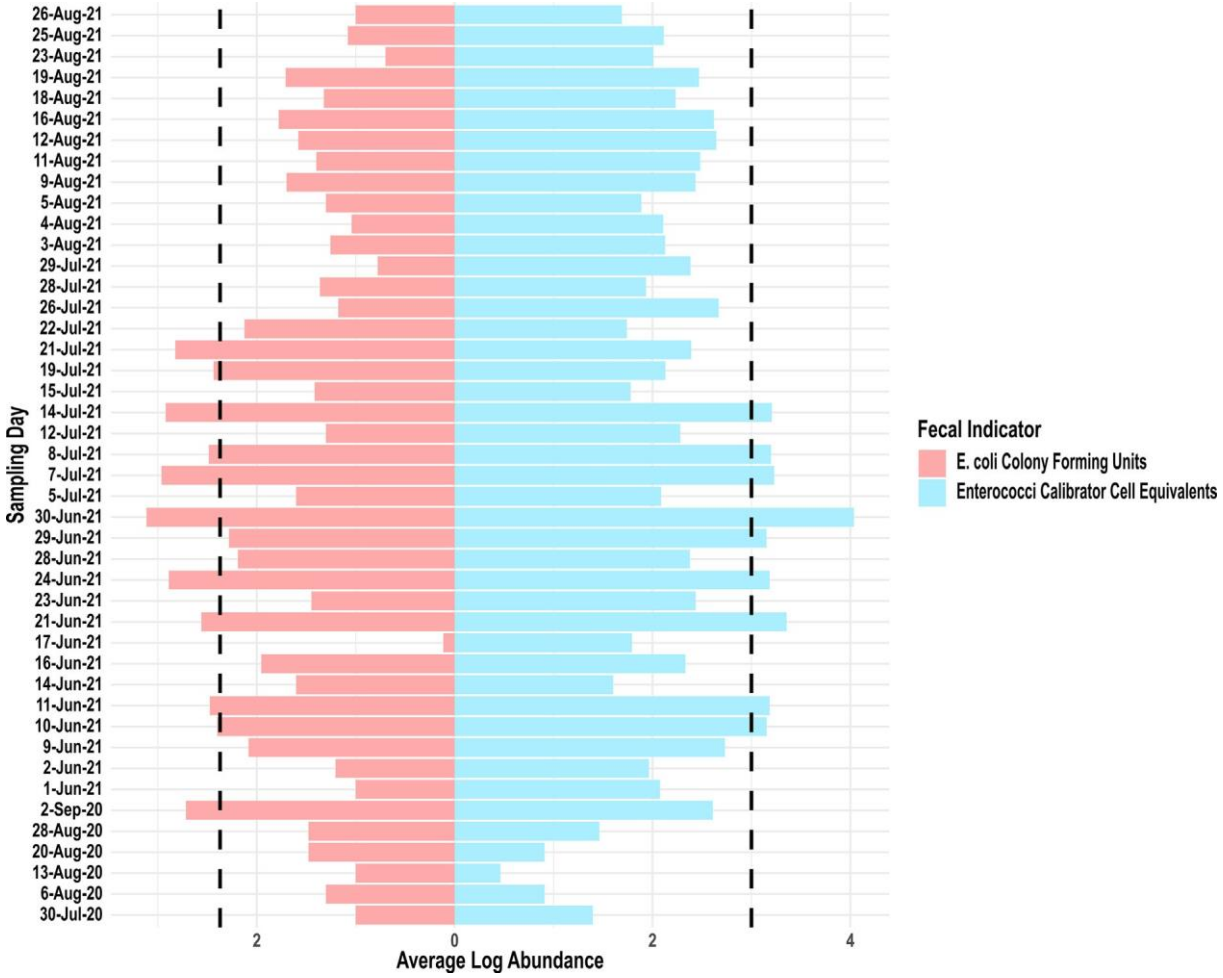


Figure 5. Divergence plot of log-transformed E. coli enumeration and qPCR-based Enterococcus average abundances for all the sampling days of Marie Curtis East Beach. Dashed line represents log-transformed E. coli enumeration Beach Action Value (235 CFUs/100m).

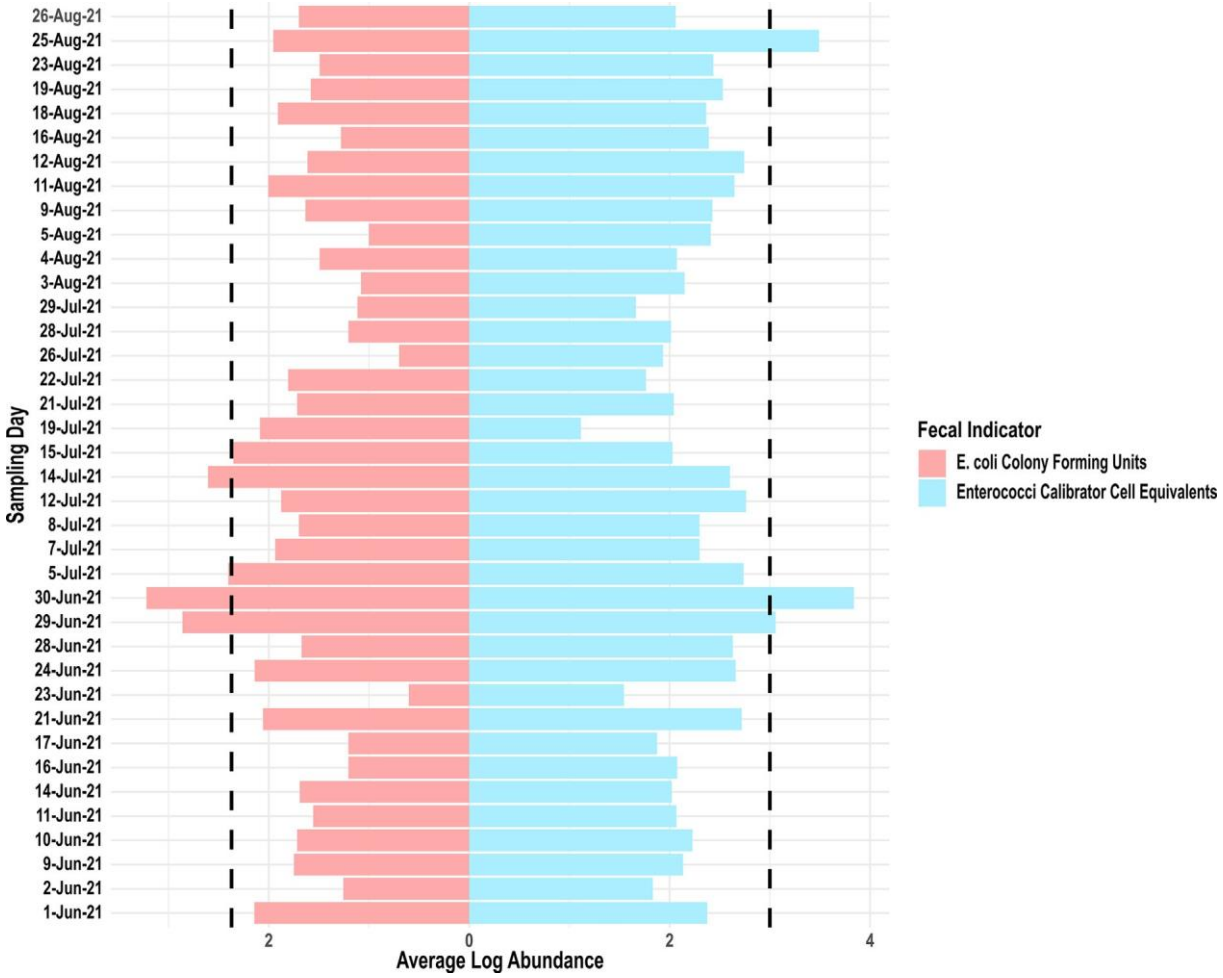


Figure 6. Divergence plot of log-transformed E. coli enumeration and qPCR-based Enterococcus average abundances for all the sampling days of Sunnyside Beach. Dashed line represents log-transformed E. coli enumeration Beach Action Value (235 CFUs/100 mL).

Quantification variability among biological replicates

Replicate water samples from both Marie Curtis and Sunnyside Beaches showed a strong positive correlation (R: 0.74–0.97, p-value <0.01) for qPCR-based and culturing-based quantification methods (Table 2). Replicates from Marie Curtis and Sunnyside Beaches demonstrated strong positive correlation coefficient values for qPCR and *E. coli* enumeration methods, and the coefficient of variation for both quantification methods was

within the range of 5–6%. (Table 2). However, enterococci by the culture at Sunnyside showed a comparatively lower correlation (R = 0.74, p-value <0.01) and higher mean coefficient of variation (9.92) for replicates. In Beach Action Value exceedance, replicates for *E. coli* and enterococci culture-based enumeration remained in agreement for all sampling days. However, on June 10, 2021, one Marie Curtis East Beach replicate (30 W) exceeded the qPCR Beach Action Value (1000 Calibrator cell equivalents CCE/100 mL), while the other (30 W-replicate) was below the Beach Action Value.

Table 2. Correlation and mean coefficient of variation between the biological replicates of water samples at two sampling sites for qPCR and culturing-based methods (n = 76 replicates/ site).

Replicate Sampling Site	Quantification Method	Pearson Correlation Coefficient (p-value <0.01)	Mean Coefficient of Variation (%)
Marie Curtis East Beach 30W	USEPA 1609.1 qPCR-Based	0.91	6.61
	E. coli Culturing	0.97	6.67
	Enterococci Culturing	0.96	6.55
Sunnyside Beach 21W	USEPA 1609.1 qPCR-Based	0.93	5.22
	E. coli Culturing	0.93	8.32
	Enterococci Culturing	0.74	9.92

Limitations of USEPA 1609.1 qPCR-based method

USEPA 1609.1 method utilizes a crude DNA extract for the qPCR amplification protocol, making it prone to potential amplification problems such as poor DNA recovery and qPCR inhibition. The amplification plot generated at the end of the qPCR run estimates data amplification quality and is a deciding factor for accepting Ct values. Our beach water

samples had good DNA recovery without qPCR inhibition, although qPCR inhibition was observed in two samples from the Humber River and one sample from Etobicoke Creek. However, our challenge testing of the *Enterococcus* qPCR method on the fifty water samples (frozen for one year at -20 °C) from the mouth of the Don River in Toronto found these samples showed poor DNA recovery. An ideal sample processing or DNA recovery provided Ct values in the range of 18–22 for Salmon DNA qPCR (Sample Processing Control) for water samples and quality control reactions (ESM Fig. S1). Water samples from the Don River challenge test that were frozen and then thawed for processing showed poor DNA recovery with a difference of up to 6 Ct compared to positive calibrator control (ESM Fig. S2). The internal amplification control in USEPA 1609.1 method is a non-target enterococci-specific sequence that utilizes the same primer set as the target sequence but with a different probe/fluorophore (ESM Fig. S1). Two of the challenge tests (Don River samples showed qPCR inhibition potentially due to water matrix interference. (ESM Fig. S3).

Consumables cost comparison between *E. coli* enumeration by culture and USEPA 1609.1 qPCR-based method

We found the USEPA 1609.1 qPCR-based method to have consumable costs about 6 times more per sample than *E. coli* enumeration by culture (ESM Table S3). The higher cost of the qPCR-based method is mainly due to TaqMan Probes and specialized environmental master mix solution. In addition to these consumable costs, the qPCR-based method requires use of a suitable thermocycler, a person with laboratory expertise in molecular methods and procedures to prepare master standard curves and quality control solutions. However, in the interest of public health and safety, the qPCR-based method allows more timely beach

posting decision-making to protect public health, potentially alleviating costs associated with illnesses from exposure to poor recreational water quality. In a study at Chicago beaches, it was estimated that under some beachgoer scenarios, these healthcare savings would exceed the cost of implementing a qPCR water quality monitoring program (Shrestha and Dorevitch, 2020).

Discussion

This study aimed to test the feasibility of using the USEPA 1609.1 qPCR method of enterococci quantification for water monitoring across different Toronto recreational water settings and water sampling regimes. For method validation, 405 water samples were collected over 44 beach days in the summer of years 2020 and 2021. Furthermore, to test the qPCR method, comparative *E. coli* and enterococci culturing assays were performed by conventional membrane filtration methods from the same water samples used for qPCR.

Application of the USEPA 1609.1 qPCR method at both Toronto beaches found the method met or exceeded all qPCR quality control criteria and provided results within 3.5 h of sample collection. *Enterococcus* qPCR results from biological replicate water samples taken at beach transects were highly correlated and had low coefficients of variation indicating high method reproducibility. The *Enterococcus* qPCR results were consistently reported to Toronto Public Health by 11:30 am or 12:00 noon each sampling day. This rapid reporting was consistent with another study that also found the method capable of providing same-day results within about 3.5 h and provided the basis for the City of Chicago to transition to qPCR method USEPA 1609.1 for monitoring water quality at its beaches (Dorevitch et al., 2017; Shrestha and Dorevitch, 2020).

Enterococci qPCR quantification results showed a significantly strong positive correlation with *E. coli* and enterococci culturing results at all our beach sites and the Etobicoke Creek and Humber River recreational settings. However, our beach results differ from the study performed on Chicago beaches, which showed a more moderate positive correlation between enterococci qPCR and *E. coli* culturing results (Dorevitch et al., 2017). Marie Curtis East and Sunnyside Beaches are impacted directly and quickly by river sources as both beaches are present at the mouths of Etobicoke Creek and Humber River, respectively. This might explain the higher correlation we obtained between enumeration and qPCR-based methods.

Regarding Beach Action Value exceedance, qPCR-based results showed 80–90% agreement with *E. coli* enumeration and >90% with enterococci enumeration. Differences may be due to differences in survival characteristics between enterococci and *E. coli*, and that qPCR results are enumerating both viable and dead enterococci cells. Our results are consistent with previous studies that demonstrated an >80% level of agreement between same-sample *Entero- coccus* qPCR-based and *E. coli* culturing-based Beach Action Value exceedances (Haugland et al., 2014; Lavender and Kinzelman, 2009). In addition, consistent with local river impacts at these beaches, we observed that on four sampling days, beach transects closer to river mouths showed exceedance in enterococci qPCR Beach Action Values, while more distal transects did not. Beaches impacted by rivers (particularly urban ones) are known to respond quickly to increased fecal indicator levels (Aragonés et al., 2016; Molina et al., 2014; Zhang et al., 2020), and microbial source tracking research has shown the prominence of fecal pollution from the Humber River impacting Sunnyside Beach closer to the river mouth at times (Edge et al., 2010). In this study, biological replicates taken from the same

beach transects had low coefficients of variation, indicating high reproducibility. However, careful consideration should be taken when reporting results on average from different sampling points or replicates as the beach transects closer to the river's mouth can result in a higher average fecal indicator count than ones at more distance.

A further method challenge assessed whether a water sampling regime that froze (-20 C) recreational water samples before qPCR analysis could be used. Analysis of these water samples that were frozen before processing found poor DNA recovery. However, water samples can be filtered and then the filters can be stored at -80 °C until processed (Aw et al., 2019). These results were consistent with guidance in the USEPA 1609.1 qPCR method and indicated that water sampling regimes should not freeze water samples prior to using this method.

Poor DNA recovery and PCR inhibition can be challenges associated with qPCR-based methods (Green and Field, 2012). While DNA recovery from all beach water samples was efficient in our study, several water samples from the mouths of Etobicoke Creek and the Humber River were inhibited and all previously frozen samples collected from the Don River mouth were not amenable to PCR. USEPA 1609.1 method recommends against frozen samples for qPCR testing, which our results support when considering the design of recreational water sampling regimes. Our inhibition results from several river mouth water samples are probably indicative of these recreational water settings being exposed to higher concentrations of fecal and other contaminants that could compromise PCR assays. Poor DNA recovery can be attributed to interferences caused by water matrix contaminants (Haugland et al., 2016). The absence of purification step in DNA extraction protocol of

USEPA 1609.1 method makes it more prone to qPCR inhibition by water matrix contaminants. Water matrix contaminants including humic acids, tannic acids, complex carbohydrates, and kaolinite clay particles can sequester DNA from polymerase, thus leading to qPCR inhibition (Albers et al., 2013; Noble et al., 2010; Schang et al., 2016; Seruge et al., 2019). Therefore, the reference matrix spike should be tested for each new water source to be analyzed using the USEPA 1609.1 method. Water samples with an offset of >1.5 Ct for internal amplification qPCR compared to the method blank should be marked positive for qPCR inhibition. In this case, a five-fold dilution of DNA extract is recommended in USEPA 1609.1 protocol alongside undiluted DNA extract for each new water source to check for qPCR inhibition due to water matrix contaminants (USEPA, 2015).

A previous study from Chicago beaches concluded that no meaningful agreement (beyond that expected by chance) was observed between beach posting decisions driven by *Enterococcus* qPCR data and 24 h old *E. coli* culture-based data. The goal of our study was to conduct a comparative method analysis for the beach water samples based on same-sample results rather than assessing beach posting decision making, which is based on 24 h delayed enumeration results. While our *Enterococcus* qPCR results are comparable to our culture-based *E. coli* and enterococci results from the same water sample, comparisons to Toronto Public Health's 24 h old culture *E. coli* data would likely reveal more different beach posting implications.

Conclusions

1. The USEPA 1609.1 qPCR-based method for enterococci quantification provided rapid same-day water quality data for two Toronto beaches.
2. The *Enterococcus* qPCR method met all qPCR quality control criteria and biological water sample replicates at beach sites had coefficients of variation as low as or lower than *E. coli* and enterococci culture-based methods. PCR inhibition was a problem for several river mouth recreational water samples, and the qPCR method was found to be not applicable for numerous previously frozen (-20 C) river mouth water samples.
3. The *Enterococcus* qPCR method provided results within 3.5 h of sample reception in the lab.
4. *Enterococcus* qPCR results were strongly correlated with both *E. coli* and enterococci culturing-based results from the same water samples.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Alex Chan, James Naphtali, Ada Cheung and Enze Li for their help during method testing. Cheryl Callen delivered samples from the collection site to the lab. Mahesh Patel from Toronto Public Health provided guidance and support in data acquisition. Albert

Simhon, Ontario Ministry of Environment, Conservation, and Parks (MECP), provided support and suggestion throughout the project. The MECP provided funding for this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jglr.2022.02.008.

References

Albers, C.N., Jensen, A., Bælum, J., Jacobsen, C.S., 2013. Inhibition of DNA polymerases used in Q-PCR by structurally different soil-derived humic substances. Geomicrobiol. J. 30 (8), 675–681.

APHA, 2012. Standard Methods for the Examination of Water and Wastewater. American Public Health Association. American Water Works Association, Water Environment Federation.

Aragonés, L., López, I., Palazón, A., López-Úbeda, R., García, C., 2016. Evaluation of the quality of coastal bathing waters in Spain through fecal bacteria *Escherichia coli* and *Enterococcus*. Sci. Total Environ. 566-567, 288–297.

Aw, T.G., Sivaganesan, M., Briggs, S., Dreelin, E., Aslan, A., Dorevitch, S., Shrestha, A., Isaacs, N., Kinzelman, J., Kleinheinz, G., Noble, R., Rediske, R., Scull, B., Rosenberg, S., Weberman, B., Sivy, T., Southwell, B., Siefring, S., Oshima, K., Haugland, R., 2019. Evaluation of multiple laboratory performance and variability in analysis of recreational

freshwaters by a rapid *Escherichia coli* qPCR method (Draft Method C). Water Res. 156, 465–474.

Converse, R.R., Griffith, J.F., Noble, R.T., Haugland, R.A., Schiff, K.C., Weisberg, S.B., 2012. Correlation between quantitative PCR and culture-based methods for measuring *Enterococcus* spp. over various temporal scales at three California marine beaches. Appl. Environ. Microbiol. 78 (4), 1237–1242.

Dorevitch, S., Shrestha, A., DeFlorio-Barker, S., Breitenbach, C., Heimler, I., 2017. Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: Implications for public notification. Environ. Health 16, 1–10.

Edge, T., Hill, S., Seto, P., Marsalek, J., 2010. Library-dependent and library- independent microbial source tracking to identify spatial variation in faecal contamination sources along a Lake Ontario beach (Ontario, Canada). Water Sci. Technol. 62, 719-727.

Edge, T.A., Boyd, R.J., Shum, P., Thomas, J.L., 2021. Microbial source tracking to identify fecal sources contaminating the Toronto Harbour and Don River watershed in wet and dry weather. J. Great Lakes Res. 47 (2), 366–377.

EPA, U.S., 2012. Recreational water quality criteria, in: Agency, U.S.E.P. (Ed.). Ferguson, D.M., Griffith, J.F., McGee, C.D., Weisberg, S.B., Hagedorn, C., 2013.

Comparison of *Enterococcus* species diversity in marine water and wastewater using Enterolert and EPA Method 1600. J. Environ. Public Health 2013.

Green, H.C., Field, K.G., 2012. Sensitive detection of sample interference in environmental qPCR. Water Res. 46 (10), 3251–3260.

Haugland, R.A., Siefring, S., Varma, M., Oshima, K.H., Sivaganesan, M., Cao, Y., Raith, M., Griffith, J., Weisberg, S.B., Noble, R.T., Blackwood, A.D., Kinzelman, J., Anan'eva, T., Bushon, R.N., Stelzer, E.A., Harwood, V.J., Gordon, K.V., Sinigalliano, C., 2016. Multi-laboratory survey of qPCR enterococci analysis method performance in US coastal and inland surface waters. J. Microbiol. Methods 123, 114–125.

Haugland, R.A., Siefring, S.C., Wymer, L.J., Brenner, K.P., Dufour, A.P., 2005. Comparison of *Enterococcus* measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis. Water Res. 39 (4), 559–568.

Haugland, R.A., Siefring, S.D., Varma, M., Dufour, A.P., Brenner, K.P., Wade, T.J., Sams, E., Cochran, S., Braun, S., Sivaganensan, M., 2014. Standardization of enterococci density estimates by EPA qPCR methods and comparison of beach action value exceedances in river waters with culture methods. J. Microbiol. Methods 105, 59–66.

Heasley, C., Sanchez, J.J., Tustin, J., Young, I., Yaseen, Z.M., 2021. Systematic review of predictive models of microbial water quality at freshwater recreational beaches. PLoS ONE 16 (8), e0256785.

Kidd, J., 2015. Toronto and Region Remedial Action Plan Progress Report. Toronto and Region Conservation Authority.

Lavender, J.S., Kinzelman, J.L., 2009. A cross comparison of QPCR to agar-based or defined substrate test methods for the determination of *Escherichia coli* and enterococci in municipal water quality monitoring programs. Water Res. 43 (19), 4967–4979.

Molina, M., Hunter, S., Cyterski, M., Peed, L.A., Kelty, C.A., Sivaganesan, M., Mooney, T., Prieto, L., Shanks, O.C., 2014. Factors affecting the presence of human- associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. Water Res. 64, 196–208.

Nappier, S.P., Ichida, A., Jaglo, K., Haugland, R., Jones, K.R., 2019. Advancements in mitigating interference in quantitative polymerase chain reaction (qPCR) for microbial water quality monitoring. Sci. Total Environ. 671, 732–740.

Noble, R.T., Blackwood, A.D., Griffith, J.F., McGee, C.D., Weisberg, S.B., 2010. Comparison of rapid quantitative PCR-based and conventional culture-based methods for enumeration of *Enterococcus* spp. and *Escherichia coli* in recreational waters. Appl. Environ. Microbiol. 76 (22), 7437–7443.

Schang, C., Henry, R., Kolotelo, P.A., Prosser, T., Crosbie, N., Grant, T., Cottam, D., O'Brien, P., Coutts, S., Deletic, A., McCarthy, D.T., Bilotta, G.S., 2016. Evaluation of techniques for measuring microbial hazards in bathing waters: A comparative study. PLoS ONE 11 (5), e0155848.

Seruge, J., Wong, M., Noble, R.T., Blackwood, A.D., Moravcik, P.S., Kirs, M., 2019. Application of a rapid qPCR method for enterococci for beach water quality monitoring

purposes in Hawaii: Loss of DNA during the extraction protocol due to coral sands. Mar. Pollut. Bull. 149, 110631.

Shrestha, A., Dorevitch, S., 2019. Evaluation of rapid qPCR method for quantification of *E. coli* at non-point source impacted Lake Michigan beaches. Water Res. 156, 395–403.

Shrestha, A., Dorevitch, S., 2020. Slow adoption of rapid testing: Beach monitoring and notification using qPCR. J. Microbiol. Methods 174, 105947.

USEPA, 2009. Report on 2009 National Epidemiologic and Environmental Assessment of Recreational Water Epidemiology Studies.

USEPA, 2012. Recreational water quality criteria. United States (US) Environmental Protection Agency (EPA), 1200 Pennsylvania .. ., p. 42.

USEPA, 2015. Method 1609.1: Enterococci in Water by TaqMan® Quantitative

Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay. US EPA.

Wade, T.J., Calderon, R.L., Brenner, K.P., Sams, E., Beach, M., Haugland, R., Wymer, L., Dufour, A.P., 2008. High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. Epidemiology, 375-383.

Wade, T.J., Sams, E., Brenner, K.P., Haugland, R., Chern, E., Beach, M., Wymer, L., Rankin, C.C., Love, D., Li, Q., Noble, R., Dufour, A.P., 2010. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: a prospective cohort study. Environ. Health 9 (1), 1–14.

Zhang, X., Zhi, X., Chen, L., Shen, Z., 2020. Spatiotemporal variability and key influencing factors of river fecal coliform within a typical complex watershed. Water Res. 178, 115835

Chapter 3. Same-day *Enterococcus* qPCR strategy provides added public health protection and reduced beach days lost

This chapter evaluates the impacts of implementing a rapid, same-day qPCR method for *Enterococcus* quantification at recreational freshwater beaches. This chapter provides a compelling case for the transition from conventional culture-based methods to rapid, precise molecular techniques, which is essential for achieving timely and effective water quality monitoring, ultimately contributing to the sustainability and safe enjoyment of recreational waters. Integrating the insights from this study into the broader context of the thesis enhances the argument for adopting advanced molecular strategies in recreational water management.

*The text I present here is a **peer-reviewed manuscript published** in the Canadian Journal of Public Health. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Schellhorn, H.E., Simhon, A. and Edge, T.A., 2023. Same-day Enterococcus qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. Canadian Journal of Public Health, 114(4), pp.676-687.*

Author Contributions

FS performed the experiments, analyzed the data, communicated the findings with Public Health Agencies, Wrote the original manuscript, and revised it as required by the reviewers. TAE collected the water samples and reviewed the manuscript before submission to the

journal. HES supervised the study and reviewed the manuscript before submission. AAS reviewed the manuscript and overlooked the project.

Abstract:

Objectives:

We evaluated the potential for using a rapid same-day qPCR monitoring method for beach posting outcomes at two Toronto beaches.

Methods:

In total, 228 water samples were collected at Marie Curtis Park East and Sunnyside Beaches over the 2021 summer season. Water samples were processed using the USEPA 1609.1 *Enterococcus* qPCR-based method. *E. coli* culture data and daily beach posting decisions were obtained from Toronto Public Health.

Results:

No significant correlation was observed between Previous-Day and Same-Day (retrospective) *E. coli* enumeration results at any Sunnyside Beach site, and only relatively low (R= 0.41-0.56) or no significant correlation was observed at sampling sites for Marie Curtis Park East Beach sites. Comparing our Same-day enterococci qPCR data to Toronto's two-day *E. coli* geomean beach posting decisions, we noted postings for 1 (2%) and 3 (8%) missed health-risk days at Sunnyside and Marie Curtis Park East Beaches respectively. The qPCR data also pointed to incorrect postings for 12 (31%) and 6 (16%) lost beach days at Sunnyside and Marie Curtis Park East Beaches respectively.

Conclusions

Application of a rapid enterococci qPCR method at two Toronto beaches revealed 5% of beach posting decisions were false negatives that missed health-risk days, while 23% of

decisions were false positives resulting in lost beach days. Thus, deployment of the rapid Same-day qPCR method offers the potential to reduce both health risks and unnecessary beach postings.

Keywords: Beach Monitoring, qPCR, Fecal Indicator, *E. coli*, enterococci, Toronto.

Introduction:

Fecal indicator bacteria such as *E. coli* and *Enterococcus* are commonly used in North America for water quality monitoring of freshwater and marine beaches, respectively (Health Canada, 2012; USEPA, 2012), and their elevated levels are associated with increased gastrointestinal illnesses in swimmers (USEPA 2012). Public health authorities monitor beaches for fecal indicator bacteria and make beach posting decisions according to water quality thresholds and Beach Action Values defined in recreational water quality criteria and guidelines (US EPA, 2012; Health Canada, 2012). Canadian public health authorities currently rely mainly on culture-based *E. coli* enumeration for beach monitoring, and beaches are posted to advise against swimming if the geometric mean of *E. coli* colony forming units (CFUs) exceeds 200 *E. coli* CFU/100mL (Health Canada, 2012). However, some jurisdictions (e.g., Toronto Public Health) continue to use a more stringent geometric mean of 100 *E. coli* CFUs/100mL as a beach posting threshold.

Although *E. coli* enumeration by culture is practical and cost-effective for beach monitoring, the results are available at the earliest 18-24h after sample collection (Shrestha and Dorevitch 2019). As a result, beach posting decisions are delayed and are at best made using previous-day's *E. coli* enumeration data. Assumptions regarding recreational water quality over 24 hours can lead to two types of errors: fecal indicator density may have increased (false-negatives) or decreased (false-positives), which can impact beach postings, respectively. False negatives results can lead to an incorrect assumption that fecal indicator densities are lower than the beach quality threshold and the beaches remain open for public. Rapid methods such as qPCR (Quantitative PCR) provide same-day results have been shown to

provide valuable same-day fecal indicator density for beach management *Enterococcus* (Dorevitch et al. 2017).

In 2015 USEPA developed a standard qPCR-based rapid method that can be used for fecal indicator monitoring of recreational water ecosystems (USEPA 2015). USEPA method 1609.1 is a qPCR-based method for *Enterococcus* quantification in both marine and freshwater. It offers several advantages over culture-based enumeration-based methods. First, the method can provide results within only 3.5-4h of sample receipt in the lab. Secondly, beach managers can make timely beach posting decisions using qPCR data on the same day of water sample collection. Third, the National Environmental and Epidemiological Assessment of Recreational water (NEEAR) study has shown that qPCR-based *Enterococcus* quantification measures can be better predictors of gastrointestinal illnesses for freshwater and marine beaches (Wade et al. 2008; Wade et al. 2010; USEPA 2009). Importantly, Method 1609.1 also provides minimum standards and rigorous quality control measures to ensure reliability of data for beach posting decision-making. Based on the NEEAR study, the USEPA has established a Beach Action Value (BAV) of ≥1000 calibrator cell equivalents (CCEs)/100mL for beach posting decisions using the 1609.1 qPCR-based method. However, the USEPA qPCR method 1609.1 for *Enterococcus* has yet to be widely adopted in beach water quality monitoring programs (Shrestha and Dorevitch, 2020).

Despite the growing interest in rapid beach monitoring methods, and the Canadian Province of Alberta adopting qPCR for *Enterococcus* at beaches (Alberta, 2021), most public health authorities in Canada still rely upon 24 hour-old results from culture-based *E. coli*

enumeration for beach posting decision-making. The city of Toronto has ten freshwater beaches along a 42 km Lake Ontario waterfront, among which, Sunnyside and Marie Curtis Park East beaches are posted more frequently Toronto Public Health oversees daily water quality testing at these beaches and makes beach posting decisions using an *E. coli* culture-based enumeration method. As a result, beach posting decisions are delayed about 24h and are based on previous-day water quality conditions. The city is unique in using a two-day rolling geometric mean of 100 *E. coli* CFU/100mL for beach posting decisions. This study builds upon a previous study that provided a validation for applying method 1609.1 at Toronto beaches (Saleem et al. 2022). The primary objectives for our study were: i) Testing the correlation between Previous-day and Same-day (retrospective) *E. coli* enumeration results at two Toronto beaches; ii) Assessing the potential impacts of Same-day *Enterococcus* qPCR results compared to Toronto Public Health's two-day *E. coli* geometric mean beach posting decisions in 2021.

Methods:

Water Sample Collection and Filtration Processing:

Water samples were collected at chest depth at two freshwater beaches during the summer of 2021. Marie Curtis Park East Beach is located in the west end of Toronto at the mouth of Etobicoke Creek and is exposed to the open waters of Lake Ontario. Sunnyside Beach is also in the west end of Toronto at the mouth of the Humber River, though a break-wall protects it from wave action on Lake Ontario. At each beach, two representative transects were selected from the City of Toronto Public Health's beach sampling transects that are regularly tested for *E. coli* enumeration by public health authorities (Supplementary Table 1). Water

samples were collected on three days (two consecutive days) each week (Mondays, Wednesdays and Thursdays) over the bathing season such that 228 water samples on 38 beach days were collected from Marie-Curtis Park East Beach (30W and 32W transects, Figure 1 and supplementary table 1), and Sunnyside Beach (18W and 21W transects, Figure 1 and supplementary table 1). Grab water samples were collected just below the water surface in 1L sterile screw-capped polyethylene terephthalate (PET) bottles. Water samples were collected between about 5:30am and 7:00am, stored on ice, and transported to the laboratory by 8:00am. Upon receipt, water samples were processed by USEPA Method 1609.1 for *Enterococcus* qPCR (USEPA, 2015). Briefly, 100 mL (0.1 L) of water sample was passed through 0.45 µm polycarbonate membrane filter (Millipore Corp., Bedford, MA) for collection of bacterial biomasses. Membrane filter was subsequently bead beaten with extraction buffer (0.2 µg/mL Salmon sperm DNA in Acetate-EDTA buffer pH 9) and centrifuged at 12,000 g for 1 minute, followed by collection of 400 µL supernatant. Collected supernatant was centrifuged again at 12000 g for 5 minutes and 350 µL of DNA extract was collected for qPCR.

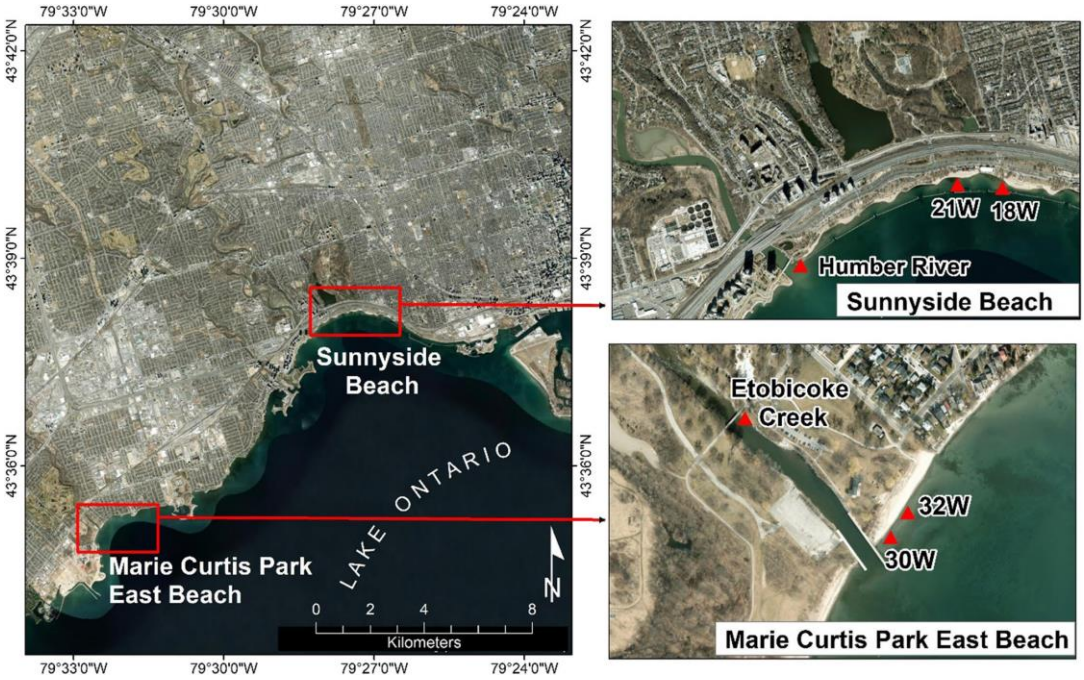


Figure 1. Geographical map of western Toronto, Ontario, and beach sampling sites and water sources (rivers) for Marie Curtis Park East and Sunnyside Beaches.

qPCR Quality Controls:

A standard curve was prepared following USEPA 1609.1 by performing ten-fold dilutions of DNA extracted from 10^9 *Enterococcus faecalis* CFUs (ATCC 29212). The DNA standard with the highest concentration comprised 40,000 target sequence copies (TSC)/5µL, while the DNA standard with the lowest TSC represented the lowest limit of standard curve quantification (LLOSQ = 10 TSC/5µL). A calibrator positive control for qPCR was prepared by diluting *E. faecalis* stock culture with sterile phosphate buffer saline (pH:7.5) to 10^4 *Enterococcus faecalis* CFUs. The phosphate buffer saline matrix spike was prepared by dispensing 1000 µL of calibrator positive control solution into 100mL buffer. The water reference matrix spike for each sampling site included 100mL of water sample spiked with positive calibrator control. The non-template control included a qPCR reaction mixture with

5µL of sterile water instead of sample DNA. Detailed descriptions of quality controls can also be accessed in our publication (Saleem et al. 2022).

qPCR Reaction Mixture and Amplification Protocol:

USEPA 1609.1 method was used for the qPCR analysis of *Enterococcus* quantification in water samples (USEPA 2015). The reaction mixture (25 µL) included 12.5 µL of TaqMan Environmental Master mix (TaqMan^{TM}), 5 µL of template DNA, 2.5 µL of bovine serum albumin (2.0 mg/mL), 3.0 µL of primer/probe working solution (Final concentration of primers and probe was 1.0 µM and 80.0 nm, respectively), and 2.0 µL of sterile nuclease-free water. The primer/probe working solution was prepared by diluting stock solutions (Primer stock: 500µM and Probe stock: 100 µM) with sterile nuclease-free water. Each qPCR run comprised eight water samples, positive calibrator controls, non-template controls, and phosphate buffer saline matrix spike was used once a week for ongoing precision recovery analysis. qPCR reactions were carried out in 96-well qPCR plates (Corning Inc., USA) on a Bio-Rad CFX96 Touch Real-Time PCR Detection System thermocycler (Bio-Rad Inc. USA). Detailed description of the qPCR protocol can be accessed (Saleem et al. 2022).

Data Analysis:

Daily culture-based *E. coli* enumeration data from June 01, 2021, to August 31, 2021, was kindly shared by Toronto Public Health for all sampling transects at Sunnyside and Marie Curtis Park East Beaches. Data received from the public health authority also included the information about which beach days were posted for summer 2021. USEPA 1609.1

Enterococcus qPCR protocol quantifies the ratio of DNA target sequence from calibrator positive control and water samples which normalizes for differences in DNA recovery using sample processing control (Salmon sperm DNA; spiked in samples before DNA extraction) (USEPA, 2015; Haugland et al., 2016). All the qPCR calculations were performed using standard excel sheet provided by USEPA (https://www.epa.gov/sites/default/files/2015-08/methods_1609-1-1611-1-calculation-spreadsheet-april-2015.xlsx). *E. coli* enumeration and *Enterococcus* qPCR data were log-transformed. Toronto Public Health's culture-based *E. coli* enumeration data (after cleanup by removal of data points lower than or equal to lower limit of quantification; ≥10 CFUs/100 mL) was used for correlation testing between their Previous-day (Results from first day in two consecutive days of sample collection) and Same-day (Results from second day in two consecutive days) *E. coli* results (recognizing that Same-day culture-based *E. coli* data is only possible from a retrospective perspective) were plotted for correlation using Shapiro-wilk's for normality testing ($p>0.05$), followed by Pearson correlation analysis with p-value cutoff ≤0.05 at 95% confidence interval (http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r). GGpubr (ggscatter) (Kassambara and Kassambara, 2020) and GGplot2 (Wikham et al., 2016) packages were used for the construction of correlation plots, followed by adjustment of p-values using p.adjust function in R (Jafari and Ansari-Pour, 2019).

Results:

qPCR Quality Analysis:

In total, 228 water samples were collected, corresponding to thirty-eight beach days for summer 2021. Method 1609.1 met all standards and quality control criteria, and Table 1

describes the qPCR parameters for method quality assessment. All the beach samples passed quality parameters for good DNA recovery (Sample processing control <3 Ct difference in comparison to calibrator positive controls) and qPCR inhibition (Internal amplification control within 1.5 Ct difference in comparison to non-template control).

Table 1. Data Quality Measures for Quality Control Analysis for qPCR.

Data Quality Parameters		
Number of Standard Curves	**4**	
R^2 (Mean ± Standard Deviation)	0.998 ± 0.003	Standard Curves Quality Analysis
No. of Calibrators	**36**	
Enterococci qPCR Threshold Cycle Value (Mean ± Standard Deviation)	31.3 ± 0.7	
Internal Amplification Control (IAC) qPCR Threshold Cycle Value (Mean ± Standard Deviation)	31.2 ± 0.5	
Sample Processing Control qPCR Threshold Cycle Value (Mean ± Standard Deviation)	21.5 ± 0.9	Calibrator Positive Controls Quality Analysis
Number of Samples	**228**	
Enterococci qPCR Threshold Cycle Value (Mean ± Standard Deviation)	34.8 ± 2.3	
Internal Amplification Control qPCR Threshold Cycle Value (Mean ± Standard Deviation)	31.1 ± 0.6	
Sample Processing Control (SPC) qPCR Threshold Cycle Value (Mean ± Standard Deviation)	22.2 ± 0.3	Beach Water Samples Quality Analysis
qPCR Inhibition (IAC Ct > 1.5)	0	
Poor DNA Recovery (SPC Ct > 3)	0	Beach Water Samples qPCR Analysis

Correlation Between Previous-day and Retrospective Same-day *E. coli* Enumeration:

E. coli culture-based enumeration results obtained from Toronto Public Health for summer 2021 were used to test the correlation between Previous, and Same-day *E. coli* results. Figures 2 and 3 represent the correlation scatter plots for Sunnyside and Marie Curtis Park

East Beaches, respectively. All six sampling transects of Sunnyside Beach, and one transect at Marie Curtis Park East Beach, revealed no significant correlation between Previous-day and retrospective Same-day *E. coli* results ($p > 0.05$). Four of five sampling transects (29W, 30W, 31W, and 33W) at Marie Curtis Park East Beach presented only low to moderate correlation ($R = 0.41\text{-}0.56$, $p \leq 0.05$) between Previous-day and Same-day *E. coli* results.

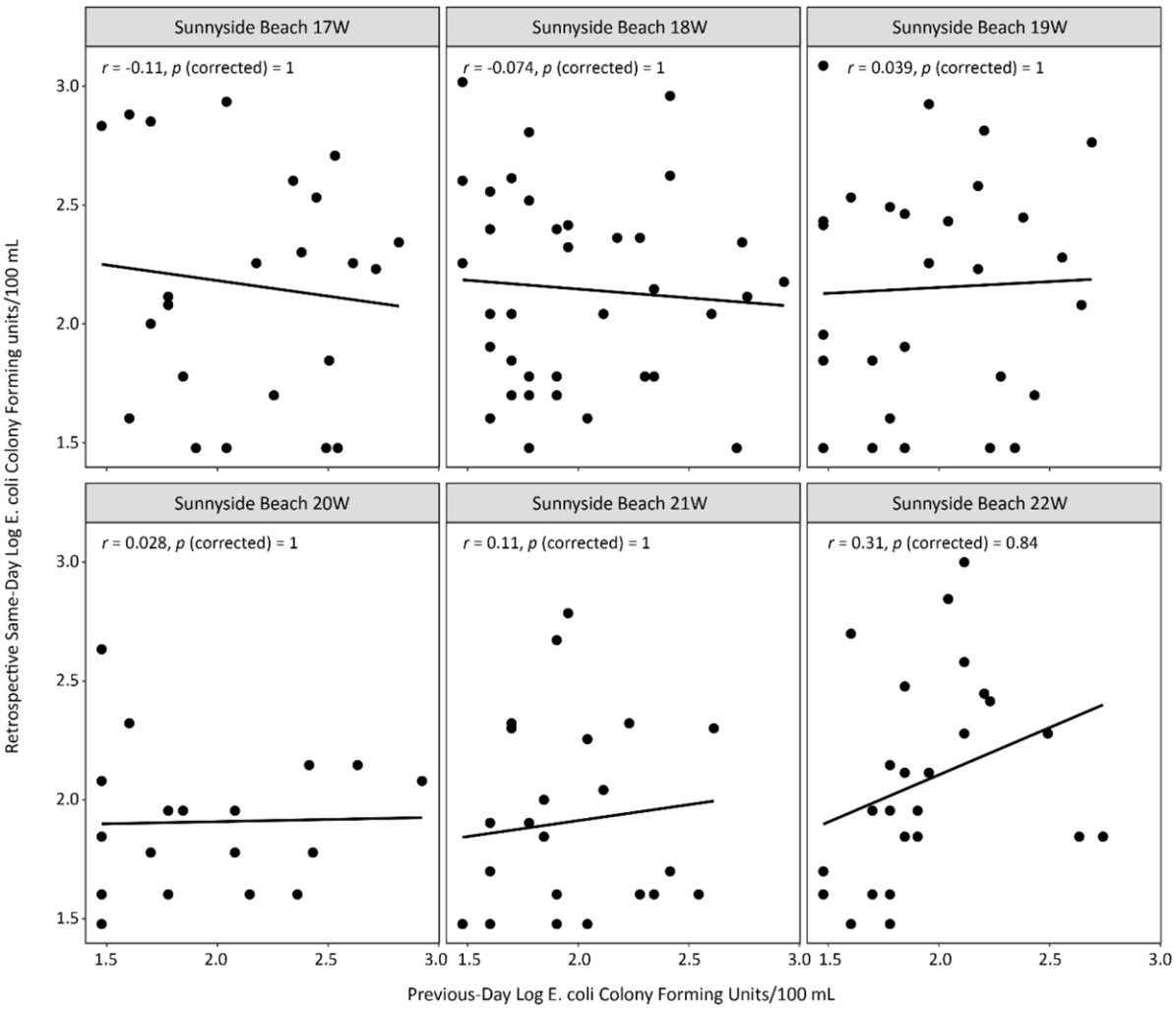


Figure 2. Correlation between previous-day and retrospective same-day log E. coli CFU/100 mL for Sunnyside Beach sampling transects. Correlation analysis was performed by using Pearson correlation at 95% confidence interval. E. coli data were obtained from Toronto Public Health.

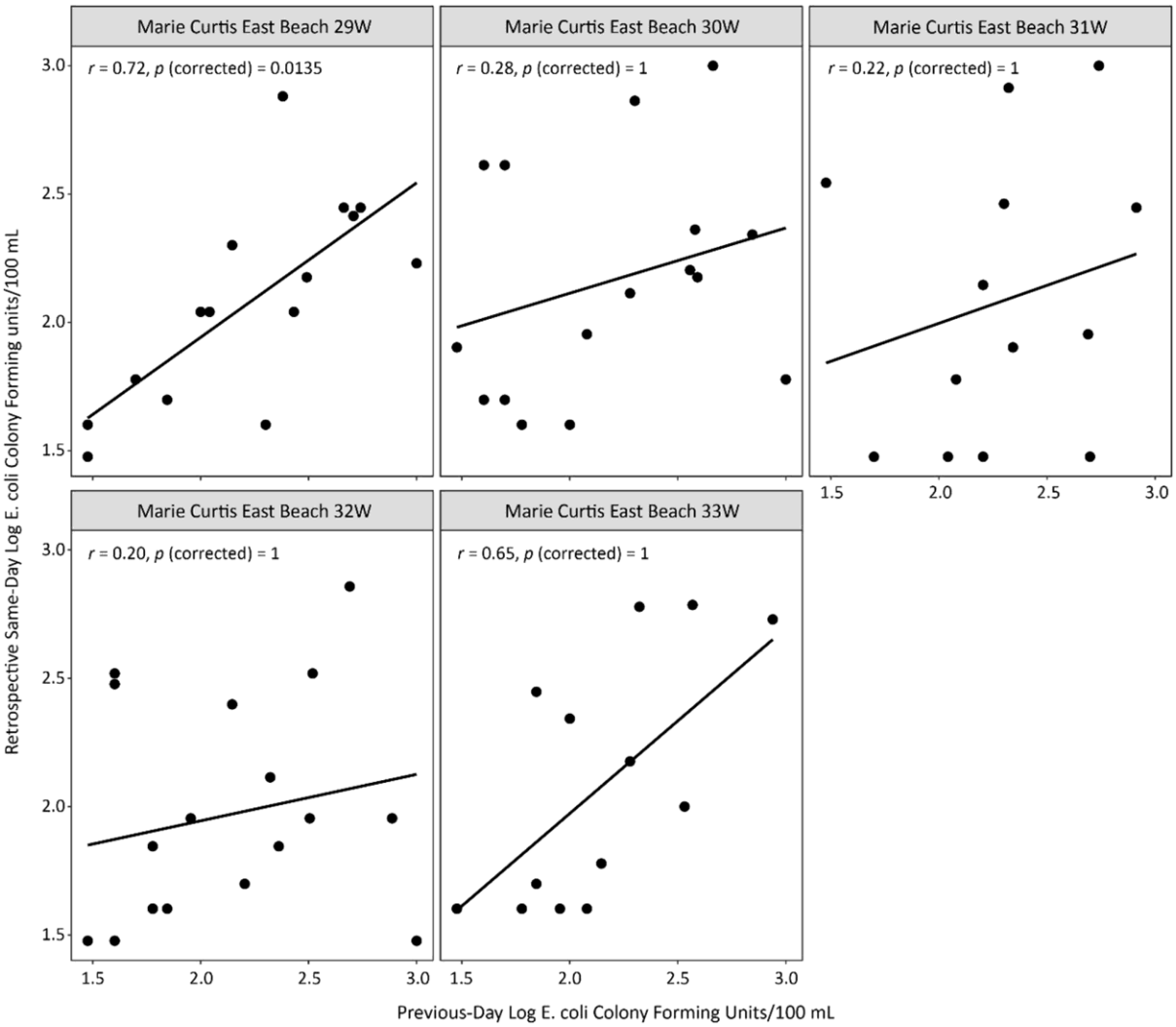


Figure 3. Correlation between previous-day and retrospective same-day log E. coli CFU/100 mL for Marie Curtis Park East Beach sampling transects. Correlation analysis was performed by using Pearson correlation at 95% confidence interval. E. coli data were obtained from Toronto Public Health.

Comparison of Same-day (retrospective) and Two-day rolling *E. coli* geometric mean for Beach Postings Outcomes:

Categorization of Toronto Public Health's data into previous-day and same-day (two consecutive days of sample collection) revealed quite different results for beach postings if Toronto Public Health could have obtained, hypothetically, Same-day culture-based *E. coli*

data (Supplementary figure 1 and 2). Same-day *E. coli* results would have indicated that Marie Curtis Park East Beach and Sunnyside Beach were either incorrectly posted or kept open 40% and 30% of the time, respectively, over the 2021 bathing season. In particular, Same-day *E. coli* results would have indicated that Marie Curtis Park East Beach and Sunnyside Beach should have been posted on an extra 17 and 15 days respectively in 2021, that were not indicated by two-day geometric *E. coli* data.

Impact of Same-day qPCR-based Monitoring on Beach Posting Outcomes:

Considerable differences were observed in beach posting outcomes when qPCR results were compared to Toronto Public Health's two-day rolling *E. coli* geometric mean. Beach posting differences were categorized as false positives (When Toronto Public Health posted the beach on the basis of two day rolling geometric mean for *E. coli* CFUs, but same-day *Enterococcus* qPCR identified beach opening) and false negative (When Toronto Public Health did not post the beach on the basis of two day rolling geometric mean for *E. coli* CFUs but same-day *Enterococcus* qPCR identified beach closure). For Marie Curtis Park East Beach (Figure 4), twelve out of thirty-eight tested beach days (32%) were posted according to Toronto Public Health's two-day rolling *E. coli* geometric mean, while our Same-day qPCR-based *Enterococcus* quantification identified nine beach days (24%) when the beach should have been posted. Of the thirty-eight tested beach days (Figure 5), our Same-day qPCR data identified three false negatives/health-risk days (8%) and six false positives/lost beach days (16%).

For Sunnyside Beach (Figure 6), fourteen out of thirty-eight tested beach days (37%) were posted according to Toronto Public Health's two-day rolling *E. coli* geometric mean, and

Same-day qPCR-based *Enterococcus* quantification also identified fourteen beach days (37%) when the beach should have been posted. Of the thirty-eight tested beach days (Figure 7), our Same-day qPCR data identified one false negative/health-risk day (2%) and twelve false positives/lost beach days (31%).

Using our Same-day qPCR results for beach monitoring would have resulted in a gain of six and twelve beach open days for Marie Curtis Park East and Sunnyside Beaches, respectively on our sampling days. Additionally, qPCR monitoring would have prevented three and one health-risk days for Marie Curtis Park East and Sunnyside Beaches, respectively on our sampling days.

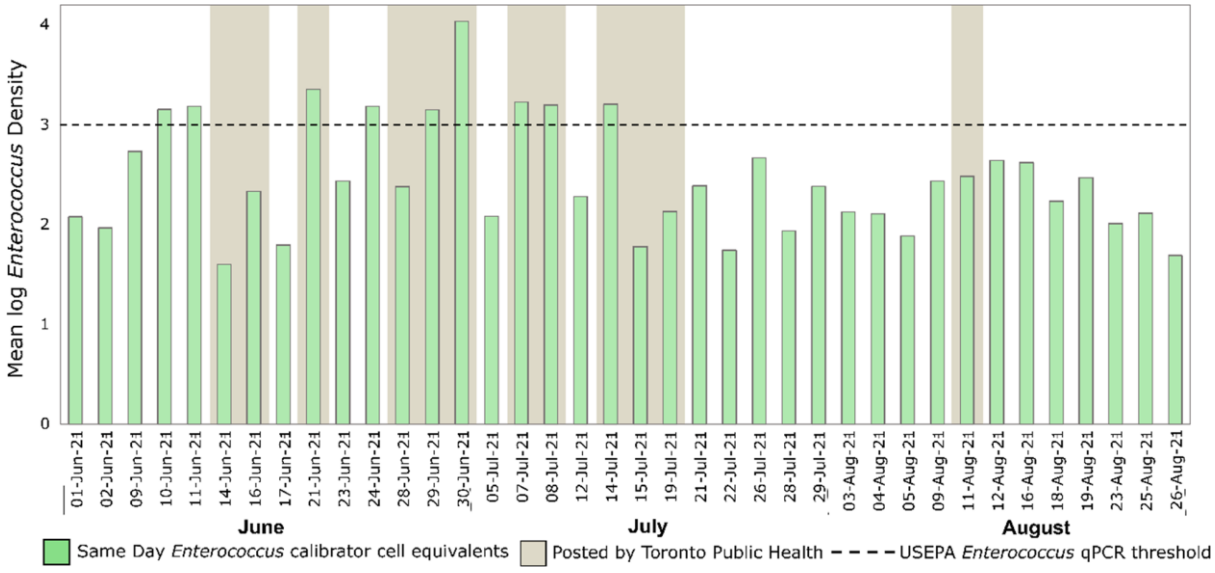


Figure 4. Comparison between Toronto Public Health's beach posting decisions and same-day Enterococcus qPCR quantification for Marie Curtis Park East Beach. Dashed line shows USEPA beach action value for Enterococcus quantification by qPCR (≥ 1000 CCE/100 mL), while gray highlighted days represent beach days posted by Toronto Public Health based on their rolling 2-day E. coli geometric mean.

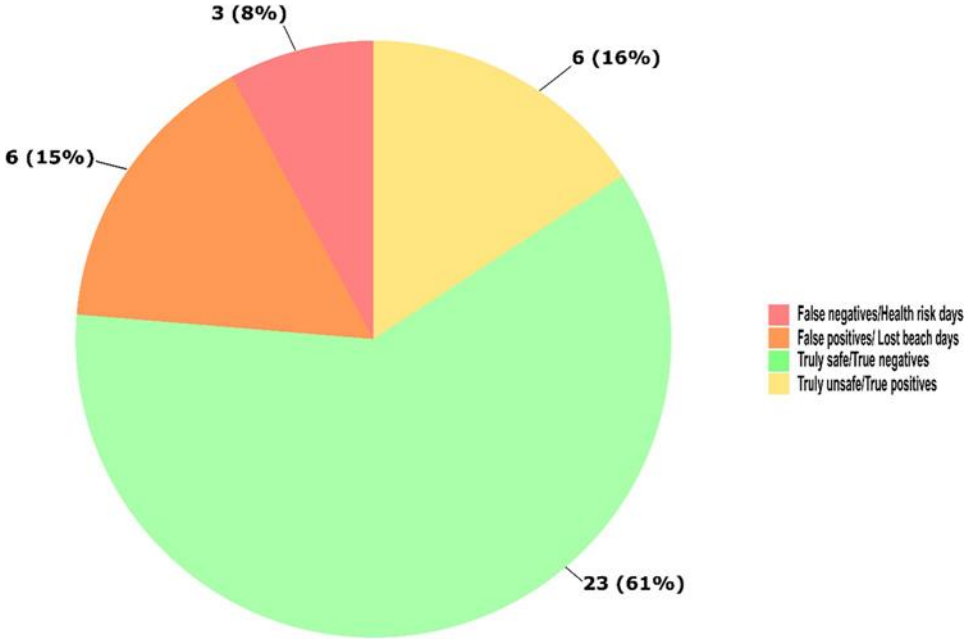


Figure 5. Analysis of impacts on Marie Curtis Park East Beach posting decision-making using the Enterococcus qPCR and compared to Toronto Public Health's reported 2-day rolling E. coli geometric mean for summer 2021 beach days (n = 38).

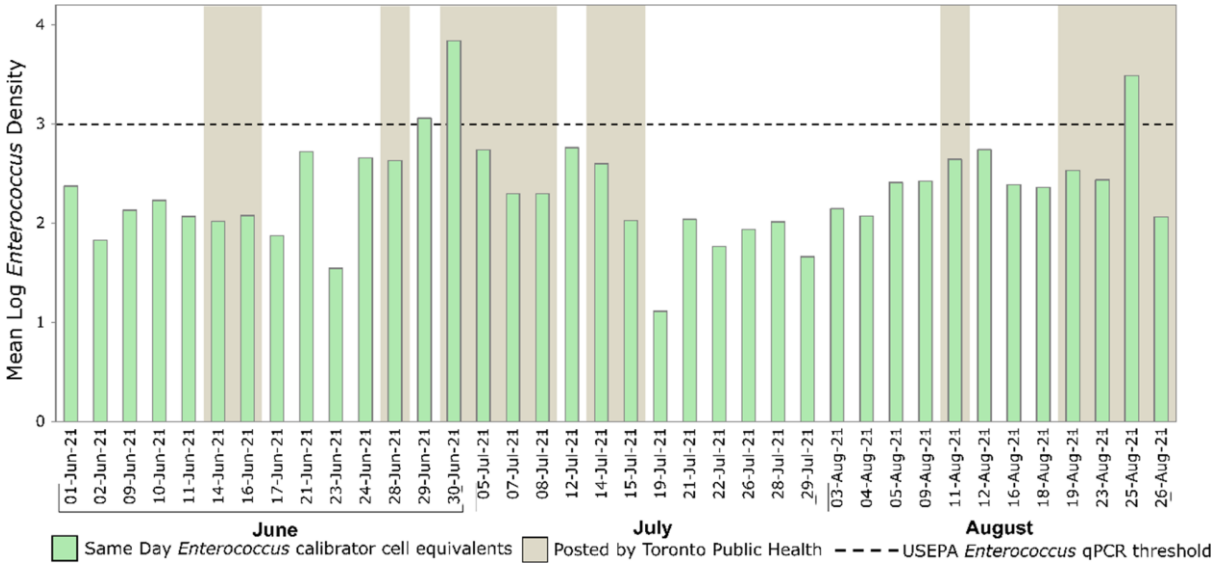


Figure 6. Comparison between Toronto Public Health's beach posting decisions and same-day Enterococcus qPCR quantification for Sunnyside Beach. Dashed line shows USEPA beach action value for Enterococcus quantification by qPCR (≥ 1000 CCE/100 mL), while gray highlighted days represent beach days posted by Toronto Public Health based on its rolling 2-day E. coli geometric mean.

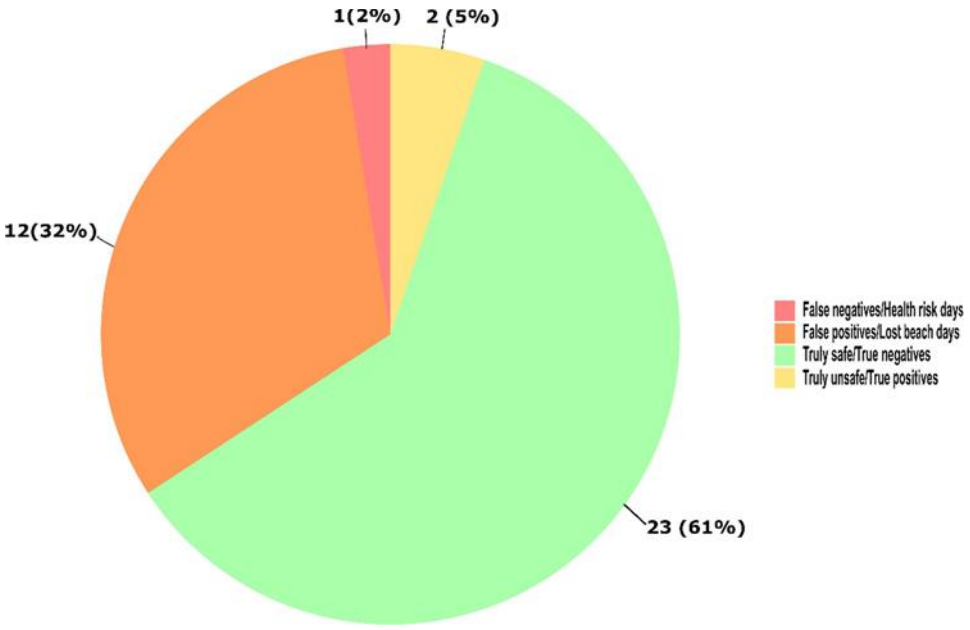


Figure 7. Analysis of impacts on Sunnyside Beach posting decision-making using the Enterococcus qPCR and compared to Toronto Public Health's reported 2-day rolling E. coli geometric mean for summer 2021 beach days (n = 38).

Discussion:

Microbial monitoring methodologies for freshwater recreational sites are being revised as new more robust measures can provide more timely results (Campbell and Kleinheinz 2020). Our study provides a comparison of beach posting outcomes between Toronto's previous-day culture-based *E. coli* method and our same-day *Enterococcus* qPCR method for Marie Curtis Park East and Sunnyside Beaches. In total, 228 water samples corresponding to 38 beach days were collected over the summer 2021 season. Water samples were collected each day by 7:00am, transported within one hour to the lab, and processed in the lab within 3.5-4h of sample receipt. qPCR quantification results were typically provided to Toronto Public

Health by noon on a given day. Our experience suggests that this qPCR approach can work for beaches located within about a one-hour drive from a qPCR lab, provided that the water sampling is completed by about 7:00am. A practical limitation will be the flexibility to have water samples collected earlier in the morning than typically now done.

Toronto Public Health's monitoring program uses a two-day rolling *E. coli* geometric mean for beach posting decision making. However, *E. coli* counts are subject to change in beaches during a 24h delay (Whitman et al., 1999; Whitman et al., 2004; Enns et al., 2012). Daily *E. coli* enumeration data from Toronto was obtained from June 01 to August 31, 2021, to assess the correlation between Previous-day and Same-day (retrospective) *E. coli* enumeration results. No significant correlation was observed between the Previous and Same-day *E. coli* results. A previous two-day *E. coli* geometric mean would likely even further reduce any potential for correlation with Same-day (retrospective) *E. coli* results. In addition, our analysis of Previous-day and Same-day (retrospective) *E. coli* results often resulted in different beach posting outcomes. These findings are consistent with two previous studies on US beaches that reported significant changes in fecal indicator bacteria numbers over a time scale of hours (Converse et al. 2012; Dorevitch et al. 2017). Additionally, it was observed that for Chicago beaches, Previous-day *E. coli* results had no better than a fifty percent chance at correctly predicting Next-day's water quality (Dorevitch et al. 2017). Therefore, reliance upon 24 hour-old *E. coli* results may not provide a true representation of beach water quality and can result in many erroneous beach postings.

The qPCR-based monitoring method (USEPA 1609.1) is designed to overcome the time limitation of culture-based enumeration methods and can lead to more rapid beach posting

decision-making for beach managers. *E. coli* Based on our qPCR-based results, *Enterococcus* qPCR quantification would have allowed the beaches to remain open for additional days. Perhaps more importantly, qPCR-based results indicated the need for beach postings to protect public health on additional beach days.

This study found substantial differences in beach posting outcomes using our Same-day qPCR method compared to the City's previous two-day *E. coli* geometric mean data even though our study was limited to two beaches and only 38 beach days for 2021 summer season. These differences might be partially explained by our two beaches being located near river mouths and more rapidly impacted by fluctuating river water quality. The difference in Toronto Public Health beach postings outcomes using a rapid qPCR method could also be because Toronto Public Health beach posting decisions are based on a two-day rolling *E. coli* geometric mean that incorporates culture-based *E. coli* results from two days before a posting decision, in addition to previous-day results. *Enterococcus.* Our results provide guidance for future studies to investigate implications of adopting more rapid qPCR testing methods for beaches. Our results indicate rapid qPCR methods can enable more timely beach posting decisions to better protect public health and reduce adverse social, tourism and economic impacts from incorrect postings.

Conclusions:

1. USEPA Method 1609.1 for *Enterococcus* by qPCR provided same-day results, within 5.5 to 6 h of recreational water sample collection and within 3.5 to 4 h of sample receipt in the laboratory, and it can provide a more rapid beach monitoring approach for beach managers to avoid erroneous beach postings based on day-old *E. coli* data.

2. In comparison to the City's two-day rolling *E. coli* geometric mean data over 38 tested beach days, our Same-day qPCR data identified 12 False positives/lost beach days at Sunnyside Beach and 6 false positives/lost beach days at Marie Curtis Park East Beach.

3. In comparison to the City's two-day rolling *E. coli* geometric mean data over 38 tested beach days, our Same-day qPCR data identified 3 false negatives/health risk days at Marie Curtis Park East Beach and 1 false negative/health risk day at Sunnyside Beach.

4. Use of the same-day *Enterococcus* qPCR method would probably contribute to many more differences in beach posting decisions than the previous-day *E. coli* method if applied every day for the complete summer season across all 10 Toronto beaches.

Contributions to Knowledge:

- What does this study add to existing knowledge?
    - This study presents the *Enterococcus* qPCR-based beach monitoring method as a potential alternative to traditional culture-based enumeration methods being used. In addition, it shows the *E. coli* same-day *Enterococcus* qPCR can help to avoid erroneous beach postings and can allow the beach managers to have beach quality results on the same day of sample collection. The fact that the assay was developed in 2015 and is recognized by both USEPA and Health Canada (Health Canada, 2021), means that is a valid alternative.
- What are the key implications for public health interventions, practice, or policy?

- The qPCR-based method provides results within 3.5-4h of sample receipt and can allow beach managers to provide more timely and reliable beach postings decisions to protect public health and reduce adverse social and economic impacts from incorrect postings. Our results can allow the Public Health Agencies to adopt the rapid qPCR-based method for beach monitoring.

References:

Alberta (2021). Alberta Safe Beach Protocol. Available at:https://open.alberta.ca/publications/9781460145395.

Campbell, A., and Kleinheinz, G. (2020). Comparison of qPCR and traditional beach monitoring methods at select Great Lakes beaches and the impact on beach management. *Lake Reservoir Management, 36*(2), 155-168.

Converse, R. R., Griffith, J. F., Noble, R. T., Haugland, R. A., Schiff, K. C., and Weisberg, S. B. (2012). Correlation between quantitative PCR and culture-based methods for measuring *Enterococcus* spp. over various temporal scales at three California marine beaches. *Applied Environmental Microbiology, 78*(4), 1237-1242.

Dorevitch, S., Shrestha, A., DeFlorio-Barker, S., Breitenbach, C., and Heimler, I. (2017). Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: Implications for public notification. *Environmental Health, 16*(1), 1-10.

Enns, A.A., Vogel, L.J., Abdelzaher, A.M., Solo-Gabriele, H.M., Plano, L.R., Gidley, M.L., Phillips, M.C., Klaus, J.S., Piggot, A.M., Feng, Z. and Reniers, A.J. (2012). Spatial and

temporal variation in indicator microbe sampling is influential in beach management decisions. *Water research*, 46(7), pp.2237-2246.

Haugland, R.A., Siefring, S., Varma, M., Oshima, K.H., Sivaganesan, M., Cao, Y., Raith, M., Griffith, J., Weisberg, S.B., Noble, R.T. and Blackwood, A.D. (2016). Multi-laboratory survey of qPCR *Enterococcus* analysis method performance in US coastal and inland surface waters. *Journal of microbiological methods*, 123, pp.114-125.

Health Canada (2012). Guidelines for Canadian Recreational Water Quality, Third Edition. Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada, Ottawa, Ontario. (Catalogue No H129-15/2012E).

Health Canada (2021). Guidelines for Canadian Recreational Water Quality: Indicators of fecal contamination. Guideline Technical Document for Public Consulation. Health Canada, Ottawa, Ontario. Available at: https://www.canada.ca/content/dam/hc-sc/documents/programs/consultation-guidelines-recreational-water-quality-fecal-contamination/consultation-guidelines-recreational-water-quality-fecal-contamination-eng.pdf

Jafari, M. and Ansari-Pour, N. (2019). Why, when and how to adjust your P values?. Cell Journal (Yakhteh), 20(4), p.604.

Kassambara, A. and Kassambara, M.A. (2020). Package 'ggpubr'. R package version 0.1, 6.

Kleinheinz, G. T., McDermott, C. M., and Chomeau, V. (2006). Evaluation of avian waste and bird counts as predicators of *Escherichia coli* contamination at Door County, Wisconsin beaches. *Journal of Great Lakes Research, 32*(1), 117-123.

Saleem, F., Edge, T. A., and Schellhorn, H. E. (2022). Validation of qPCR method for enterococci quantification at Toronto Public Health beaches: Application for rapid recreational water monitoring. *Journal of Great Lakes Research* 48(3), 707-716.

Shrestha, A., and Dorevitch, S. (2019). Evaluation of rapid qPCR method for quantification of *E. coli* at non-point source impacted Lake Michigan beaches. *Water Research, 156*, 395-403.

USEPA (2009). National Epidemiologic and Environmental Assessment of Recreational Water Epidemiology Studies. EPA/600/R-10/168. Available at: http://www.epa.gov/neear/files/Report2009v5_508comp.pdf

USEPA (2012). Recreational water quality criteria. (pp. 42): United States (US) Environmental Protection Agency (EPA). EPA-820-F-12-058. Available at: https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf

USEPA (2015). Method 1609.1: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay. EPA-820-R-15-099. Available at: https://www.epa.gov/sites/default/files/2015-08/documents/method_1609-1-*Enterococcus*-iac_2015_3.pdf

Wade, T.J., Calderon, R.L., Brenner, K.P., Sams, E., Beach, M., Haugland, R., Wymer, L. and Dufour, A.P. (2008). High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. Epidemiology, pp.375-383.

Wade, T.J., Sams, E., Brenner, K.P., Haugland, R., Chern, E., Beach, M., Wymer, L., Rankin, C.C., Love, D., Li, Q. and Noble, R. (2010). Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: a prospective cohort study. Environmental Health, 9(1), pp.1-14.

Whitman, R.L., Nevers, M.B. and Gerovac, P.J. (1999). Interaction of ambient conditions and fecal coliform bacteria in southern Lake Michigan beach waters: monitoring program implications. *Natural Areas Journal*, 19(2), pp.166-171.

Whitman, R.L., Nevers, M.B., Korinek, G.C. and Byappanahalli, M.N. (2004). Solar and temporal effects on *Escherichia coli* concentration at a Lake Michigan swimming beach. *Applied and environmental microbiology*, 70(7), pp.4276-4285.

Wickham, H., Chang, W. and Wickham, M.H. (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version, 2(1), pp.1-189.

Chapter 4. Application of Same-Day *Enterococcus* qPCR-Based Analyses for Quality Assessment of Shorelines (Water and Sand) at Recreational Beaches

This chapter continues with the overarching theme of this thesis by demonstrating how same-day qPCR analyses can improve the accuracy and timeliness of water quality monitoring. This peer-reviewed study builds on the ability of a rapid qPCR-based strategy to accurately monitor both water and sand quality, ensuring a more comprehensive assessment of beach environments and addressing the risk of fecal contamination in both mediums rather than just shoreline waters.

*The text I present here is a **peer-reviewed manuscript published** in Water. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Edge, T.A. and Schellhorn, H.E., 2023. Application of Same-Day Enterococcus qPCR-Based Analyses for Quality Assessment of Shorelines (Water and Sand) at Recreational Beaches. Water, 15(13), p.2338.*

Author Contributions

FS performed the experiments, analyzed the data, communicated the findings with Public Health Agencies, Wrote the original manuscript, and revised it as required by the reviewers. TAE collected the water samples and reviewed the manuscript before submission to the journal. HES supervised the study and reviewed the manuscript before submission to the journal.

Abstract

Rapid water-quality monitoring methods for beach water and sand can be helpful for public health authorities to develop comprehensive beach monitoring programs. In this study, we evaluated the performance of the *Enterococcus* qPCR (USEPA 1609.1) method for quality monitoring of beach water and sand-porewater at two Niagara region beaches in Ontario, Canada (Lakeside and Sunset Beaches). While the USEPA 1609.1 method has been validated for beach water, its potential for assessing associated beach sands (which may function as a microbial reservoir) has not been fully explored. All beach water samples (n = 168) passed the qPCR quality control (QC). However, only 20 out of 48 (41.7%) sand-porewater samples passed the qPCR QC, potentially due to interference by soluble inhibitors. The proportion of the sand-porewater samples passing QC improved slightly to 63 out of 120 (52.5%) with a prefiltration step to remove sand and other large particles. The faecal indicator density in the sand-porewater, tested in parallel, did not correlate with the beach water faecal indicator density. Comparing beach water data for the same-day *Enterococcus* qPCR threshold with the previous-day *E. coli* culturing-based threshold across all beach days tested, *Enterococcus* qPCR analyses identified 3 (7%) and 7 (16%) false positive/lost beach days for Lakeside and Sunset Beaches, respectively. Additionally, of the total beach days tested, *Enterococcus* qPCR analyses identified 2 (5%) and 1 (2%) false negative/health-risk days for Lakeside and Sunset Beaches, respectively. Sand-porewater testing analyses identified days when faecal indicators (in the sand) exceeded beach water quality thresholds. Compared with conventional *E. coli* culturing, use of the same-day qPCR method would result in fewer beach postings and could identify several additional health-risk days (when the beaches may not be posted). Future studies could include additional

prefiltration steps or modifications in the *Enterococcus* qPCR protocol to improve the method's applicability for sand quality monitoring.

Keywords: *Enterococcus*; qPCR; beach water monitoring; Niagara; beach sand monitoring

Introduction

Water quality for Ontario freshwater beaches is currently determined using conventional *E. coli* culturing-based enumeration methods to mark the beach water quality as safe or unsafe for recreational purposes [1]. Two major practical issues associated with beach monitoring programs are (1) Delayed results associated with culturing-based methods and (2) Lack of routine sand quality monitoring [2,3]. Faecal indicator levels in freshwater beaches can change overnight [4], and thus the faecal indicator bacteria (FIB) counts based on samples collected on the previous day may not be a good indicator of the next day's beach quality [5,6]. Additionally, beachgoers' recreational use may involve direct contact with beach water and sand, whereas public health agencies currently actively test beach water quality only [7,8]. If a beach is posted to prevent swimming, this may result in more beachgoer time spent onshore and more human exposure to beach sand.

Beach sand can be a reservoir for faecal indicators and pathogenic bacteria [9–12]; bacterial accumulation and long-term persistence are possible because of the sand's ultraviolet and temperature shielding effects [3]. Epidemiology studies [13,14] also indicate that the direct contact of beachgoers with shore sand positively correlates with gastrointestinal disease morbidity. In addition, the wave action in beach water may mobilize sand-particle-attached bacteria, which can be transported to the adjacent shoreline, resulting in lower beach water quality [4]. However, beach sand monitoring is not typically performed, and when it is, it

mainly relies on culturing-based 24 h-delayed methods such as Colilert® and Enterolert® (IDEXX, Westbrook, ME, USA) [15]. Therefore, incorporating sand quality assessment methods may result in more effective, improved beach quality monitoring strategies.

Culturing-based methods for faecal indicator bacteria have inherent limitations for assessing beach water quality and may not be suitable for dormant or non-culturable enterococci [16]. Compared with culturing-based beach monitoring, qPCR-based methods, such as the USEPA 1609.1 *Enterococcus* testing protocol, are much more robust and can provide results to the beach managers within 3–4.5 h of sample collection [5,17]. In addition, qPCR-based methods can more comprehensively estimate faecal indicator densities by quantifying the culturable and non-culturable indicators [15]. National Environmental Epidemiological Assessment of Recreational Water (NEEAR) studies indicate that qPCR-based enterococci measures are a potentially more reliable indicator of gastrointestinal diseases among beachgoers [18,19]. While the applicability of the qPCR-based method has been tested for beach water, it is not generally used for monitoring beach sands (FIB may be transported onto beaches by wave action or rainwater runoff). Compared with beach water, sand can potentially harbour higher levels of qPCR inhibitory compounds, such as humic acids, that can inhibit PCR reactions [2]. Testing the applicability of the qPCR-based method for more rapid monitoring of both beach water and beach sand may inform public health agencies in revising beach monitoring regimens.

Despite the growing adoption of rapid beach monitoring methods [4,20], many public health agencies still depend on 24 h-delayed *E. coli* culturing from water samples, and sand-porewater is not routinely monitored. Our study builds on previous studies on the microbial

analysis of important freshwater beaches [5,6,10] and focuses on the applicability/comparative analysis of the *Enterococcus* qPCR-based method for both beach water and sand environments at Niagara beaches. Compared with the Toronto beaches (Marie Curtis Park East and Sunnyside Beaches) [5,6], the Niagara beaches (Lakeside and Sunset Beaches) are posted less frequently and are not as heavily contaminated by industrial/faecal contamination sources. For this study, we selected the Niagara beaches as part of our goal to extend testing of the applicability of an *Enterococcus* qPCR-based beach water testing method for different geographical locations (with different sources or levels of contamination) and different environments (shore water and sand). Objectives for our study were: (1) Testing the applicability of the USEPA 1609.1 *Enterococcus* protocol for assessing beach water and sand quality at two different Niagara region beaches, (2) Comparative analysis of the FIB densities in beach water and sand between beach open and posted days, (3) Analysis of beach posting outcomes using *Enterococcus* qPCR compared with Public Health's *E. coli* culturing data, and (4) Testing the correlation between the FIB densities (as measured by culture and qPCR) and environmental variables such as temperature and wave height.

Materials and Methods

Study Plan

Our study design included beach water and sand-porewater sampling of two Niagara beaches on Lake Ontario in Ontario, Canada (Lakeside and Sunset Beaches). The sample collection plan included the 2022 summer season from 31 May 2022 to 1 September 2022. Figure 1 and Supplementary Table S1 show the geographical locations/coordinates of the sampling

sites corresponding to Lakeside and Sunset Beaches. Water and sand samples were collected from two sampling sites for each beach (LK3 and LK5 for Lakeside Beach; SS3 and SS5 for Sunset Beach). Samples were collected in the morning between 8 a.m. and 9 a.m. for three consecutive days a week and were delivered to the lab for processing before 10 a.m. For each sampling day, eight samples were collected, including four samples from beach water and four from beach sand-porewater. The beach season for summer 2022 consisted of 94 beach days from 31 May 2022 to 1 September 2022. In total, 336 samples (168 beach water samples and 168 sand-porewater samples) corresponding to 42 beach days were collected and processed for the 2022 summer season.

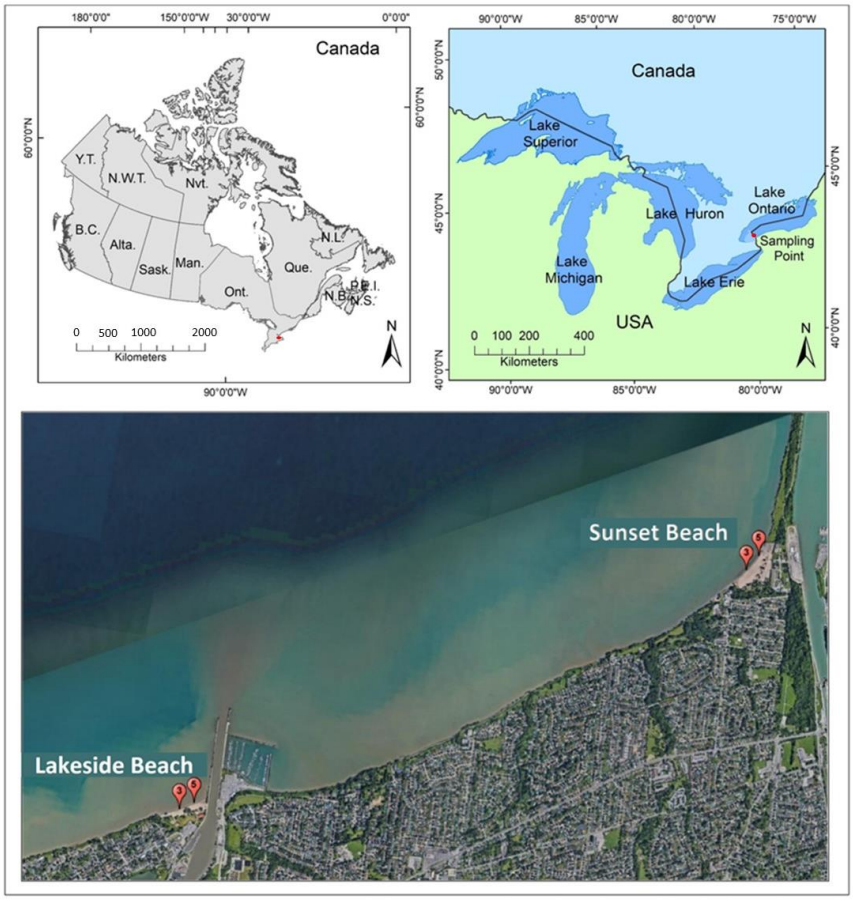


Figure 1. Geographical map of sampling sites from Lakeside and Sunset Beaches (by Ontario Geohub).

Water Sample Collection and Filtration

Beach water samples were collected in sterile 1000 mL polyethene terephthalate (PET) water bottles 30 cm below the water surface at a water depth of 1.0 m. For sand-porewater samples, a 30 cm × 30 cm × 30 cm hole was dug 1 m inland from the lake in the foreshore sand region of each sampling site, followed by water collection (500 mL) by allowing surrounding groundwater to seep into the bottom of the hole (Figure 2) [10]. After the sample receipt in the lab, a 100 mL beach water sample was processed through a 0.45 µm polycarbonate

membrane filter (Millipore Corp., Bedford, MA, USA). For beach sand-porewater, initially (31 May 2022 to 30 June 2022), a 50 mL water sample was directly passed through a 0.45 µm membrane filter. For the samples from July 01 onwards, 50 mL sand-porewater samples were first passed through a prefiltration assembly made by enclosing a 50 µm nylon filter (Dynamic Aqua Supply, Surrey, Canada) inside a plastic garden mesh (Supplementary Figure S1) to remove larger sand and dirt particles, followed by passing the filtrate through a 0.45 µm filter.

***E. coli* Culture and Enterococcus qPCR**

For *E. coli* culture enumeration from sand-porewater samples, 1:10 dilutions were prepared as required before filtration. Each sand-porewater sample was filtered in triplicate, and membrane filters were placed on differential coliform agar plates (Oxoid^{TM}), followed by incubation for 24 h at 44.5 °C. *Enterococcus* qPCR was performed as described in the USEPA 1609.1 method: *Enterococci in Water by TaqMan^{TM} Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay* [17], as recently validated [5,6]. In brief, the qPCR assay included calibrator positive control, internal amplification control (to check for qPCR inhibition), sample processing control (salmon sperm DNA qPCR to check for extraction efficiency), water matrix spike (to assess for interference effects from the water matrix), and non-template control (to check for contamination) as quality controls. DNA extraction for the standard curve was performed using Norgen soil plus DNA extraction kit (Norgen Biotek, Ontario, Canada) while, for routine sample analysis, the standard DNA extraction protocol was used as described in the USEPA 1609.1 method [17]. Four individual standard curves ranging from 40,000 to 10 target sequence

copies (TSC)/5 µL were analyzed using *Enterococcus faecalis* (ATCC 2921) cell suspension (10^9 cells) to obtain a composite standard curve. For each sampling day, two calibrator positive controls, two method blanks, and one non-template control were analyzed along with the beach water and sand-porewater samples. All the samples and quality controls were analyzed in duplicate on the Bio-Rad CFX96 Touch Real-Time PCR Detection System thermocycler (Bio-Rad Inc. Mississauga, Ontario, Canada.

Data Analysis

Calculations for qPCR inhibition, DNA recovery, and calibrator cell equivalents (CCEs/100mL) were performed using a USEPA 1609.1 Excel sheet (https://www.epa.gov/cwa-methods/other-clean-water-act-test-methods-microbiological#file-183743) (Accessed on 01 June 2022). The arithmetic means of the density of *Enterococcus* calibrator cell equivalents were calculated using counts from both sampling sites for each beach. The *E. coli* culturing counts from the sand-porewater samples included calculating the arithmetic mean of three replicates for each sample and two sampling sites for each beach (3 replicates/sample × 2 sampling sites/beach). The current Health Canada and Ontario quality threshold for *E. coli* culturing (beach water) is 200 CFUs/100 mL with a single-sample max of 400 CFUs/100 mL [8]. For *Enterococcus* qPCR and *E. coli* culturing, the Beach Action Value and water quality thresholds used for data interpretation were 1000 calibrator cell equivalents (CCEs)/100 mL and 200 colony-forming units (CFUs)/100 mL, respectively. For the sand quality comparison, we used faecal indicator beach water quality thresholds due to the unavailability of quality thresholds for beach sand. Culturing and qPCR data were log-transformed for correlation analysis and

analyzed for normality distribution using the Shapiro–Wilk test (http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r) (Accessed on 15 September 2022). The Pearson correlation with a 95% confidence interval and $p < 0.01$ was used to evaluate the degree of correlation between the environmental variables and faecal indicator counts. We also obtained *E. coli* data and environmental variable data such as wave height, air/water temperature, and wind speed from Niagara Public Health (referred to as Public Health in the later sections). The beach postings comparison analysis was performed using Public Health's beach posting decisions for the summer of 2022 based on an *E. coli* geomean of samples at a beach. Retrospective same-day *E. coli* data from Public Health enabled an assessment of the hypothetical possibility of obtaining culturing results on the same day as sample collection. For the beach postings comparison, beach days were categorized as false negative days when the beaches were not posted according to *E. coli* culturing and the same-day *Enterococcus* qPCR results identified Beach Action Value threshold exceedances. Conversely, beach days were categorized as false positive beach days when the beaches were posted according to *E. coli* culturing and the same-day *Enterococcus* qPCR results did not identify Beach Action Value threshold exceedances. Due to common qPCR quality control failures, beach posting comparisons and correlations using sand-porewater qPCR results often included enterococci calibrator cell equivalents from a single sampling site for each beach (no sampling site replicate). All the plots were generated using the ggplot2 package in R [21].

Results

3.1. qPCR Quality Control Analysis

The coefficient of determination (R^2) for the composite standard curve was 0.998, while the slope and intercept values were observed as −3.28 and 39.55, respectively (Table 1). In total, 336 water samples were collected/processed, including 168 samples each for beach water and sand-porewater, corresponding to 42 beach sampling days. All the beach water samples passed the quality control criteria recommended in the USEPA 1609.1 protocol [17]. For sand-porewater samples processed without nylon mesh prefiltration, the quality control (QC) qPCR pass rate was only 20 of 48 (41.7%). The pass rate improved slightly with nylon mesh prefiltration in 63 out of 120 (52.5%) (Table 1). In comparison with beach water and sand-porewater samples that passed the QC criteria, most (74/85; 87%) of the QC failed samples showed a highly significant ($p < 0.01$) increase in the sample processing control Ct value (Supplementary Figure S2).

Table 1. qPCR quality control analytics for beach and sand-porewater samples from Lakeside and Sunset Beaches.

Data Quality Parameters			
Standard Curve			
Number of Standard Curves	4		
Coefficient of Determination	0.998		
Amplification Efficiency	2.02		
Slope	−3.28		
Intercept	39.55		
Beach Water Samples			
Number of Samples	168		
qPCR Quality Pass	168 (100%)		
Sand-Porewater Samples			
Before Nylon Mesh		**After Nylon Mesh**	
Number of Samples	48	Number of Samples	120
qPCR Quality Pass	20 (41.7%)	qPCR Quality Pass	63 (52.5%)

3.2. Association between Sand-Porewater and Beach Water Faecal Indicator Densities

No significant correlation ($p > 0.05$) was observed between the sand-porewater and beach

water *Enterococcus* calibrator cell equivalents (CCEs) or between the sand-porewater *E. coli* colony-forming units (CFUs) and beach water *Enterococcus* CCE (Supplementary Figures S3 and S4). However, there was a significant positive correlation ($p < 0.01$, R = 0.5–0.6) between the sand-porewater *Enterococcus* CCEs and *E. coli* CFUs (Supplementary Figure S5). Among the environmental variables, there was a significant positive correlation between *Enterococcus* CCE and wave height (Supplementary Figure S6), while there was no significant correlation between either air/water temperature or wind speed and the faecal indicator counts.

3.3. Comparison of Same-Day qPCR-Based Monitoring Method with 24 h-Delayed E. coli Enumeration by Culture for Beach Posting Outcomes

According to the *E. coli* CFUs data for summer 2022 (Public Health's data), out of our 42 tested beach days, Lakeside and Sunset Beaches were posted for 3 and 13 beach days, respectively. Compared with Public Health's data, our same-day *Enterococcus* qPCR identified considerable differences in the beach postings. For Lakeside Beach (Figure 2), of 42 tested beach days, our same-day *Enterococcus* qPCR results identified three (7%) false positives, or lost beach days, and two (5%) false negatives, or health-risk days (Figure 3). For Sunset Beach (Figure 4), our same-day *Enterococcus* qPCR results identified seven (16%) false positives, or lost beach days, and one (2%) false negative, or health-risk day (Figure 5).

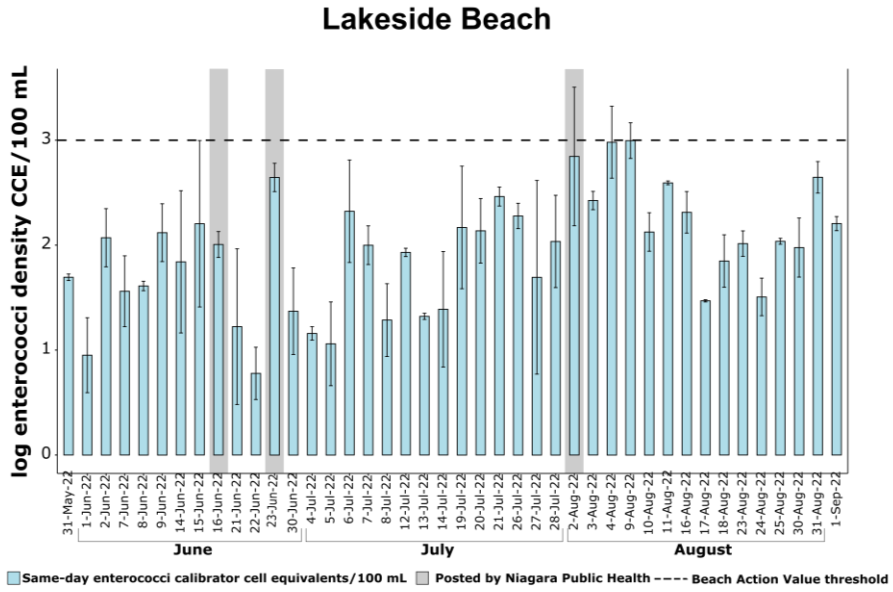


Figure 2. Comparison between Public Health's beach posting decisions and same-day enterococci qPCR quantification for Lakeside Beach. The dashed line represents the USEPA Beach Action Value for Enterococcus qPCR quantification method (≥1000 CCEs/100 mL), while the grey highlighted days represent beach days posted by Public Health based on their 24 h-delayed E. coli geometric mean. Error bars represent the standard deviation of log-transformed Enterococcus calibrator cell equivalents.

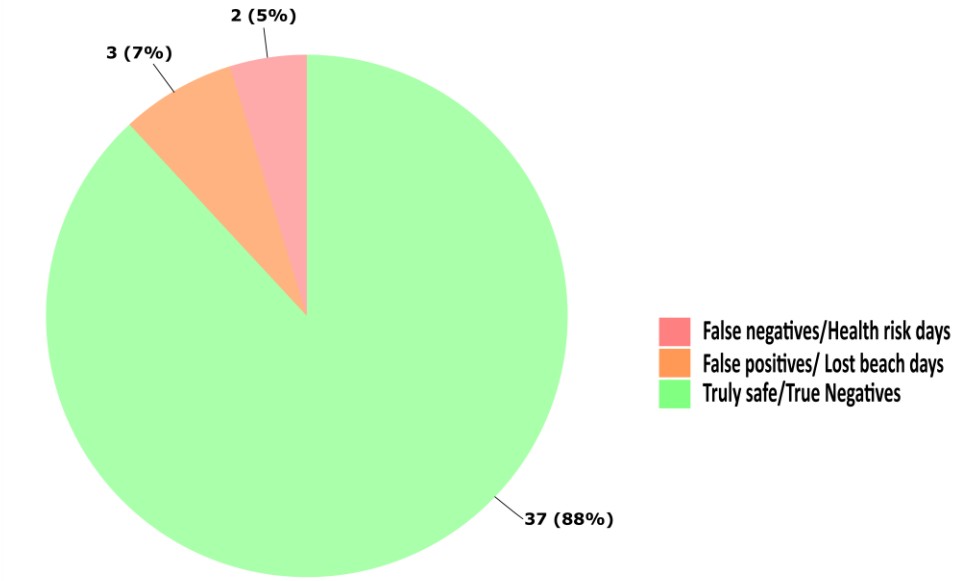


Figure 3. Analysis of impacts on Lakeside Beach postings using the same-day Enterococcus qPCR and compared with Public Health's reported 24 h-delayed E. coli geometric mean for our summer 2022 beach sampling days (n = 42).

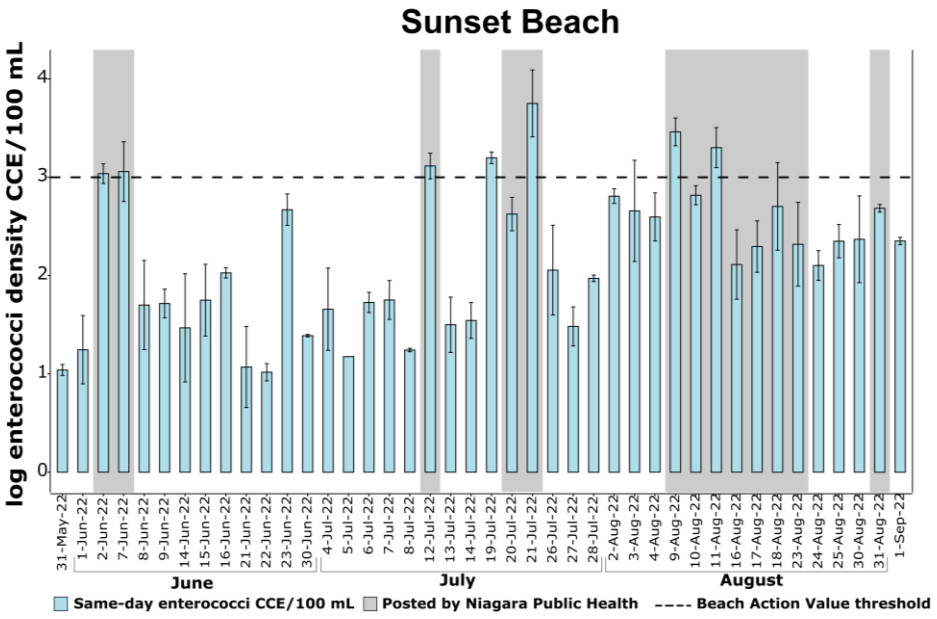


Figure 4. Figure 4. Comparison between Public Health's beach posting decisions and same-day Enterococcus qPCR quantification for Sunset Beach. The dashed line represents

the USEPA Beach Action Value for the Enterococcus qPCR quantification method (≥1000 CCEs/100), while the grey highlighted days represent beach days posted by Public Health based on their 24 h-delayed E. coli geometric mean. Error bars represent the standard deviation of log-transformed Enterococcus calibrator cell equivalents.

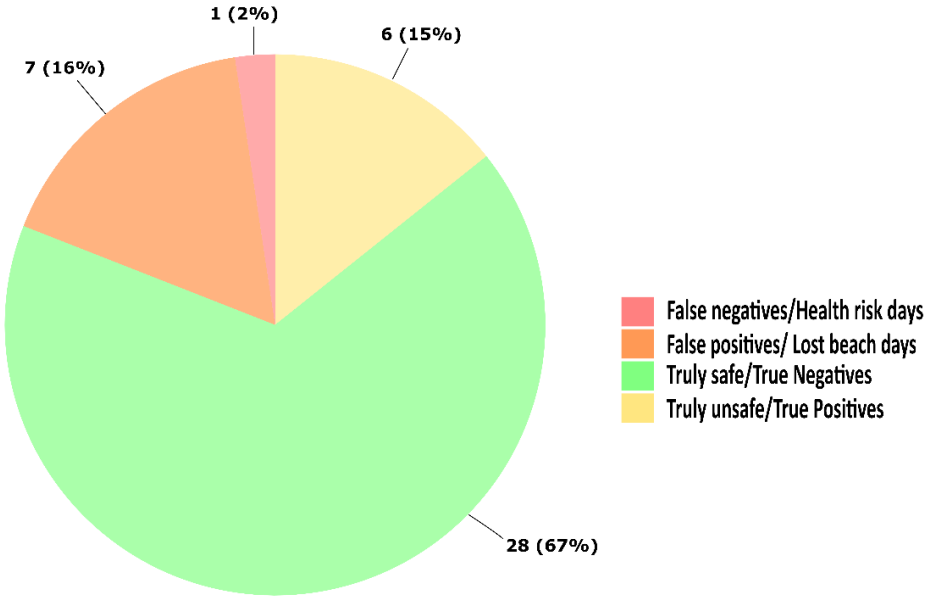


Figure 5. Analysis of impacts on Sunset Beach postings using the same-day Enterococcus qPCR and compared with Public Health's reported 24 h-delayed E. coli geometric mean for our summer 2022 beach sampling days (n = 42).

Same-day *Enterococcus* qPCR provided results for 27 and 32 beach days for sand-porewater samples from Lakeside and Sunset Beaches, respectively. Of the 27 beach days for Lakeside Beach, same-day qPCR identified 7 (26%) beach days when *Enterococcus* density in beach sand was higher than the USEPA Beach Action Value for beach water quality (BAV; 1000 CCEs/100 mL) and the beach remained open for recreational activities (Figure 6 and Supplementary Table S2). For Sunset Beach, same-day *Enterococcus* qPCR identified 13 (41%) beach days when the beach was not posted and sand *Enterococcus* levels exceeded the USEPA BAV for beach water quality (Figure 7 and Supplementary Table S2).

Enterococcus qPCR for sand-porewater also identified two (7%) and three (9%) beach days for Lakeside and Sunset sand-porewater, respectively, when the beaches were posted (while *Enterococcus* densities remained below BAV) (Figures 6 and 7).

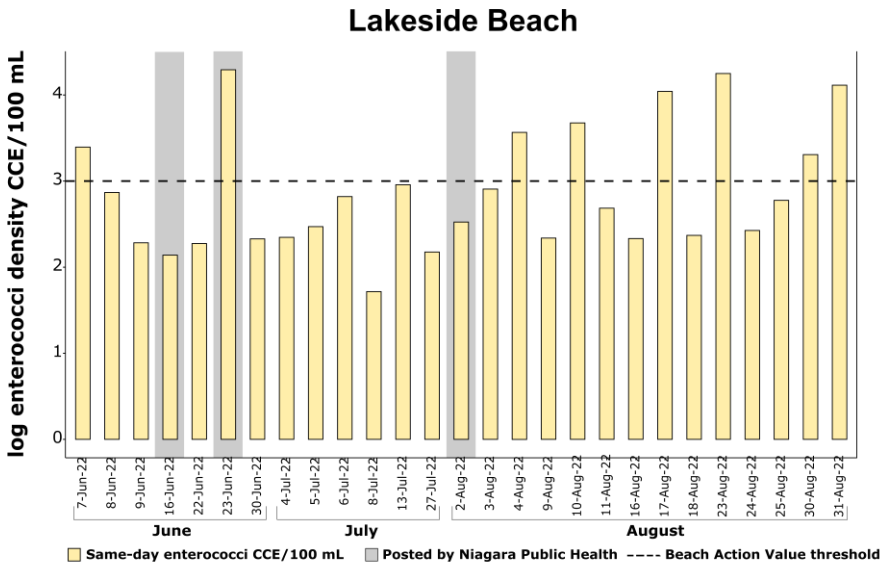


Figure 6. Comparison of Lakeside Beach sand Enterococcus exceedances of beach water Beach Action Value with Niagara Public Health's reported 24 h-delayed E. coli posting results for our summer 2022 beach sampling days (n = 27).

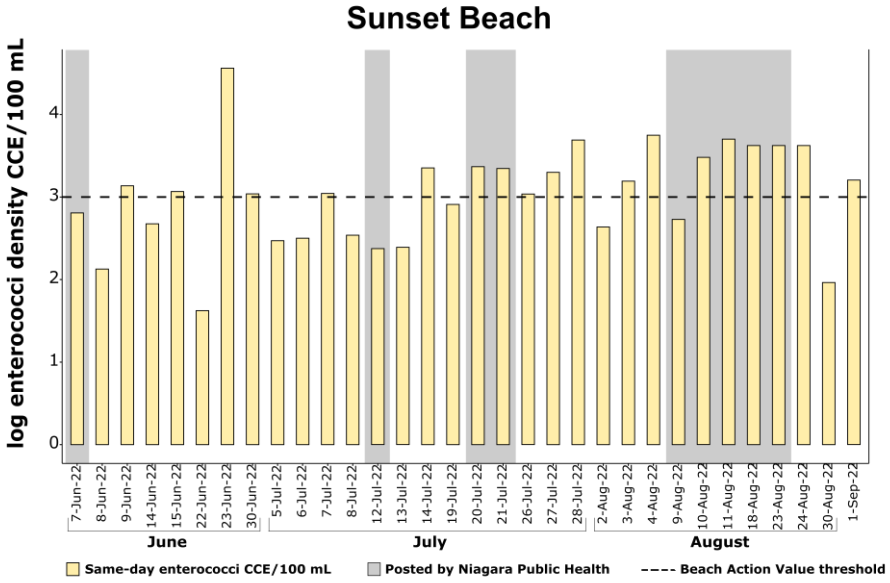


Figure 7. Comparison of Sunset Beach sand Enterococcus exceedances of beach water Beach Action Value with Niagara Public Health's reported 24 h-delayed E. coli posting results for our summer 2022 beach sampling days (n = 32)

E. coli culturing results were obtained for all 42 tested beach days. For sand-porewater from Lakeside Beach (Figure 8 and Supplementary Table S2), 25 (60%) beach days were identified with retrospective same-day *E. coli* density in sand higher than the Ontario Public Health posting threshold for beach water (200 CFUs/100 mL), and the beach remained open to the public. For sand-porewater samples from Sunset Beach (Figure 9 and Supplementary Table S2), we observed 22 (53%) beach days when retrospective same-day *E. coli* density in the sand was higher than the Ontario Public Health posting threshold for beach water and the beach remained open to the public. *E. coli* culturing also identified one (2%) and three (7%) beach days for Lakeside and Sunset Beaches, respectively, when beaches were posted while sand-porewater did not show BAV exceedance.

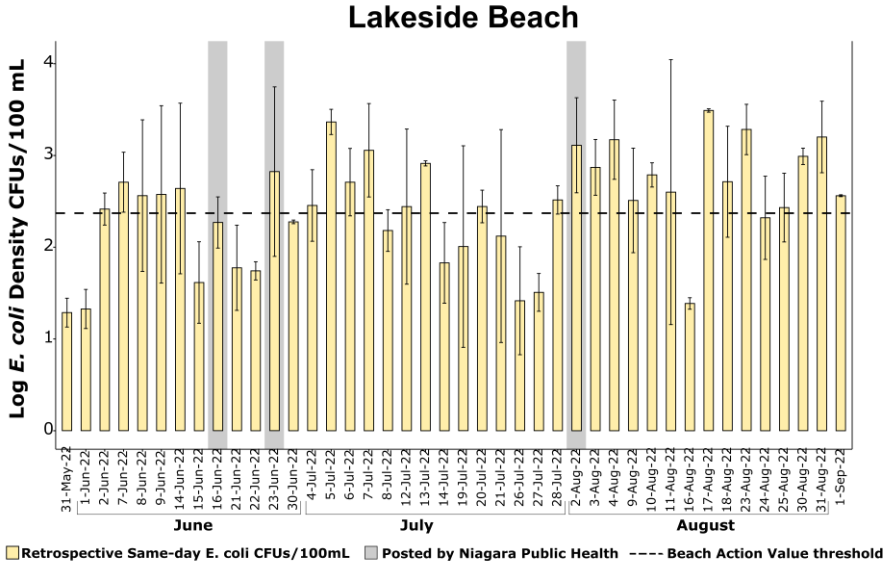


Figure 8. Comparison of Lakeside Beach sand retrospective same-day E. coli exceedances of Niagara Public Health's reported 24 h-delayed E. coli beach water postings for our summer 2022 beach sampling days (n = 42). Error bars represent the standard deviation of log-transformed E. coli colony-forming units.

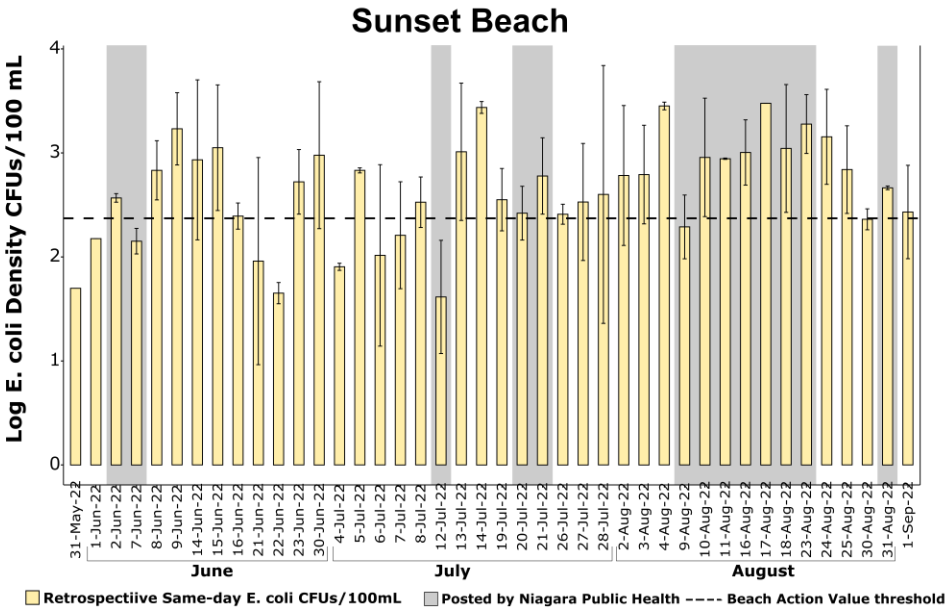


Figure 9. Comparison of Sunset Beach sand retrospective same-day E. coli exceedances of Niagara Public Health's reported 24 h-delayed E. coli beach water postings for our summer 2022 beach sampling days (n = 42). Error bars represent the standard deviation of log-transformed E. coli colony-forming units.

Discussion

Current beach monitoring practices rely heavily on culturing methodology, which has the well-known limitation of delayed data availability [9]. However, faecal indicator densities change within 24 h [4], and results obtained from the previous day's samples may not reflect the change in water quality. Rapid molecular methods, including *Enterococcus* qPCR, can provide results for beach quality monitoring within 3.5–4 h and can overcome the delay associated with culturing-based methods. This study continues our goal to evaluate the use and applicability of rapid molecular methods, including *Enterococcus* qPCR (USEPA 1609.1), for rapid quality assessment of different environments (beach water and sand) and geographical locations. Prior use of the *Enterococcus* qPCR-based method identified

additional beach open/health-risk days for the Toronto beaches (Marie Curtis Park East and Sunnyside Beaches) and that the beach sands are potential reservoirs of bacterial pathogens [5,6,10]. To further demonstrate the adaptability and applications of the *Enterococcus* qPCR-based method, in this study, we expanded our sampling regime to the Niagara beaches (Lakeside and Sunset Beaches), which have less impact from high-density human activity compared with the Toronto beaches. We systematically analysed sand-porewater as often as beach water for the same beach days to provide a comparative estimate of the faecal indicator densities between the two environments.

The *Enterococcus* qPCR analyses for the Niagara beaches revealed that all the beach water samples passed the USEPA 1609.1 qPCR quality control (QC) parameters (no significant qPCR inhibition or poor DNA recovery), and the results were reported to Public Health within 3.5 h after sample receipt in the lab. This was similar to our validation study at two Toronto beaches [5,6], further supporting the applicability of the same-day qPCR-based method for water quality monitoring at Ontario beaches. However, there remain challenges in adapting USEPA method 1609.1 for assessing sand-porewater quality. For sand-porewater samples, with the inclusion of the prefiltration step using 50 µm nylon mesh, we observed a higher proportion of samples passing the qPCR QC. Most sand-porewater samples that failed the qPCR QC criteria showed significantly higher Ct values for the sample processing control than the beach water samples, possibly due to DNA loss during the extraction (bead beating) step. Coral sands have been identified to interfere with DNA extraction using rapid DNA extraction protocols [2]. Similar to the coral sands, our results indicate that using the rapid DNA extraction protocol (USEPA 1609.1) in the presence of

sand particles can result in poor DNA recovery. However, the *Enterococcus* qPCR method can be modified further for sand quality monitoring by acidifying the water samples before filtration [2] or including additional prefiltration steps before sample processing.

Although beach sand can be a reservoir for faecal indicators and pathogenic bacteria [10–12], beach monitoring protocols of public health agencies currently rely on beach water quality analysis. In addition, factors such as erosion, rainwater runoff, and wave swash allow the bacterial communities in beach sand to enter beach water. Thus, beach sand can serve as a non-point source of faecal contamination [22,23]. Our results indicate little correlation between the samples from the different environments (beach water vs. sand faecal indicator densities). Similar to our study on Toronto beach waters [5], a significant positive correlation was observed between retrospective same-day *E. coli* culturing and *Enterococcus* qPCR for beach water samples taken from the two Niagara beaches. In this Niagara study, these two beach monitoring methods were also correlated with sand-porewater samples. However, additional research is needed to understand better the potential for health risks associated with faecal indicator bacteria in beach sand, as epidemiology studies have largely focused on beach water. Additional research is also needed to further advance the methods and safety thresholds for assessing faecal indicator bacteria levels in beach sand, as there are no standard field sampling and lab processing methods for beach sand.

Enterococcus qPCR analyses identified additional beach loss days and health-risk days for both Lakeside and Sunset Beaches. In addition, the *E. coli* culturing method posted the beaches more frequently than the *Enterococcus* qPCR. Our results concur with a Chicago beaches (U.S.) study [4], which concluded that there is no better than a 50% chance of

agreement between the previous-day's (first sampling day in two consecutive beach days) and next-day's (second sampling day in two consecutive beach days) faecal indicator densities, and that culturing-based methods can generate about three times more beach postings. Faecal indicator densities in recreational waters can change markedly within hours [4,24], leading to the potential for less reliable beach posting decision-making if the 24 h-delayed culturing-based methods are used alone for beach monitoring.

Comparing beach sand-porewater faecal indicator densities (*Enterococcus* qPCR and *E. coli* culturing) with Public Health's beach postings revealed that a substantial number of our beach testing days had higher faecal indicator densities in the beach sand than the shoreline water. Noticeably higher faecal indicator densities in sand-porewater than in beach water suggest that faecal contamination sources may significantly impact the sand but have less or no contact with neighbouring beach water [25]. Similarly, studies performed from beaches in the U.S. [26,27], Australia [28], and France [29] have demonstrated the pattern of comparatively higher faecal indicator densities in beach sand than in shoreline water. Alternatively, wave-induced resuspension of sand bacteria into beach water can deteriorate the water quality near the beach shoreline [30,31] and, for some beaches, can serve as a source of enterococci dissemination into beach water [7,32]. Therefore, standard water quality testing at a distance out from the shoreline may not provide real-time or comprehensive information about faecal bacteria dynamics for a complex beach water/sand interface.

This study identified differences in beach posting interpretations using the same-day *Enterococcus* qPCR method compared with Public Health's 24 h-delayed *E. coli* culturing-

based results. Differences in beach posting outcomes might be particularly due to changing water conditions over the 24 h delay associated with the culturing-based *E. coli* method. The significance of these differences could have been much higher if our sampling had been performed for the whole summer and extended to all 19 Niagara region beaches (www.niagararegion.ca/living/water/beaches/default.aspx) (accessed on 15 September 2022)Our results provide further support for public health agencies to consider adopting the same-day USEPA *Enterococcus* qPCR-based method for beaches.

5. Conclusions

1. The *Enterococcus* qPCR-based beach monitoring method provided same-day results within 3.5 h of sample processing for the Niagara beaches (Lakeside and Sunset Beaches).
2. There was no correlation between the sand-porewater and beach water faecal indicator densities for either the culturing- or qPCR-based methods.
3. Of the 33 beach days tested, *Enterococcus* qPCR analyses identified 3 (7%) and 7 (16%) false positive/lost beach days for Lakeside and Sunset Beaches, respectively.
4. Sand-porewater testing indicated that up to 60% of the tested days exceeded the beach water quality posting thresholds. However, beach water quality thresholds may not be applicable for sand-porewater quality, and therefore further evaluation is required to identify suitable safety thresholds for sand-porewater.
5. While a prefiltering screen improved the *Enterococcus* qPCR method by ~10% for assessing *Enterococcus* levels in beach sand-porewater, additional research is required to advance the application of this method for evaluating beach sands.

6. The *Enterococcus* qPCR-based method provides rapid/robust beach water quality assessment and can augment current beach monitoring methods.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1. Figure S1: Collection of sand-porewater samples from foreshore sand regions of sampling sites. (a) Hole for sand-porewater collection, (b) Seeped groundwater inside sand-porewater collection hole, (c) Sand-porewater collected in 500 mL sterile PET bottles, and (d) Prefiltration assembly with 50 µ nylon filter enclosed in plastic mesh for sand-porewater samples. Figure S2: Comparison between sample processing control (SPC; salmon sperm DNA) Ct values from beach water, QC-passed sand-porewater, and QC-failed sand-porewater samples. Figure S3: Correlation between sand-porewater log *E. coli* CFUs/100 mL and beach water enterococci calibrator cell equivalents (CCEs/100 mL) for Lakeside and Sunset Beaches. Correlation analysis was performed using Pearson correlation at a 95% confidence interval. Figure S4: Correlation between enterococci calibrator cell equivalents (CCEs/100mL) of beach water and sand-porewater from Lakeside and Sunset Beaches. Correlation analysis was performed using Pearson correlation at a 95% confidence interval. Figure S5: Correlation between sand-porewater log *E. coli* CFUs/100 mL and enterococci calibrator cell equivalents (CCEs/100 mL) for Lakeside and Sunset Beaches. Correlation analysis was performed using Pearson correlation at a 95% confidence interval. Figure S6: Correlation between wave height and log enterococci calibrator cell equivalents (CCEs/100 mL) for sand-porewater and beach water from Lakeside and Sunset Beaches. Correlation analysis was performed using Pearson correlation at a 95% confidence interval. Wave height data were obtained from Niagara Public Health for analysis. Figure

S7: Correlation between wave height and sand porewater log *E. coli* CFUs/100 mL for Lakeside and Sunset Beaches. Correlation analysis was performed using Pearson correlation at 95% confidence interval. Wave height data were obtained from Niagara Public Health for analysis. Table S1: Geographical coordinates for sampling sites from Lakeside and Sunset Beaches. Table S2: Beach posting status in comparison with *E. coli* culturing beach water threshold exceedances for sand-porewater samples from Lakeside and Sunset Beaches.

Author Contributions: validation, F.S.; methodology, F.S.; writing—original draft preparation, F.S.; writing—review and editing F.S., T.A.E., and H.E.S.; supervision, H.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Government of Ontario (Canada-Ontario Agreement grant agreement #2607). Such support does not indicate endorsement by the Government of Ontario of the contents or conclusions of this contribution.

Data Availability Statement: Data is contained within the article or supplementary material

Acknowledgments: We would like to thank Glen Hudgin and Jeremy Kelly from Niagara Public Health, who provided guidance and support in data acquisition. Albert Simhon, Ontario Ministry of Environment, Conservation, and Parks (MECP), provided support and suggestions throughout the project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. OMHLTC (Ontario Ministry of Health and Long-Term Care, Population and Public Health Division). Operational Approaches for Recreational Water Guideline. 2018.

Available online: https://www.health.gov.on.ca/en/pro/programs/publichealth/oph_standards/docs/protocols_guidelines/Operational_Approaches_to_Rec_Water_Guideline_2018_en.pdf (accessed on 1 June 2022).

2. Seruge, J.; Wong, M.; Noble, R.T.; Blackwood, A.D.; Moravcik, P.S.; Kirs, M. Application of a rapid qPCR method for enterococci for beach water quality monitoring purposes in Hawaii: Loss of DNA during the extraction protocol due to coral sands. *Mar. Pollut. Bull.* **2019**, *149*, 110631.
3. Sabino, R.; Rodrigues, R.; Costa, I.; Carneiro, C.; Cunha, M.; Duarte, A.; Faria, N.; Ferreira, F.C.; Gargaté, M.J.; Júlio, C.; et al. Routine screening of harmful microorganisms in beach sands: Implications to public health. *Sci. Total Environ.* **2014**, *472*, 1062–1069.
4. Dorevitch, S.; Shrestha, A.; DeFlorio-Barker, S.; Breitenbach, C.; Heimler, I. Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: Implications for public notification. *Environ. Health* **2017**, *16*, 45.
5. Saleem, F.; Edge, T.A.; Schellhorn, H.E. Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring. *J. Great Lakes Res.* **2022**, *48*, 707–716.
6. Saleem, F.; Schellhorn, H.E.; Simhon, A.; Edge, T.A. Same-day *Enterococcus* qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. *Can. J. Public Health* **2023**, 1–12. https://doi.org/10.17269/s41997-023-00763-8.

7. Phillips, M.C.; Solo-Gabriele, H.M.; Piggot, A.M.; Klaus, J.S.; Zhang, Y. Relationships between sand and water quality at recreational beaches. *Water Res.* **2011**, *45*, 6763–6769.
8. Health Canada. *Guidelines for Canadian Recreational Water Quality*, 3rd ed.; Catalogue No. H129–15/2012E; Water, Air and Climate Change Bureau, Healthy Environments and Consumer Safety Branch, Health Canada: Ottawa, ON, Canada, 2012.
9. Brandão, J.; Weiskerger, C.; Valério, E.; Pitkänen, T.; Meriläinen, P.; Avolio, L.; Heaney, C.D.; Sadowsky, M.J. Climate Change Impacts on Microbiota in Beach Sand and Water: Looking Ahead. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1444.
10. Mohiuddin, M.M.; Salama, Y.; Schellhorn, H.E.; Golding, G.B. Shotgun metagenomic sequencing reveals freshwater beach sands as reservoir of bacterial pathogens. *Water Res.* **2017**, *115*, 360–369.
11. Whitman, R.L.; Harwood, V.J.; Edge, T.A.; Nevers, M.B.; Byappanahalli, M.; Vijayavel, K.; Brandão, J.; Sadowsky, M.J.; Alm, E.W.; Crowe, A.; et al. Microbes in beach sands: Integrating environment, ecology and public health. *Rev. Environ. Sci. Bio/Technol.* **2014**, *13*, 329–368.
12. Solo-Gabriele, H.M.; Harwood, V.J.; Kay, D.; Fujioka, R.S.; Sadowsky, M.J.; Whitman, R.L.; Wither, A.; Caniça, M.; Da Fonseca, R.C.; Duarte, A.; et al. Beach sand and the potential for infectious disease transmission: Observations and recommendations. *J. Mar. Biol. Assoc. United Kingd.* **2016**, *96*, 101–120.
13. Heaney, C.D.; Sams, E.; Wing, S.; Marshall, S.; Brenner, K.; Dufour, A.P.; Wade, T.J. Contact with beach sand among beachgoers and risk of illness. *Am. J. Epidemiol.* **2009**, *170*, 164–172.

14. Heaney, C.D.; Sams, E.; Dufour, A.P.; Brenner, K.P.; Haugland, R.A.; Chern, E.; Wing, S.; Marshall, S.; Love, D.C.; Serre, M.; et al. Fecal indicators in sand, sand contact, and risk of enteric illness among beachgoers. *Epidemiology* **2012**, *23*, 95.
15. Brandão, J.; Valério, E.; Weiskerger, C.; Veríssimo, C.; Sarioglou, K.; Babič, M.N.; Solo-Gabriele, H.M.; Sabino, R.; Rebelo, M.T. Strategies for Monitoring Microbial Life in Beach Sand for Protection of Public Health. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5710.
16. Halliday, E.; Griffith, J.F.; Gast, R.J. Use of an exogenous plasmid standard and quantitative PCR to monitor spatial and temporal distribution of *Enterococcus* spp. in beach sands. *Limnol. Oceanogr. Methods* **2010**, *8*, 146–154.
17. USEPA. Method 1609.1: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay. EPA-820-R-15–099. 2015. Available online: https://www.epa.gov/sites/default/files/2015–08/documents/method_1609–1-*Enterococcus*-iac_2015_3.pdf (accessed on 1 May 2022).
18. Wade, T.J.; Calderon, R.L.; Brenner, K.P.; Sams, E.; Beach, M.; Haugland, R.; Wymer, L.; Dufour, A.P. High sensitivity of children to swimming-associated gastrointestinal illness: Results using a rapid assay of recreational water quality. *Epidemiology* **2008**, *19*, 375–383.
19. Wade, T.J.; Sams, E.; Brenner, K.P.; Haugland, R.; Chern, E.; Beach, M.; Wymer, L.; Rankin, C.C.; Love, D.; Li, Q.; et al. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: A prospective cohort study.

Environ. Health **2010**, *9*, 66.

20. Alberta. Alberta Safe Beach Protocol. 2021. Available online: https://open.alberta.ca/publications/9781460145395 (accessed on 15 September 2022).
21. Villanueva, R.A.M.; Chen, Z.J. Ggplot2: Elegant graphics for data analysis. *Meas. Interdiscip. Res. Perspect.* **2019**, *17*, 160–167. https://doi.org/10.1080/15366367.2019.1565254.
22. Imamura, G.J.; Thompson, R.S.; Boehm, A.B.; Jay, J.A. Wrack promotes the persistence of fecal indicator bacteria in marine sands and seawater. *FEMS Microbiol. Ecol.* **2011**, *77*, 40–49.
23. Torres-Bejarano, F.; González-Márquez, L.C.; Díaz-Solano, B.; Torregroza-Espinosa, A.C.; Cantero-Rodelo, R. Effects of beach tourists on bathing water and sand quality at Puerto Velero, Colombia. *Environ. Dev. Sustain.* **2018**, *20*, 255–269.
24. Converse, R.R.; Griffith, J.F.; Noble, R.T.; Haugland, R.A.; Schiff, K.C.; Weisberg, S.B. Correlation between quantitative PCR and culture-based methods for measuring *Enterococcus* spp. Over various temporal scales at three California marine beaches. *Appl. Environ. Microbiol.* **2012**, *78*, 1237–1242.
25. Nevers, M.B.; Byappanahalli, M.N.; Nakatsu, C.H.; Kinzelman, J.L.; Phanikumar, M.S.; Shively, D.A.; Spoljaric, A.M. Interaction of bacterial communities and indicators of water quality in shoreline sand, sediment, and water of Lake Michigan. *Water Res.* **2020**, *178*, 115671.
26. Ishii, S.; Hansen, D.L.; Hicks, R.E.; Sadowsky, M.J. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. *Environ. Sci. Technol.*

2007, *41*, 2203–2209.

27. Halliday, E.; Gast, R.J. Bacteria in beach sands: An emerging challenge in protecting coastal water quality and bather health. *Environ. Sci. Technol.* **2011**, *45*, 370–379.
28. Whiley, H.; Austin, J.; da Silva, G.M.; Ross, K. Faecal indicator bacteria present in sand at south port beach, South Australia. *J. Coast. Res.* **2018**, *34*, 215–219.
29. Toubiana, M.; Salles, C.; Tournoud, M.-G.; Licznar-Fajardo, P.; Zorgniotti, I.; Trémélo, M.-L.; Jumas-Bilak, E.; Robert, S.; Monfort, P. Monitoring urban beach quality on a summer day: Determination of the origin of fecal indicator bacteria and antimicrobial resistance at Prophète Beach, Marseille (France). *Front. Microbiol.* **2021**, *12*, 710346.
30. Wu, M.Z.; O'Carroll, D.M.; Vogel, L.J.; Robinson, C.E. Effect of low energy waves on the accumulation and transport of fecal indicator bacteria in sand and pore water at freshwater beaches. *Environ. Sci. Technol.* **2017**, *51*, 2786–2794.
31. Vogel, L.J.; O'Carroll, D.M.; Edge, T.A.; Robinson, C.E. Release of *Escherichia coli* from foreshore sand and pore water during intensified wave conditions at a recreational beach. *Environ. Sci. Technol.* **2016**, *50*, 5676–5684.

Feng, Z.; Reniers, A.; Haus, B.K.; Solo-Gabriele, H.M.; Wang, J.D.; Fleming, L.E. A predictive model for microbial counts on beaches where intertidal sand is the primary source. *Mar. Pollut. Bull.* 2015, *94*, 37–47

the content.

Chapter 5. Utilizing novel *Escherichia coli*-specific conserved signature proteins for enhanced monitoring of recreational water quality

This chapter aligns with one of the major goals of my thesis to develop new and more robust molecular methods for diagnostic purposes. By integrating the revolutionary concept of Conserved Signature Proteins (CSPs)/Genes, I developed a highly specific qPCR diagnostic method for monitoring *E. coli* concentrations in recreational waters. This chapter describes the use of CSPs for developing a practical diagnostic strategy and highlights the significance of taxonomically specific proteins or genes for targeted remediation purposes.

*The text I present here is a **peer-reviewed manuscript published** in the MicrobiologyOpen journal. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Li, E., Tran, K.L., Rudra, B., Edge, T.A., Schellhorn, H.E. and Gupta, R.S., 2024. Utilizing novel Escherichia coli-specific conserved signature proteins for enhanced monitoring of recreational water quality. MicrobiologyOpen, 13(3), p.e1410.*

Author Contributions

FS performed the experiments, analyzed the data, wrote the original manuscript, and revised it as required by the reviewers. EL, KT, and BR helped with the experiments. RSG provided the Conserved Signature Protein sequences. HES, TAE, and RSG supervised the study and reviewed the manuscript before submission to the journal.

Abstract

Escherichia coli is a proxy indicator of fecal contamination in aquatic ecosystems, but identification using traditional culturing can take up to 24 h. The use of DNA markers such as Conserved Signature Proteins (CSPs) genes (unique to all species/strains of a specific taxon) can form the basis of novel PCR tests for the unambiguous identification/detection of targeted bacterial taxa of interest. We report the identification of three new highly-conserved CSPs (genes), *YahL*, *YdjO*, and *YjfZ*, found only within *E. coli/Shigella*. Using PCR primers based on highly-conserved regions within these CSPs, we developed qPCR assays for the assessment of *E. coli/Shigella* species in water ecosystems. In-silico and experimental PCR testing confirmed the absence of sequence match when tested against other bacteria, confirming 100% specificity of the tested CSPs for *E. coli/Shigella*. qPCR assays for each of the three CSPs provided reliable quantitation for all tested enterohaemorrhagic and environmental *E. coli*, which is required for water testing. For recreational water samples, CSP-based quantification was highly correlated ($r > 7$, p-value < 0.01) with conventional viable *E. coli* enumeration, indicating that novel CSP-based qPCR assays for *E. coli* can be robust tools for water ecosystem monitoring and other important areas, including food monitoring.

Keywords: *Escherichia coli*, *Shigella*, Conserved Signature Proteins, qPCR, Fecal Indicators, Recreational beaches, Water ecosystems.

Introduction:

Escherichia coli (*E. coli*) is a common intestinal inhabitant of homeotherms, including humans and may be found in environmental waters due to fecal contamination (Kostyla et al., 2015). Water quality monitoring agencies commonly use *E. coli* density as a fecal contamination proxy indicator of freshwater quality, and *E. coli* Beach Action Values (BAVs) are used as posting guidelines for recreational waters and protecting beachgoers from gastrointestinal illnesses (USEPA, 2012; Health Canada, 2023). Although bacteria from non-fecal sources typically predominate in water ecosystems (Becerra-Castro et al., 2016; Sun et al., 2019), enteric pathogens, even at low levels, can be detrimental to public health (Griffith et al., 2016; Karojkic et al., 2018). Therefore, any *E. coli* enumeration methodology should be specific only to this fecal indicator and sensitive enough to detect/quantify lower densities, even in the case of stochastic (outlier) contamination events.

Conventional methods of *E. coli* detection for water ecosystems usually rely on culturing-based enumeration techniques, which have an inherent limitation of an 18-24 h delay before results are known (Dorevitch et al., 2017; Saleem et al., 2022). Furthermore, *E. coli* counts of the prior day may not be a good estimator following day's water quality (Saleem et al., 2023), suggesting that the changing *E. coli* concentrations within 24 h of sample collection may impact the reliability of water posting decisions. *E. coli*-specific chromogenic media is a method of choice for culturing-based enumeration methods, but some of these media can have high false-positive rates due to chemical constituents in complex water samples (McLain et al., 2008; McLain et al., 2011). Commonly used Most Probable Number (MPN)

methods, including Colilert-18 for *E. coli* enumeration, can have >3% false positive/negative rate for freshwater *E. coli* isolates (Chao et al., 2004). Chromogenic media tests for FIB mainly rely on the activity of enzymes, including β-galactosidase, which some environmental isolates may not express under temperature (44.5°C) required for *E. coli* testing (Alonso et al., 1998). To overcome the limitations associated with culturing-based methods, nucleic acid amplification-based methods have been developed for FIB like *E. coli* and *Enterococci*, including qPCR (Chern et al., 2011; USEPA, 2015; Haugland et al. 2021), Droplet Digital PCR (Cao et al., 2015; Ibekwe et al., 2020), and RNA-based RT-PCR (Heijnen and Medema, 2009) have been developed for water quality monitoring.

Culture-independent molecular methods can serve as a potential alternative to conventional culturing-based methods (Ricchi et al., 2017). In PCR-based methods, quantification or detection of specific FIB depends on genetic markers that are taxonomically conserved only in the target and universal across all target strains. For example, *E. coli* detection PCR assays have commonly targeted hypervariable regions in universally distributed genes, including 16S rRNA gene (Clifford et al., 2012) and 23S rRNA gene (Ahmed et al., 2012). Although universal taxonomic markers have been developed for environmental water testing, inherent limitations related to the specificity and sensitivity of assays can lead to false positives/negatives (Maheux et al., 2009; Gensberger et al., 2014; Zhang et al., 2015). Additionally, universal taxonomic markers in environmental variants/isolates can have differing gene copy numbers (Kembel et al., 2012) or nucleotide polymorphisms (Hakovirta et al., 2016), which can impact the reliability and accuracy of qPCR assays for targeted taxa.

Therefore, it is important to identify conserved genes that are uniquely found in a fecal indicator bacteria, based on which specific qPCR assays can be designed for environmental water testing.

In this study, we report the identification of several conserved signature proteins (CSPs) whose gene sequences are uniquely found in different *E. coli* strains. Extensive earlier work on CSPs, specific for other microbial taxa, shows that the sequences of these molecular markers provide reliable means for the demarcation of diverse microbial taxa at multiple phylogenetic depths (Gao et al., 2009; Gao and Gupta, 2012; Naushad et al., 2014). In view of their taxon-specificity and predictive ability to be found in other members of a specific taxon, the sequences of these taxon-specific CSPs also provide highly specific means for developing novel diagnostic tests for qualitative/quantitative assessment of specific microorganisms in biological samples, including water ecosystems (Gupta and Griffiths, 2006; Wong et al., 2014). Because of the specificity of these CSPs for a particular taxon, qPCR protocols utilizing them can overcome specificity limitations associated with other conventional universal markers, such as the 16S rRNA gene. This proof of concept study aims to identify new *E. coli/Shigella*-specific CSPs and explore their potential use in the development of robust qPCR assays for water quality monitoring. Specific questions we address in this study are: 1) Can conserved signature proteins/genes unique to *E. coli* (and *Shigella*) be identified? 2) Can CSPs/genes be used to develop a qPCR protocol for potential water monitoring strategies? 3) Is there a good correlation between *E. coli/Shigella*-specific

CSPs gene copies and *E. coli* Colony Forming Units (CFUs) from recreational water samples?

Experimental Procedures

Identification of *E. coli* and *Shigella* spp. Conserved Signature Proteins/DNA sequences:

E. coli and *Shigella*-specific Conserved Signature Proteins/DNA sequences were identified by methods used in our previous studies (Gao and Gupta, 2007; Gupta and Mok, 2007; Gupta and Mathews, 2010). Local BLASTp (Altschul et al., 1990) searches were initially conducted on individual proteins from *Escherichia coli* str. K-12 substr. MG1655 against a database of >2000 different genomes, including>500 genomes for available *Enterobacterales* species and >200 genomes for diverse *E. coli/Shigella* strains. Based on these BLASTp searches, candidate *E. coli* signature proteins were identified for which all significant BLASTp hits were for *E. coli/Shigella* strains, and the homologs for these proteins were either not found in other bacteria or their E values were <1e-3. Additional BLASTp searches were conducted on the protein sequences of candidate *E. coli* CSPs against the NCBI non-redundant (nr) database without the low-complexity filter, and the top 5000 hits were examined. Based on these BLASTp searches, those proteins were identified where all significant BLASTp hits (E value <1e-3) were for *E. coli/Shigella* strains and the protein was broadly found in >1000 *E. coli*/*Shigella* strains (Gao and Gupta, 2007; Gupta and Mok, 2007; Gupta and Mathews, 2010). The genes for three of the proteins identified by these searches (*YahL*, *YdjO*, and *YjfZ*) were chosen for these studies.

Primer/Probe Design and In-Silico Specificity Testing:

For the qPCR assays, PCR primers sets were designed for the three *E. coli*/*Shigella* CSPs to be less than 120 bp in size for efficient PCR amplification. The sequences of the PCR primers and qPCR probes for the three CSPs are indicated in Table 1. Specifically, the primers for *YahL*, *YdjO*, and *YjfZ* qPCR assays generated amplicon sizes of 112, 98, and 114 bases, respectively. The in-silico specificity of these primers was tested using Primer-BLAST (Ye et al., 2012) against NCBI nr and RefSeq genome databases with default parameters and specifying organism type as 'bacteria.' Additionally, in-silico amplification specificity was also tested on the phylum level by performing separate searches against each bacterial phylum in the RefSeq genomes database. Probes were designed using specific quality criteria (Lim et al., 2011): 1) Location of the probes was kept in close proximity to one of the primers, 2) Melting temperature of the probes was kept at 5-10 degrees higher than the primers, and 3) GC content was kept between 35-65%. Probes were aligned against the NCBI RefSeq databases to check for specificity.

Table 1. Primers and Probes targeting E. coli/Shigella-specific CSPs for qPCR assays.

Gene	Primer/ Probe	Sequence (5′ to 3′)	Amplicon Size (bp)
YahL	Forward Reverse Probe	ACAGACGCGCCCATTAAGC CGTCCAGAACAGAGAGCAATAA (FAM*)-AGGCGCTTGCGCAT GGATTATT-(MGBNFQ*)	112
YdjO	Forward Reverse Probe	TTCTCGCTACAGGCACATTC GGCGATGCATACTGACTCAT (FAM*)-TGAGCCAGGAATGTATTG ATAAGTTGGACA-(MGBNFQ*)	98
YjfZ	Forward Reverse Probe	CAACAGGACGTATGCTCTATCG GCCGTAAACCTTCTGCTAACTC (FAM*)-ACCTCAGCTTTAGACGA AATATATGGTGGT-(MGBNFQ*)	114

*[FAM]: 6-Carboxyfluorescein (Fluorophore); *[MGBNFQ]: Minor groove binding and non-fluorescent quencher

Bacterial Strain Growth and In-vitro Specificity Testing:

The experimental specificity of the PCR primers was examined using negative controls, including *Citrobacter rodentium (Enterobacteriaceae)* and *Serratia marcescens* (*Pseudomonadota*) as in-group negative controls, and *Micrococcus luteus* (*Actinomycetota*), *Bacillus subtilis* (*Bacillota*), and *Staphylococcus epidermis* (*Bacillota*) as non-*Pseudomonadota* or out-group negative controls. Bacterial strains were grown on LB (Luria–Bertani) agar plates overnight at 37°C. DNA was extracted from single colonies by incubation at 98°C for 5 min in 30 µL of 0.2% SDS (Tris-EDTA) lysis buffer (Packeiser et al., 2013). The DNA concentration in lysate was measured using a QUBIT Fluorometer

(dsDNA High-Sensitivity Assay kit, Thermo Fisher Scientific, USA) to ensure successful DNA extraction. Initial primer pair specificity testing was performed using cell lysate from *E. coli* K12, *C. rodentium*, *S. marcescens* and *M. luteus*, with 16S rRNA amplification as PCR reaction positive control and non-template control. The primers and probes (qPCR assays) for the three CSPs were tested with *E. coli* K12 as positive control, negative controls (Discussed earlier), and wastewater DNA as environmental sample positive control. To test for the PCR specificity and nucleotide identity of CSPs in microbially complex samples (wastewater samples), larger amplicon fragments (600-700 bp) for each CSP (Supplementary Table 1) were amplified in a total of 25 µL PCR reaction mix containing 1 µL of each primer (10 µm), 12.5 µL of Environmental master mix 2.0 (Thermo Scientific USA), and 10.5 µL of nuclease-free water and 1 µL of DNA extracted from wastewater samples. PCR amplification cycle consisted of initial denaturation at 98°C for 10 min, followed by 35 cycles of 98°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 60 s, followed by final extension at 72°C for 5 min. Amplified PCR fragments for each CSP were then purified using Monarch DNA Gel Extraction Kit (New England Biolabs), followed by Sanger sequencing on the SeqStudio Flex Genetic Analyzer at Farncombe Sequencing Institute (McMaster University). Sequenced CSP fragments were aligned against NCBI nr and RefSeq genomes databases to validate that the amplified fragments correspond to *E. coli/Shigella*.

qPCR Assay Development and Sensitivity Testing:

PCR fragments (98-114 bp) for each assay were purified using Monarch DNA Gel Extraction Kit (New England Biolabs), followed by DNA quantification and copy number calculations. To avoid inherent *E. coli* DNA contamination from master mixes (Palomino-Kobayashi et al., 2022), Environmental master mix 2.0 (Thermo Scientific, USA), which contains ultra-purified Taq Polymerase was used (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_079133.pdf). Purified DNA was diluted 10-fold to generate DNA standards ranging from 10^7 to 10^1 gene copies/µL. Standard curves for each assay were generated in 25 µL total qPCR reaction containing 1 µL of each primer (10 µm), 1 µL of probe (100µm), 12.5 µL of Environmental master mix 2.0 (Thermo Scientific USA), and 9.5 µL of nuclease-free water. qPCR program included initial denaturation at 98°C for 10 min, followed by 40 cycles of 98°C for 30 sec and 60°C for 30 sec. Standard curves were only accepted if the coefficient of determination (R^2) was higher than 0.95 and amplification efficiency was >90%. qPCR assay sensitivity was determined by analyzing DNA extracted from ten enterohemorrhagic *E. coli* strains (Karmali et al., 2003; Riley et al., 1983), including O98:H25-EC3, O84:NM-EC2, O172:NM-EC6, O103:H25-N00, O121:NM-N99, O113:H21-CL3, O5:NM-N00, O111:NM, O121:H19, O157:H7, and seven environmental *E. coli* isolates from aquatic ecosystems (obtained from Environment Canada). Gene copies/ng of DNA were calculated using slope and intercept values from standard curves generated for each assay. Each CSP gene target exists as a single gene copy per genome, and the single gene copy targets are an equivalent measure of the number of microorganisms

(Harwood et al., 2014). Therefore, the lower limit of detection (LLOD) for each assay was calculated as the lower limit of quantification (LLOQ), defined as the minimum number of gene copies that can be reliably detected per reaction (Klymus et al., 2020). The lower limit of quantification (LLOQ) was calculated by analyzing dilutions of standards in the range of 2 to 10 gene copies/reaction, and the coefficient of variation between the replicates of each qPCR assay was less than 15%

Recreational Water Sample Collection and *E. coli* enumeration by culture:

Water sample collection from recreational beaches and *E. coli* enumeration were performed as described earlier (Saleem et al., 2022; Saleem et al., 2023). In brief, 309 water samples were collected from two freshwater beaches (Marie Curtis Park East and Sunnyside beaches) and their adjacent river mouths (Etobicoke Creek and Humber River) between May 31, 2022 and August 26, 2022. Water samples were delivered to the lab within 1 h of sample collection and processed for *E. coli* enumeration by filtering 100 mL of water sample through a 0.45 µm polycarbonate membrane filter (Millipore Corp., Bedford, MA) and incubating filters on differential coliform agar (OxoidTM) for 24h at 44.5°C. Only the samples exceeding the USEPA *E. coli* enumeration beach action value (≥235 CFUs/100mL, n = 30) (USEPA, 2012) were used for qPCR testing.

DNA Extraction, Application of CSP qPCR assays for Recreational Waters and Data Analysis:

Approximately 100 mL of water sample was filtered through a 0.22 µm nitrocellulose membrane filter (Millipore Corp., Bedford, MA), followed by DNA extraction using the Norgen Soil Plus DNA Extraction kit (Norgen Biotek Corp., Canada), as described previously (Saleem et al., 2024). The final eluate volume of DNA was 50 µL. The DNA concentration was measured using the QUBIT fluorometer (Thermo Scientific, USA). qPCR assays and gene copy estimation for DNA from water samples were performed as described in an earlier section. For correlation analysis, data was log-transformed, and Shapiro-Wilk's normality testing (Stats v3.6.2 R package) was used to determine the normal distribution, followed by either Spearman's or Pearson's methods for correlation analysis.

Results:

***E.coli*/*Shigella*-specific Conserved Signature Proteins/Genes**:

Conserved Signature Proteins/DNA sequences (CSPs) specific for *E. coli* and *Shigella* spp. were identified as described in the Methods section. Based on these studies, the genes for three CSPs (*YahL*, *YdjO* and *YjfZ*) found uniquely in *E. coli* and *Shigella* spp. were chosen for the present work. The sequences for these three CSPs matched only *E. coli* and *Shigella* spp. when aligned against NCBI(nr/nt) and RefSeq Genomes Databases. Some characteristics of these CSPs are indicated in Table 2. Of these three CSPs, two (*YdjO* and

YjfZ) are annotated as hypothetical/uncharacterized proteins as their cellular functions are yet to be determined.

Table 2. Conserved Signature Proteins/Genes specific for E. coli and Shigella spp.

Protein Name (Gene Symbol)	Gene ID (NCBI)	Protein Length (aa)	Gene Length (bp)
Uncharacterized Protein (*YahL*)	944970	271	816
Hypothetical Protein (*YdjO*)	917061	267	804
DUF2686 domain-containing protein (*YjfZ*)	948719	264	795

In-silico and Experimental Validation of Primer/Probe Specificity Based on the Conserved Signature Proteins/Genes:

In-silico PCR against NCBI Non-Redundant (Supplementary Table 2) and ResSeq genome databases (Supplementary Figures 1, 2 and 3) was used as a first step to assess the specificity of PCR amplification/detection. At the species level, in-silico PCR hits matched with only *E. coli* for the genus *Escherichia*, while *Shigella* hits corresponded to three species (*S. dysentriae*, *S. flexneri* and *S. sonnei*). To validate the specificity of the designed PCR primers for these three CSPs, colony PCR was performed using *E. coli* as a positive control, *C. rodentium* and *S. marcescens* as in-group negative controls, *M. luteus* as an out-group negative control, and 16S rDNA as PCR reaction positive control (Supplementary Figure 4). Similar to in-silico PCR, the primer sets for all three CSPs amplified DNA fragments at the

expected sizes for *E. coli*, and no amplification was observed in the examined in-group or out-group negative-control species. Further, probe-based qPCR assays for the CSPs were tested for specificity using *E. coli* DNA, wastewater samples and negative control species (*C. rodentium*, *S. marcescens*, *M. luteus*, *S. epidermis*, and *B. subtilis*) (Supplementary Table 3). Similar to the results for primer specificity tests, no amplification/fluorescence was observed for non-target species, while the probes for all three CSPs generated positive fluorescence for *E. coli* and wastewater DNA in qPCR assays. To test the PCR specificity and nucleotide identity of CSPs from complex microbial community samples (wastewater DNA), we amplified a larger (500-700 bp) PCR fragment for each CSP, which was then sequenced (Sanger) and aligned against RefSeq reference sequences (Supplementary Figures 5, 6 and 7). As expected, all three sequenced CSP fragments from wastewater DNA matched only to *E. coli* and *Shigella* species when tested against the NCBI RefSeq Genome Database. Query coverage for each CSP fragment ranged between 98% and 100 %, while percentage identity was 94-99.7 %.

qPCR Primer/Probe Testing and Quality Control Analytics:

qPCR assays for three CSPs were first validated on *E. coli* genomic DNA, wastewater DNA as positive controls, and negative/non-template controls (Supplementary Table 3). For three CSP assays, *E. coli* and wastewater DNA showed comparable threshold cycle (Cq) values. Following qPCR primer/probe testing, standard curves were generated for three CSP-based qPCR assays (Table 3). The coefficients of determination for all three qPCR assays were above 0.99, and the efficiency of amplification ranged between 92 and 101%. Lower limits

of quantification for *YahL*, *YdjO* and *YjfZ* qPCR assays were determined as 2, 6 and 2 gene copies, respectively.

Table 3. Standard curve quality parameters for each qPCR assay from three CSPs.

Quality Control Parameter	***YahL***	***YdjO***	***YjfZ***
Coefficient of Determination (R^2)	0.999	0.996	0.996
Slope	-3.2	-3.5	-3.3
Intercept	40.5	38.6	38.3
Efficiency (%)	101	92	99
Lower Limit of Quantification (LLOQ)	2	6	2

Sensitivity Testing using Pathogenic and Environmental *E. coli* strains:

The sensitivity of the qPCR assays was tested against ten hemorrhagic and seven environmental *E. coli* strains (Table 4). All three qPCR assays provided positive amplification for pathogenic and non-pathogenic *E. coli* strains, with gene copies per nanogram of genomic DNA ranging between 1.5 and 5.5. Gene copies for each strain were comparatively similar between the three qPCR assays. Additionally, a significant (p-value <0.001) positive correlation (r_p >0.7) was observed between three qPCR assays for gene copies obtained from *E. coli* strains (Supplementary Table 4). Specifically, a strong positive correlation was observed between *YahL* and *YdjO* (r_p = 0.92, p-value = 2.79E-04), followed by *YdjO*-*YjfZ* (r_p = 0.73, p-value = 9.40E-04), and *YahL*-*YjfZ* (r_p = 0.74, p-value = 9.40E-04).

Table 4. Sensitivity testing of YahL, YdjO and YjfZ qPCR assays for hemorrhagic and non-hemorrhagic (environmental isolates) E. coli strains.

***E. coli* Strain-Serotype-Seropathotype**	**Host**	**Source**	**Log Gene Copies/ng of DNA**		
			YahL	***YdjO***	***YjfZ***
O98:H25-EC3-377-E	Bovine	Karmali et al., 2003	4.6	3.9	4.3
O84:NM-EC2-044-E	Bovine	"	4.8	4.0	4.4
O172:NM-EC6-484-E	Bovine	"	4.9	4.1	1.5
O103:H25-N00-4859-D	Human	"	5.1	4.2	4.8
O121:NM-N99-4390-C	Human	"	4.8	4.0	4.5
O113:H21-CL3-C	Human	"	4.9	4.2	4.6
O5:NM-N00-4067-C	Human	"	5.3	4.6	5.2
O111:NM-R82F2-B	Human	"	5.0	4.3	4.2
O121:H19-CL106-B	Human	"	5.4	4.5	5.1
O157:H7-EDL933-A	Human	Riley et al., 1983	5.3	4.6	5.1
Environmental Isolate	-	Environ. Canada	4.4	3.7	4.1
Environmental Isolate	-	"	4.3	3.9	4.0
Environmental Isolate	-	"	5.2	4.5	4.3
Environmental Isolate	-	"	4.4	3.9	3.7
Environmental Isolate	-	"	4.5	4.0	3.6
Environmental Isolate	-	"	5.1	4.5	4.6
Environmental Isolate	-	"	4.8	4.1	4.5

Application of qPCR protocol for Beach Quality Monitoring:

Thirty recreational water samples collected from two different beaches and associated rivers were tested for each of the *E. coli/Shigella*-specific CSP-based qPCR assays (probe-based) to assess the applicability of these assays for beach monitoring applications. The detection rate of the *YahL* qPCR assay was 100% for the tested sites, followed by 96% for *YjfZ* and 93% for *YdjO* qPCR assays (Supplementary Table 5). Additionally, gene copy data from three qPCR assays were compared against culturing-based *E. coli* colony forming units (CFUs) data to assess the relationship between methods (Figure 1). Gene copies from all three qPCR assays showed a significant (p-value <0.001) positive (r >0.7) correlation with *E. coli* CFUs. Specifically, a strong correlation was observed between *YjfZ* and *E. coli* CFUs (r_p = 0.84, p-value = 3.99E-15), followed by *YahL*-*E. coli* CFUs (r_s = 0.78, p-value = 2.74E-12), and *YdjO*-*E. coli* CFUs (r_p = 0.65, p-value = 2.10E-07). Correlation analysis was also performed to test the relationship between quantification results from the three CSP-based qPCR assays for the same recreational waters (Figure 2). Similar to the culturable *E. coli* comparison, a strong significant (p-value <0.001) positive correlation (r_p >0.7) was observed for the gene copies obtained from the three different assays for the recreational water samples. Gene copies from *YahL* and *YjfZ* showed the strongest correlation (r_p = 0.84, p-value = 4.40E-16), followed by *YahL*-*YdjO* (r_s = 0.79, p-value = 2.9E-4), and *YdjO*-*YjfZ* (r_s = 0.74, p-value = 3.71E-4).

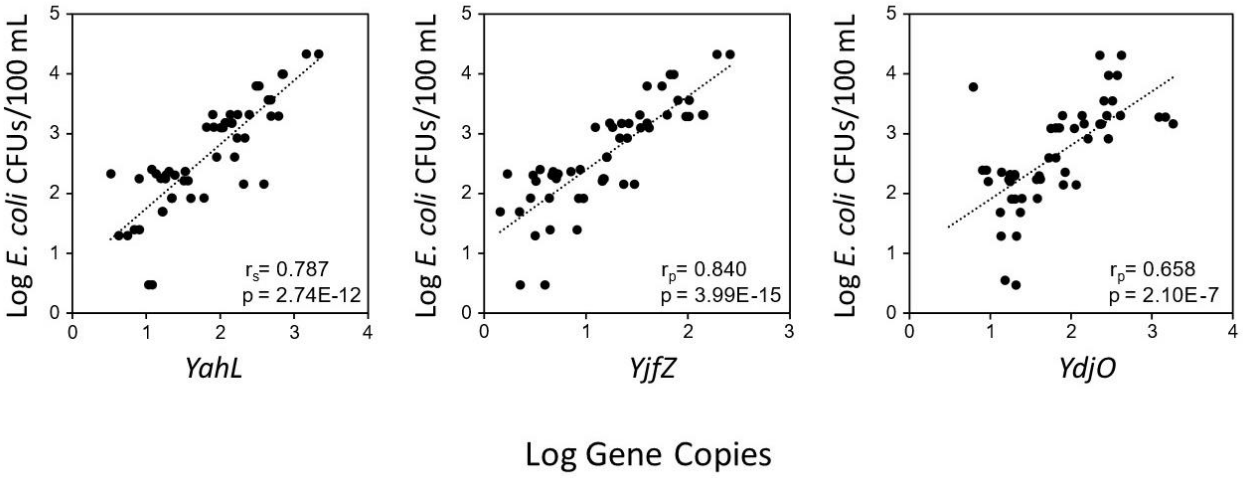


Figure 1. Correlation analysis between E. coli/Shigella-specific conserved signature DNA sequences-based qPCR assays and culture-based E. coli colony forming units for recreational beaches and rivers.

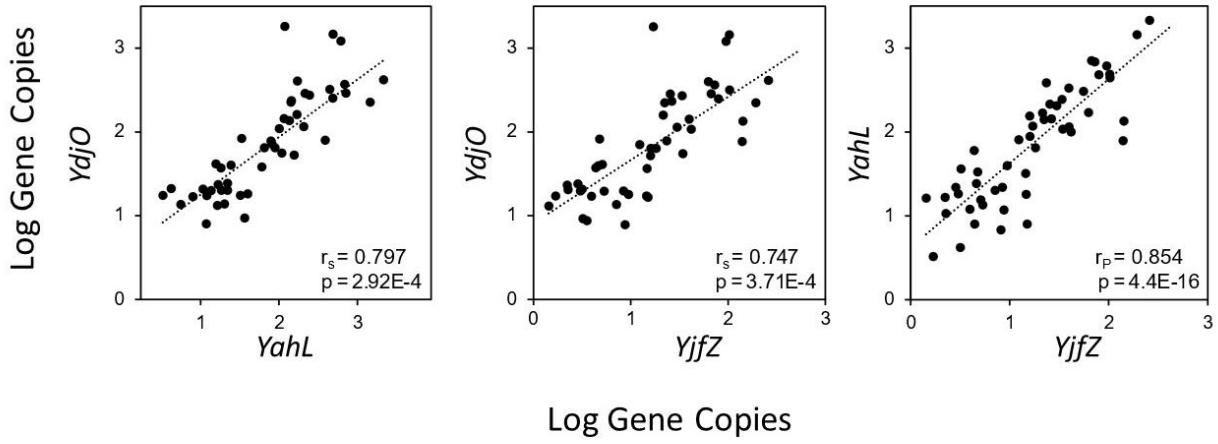


Figure 2. Correlation analysis between gene copies from three E. coli/Shigella-specific conserved signature DNA sequences-based qPCR assays for recreational beaches and associated rivers.

Discussion:

Beach water quality monitoring strategies mainly rely on testing fecal indicator bacteria, including *E. coli*, using culture-based methods, which can take up to 18-24 h and lead to delays in beach posting decisions (Dorevitch et al., 2017; Saleem et al., 2023). To date, universal taxonomic genes, including the 16S rRNA gene (Clifford et al., 2012) and the 23S rRNA gene (Ahmed et al., 2012), have been the primary targets of rapid qPCR-based methods for *E. coli* detection. However, due to the occurrence of environmental variants of these targeted sequences, qPCR assays targeting universal taxonomic markers may lack specificity (Maheux et al. 2009; Hakovirta et al., 2016) and sensitivity (Kembel et al., 2012). Compared to conventional universal taxonomic DNA markers, Conserved Signature Proteins (CSPs)/DNA sequences represent conserved genes which are unique to specific taxonomic groups (Gao and Gupta, 2007; Gupta and Mok, 2007; Gupta and Mathews, 2010; Naushad et al., 2014). Because of their taxonomic specificity and sequence conservation, the DNA sequences of these CSPs can be targeted to detect specific taxa of interest (Gupta and Griffiths, 2006). In this proof-of-principle study, we identified three *E. coli/Shigella*-specific CSPs and used them to develop a qPCR-based protocol for testing fecal pollution in recreational freshwater beaches.

In-silico and in-vitro primer testing validated that all three CSP sequences (*YahL*, *YdjO* and *YjfZ*) were specific for *E. coli* and *Shigella* species, highlighting their taxonomic/evolutionary conservation among the two taxa. As *Shigella* species are phylogenetically not distinct from *E. coli* (branch in between different *E. coli* strains) (Sims

and Kim, 2011; Meier-Kolthoff et al., 2014), the shared presence of these CSPs in both *Shigella* and *E. coli* is expected. A previous study (Walker et al., 2017) developed a qPCR method targeting the *ybbW* gene, which is purportedly specific to *E. coli* and thus not present in *Shigella* or other bacteria. However, when tested (BLASTn and in-silico PCR) against the NCBI RefSeq Representative Genome Database (Supplementary figures 8 and 9), we found *ybbW* to be also present in non-*E. coli* species, including *Escherichia marmotae, Shigella* and multiple non-*Escherichia* taxa *(*>70% percentage similarity, nucleotide matched >1000 bp). Similarly, non-specific in-silico matches for primers and probes were observed against *Klebsiella*, *Citrobacter* and other *Escherichia* species for *E. coli*-specific qPCR method based on the detection of the 23S rRNA gene (Chern et al., 2011; Lane et al., 2020). In comparison, *YahL*, *YdjO*, and *YjfZ* were only found in *E. coli* and *Shigella* species, which signifies their value for *E. coli/Shigella* detection.

Universal taxonomic markers (including 16S and 23S rRNA genes) typically rely on a few conserved nucleotides for taxonomic characterization, but the potential for diverse single nucleotide polymorphisms within or environmental, genetic variants in conserved DNA nucleotides can lead to false positive detection of target taxa (McIlroy et al., 2011; Thorson et al., 2016). In contrast to the other universal molecular markers used for taxonomic characterization or identification of species in environmental samples, where only a few nucleotides discriminate among different taxa, the entire coding sequences of the CSPs, which are generally quite large (in the present case ~800 bp), are specific for the members of a given taxon (*E. coli* and *Shigella* spp.). Hence, the PCR primers and qPCR probes based

on these sequences provide more reliable and highly specific means for the identification/characterization of genetically diverse species such as *E. coli* in complex/ever-evolving microbial environments such as water ecosystems.

CSP-based qPCR assays provided positive results for all ten pathogenic (including O157:H7) and seven environmental *E. coli* isolates with comparable gene copies between tested isolates, which signifies the potential of using these assays for broad-range environmental testing. False negative detection associated with conventional culture-based enumeration methods is a well-known problem (Kibbee and Örmeci, 2017; Ding et al., 2017). Specifically, *E. coli* O157:H7 can exist in a viable but not culturable (VBNC) state (Liu et al., 2020; Li et al., 2020) and cannot be detected using conventional culturing-based water methods at 44.5°C, which can lead to underestimation of health risks. Differences in correlation strengths can be due to the inability of culturing-based methods to culture all environmental *E. coli* isolates, including viable but not culturable cells and *E. coli* isolates, which may not grow at a specific incubation temperature (44.5 °C) recommended for culturing-based analysis (Pommepuy et al., 1996; Servais et al., 2009). Additionally, β-glucuronidase activity-based *E. coli* enumeration methods, including COLIFAST and COLIMINDER, can generate false positives by detecting β-glucuronidase-positive phenotypes belonging to *Klebsiella*, *Citrobacter*, *Aeromonas* and *Enterobacter, Yersinia and Salmonella* species (Feng and Hartman, 1992; Frampton and Restaino, 1993; Ciebin et al., 1995). if β-D-glucuronidase activity is either lacking (Maheux et al., 2008) or is present in lower levels (Fricker et al., 2010) in some environmental *E. coli* isolates, this could also

result in underestimated quantification. In comparison, qPCR assays can also detect VBNC *E. coli* and *Shigella* species, allowing estimation of the whole spectrum of targeted taxa in complex environmental samples. However, factors including the detection of genetic material from non-viable cells (Gedalanga et al., 2009), environmental nucleotide variants/polymorphisms (Boyle et al., 2009; Fernández-No et al., 2015) and some environmental strains carrying a different number of gene copies (Větrovský et al. 2013) can impact the PCR-based quantification methods. However, a significant positive correlation between gene copy estimates of the three assays can indicate a high level of agreement between the methods.

A previously described *E. coli*-specific RNA-based qPCR assay (Heijnen and Medema, 2009) has limited sensitivity due to a high lower limit of quantification (LLOQ = ~10^4 gene copies) (Walker et al., 2017). Environmental water samples can harbour diverse microbial communities with lower fecal indicator densities (Saleem et al., 2024), which may go unnoticed using qPCR assays with higher LLOQs (Walker et al., 2017). All three CSP-based qPCR assays tested in this study demonstrated low LLOQs for recreational water samples, featuring a sensitive detection for environmental water testing. Additionally, unlike RNA-based qPCR assays (as described previously), CSP-based assays showed high detection rate of *E. coli* gene copies from recreational waters, which can overcome the variable gene expression limitation associated with RNA-based qPCR assays. Furthermore, gene copies from three CSP-based assays showed significantly strong positive correlations with *E. coli*

Colony Forming Units, indicating the potential application of CSP-based assays as a rapid alternative to conventional culture-based methods for beach monitoring.

In this study, we developed three independent qPCR assays using three *E. coli/Shigella*-specific Conserved Signature Proteins/genes as targets. The potential of CSP-based qPCR assays to detect *E. coli/Shigella* species in complex recreational water samples was also explored. All three assays can also detect *Shigella* species (*S. sonnei, S. dysenteriae, S. boydii, and S. flexneri*) of public health concern (Health Canada, 2020), which strengthens the potential of these assays. This proof of principle study demonstrates the potential of Conserved Signature Proteins/DNA sequences for designing and developing taxonomically specific quantitative/qualitative molecular assays for other water-related and clinically important organisms. Additionally, this study can serve as the foundation for future studies to assess the relationship between CSP-based fecal indicator estimates and Beach Action Values or water quality thresholds and the development of other CSP-based tests for clinical, food and water quality surveillance.

Conclusions:

1. *YahL*, *YdjO*, and *YjfZ* proteins/genes were identified as conserved for *E. coli* and *Shigella sp* only, and in-silico/in-vitro testing validated the conservation of three CSPs, and their potential as molecular markers for developing PCR-based assays.

2. Positive amplification was observed for Enterohemorrhagic and environmental *E. coli* strains, indicating high detection sensitivity across a range of clinical and environmental isolates of *E. coli* for CSP-based qPCR assays.

3. *E. coli* Colony Forming Units (CFUs) from culturing-based tests and CSP gene copies (qPCR) from recreational water samples showed a significant positive correlation, indicating the potential of CSP-based qPCR assays for water monitoring applications.

4. CSP-based qPCR assays can be a rapid testing alternative to traditional culture-based testing methods for *E. coli* and offer a more phylogenetically targeted approach to the detection of *E. coli* and *Shigella* for water quality monitoring strategies.

Author Contributions

Faizan Saleem: Data curation; formal analysis; investigation; methodology; writing—original draft (lead); writing—review and editing. Enze Li: methodology; writing—review and editing. Kevin L. Tran: Formal analysis; methodology; writing—review and editing. Bashudev Rudra: Formal analysis; methodology. Thomas A. Edge – Conceptualization; writing—review and editing. Herb E. Schellhorn: Supervision; conceptualization; writing—review and editing. Radhey S. Gupta: Supervision; conceptualization; writing—review and editing.

Conflict of Interest Statement

The authors declare no competing financial interest.

Data Availability Statement:

All the analytical data supporting the findings in this study is provided in the figures and tables in the main manuscript and supplementary data files.

Funding information

This research was funded by an Alliance grant (#554507-20) by the Natural Sciences and Engineering Research Council of Canada

References:

Ahmed, W., Richardson, K., Sidhu, J.P. and Toze, S., 2012. *Escherichia coli* and *Enterococcus* spp. in rainwater tank samples: comparison of culture-based methods and 23S rRNA gene quantitative PCR assays. Environmental science & technology, 46(20), pp.11370-11376.

Alonso, J.L., Soriano, A., Amoros, I. and Ferrus, M.A., 1998. Quantitative determination of *E. coli*, and fecal coliforms in water using a chromogenic medium. Journal of Environmental Science & Health Part A, 33(6), pp.1229-1248.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Becerra-Castro, C., Macedo, G., Silva, A.M., Manaia, C.M. and Nunes, O.C., 2016. Proteobacteria become predominant during regrowth after water disinfection. Science of the Total Environment, 573, pp.313-323.

Boyle, B., Dallaire, N. and MacKay, J., 2009. Evaluation of the impact of single nucleotide polymorphisms and primer mismatches on quantitative PCR. BMC biotechnology, 9, pp.1-15.

Cao, Y., Raith, M.R. and Griffith, J.F., 2015. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. Water research, 70, pp.337-349.

Chao, K.K., Chao, C.C. and Chao, W.L., 2004. Evaluation of Colilert-18 for detection of coliforms and Eschericha coli in subtropical freshwater. Applied and Environmental Microbiology, 70(2), pp.1242-1244.

Chern, E.C., Siefring, S., Paar, J., Doolittle, M. and Haugland, R.A., 2011. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Letters in applied microbiology, 52(3), pp.298-306.

Chern, E.C., Siefring, S., Paar, J., Doolittle, M. and Haugland, R.A., 2011. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Letters in applied microbiology, 52(3), pp.298-306.

Clifford, R.J., Milillo, M., Prestwood, J., Quintero, R., Zurawski, D.V., Kwak, Y.I., Waterman, P.E., Lesho, E.P. and Mc Gann, P., 2012. Detection of bacterial 16S rRNA and identification of four clinically important bacteria by real-time PCR. PloS one, 7(11), p.e48558.

Ding, T., Suo, Y., Xiang, Q., Zhao, X., Chen, S., Ye, X. and Liu, D., 2017. Significance of viable but nonculturable *Escherichia coli*: induction, detection, and control.

Dorevitch, S., Shrestha, A., DeFlorio-Barker, S., Breitenbach, C. and Heimler, I., 2017. Monitoring urban beaches with qPCR vs. culture measures of fecal indicator bacteria: Implications for public notification. *Environmental Health*, *16*, pp.1-10.

Fernández-No, I.C., Böhme, K., Caamaño-Antelo, S., Barros-Velázquez, J. and Calo-Mata, P., 2015. Identification of single nucleotide polymorphisms (SNPs) in the 16S rRNA gene of foodborne Bacillus spp. Food microbiology, 46, pp.239-245.

For health professionals: Shigellosis (Shigella). Available at: https://www.canada.ca/en/public-health/services/diseases/shigella/health-professionals.html

Gao, B. and Gupta, R.S., 2007. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. BMC Genomics, 8(1), pp.1-23.

Gao, B. and Gupta, R.S., 2012. Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. Microbiology and molecular biology reviews, 76(1), pp.66-112.

Gao, B., Mohan, R. and Gupta, R.S., 2009. Phylogenomics and protein signatures elucidating the evolutionary relationships among the Gammaproteobacteria. International journal of systematic and evolutionary microbiology, 59(2), pp.234-247.

Gensberger, E.T., Polt, M., Konrad-Köszler, M., Kinner, P., Sessitsch, A. and Kostić, T., 2014. Evaluation of quantitative PCR combined with PMA treatment for molecular assessment of microbial water quality. Water research, 67, pp.367-376.

Griffith, J.F., Weisberg, S.B., Arnold, B.F., Cao, Y., Schiff, K.C. and Colford Jr, J.M., 2016. Epidemiologic evaluation of multiple alternate microbial water quality monitoring indicators at three California beaches. Water research, 94, pp.371-381.

Gupta, R. S., & Griffiths, E. (2006). Chlamydiae-specific proteins and indels: novel tools for studies. Trends in microbiology, 14(12), 527-535.

Gupta, R.S. and Mok, A., 2007. Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. BMC Microbiology, 7, pp.1-20.

Gupta, Radhey S., and Divya W. Mathews. “Signature proteins for the major clades of Cyanobacteria.” BMC evolutionary biology 10, no. 1 (2010): 1-20.

Hakovirta, J.R., Prezioso, S., Hodge, D., Pillai, S.P. and Weigel, L.M., 2016. Identification and analysis of informative single nucleotide polymorphisms in 16S rRNA gene sequences of the Bacillus cereus group. Journal of clinical microbiology, 54(11), pp.2749-2756.

Haugland, R., Oshima, K., Sivaganesan, M., Dufour, A., Varma, M., Siefring, S., Nappier, S., Schnitker, B. and Briggs, S., 2021. Large-scale comparison of *E. coli* levels determined by culture and a qPCR method (EPA Draft Method C) in Michigan towards the implementation of rapid, multi-site beach testing. Journal of microbiological methods, 184, p.106186.

Health Canada (2023). Canadian recreational water quality guidelines-Indicators of fecal contamination: Overview Available at: https://www.canada.ca/en/health-canada/services/publications/healthy-living/recreational-water-quality-guidelines-indicators-fecal-contamination.html

Heijnen, L. and Medema, G., 2009. Method for rapid detection of viable *Escherichia coli* in water using real-time NASBA. Water Research, 43(12), pp.3124-3132.

Ibekwe, M.A., Murinda, S.E., Park, S., Obayiuwana, A., Murry, M.A., Schwartz, G. and Lundquist, T., 2020. Comparative use of quantitative PCR (qPCR), droplet digital PCR (ddPCR), and recombinase polymerase amplification (RPA) in the detection of Shiga toxin-producing *E. coli* (STEC) in environmental samples. Water, 12(12), p.3507.

Karmali, M.A., Mascarenhas, M., Shen, S., Ziebell, K., Johnson, S., Reid-Smith, R., Isaac-Renton, J., Clark, C., Rahn, K., and Kaper, J.B. (2003). Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. Journal of Clinical Microbiology, 41, 4930-4940

Kembel, S.W., Wu, M., Eisen, J.A. and Green, J.L., 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PloS computational biology, 8(10), p.e1002743.

Kibbee, R.J. and Örmeci, B., 2017. Development of a sensitive and false-positive free PMA-qPCR viability assay to quantify VBNC *Escherichia coli* and evaluate disinfection performance in wastewater effluent. Journal of microbiological methods, 132, pp.139-147.

Korajkic, A., McMinn, B.R. and Harwood, V.J., 2018. Relationships between microbial indicators and pathogens in recreational water settings. International Journal of Environmental Research and Public Health, 15(12), p.2842.

Kostyla, C., Bain, R., Cronk, R. and Bartram, J., 2015. Seasonal variation of fecal contamination in drinking water sources in developing countries: a systematic review. Science of the Total Environment, 514, pp.333-343.

Li, Y., Huang, T.Y., Ye, C., Chen, L., Liang, Y., Wang, K. and Liu, J., 2020. Formation and control of the viable but non-culturable state of foodborne pathogen *Escherichia coli* O157: H7. Frontiers in Microbiology, 11, p.1202.

Liu, Y., Kumblathan, T., Uppal, G.K., Zhou, A., Moe, B., Hrudey, S.E. and Li, X.F., 2020. A hidden risk: survival and resuscitation of *Escherichia coli* O157: H7 in the viable but nonculturable state after boiling or microwaving. Water Research, 183, p.116102.

McIlroy, S.J., Tillett, D., Petrovski, S. and Seviour, R.J., 2011. Non-target sites with single nucleotide insertions or deletions are frequently found in 16S rRNA sequences and can lead to false positives in fluorescence in situ hybridization (FISH). Environmental microbiology, 13(1), pp.33-47.

McLain, J.E. and Williams, C.F., 2008. Seasonal variation in accurate identification of *Escherichia coli* within a constructed wetland receiving tertiary-treated municipal effluent. Water research, 42(15), pp.4041-4048.

McLain, J.E., Rock, C.M., Lohse, K. and Walworth, J., 2011. False-positive identification of *Escherichia coli* in treated municipal wastewater and wastewater-irrigated soils. Canadian journal of microbiology, 57(10), pp.775-784.

Meier-Kolthoff, J.P., Hahnke, R.L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L.A. and Chertkov, O., 2014. Complete genome sequence of DSM 30083 T, the type strain (U5/41 T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Standards in genomic sciences, 9, pp.1-19.

Naushad, H.S., Lee, B. and Gupta, R.S., 2014. Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera Dickeya, Pectobacterium and Brenneria. International journal of systematic and evolutionary microbiology, 64(Pt_2), pp.366-383.

Packeiser, H., Lim, C., Balagurunathan, B., Wu, J. and Zhao, H., 2013. An extremely simple and effective colony PCR procedure for bacteria, yeasts, and microalgae. Applied biochemistry and biotechnology, 169, pp.695-700.

Palomino-kobayashi, L.A., Pons, M.J. and Ruiz, J., 2022. Estimation of inherent bacterial DNA contamination in a qPCR master mix: concerns about background DNA of reagents. Minerva Biotechnology & Biomolecular, 34(4).

Ricchi, M., Bertasio, C., Boniotti, M.B., Vicari, N., Russo, S., Tilola, M., Bellotti, M.A. and Bertasi, B., 2017. Comparison among the quantification of bacterial pathogens by qPCR, dPCR, and cultural methods. Frontiers in microbiology, 8, p.1174.

Riley, L.W., Remis, R.S., Helgerson, S.D., Mcgee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A., and Cohen, M.L. (1983). Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. The New England Journal of Medicine, 308, 681-685

Saleem, F., Edge, T.A. and Schellhorn, H.E., 2022. Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring. Journal of Great Lakes Research, 48(3), pp.707-716.

Saleem, F., Li, E., Edge, T.A., Tran, K.L. and Schellhorn, H.E., 2024. Identification of Potential Microbial Risk Factors Associated with Fecal Indicator Exceedances at Recreational Beaches. Environmental Microbiome.

Saleem, F., Schellhorn, H.E., Simhon, A. and Edge, T.A., 2023. Same-day *Enterococcus* qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. Canadian Journal of Public Health, pp.1-12.

Shrestha, A. and Dorevitch, S., 2019. Evaluation of rapid qPCR method for quantification of *E. coli* at non-point source impacted Lake Michigan beaches. Water Research, 156, pp.395-403.

Sims, G.E. and Kim, S.H., 2011. Whole-genome phylogeny of *Escherichia coli*/Shigella group by feature frequency profiles (FFPs). Proceedings of the National Academy of Sciences, 108(20), pp.8329-8334.

Sun, F., Wang, Y., Wang, C., Zhang, L., Tu, K. and Zheng, Z., 2019. Insights into the intestinal microbiota of several aquatic organisms and association with the surrounding Environment. Aquaculture, 507, pp.196-202.

Thorsen, J., Brejnrod, A., Mortensen, M., Rasmussen, M.A., Stokholm, J., Al-Soud, W.A., Sørensen, S., Bisgaard, H. and Waage, J., 2016. Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. Microbiome, 4, pp.1-14.

USEPA (2012). Recreational water quality criteria. (pp. 42): United States (U.S.) Environmental Protection Agency (EPA). EPA-820-F-12-058. Available at: https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf

USEPA. (2015). Method 1609.1: Enterococci in water by TaqMan® quantitative polymerase chain reaction (qPCR) with internal amplification control (IAC) assay. EPA-820-R-15–099. Available at: https://www.epa.gov/sites/default/files/2015-08/documents/method_1609-1-*Enterococcus*-iac_2015_3.pdf.

Větrovský, T. and Baldrian, P., 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PloS one, 8(2), p.e57923.

Walker, D.I., McQuillan, J., Taiwo, M., Parks, R., Stenton, C.A., Morgan, H., Mowlem, M.C. and Lees, D.N., 2017. A highly specific *Escherichia coli* qPCR and its comparison with existing methods for environmental waters. Water Research, 126, pp.101-110.

Wong, S. Y., Paschos, A., Gupta, R. S., & Schellhorn, H. E. (2014). Insertion/deletion-based approach for the detection of *Escherichia coli* O157: H7 in freshwater environments. Environmental science & technology, 48(19), 11462-11470.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC bioinformatics, 13, pp.1-11.

Zhang, Y., Hong, P.Y., LeChevallier, M.W. and Liu, W.T., 2015. Phenotypic and phylogenetic identification of coliform bacteria obtained using 12 coliform methods approved by the U.S. Environmental Protection Agency. Applied and environmental microbiology, 81(17), pp.6012-6023.

Harwood, V.J., Staley, C., Badgley, B.D., Borges, K. and Korajkic, A., 2014. Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. FEMS microbiology reviews, 38(1), pp.1-40.

Klymus, K.E., Merkes, C.M., Allison, M.J., Goldberg, C.S., Helbing, C.C., Hunter, M.E., Jackson, C.A., Lance, R.F., Mangan, A.M., Monroe, E.M. and Piaggio, A.J., 2020. Reporting the limits of detection and quantification for environmental DNA assays. Environmental DNA, 2(3), pp.271-282.

Lim, J., Shin, S.G., Lee, S. and Hwang, S., 2011. Design and use of group-specific primers and probes for real-time quantitative PCR. Frontiers of Environmental Science & Engineering in China, 5, pp.28-39.

Pommepuy, M., Butin, M., Derrien, A., Gourmelon, M., Colwell, R.R. and Cormier, M., 1996. Retention of enteropathogenicity by viable but nonculturable *Escherichia coli* exposed to seawater and sunlight. Applied and environmental microbiology, 62(12), pp.4621-4626.

Servais, P., Prats, J., Passerat, J. and Garcia-Armisen, T., 2009. Abundance of culturable versus viable *Escherichia coli* in freshwater. Canadian Journal of Microbiology, 55(7), pp.905-909.

Feng, P.C. and Hartman, P.A., 1982. Fluorogenic assays for immediate confirmation of *Escherichia coli*. Applied and environmental microbiology, 43(6), pp.1320-1329.

Frampton, E.W. and Restaino, L., 1993. Methods for *Escherichia coli* identification in food, water and clinical samples based on beta-glucuronidase detection.

Ciebin, B.W., Brodsky, M.H., Eddington, R., Horsnell, G., Choney, A., Palmateer, G., Ley, A., Joshi, R. and Shears, G., 1995. Comparative evaluation of modified m-FC and m-TEC

media for membrane filter enumeration of *Escherichia coli* in water. Applied and environmental microbiology, 61(11), pp.3940-3942.

Maheux, A.F., Huppé, V., Boissinot, M., Picard, F.J., Bissonnette, L., Bernier, J.L.T. and Bergeron, M.G., 2008. Analytical limits of four β-glucuronidase and β-galactosidase-based commercial culture methods used to detect *Escherichia coli* and total coliforms. Journal of microbiological methods, 75(3), pp.506-514.

Fricker, C.R., Warden, P.S. and Eldred, B.J., 2010. Understanding the cause of false negative β-d-glucuronidase reactions in culture media containing fermentable carbohydrate. Letters in applied microbiology, 50(6), pp.547-551.

Lane, M.J., Rediske, R.R., McNair, J.N., Briggs, S., Rhodes, G., Dreelin, E., Sivy, T., Flood, M., Scull, B., Szlag, D. and Southwell, B., 2020. A comparison of *E. coli* concentration estimates quantified by the EPA and a Michigan laboratory network using EPA Draft Method C. Journal of Microbiological Methods, 179, p.106086.

Chern, E.C., Siefring, S., Paar, J., Doolittle, M. and Haugland, R.A., 2011. Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes. Letters in applied microbiology, 52(3), pp.298-306.

Gedalanga, P.B. and Olson, B.H., 2009. Development of a quantitative PCR method to differentiate between viable and non-viable bacteria in environmental water samples. Applied microbiology and biotechnology, 82, pp.587-596.

Maheux, A.F., Picard, F.J., Boissinot, M., Bissonnette, L., Paradis, S. and Bergeron, M.G., 2009. Analytical comparison of nine PCR primer sets designed to detect the presence of *Escherichia coli*/Shigella in water samples. Water research, 43(12), pp.3019-3028.

Chapter 6. Identification Of Potential Microbial Risk Factors Associated with Fecal Indicator Exceedances at Recreational Beaches

This chapter provides a comprehensive evaluation of recreational water microbiome changes associated with water quality and characterizes the potential microbial risk factors that are independent of changes in conventional bacterial fecal markers. The comprehensive view presented in this chapter provides a robust framework for developing targeted and effective beach monitoring strategies, which is a central theme of my research.

*The text I present here is a **peer-reviewed manuscript published** in the Environmental Microbiome journal. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Li, E., Edge, T.A., Tran, K.L. and Schellhorn, H.E., 2024. Identification of potential microbial risk factors associated with fecal indicator exceedances at recreational beaches. Environmental Microbiome, 19(1), p.4.*

Author Contributions

FS performed the experiments, analyzed the data, wrote the original manuscript, and revised it as required by the reviewers. EL and KT helped with the experiments. HES and TAE supervised the study and reviewed the manuscript before submission to the journal.

Abstract:

Background

Fecal bacterial densities are proxy indicators of beach water quality, and beach posting decisions are made based on Beach Action Value (BAV) exceedances for a beach. However, these traditional beach monitoring methods do not reflect the full extent of microbial water quality changes associated with BAV exceedances at recreational beaches (including harmful cyanobacteria). This proof of concept study evaluates the potential of metagenomics for comprehensively assessing bacterial community changes associated with BAV exceedances compared to non-exceedances for two urban beaches and their adjacent river water sources.

Results:

Compared to non-exceedance samples, BAV exceedance samples exhibited higher alpha diversity (diversity within the sample) that could be further differentiated into separate clusters (Beta-diversity). For Beach A, Cyanobacterial sequences (resolved as *Microcystis* and *Pseudanabaena* at genus level) were significantly more abundant in BAV non-exceedance samples. qPCR validation supported the Cyanobacterial abundance results from metagenomic analysis and also identified saxitoxin genes in 50% of the non-exceedance samples. *Microcystis* sp and saxitoxin gene sequences were more abundant on non-exceedance beach days (when fecal indicator data indicated the beach should be open for water recreational purposes). For BAV exceedance days, Fibrobacteres, *Pseudomonas*, *Acinetobacter*, and *Clostridium* sequences were significantly more abundant (and positively

correlated with fecal indicator densities) for Beach A. For Beach B, Spirochaetes (resolved as *Leptospira* on genus level) *Burkholderia* and *Vibrio* sequences were significantly more abundant in BAV exceedance samples. Similar bacterial diversity and abundance trends were observed for river water sources compared to their associated beaches. Antibiotic Resistance Genes (ARGs) were also consistently detected at both beaches. However, we did not observe a significant difference or correlation in ARGs abundance between BAV exceedance and non-exceedance samples for ARGs.

Conclusion

This study provides a more comprehensive analysis of bacterial community changes associated with BAV exceedances for recreational freshwater beaches. While there were increases in bacterial diversity and some taxa of potential human health concern associated with increased fecal indicator densities and BAV exceedances (e.g. *Pseudomonas*), metagenomics analyses also identified other taxa of potential human health concern (e.g. *Microcystis*) associated with lower fecal indicator densities and BAV non-exceedances days. This study can help develop more targeted beach monitoring strategies and beach-specific risk management approaches.

Keywords: Metagenomics, Freshwater Beaches, Recreational Beaches, Microbiome, Beach Action Values, Fecal Indicators.

Introduction:

Recreational water ecosystems, such as freshwater beaches, are subject to fecal contamination, resulting in beach postings deeming beaches unsuitable for public recreational activities [1,2]. Deterioration of water quality can cause gastrointestinal illnesses among beachgoers and may be caused by fecal contamination sources, including wastewater treatment plants, septic tank systems, combined sewer overflows and animal/bird feces [3,4]. Fecal indicator bacteria, including *E. coli* and *Enterococcus,* are correlated with gastrointestinal illness and are thus proxy indicators of water quality for recreational water ecosystems [2,5]. Despite their common use, fecal indicator bacteria (FIB) are limited by a lack of host specificity and sensitivity [6,7] and may not necessarily be correlated with the broad spectrum of enteric and non-enteric pathogens of health concern [8]. Additionally, FIB density can be biased due to environmental and physicochemical factors, including water temperature [9], the persistence of FIB for an extended period of time outside the host cell [10], and potential to regrow in beach sediments [11].

Traditional methods for fecal indicator testing in recreational water ecosystems, such as beaches, include culture-based enumeration [12]. *E. coli* and *Enterococcus* are commonly used fecal indicators for beach quality monitoring, and beaches are posted/closed for public visits based on exceedance above Beach Action Value (BAV) thresholds recommended by the United States Environmental Protection Agency [13] and Health Canada [2]. These Beach Action Values are determined based on epidemiological studies (mainly from US Beaches) on the correlation between gastrointestinal diseases among beachgoers/swimmers

and fecal indicator densities [1,14,15]. However, the occurrence and abundance of fecal indicator bacteria can be variable from different fecal pollution sources [16,17,18], and may not be indicative of many pathogens of human health concern from non-fecal sources (e.g. toxigenic cyanobacteria).

Molecular methods such as quantitative PCR and digital PCR have been tested as alternatives to augment culturing-based methods for monitoring beach water quality [19,20]. However, these PCR methods typically rely on a limited number of microbial markers, such as *Enterococcus* for FIB monitoring and HF183 for human fecal source tracking, which may not provide a comprehensive overview of the microbial communities in recreational waters. Additionally, obtaining broad microbial information from PCR-based methods requires multiple assays, which can be laborious and costly for complex microbial environments such as water ecosystems. In comparison, a DNA sequencing-based metagenomic approach can simultaneously characterize most of the taxa in a beach water sample and provide gene profiles of the identified organisms [21,22]. Therefore, a broader range of information obtained from the metagenomics-based approach can potentially provide a valuable screen of microbial changes and improve beach monitoring and management strategies.

The current study evaluates the potential of a metagenomics-based approach to augment beach monitoring strategies by obtaining a comprehensive overview of bacterial community changes associated with Beach Action Value exceedances for recreational freshwater beaches. The questions we focused on are: 1) Do the beach water microbial communities differ significantly between Beach Action Value Exceedance and Non-Exceedance Beach

Days? 2) Do the two different beaches show similar patterns of bacterial community changes in response to fecal indicator exceedances? 3) Can the adjacent river and creek water sources account for bacterial changes observed at the beaches? 4) Are there any bacterial taxa of human health concern that do not correlate with fecal indicator densities? 5) Do Antibiotic Resistance Genes (ARGs) abundance correlate with changes in fecal indicator densities?

Materials and Methods:

Study Design:

This study focused on microbial changes associated with fecal contamination at two Toronto recreational freshwater beaches, including Marie Curtis Park East Beach (referred to as Beach A) and Sunnyside Beach (referred to as Beach B), along with their adjacent river and creek water sources as assessed by shotgun DNA sequencing. The water sampling was performed three days a week for the summer season 2021 (June 1 to August 26). On each sampling day, eight samples were collected, including one each from Etobicoke Creek (referred to as River A and source water for MCPEB) and Humber River (referred to as River A and source water for Sunnyside Beach), and three from each beach transect 30W (43.585610 -79.540054), 30W replicate, and 32W (43.585110 -79.540560) for Marie Curtis Park East Beach, and transect 18W (43.636612 -79.452670), 21W (43.637110 -79.457530) and 21W replicate for Sunnyside Beach. These sampling sites and the names used (30W, 32W, 18W, and 21W) were selected according to Toronto Public Health's Beach Monitoring Program (Figure 1). A total of 309 samples corresponding to 38 beach days were collected. The samples were collected between 5:30 and 7 am and were transported (on ice) to the lab

within one hour for further processing.

Figure 1. Sampling sites for Marie Curtis Park East Beach, Etobicoke Creek, Sunnyside Beach and Humber River.

Water Sample Collection, Filtration and DNA Extraction:

Water samples were collected 30 cm below the water surface in sterile screw-capped polyethylene terephthalate (PET) bottles (1000 mL). For *E. coli* culturing and *Enterococcus* qPCR, a 100 mL sample was passed through a 0.45 µm mixed cellulose esters (MCE) filter (Millipore Corp., Bedford, MA), while for metagenomic DNA extraction, 100 mL of water sample was filtered through 0.22 µm nitrocellulose filters (Millipore Corp., Bedford, MA). For *Enterococcus* qPCR, we used the DNA extraction protocol as described previously [23]. Briefly, the membrane filters were placed in a 2 mL microcentrifuge tube with 0.3 gm glass beads (600 µL of 0.2 µg/mL Salmon sperm DNA as extraction buffer) and bead beaten for 60 s, followed by centrifuging for 60 s at 12, 000 g. The supernatant (400 µL) was collected

in a 1.5 mL tube and centrifuged for 5 min at 12, 000 g to collect DNA extract (~350 µL) for qPCR. For DNA Sequencing, DNA extraction was performed by using the Norgen Soil Plus DNA Extraction kit (Norgen Biotek Corp., Canada) as described previously [24], followed by DNA quantification using Qubit fluorometer (dsDNA High-Sensitivity Assay kit, Thermo Fisher Scientific, USA).

E. coli Enumeration by Culturing

E. coli enumeration was performed for all samples using Differential Coliform Agar (OxoidTM) with cefsulodin as described previously [23]. Beach water samples were directly processed (100 mL), while Creek/River samples were diluted 1:10 for *E. coli* culturing. The membrane filters were placed on 47mm agar plates and incubated for 24h at 44.5 °C [25]. The phosphate buffer saline (passed through the membrane filters) was used alongside the water sample as a negative control. Each sample was tested in triplicate, and enumeration counts for each sampling site were recorded as the mean.

Enterococcus qPCR

Enterococcus qPCR was performed as described previously [23,26] using Method 1609.1: *Enterococci in Water by TaqMan Quantitative Polymerase Chain Reaction (qPCR) with Internal Amplification Control (IAC) Assay* [27]. In brief, the stock cultures (10^9 CFUs) of *Enterococcus faecalis* (ATCC 2921) were used to prepare calibrator-positive controls (10^4 CFUs) and for DNA extraction to prepare Standard Curves. For standard curves, DNA was extracted from the stock cultures using the Norgen Soil Plus Extraction kit described in the previous section. The DNA quantification was done using a Qubit fluorometer (dsDNA

High-Sensitivity Assay kit, Thermo Fisher Scientific, USA), followed by the calculation of Target Sequence Copies. Ten-fold dilutions (10-40,000 Target Sequence Copies) were prepared from Stock Culture DNA, and four individual standard curves were used to create a composite standard curve. Alongside each batch of samples, two calibrator positive controls, two method blanks (phosphate buffer saline passed through filters), and two non-template controls (DNA extract replaced by nuclease-free water) were analyzed to test for DNA extraction efficiency and contamination. Each PCR reaction was carried out in duplicate and comprised DNA Recovery Control (Salmon DNA qPCR) and a PCR inhibition control (Internal Amplification Control). Each qPCR reaction (25 µL) comprised of 12.5 µL TaqMan Environmental Master mix (Thermo Fisher Scientific, USA) , 3.0 µL of primer-probe working solution (primers and probes concentration was 1.0 µM and 80.0 nm, respectively), 2.0 µL of internal amplification control and 2.5 µL Bovine Serum Albumin (2mg/mL). All the reactions were performed on Bio-Rad CFX96 Touch Real-Time PCR (Bio-Rad Inc. USA). The *Enterococcus* Calibrator Cell Equivalents were calculated using an Excel sheet (https://www.epa.gov/cwa-methods/other-clean-water-act-test-methods-microbiological#file-183743) provided by USEPA.

Sample Selection Criteria for Shotgun DNA Sequencing

The following considerations were used for sample selection: 1) samples having *E. coli* >235 CFU/mL by plate counting or having *Enterococcus* >1000 calibrator cell equivalents by qPCR as determined previously [23,26] were defined as BAV exceedances, 2) Sampling Days showing exceedance or non-exceedance for all the sampling sites for a beach were

prioritized, 3) an equal number of beach samples were selected for BAV Exceedance and Non-exceedance groups, 4) for a single Beach Day, DNA from all the sampling sites for a beach was pooled and, 5) For creek/river water sources (Etobicoke Creek and Humber River), sampling dates matching to the selected beach samples were sequenced for comparison of bacterial trends with associated beaches. In total, 48 pooled water samples were selected for shotgun sequencing (Supplementary Table 1).

Shotgun DNA Sequencing and Quality Control Analysis

To avoid DNA concentration bias during library pooling, input DNA from each sample was normalized to 200 ng. The library was prepared using NEBNext® Ultra™ II DNA library preparation kit with TruSeq3 paired-end adapters. The fragment size and read length were 500 bp and 150 bp, respectively. DNA sequencing was performed on Illumina NextSeq 2000 (2x150) at the Farncombe Sequencing Institute at McMaster University. The quality of raw reads was analyzed using FASTQC [28]. Adapter trimming, decontamination (removal of reads mapping to human), quality filtration (Quality Score >30, by the sliding window algorithm, windowsize = 4 bases), Length Filtration (> 75 bp) and removal of tandem repeats were performed using the KneadData pipeline (Available at: http://huttenhower.sph.harvard.edu/kneaddata).

Cyanobacteria and Cyanotoxin Gene qPCR:

The number of Cyanobacterial and Cyanotoxin (Microcystin, Saxitoxin and Cylindrospermopsin) gene copies from metagenomics sequencing were validated using CyanoDTec Total Cyanobacteria and Toxin Kit (Phytoxigene^{TM}) according to the

manufacturer's instructions. In total, 17 (8 for non-exceedance and 9 for exceedance beach days) pooled DNA samples from Marie Curtis Park East Beach were used for Cyanobacteria/toxin qPCR. Four standard curves were run separately for each assay (Total Cyanobacteria and Toxin), followed by preparing a composite standard curve. The standard curve range for each assay was 10- 100,000 gene copies. Each reaction comprised 20 µL of mastermix/primer-probe solution and 5 µL of DNA from pooled samples (the same DNA used for shotgun sequencing). The gene copy numbers for each sample were calculated using the Slope-Intercept equation from the composite standard curve and normalized to gene copies per nanogram of DNA. A gene was only considered present in a sample if the gene copies/Threshold-cycle values were within the range of the standard curve, and the results were only accepted if the internal amplification control threshold-cycle (Ct) value for a sample was not offset more than 1.5 compared to non-template control.

Bioinformatics and Data Analysis

Clean reads were aligned against NCBI RefSeq protein Database (Accessed on February 24, 2023) using DIAMOND BLASTx on sensitive mode [29], followed by annotation using MEGAN6 (Weighted Lowest Common Ancestor (LCA) method: Minimum-Score = 50, Top-percent filter = 10%, Minimum-Support = 50) [30,31]. For lateral comparison, annotated reads were rarefied (normalized) to the sample of the smallest size (~4.8 million reads) to neutralize bias associated with sequencing depth [31,32,33]. For the Core Microbiome Analysis, bacterial genera present in ≥50% of the samples with a relative abundance of ≥0.1% were selected [34,35]. For the Alpha diversity analysis, Shannon-

Weaver and Simpson's Reciprocal diversity indices measurements were calculated [36,37]. Two alpha diversity matrices were used for cross-validation of in-sample diversity. For Antibiotic resistance genes (ARGS) analysis, quality-filtered sequences were assembled into contigs using the MEGAHIT (Metagenome Assembler: k_min+1 = 2, Min= kmer Size = 21, Max kmer Size = 99, k-step = 20, and minimum-contig size = 200) [38], followed by ARGs annotation using Pathofact pipeline (Combines DeepARG and Resistance gene identifier results for cross-validation) [39]. Statistical analysis was performed using STAMP metagenomic data statistical analysis software [40]. Analysis of Similarity (ANOSIM) was used for the Beta-Diversity assessment [24]. Shapiro-Wilk's normality testing (Stats v3.6.2 R package) was used to determine the normal distribution of the tested microbial variables, followed by either Welch's (and one-way ANOVA) or Wilcoxon-Mann-Whitney t-test for comparison between the groups [41] and Spearman's rank test (Log-transformed data) for correlation analysis between the variables [42,43].

Results:

Quality Control Analytics:

A total of 48 pooled DNA samples from recreational beaches and associated creek/river water were sequenced (Supplementary Table 1). Next-generation DNA sequencing provided 12.9 ± 3.2 and 13.7 ± 3.5 million raw reads for Beach A and B, respectively (Supplementary Table 2). A high proportion (83%) of raw reads passed the quality control criteria and were processed for downstream taxonomic/functional analysis. The R^2 value for the qPCR composite standard curves (*Enterococcus*, Total Cyanobacteria and Toxin Assays) was

within 0.992-0.999, while slope and intercept values were between -3.25 to -3.46 and 38.66 to 39.13, respectively (Supplementary Table 2). Supplementary Figure 2 shows the taxonomic identification after annotation/Least Common Ancestor (LCA) analysis. Bacterial sequences dominated all the samples (93-96%), followed by eukaryotes (3-4%), viruses (1-2%) and archaea (<1%). The results below highlight bacterial diversity/composition changes associated with the Beach Action Value exceedance and non-exceedance of fecal indicator bacteria.

Diversity Between Beach Action Value (BAV) Exceedance and Non-Exceedance Beach Days

Alpha diversity (diversity within the samples) was approximately 10% higher in Beach Action Value (BAV) Exceedance Day samples than the non-exceedance samples for both Beach A (4.4 ± 0.5 versus 4.0 ± 0.3) and Beach B (4.1 ± 0.1 versus 3.7 ± 0.7) (Table 1). Interestingly, exceedance samples from Beach A showed a ~9% higher alpha diversity than associated river source. Additionally, when concatenated at the genus level, exceedance samples were segregated into separate clusters from non-exceedance samples for both beaches on principal component analysis plots (Figure 2 and Supplementary Figure 3). Analysis of core microbiome differences identified 9 (24%) and 3 (10%) bacterial genera exclusive to BAV Exceedance beach days from Beach A and Beach B, respectively (Supplementary Figure 4).

Table 1. Alpha Diversity Analysis for beach water samples from Beach A, Beach B and Associated River Sources.

Microbial Diversity Within the Samples (Alpha Diversity)	**Beach Action Value Exceedance Day (Mean ± Standard Deviation)**	**Beach Action Value Non-Exceedance Day (Mean ± Standard Deviation)**	**Sampling Site**
Shannon-Weaver	4.4±0.5	4.0±0.3	Beach A
Diversity Matrix	4.1±0.1	3.7±0.7	Beach B
Simpson-Reciprocal	8.8±1.9	8.4±1.8	Beach A
Diversity Matrix	8.3±1	7.9±3.7	Beach B
	Creek/River Samples (Mean ± Standard Deviation)		
Shannon-Weaver	4.0±0.8		River A
Diversity Matrix	4.1±0.6		River B
Simpson-Reciprocal	7.8±3.4		River A
Diversity Matrix	7.5±3.2		River B

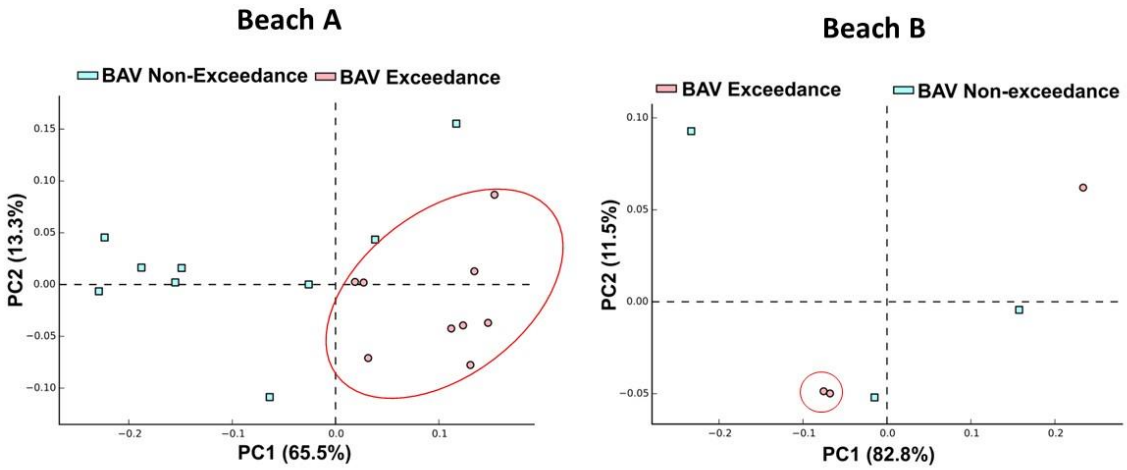


Figure 2. Differential abundance Principal Component Analysis plots for Beach Action Value (BAV) Exceedance and Non-Exceedance samples from Beach A and Beach B.

Bacterial Community Changes Associated with Fecal Indicator Exceedances/Non-Exceedances

Bacterial community composition on the phylum level was similar for both beaches (Figure 3). Proteobacteria (38-75%) were the most abundant, followed by Bacteroidetes (20-40%), Actinobacteria (10-30%), Verrucomicrobia (5-7%), Cyanobacteria (2-4%) and Firmicutes (1-2%). For Beach A, at the phylum level, Proteobacteria ($p = 1.04e\text{-}5$) and Fibrobacteres ($p = 0.049$) were significantly more abundant on Beach Action Value Exceedance days, while Actinobacteria ($p = 7.64e\text{-}4$) and Cyanobacteria ($p = 0.021$) were significantly more abundant on non-exceedance beach days (Supplementary Figure 5). Compared to

corresponding sampling dates from beach and river, Cyanobacterial sequences were more abundant for non-exceedance beach samples (Figure 4). Fibrobacteres sequences were more abundant for exceedance samples and showed a similar abundance pattern between river A and beach A samples (Supplementary Figure 6). Supplementary Figures 7 and 8 show that Cyanobacterial sequences classified into *Microcystis* and *Pseudanabaena* genera. Metagenomics analyses indicated that *Microcystis* sequences were absent in exceedance samples but detected in 6/9 = 67% of non-exceedance samples (and 100% of samples from late July through August). *Pseudanabaena* sequences were detected in both exceedance and non-exceedance samples. However, for *Pseudanabaena*, we did not observe any statistically significant ($p = 0.36$) difference between BAV exceedance and non-exceedance samples.

We further validated the Cyanobacterial findings using 16S rRNA (Cyanobacterial-specific) qPCR (Figure 5). Cyanobacterial-specific qPCR results confirmed the metagenomic analysis. qPCR results validated that BAV-non-exceedance beach days had significantly higher ($p = 9.99e\text{-}4$) cyanobacterial gene copies compared to BAV exceedance days for Beach A. Although qPCR did not detect Cyanobacteria-associated microcystin and cylindrospermopsin genes (Table 2), 50% (4/8) of the qPCR tested BAV non-exceedance samples (corresponding to samples with higher Cyanobacterial gene copies) showed the presence of saxitoxin genes.

For Beach B, at the phylum level, only Spirochaetes sequences were significantly ($p = 0.016$) more abundant in BAV exceedance samples than non-exceedance samples, and corresponding sampling dates from River B also showed higher mean abundance for

sampling dates corresponding to BAV exceedance beach samples (Supplementary Figure 9). At the genus level, Spirochaetes sequences were mainly classified into *Leptospira* and showed comparatively higher mean abundance in BAV exceedance samples (Supplementary Figure 10), though the difference was not significant ($p = 0.25$).

For Beach A, at the genus level, abundances of *Pseudomonas* ($p = 9.41e\text{-}3$), *Acinetobacter* ($p = 5.09e\text{-}3$), and *Clostridium* ($p = 0.038$) were significantly higher for BAV exceedance days than non-exceedance day samples (Figure 6). Compared with non-exceedance beach samples, the abundances of *Pseudomonas* and *Acinetobacter* were higher in both Etobicoke Creek and Marie Curtis Park East Beach for the sampling dates corresponding to Beach Action Value Exceedances (Supplementary Figures 11 and 12). Table 3 shows the correlation between fecal indicator densities and differentially abundant genera for Beach A and River A. For the Beach A samples, the fecal indicator densities (*E. coli* by culture and *Enterococcus* by qPCR) showed a significant positive correlation with *Pseudomonas* ($r_s = 0.7$, $p < 0.01$), *Acinetobacter* ($r_s = 0.6$ and 0.5, $p \leq 0.01$), and *Clostridium* ($r_s = 0.5$, $p \leq 0.05$). River A samples also showed a significant positive correlation between fecal indicator densities with *Pseudomonas* ($r_s = 0.7$ and 0.8, $p < 0.01$) and *Acinetobacter* ($r_s = 0.5$, $p = 0.01$).

For Beach B, at the genus level, 14 genera, including *Burkholderia* ($p = 0.034$) and *Vibrio* ($p = 0.05$), were significantly more abundant on BAV exceedance days than non-exceedance day samples (Figure 7). The mean abundance of *Burkholderia* and *Vibrio* was also higher in River B on the sampling dates corresponding to BAV exceedance at Sunnyside Beach

(Supplementary Figures 13 and 14). For correlation with fecal indicator densities, only the *Burkholderia* counts showed a significant positive correlation with *E. coli* density for Beach A (Table 4).

Escherichia sequences were detected in more samples than *Enterococcus* (Supplementary Figures 15 and 16). *E. coli* sequences were detected in 90% of the samples for both beaches, while *Enterococcus* sequences were detected in only one sample for each beach. Additionally, the difference in mean proportions between BAV exceedance and non-exceedance samples for *Escherichia* was not significant ($p > 0.05$) for both beaches.

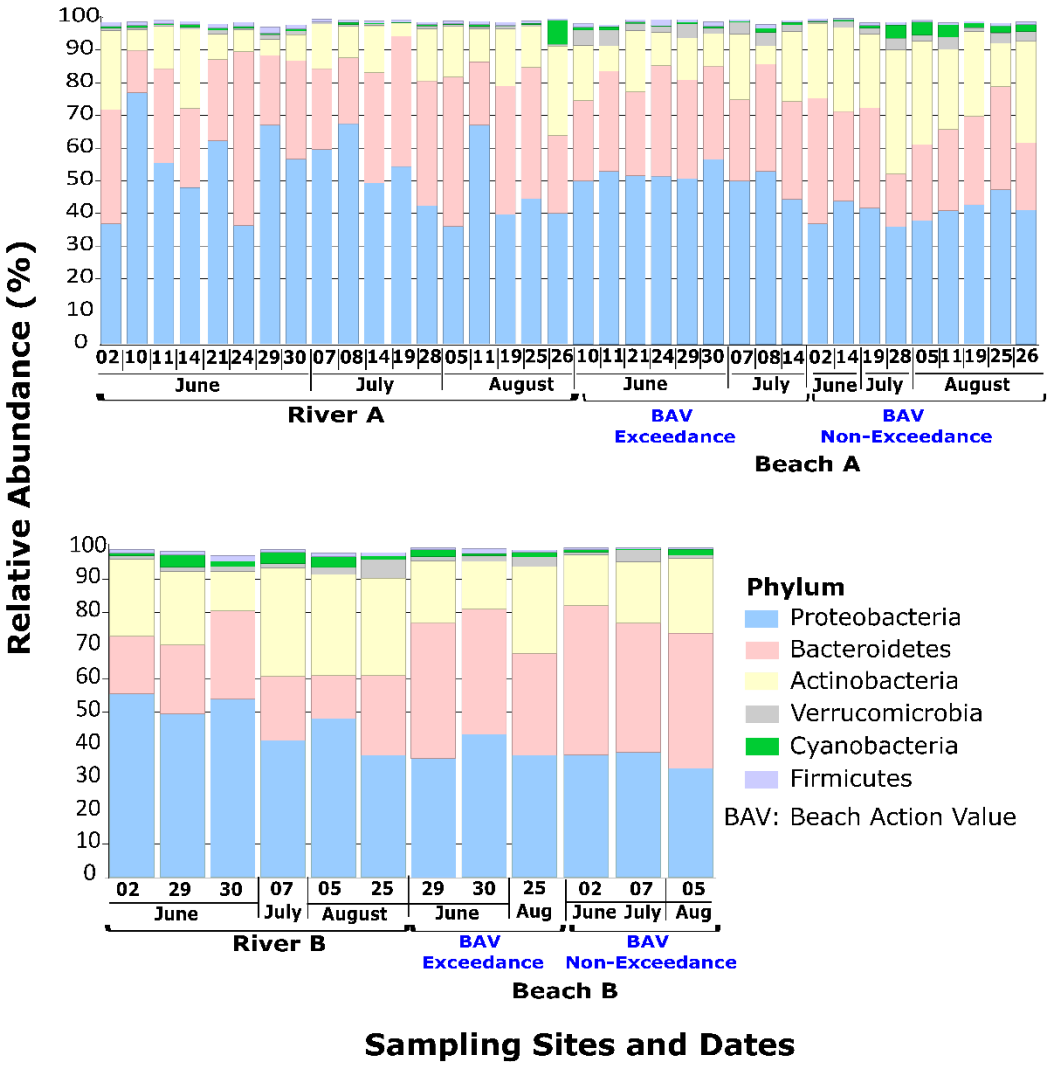


Figure 3. Relative abundance profile for six top most abundant Phyla identified in Beach Action Value Exceedance and Non-Exceedance Beach Day samples.

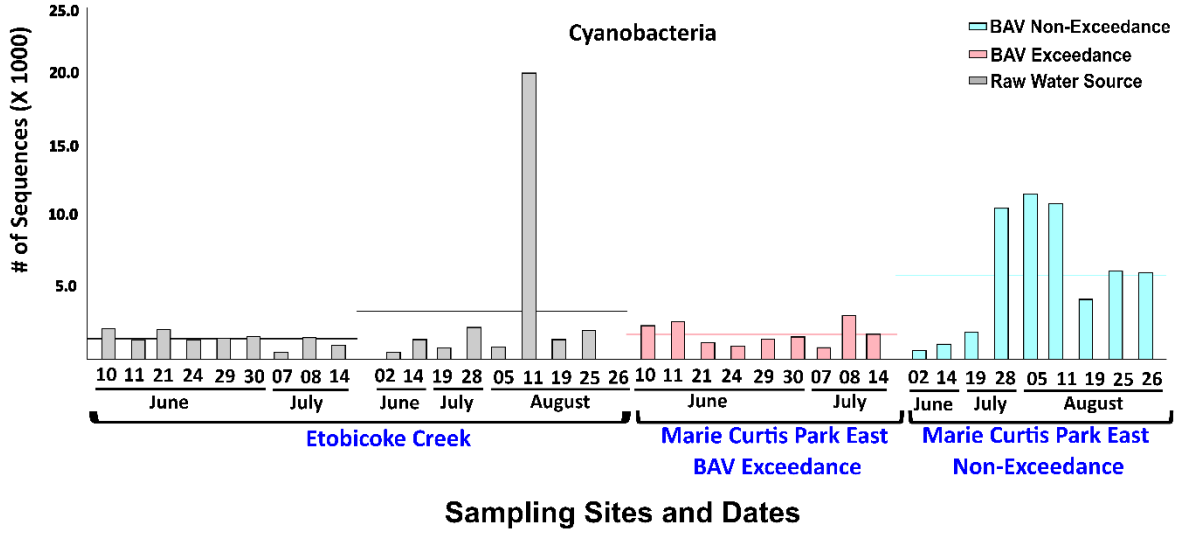


Figure 4. Bar plot of Cyanobacterial abundance for samples from Marie Curtis Park East Beach and Etobicoke Creek. The horizontal line represents the average number of normalized sequences for each group.

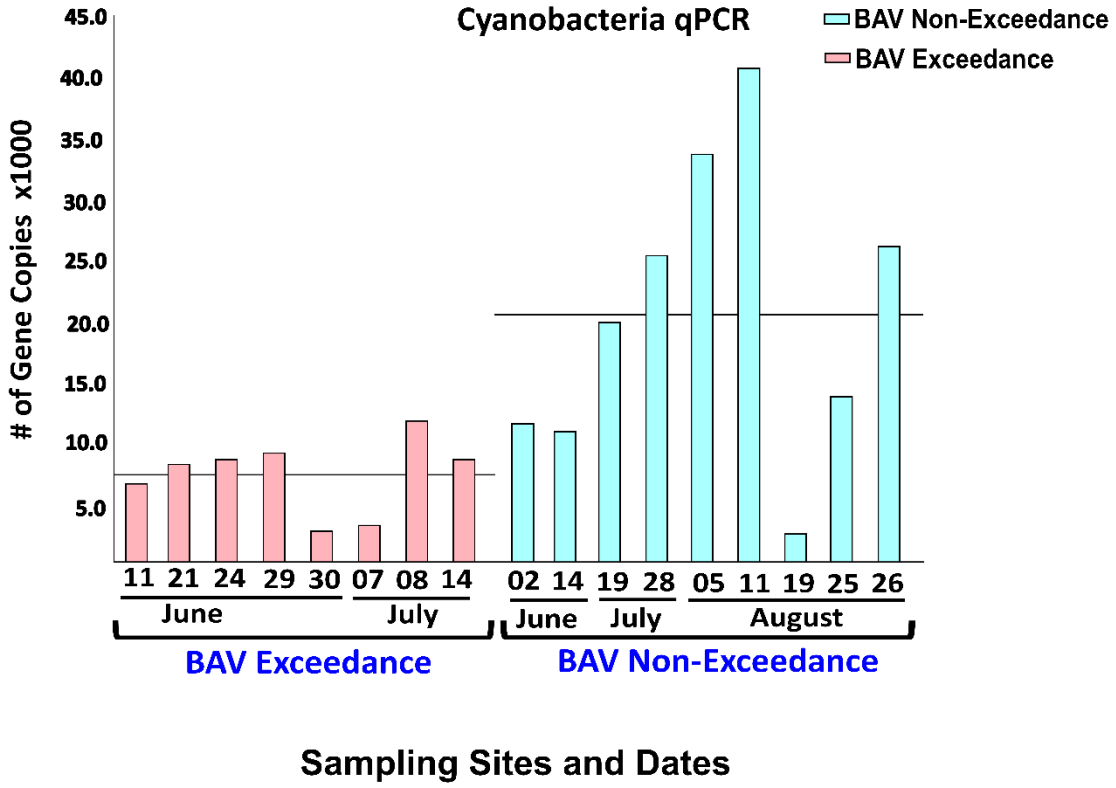


Figure 5. Bar plot of Microcystis abundance measured by qPCR for Beach A samples. The horizontal line represents the average number of gene copies for each group.

Table 2. Cyanotoxin presence/absence for BAV Exceedance and Non-exceedance Samples from Marie Curtis Park East Beach.

Date	**Beach Action Value Status**	**Microcystin/Nodularin**	**Cylindrospermopsin**	**Saxitoxin**	**Saxitoxin Gene Copies**
June 11		-	-	-	-
June 21		-	-	-	-
June 24		-	-	-	-
June 29		-	-	-	-
June 30	Exceedance	-	-	-	-
July 7		-	-	-	-
July 8		-	-	-	-
July 14		-	-	**+**	**2284**
June 2		-	-	-	-
June 14		-	-	-	-
July 19		-	-	**+**	**76**
July 28		-	-	**+**	**145**
August 5		-	-	**+**	**12079**

August 11	Non-Exceedance	-	-	+	**400**
August 19		-	-	-	-
August 25		-	-	-	-
August 26		-	-	-	-

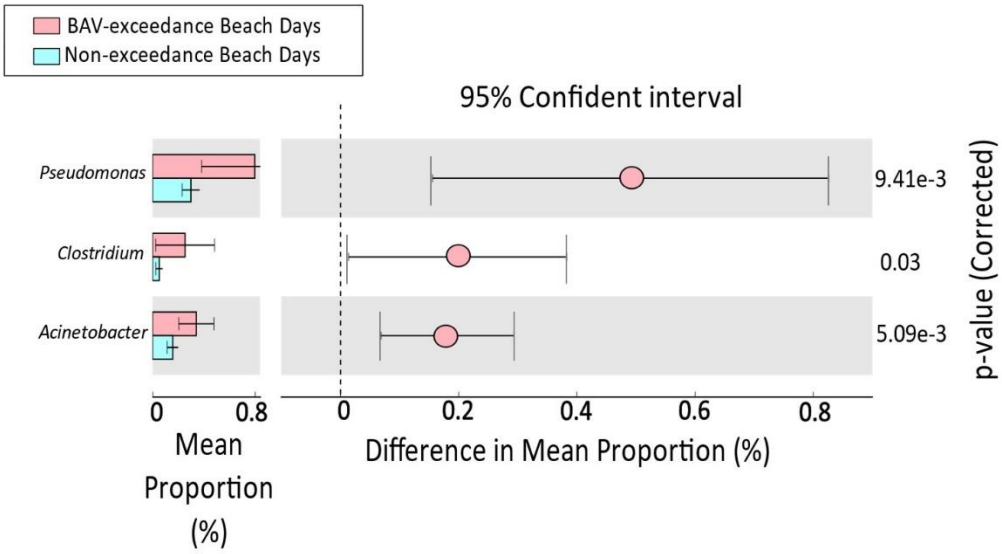


Figure 6. Differential abundance of statistically significant bacterial genera between Beach Action Value Exceedance and Non-Exceedance Beach Days from Beach A.

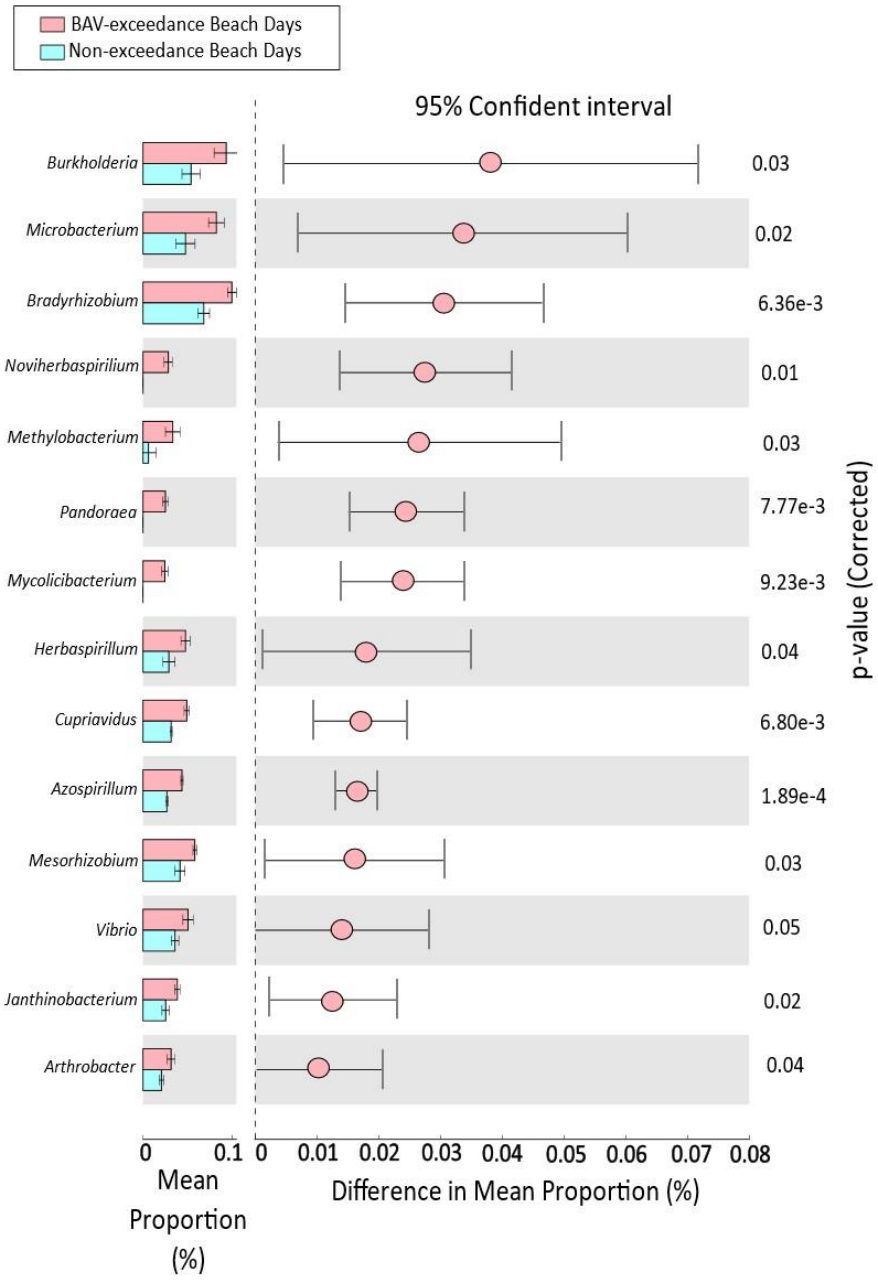


Figure 7. Differential abundance of statistically significant bacterial genera between Beach Action Value Exceedance and Non-Exceedance Beach Days from Beach B.

Table 3. Correlation analysis between fecal indicator densities and differentially abundant bacterial genera for Beach A and River A.

Fecal Indicator	Bacterial Genera	Correlation Coefficient (r_s)	P-value	Sampling Site
	Vibrio	0.1	0.5	
	Clostridium	0.5	0.03	
	Acinetobacter	0.6	0.002	
E. coli	*Pseudomonas*	0.7	5e-4	Beach A
	Vibrio	0.2	0.3	
	Clostridium	0.6	0.008	
	Acinetobacter	0.5	0.01	
Enterococcus	*Pseudomonas*	0.7	1e-4	Beach A
	Vibrio	0.3	0.1	
	Clostridium	0.3	0.1	
E. coli	*Acinetobacter*	0.5	0.01	
	Pseudomonas	0.7	5e-4	River A
	Vibrio	0.5	0.03	
	Clostridium	0.3	0.1	

	Acinetobacter	0.5	0.01	
Enterococcus	*Pseudomonas*	0.8	2.4e-5	River A

Table 4. Correlation analysis between fecal indicator densities and differentially abundant bacterial genera for Beach B and River B.

Fecal Indicator	**Bacterial Genera**	**Correlation Coefficient**	**P-value**	**Sampling Site**
	Leptospira	0.2	0.7	
	Burkholderia	0.8	0.05	
	Vibrio	0.7	0.1	
E. coli	*Clostridium*	0.4	0.3	Beach B
	Leptospira	0.3	0.4	
	Burkholderia	0.4	0.3	
	Vibrio	0.6	0.2	
Enterococcus	*Clostridium*	0.4	0.4	Beach B
	Leptospira	0.7	0.1	
	Burkholderia	0.5	0.2	
	Vibrio	0.4	0.3	
E. coli	*Clostridium*	0.8	0.03	River B
	Leptospira	0.5	0.2	
	Burkholderia	0.6	0.1	
	Vibrio	0.02	1	

Enterococcus	*Clostridium*	0.4	0.3	River B

Relationship Between Antibiotics Resistance Genes and Fecal Indicator Exceedances/Non-Exceedances

Antibiotic resistance analysis revealed that for both Marie Curtis Park East and Sunnyside Beaches, the five most abundant resistant gene groups were those involved in Beta-lactam antibiotic resistance, Multidrug resistance efflux pumps, Aminoglycoside resistance, Macrolides-Lincosamides resistance and Tetracycline resistance (Figure 8). However, the difference in mean proportions for the antibiotic-resistance genes was not statistically significant ($p > 0.05$) between the Beach Action Value Exceedance and Non-exceedance days for both beaches. Additionally, there was no significant correlation between fecal indicator densities and antibiotic resistance gene abundances (Supplementary Table 4).

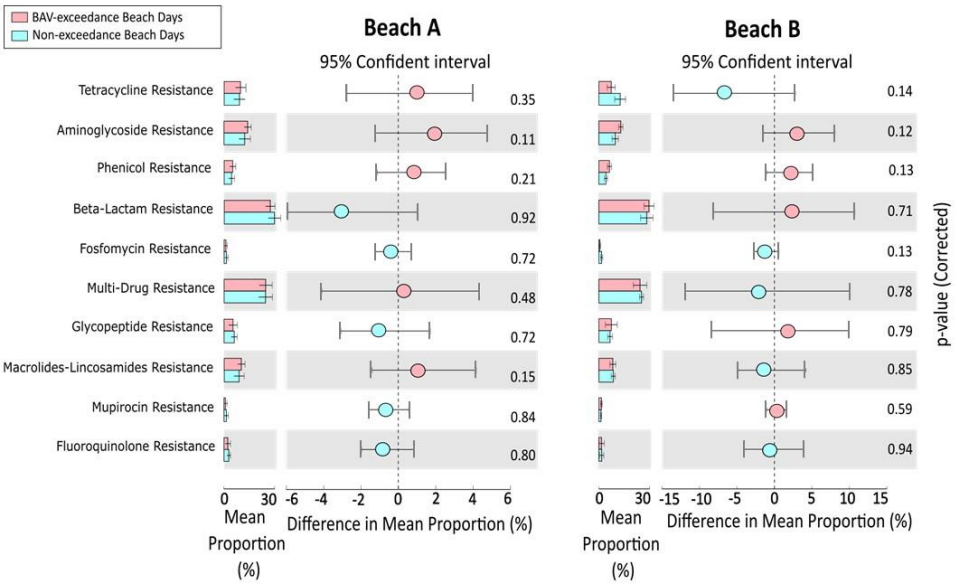


Figure 8. Differential Abundance of Antibiotic Resistance Genes in Beach Action Value Exceedance and Non-Exceedance Beach Days from Beach A and Beach B.

Discussion

Freshwater Beach monitoring programs use densities of fecal indicators, including *E. coli* and *Enterococcus,* as a reference for evaluating beach water quality. For each fecal indicator and analysis method (enumeration by culturing or quantitative PCR), a specific Beach Action Value (BAV) is recommended by the Public Health Authorities, and beaches are posted for recreational uses if the levels of fecal indicator densities are above the BAV [2,5]. However, limited experimental data is available for microbial community changes associated with Beach Action Value exceedances, and a single fecal indicator may not be indicative of diverse potential health risks from enteric and non-enteric pathogens, toxigenic cyanobacteria and aspects such as antimicrobial resistance (Ferguson et al., 2012; Li et al.,

2021). Compared to conventional analysis methods (culturing and PCR), a metagenomics-based approach provides a robust and comprehensive taxonomic and functional screening profile for water ecosystems. Additionally, metagenomic analysis can provide a foundation for targeted water quality monitoring by identifying region/site-specific microbial/functional differences [44,45]. This study aimed to provide a comprehensive profile of bacterial community changes associated with fecal indicator Beach Action Value Exceedances and Non-exceedances across two urban recreational freshwater beaches.

Alpha diversity for water samples from Great Lakes beaches has been found to range from 3.5 [46,47] to 7 [48], and our results for Marie Curtis Park East and Sunnyside Beaches were within this range. Beach Action Value Exceedance samples from both beaches showed comparatively higher alpha diversity than non-exceedance and associated creek/river water sources, indicating other bacterial groups may be present in exceedance day samples. Aside from creeks or rivers, sand/sediment resuspension and other fecal pollution sources can impact bacterial diversity in recreational waters, including birds and mammals defecating nearby around the beach ecosystem [3,49,50]. Beach Action Value Exceedance and non-exceedance samples from both beaches separated into independent clusters on principle component analysis plots, which indicates bacterial abundance and diversity differences. Bacterial genera that differed between Beach Action Value Exceedance days and Non-exceedance days differed between the two beaches, suggesting localized influences around each beach rather than regional processes drive microbial community changes. Differences in diversity and the core microbiome between BAV exceedance and non-exceedance beach

day samples could be due to diverse environmental factors, including rain events that can increase bacterial diversity on BAV exceedance beach days by increasing the bacterial load from urban runoff, creek/river plumes entering lakes, or increased flows dislodging soil and sediment-attached microbial communities [51,52].

At the phylum level, Proteobacteria, Bacteroidetes, Actinobacteria, Verrucomicrobia, and Firmicutes were abundant in all tested samples, consistent with other findings (surface water and sediment) for the Great Lakes region [24,53,54]. Interestingly, for Beach A, Cyanobacteria sequences were significantly more abundant in BAV non-exceedance samples than exceedance days or associated with the adjacent creek water source. Additionally, the increased abundance of Cyanobacteria and Saxitoxin genes was more notable in the later summer weeks (July-August 2021), corresponding to BAV non-exceedance beach days. Thus, we detected *Microcystis*, *Pseudanabaena*, and saxitoxin gene sequences on many non-exceedance beach days when *E. coli* data indicated Beach A should be open for water recreation. Both Beach A and B are not routinely monitored for cyanobacteria or harmful algal species and would typically be tested in response to visual complaints of bloom formations. Cyanobacterial genera, including *Microcystis,* can lead to deteriorating (eutrophication and toxin production) water quality for recreational purposes [55,56]. Similar to our findings, a study on recreational waters [57] identified decreased fecal indicator densities associated with higher Cyanobacterial (specifically *Microcystis*) levels. Therefore, relying solely on the fecal indicator densities for recreational water quality may provide an incomplete perception of human health risks at beaches. Fibrobacteres species

(cellulose-degrading bacteria) are specific to the rumen microbiome of ruminant animals [58], while Spirochaetes are found in farm animals (cows and pigs) but not human fecal material [59]. *Leptospira* contamination can be from domestic and wild animals [60], which may indicate that fecal contamination for Beach B can be from both water sources (Humber River) and localized (wild animals). The increased abundance of Fibrobacteres and Spirochaetes, along with increased *E. coli* levels, on BAV exceedance days at our two beaches may indicate fecal contamination from livestock or other ruminants in runoff to river sources that subsequently impacts the beaches.

On the genus level, *Pseudomonas, Acinetobacter* and *Clostridium* were significantly more abundant in BAV exceedance samples from Beach A (and River A samples for the same dates), while *Bulkholderia* and *Vibrio* were more abundant for BAV exceedance samples from Beach B (and River B samples for the same dates). Studies in the Great Lakes region [24,61] have identified *Pseudomonas, Clostridium* and *Acinetobacter* as common genera in stormwater, which may explain the influx of these genera into Beach A from the adjacent creek water source. Both River A and River B are significantly impacted by stormwater systems at times, contributing to increased fecal contamination at Marie Curtis Park East and Sunnyside Beaches [62,63]. *Burkholderia* and *Vibrio* sequences have been identified to be associated with both human and animal fecal contamination in urban recreational waters [64], which is in agreement with our findings for Beach B and may indicate fecal contamination from both associated waters (Humber River) and localized (wild animals) sources.

The metagenomics method also provided a screen for detecting a diverse range of antibiotic resistance genes. We detected numerous Antibiotic Resistance Genes (ARGs) of clinical concern in water samples from both Beaches. Similar to our beaches, other studies have also identified ARGs at beaches [65] and across ~350 lakes in Canada [66]. We found no significant association between numbers of antibiotic resistance genes (ARGs) and fecal indicators, with exceedances based on *E. coli* or *Enterococcus* numbers, this suggests the occurrence of AMR genes was not solely driven by AMR genes associated with these two FIBs. AMR genes are likely associated with diverse other bacteria, including those unrelated to fecal pollution, which is another limitation of using traditional culture (*E. coli*) or qPCR (*Enterococcus*) methods to predict overall AMR gene occurrence. However, ARG pools were consistently present at our two beaches, with the potential for horizontal gene transfer to bacterial species of human health concern. While water can play a role in routes of ARG exposure, quantifying that role and its associated human health risks requires further research [67].

Overall, this proof of concept study demonstrates the potential value of metagenomics for enabling a more comprehensive screen of bacterial community changes associated with fecal indicator Beach Action Value Exceedance and Non-exceedance conditions at freshwater beaches. The differences in bacterial diversity and abundance in response to BAV exceedances for Beach A were more pronounced (supported by correlation test) compared to Beach B. This may be due to Beach A sampling locations being closer to the mouth of River A than the proximity of Beach B sampling locations to the mouth of River B. In

addition, a break wall limits the direct influx from River B into Beach B, while Beach A is on an open coastline that receives unhindered water flow directly from River A under the right wind and current conditions. One limitation of this study is the localization of both tested beaches in the same geographical location or close proximity. However, we identified site-specific microbial differences between the two tested beaches, and future studies can build on our results/methodology to include a larger study area. Additionally, Our results provide insight into localized processes influencing bacterial community changes at freshwater beaches and further identify limitations of existing culture-based and single-gene PCR assay approaches for assessing recreational water quality. The results provide a foundation to guide more comprehensive screening for harmful microorganisms, as well as toxin and antimicrobial resistance genes, in order to improve recreational water quality monitoring and enable more targeted and site-specific risk management strategies.

Conclusion

1. Cyanobacterial sequences (*Microcystis* and saxitoxin genes in particular) were significantly more abundant in Beach Action Value Non-exceedance samples from Beach A, demonstrating that fecal indicator bacteria densities may not indicate health risks associated with harmful algal blooms and the eutrophication of recreational waters.
2. The increase of Fibrobacteres sequences in BAV exceedance days of Beach A may represent an influx of fecal contamination from livestock or other ruminant animals.
3. *Pseudomonas*, *Acinetobacter*, and *Clostridium* sequences were significantly more

abundant on the BAV exceedance days and positively correlated with fecal indicator densities at Beach A.

4. The increase of Spirochaetes (specifically *Leptospira*), *Burkholderia*, and *Vibrio* was significantly associated with Beach Action Value Exceedance samples from Beach B.
5. Similar bacterial diversity and abundance trends between beach and river samples suggest the Creek and River are primary sources of bacterial contamination at the beaches.
6. Pools of Antibiotic Resistance Genes (ARGs) were consistently detected at both beaches, indicating potential for transfer to potentially pathogenic genera by horizontal gene transfer.
7. The metagenomics approach provided the capability of extending beyond *E. coli* and single gene PCR testing to provide a comprehensive screen of beach water samples for bacterial community composition and toxin and antimicrobial resistance genes associated with changing beach water conditions.

Declarations:

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The sequencing data from this study is available from the corresponding author upon reasonable request. All the analytical data supporting the findings in this study is provided in the figures and tables in the main manuscript and supplementary data files.

Competing interests

The authors declare that they have no competing interests

Funding

This project has received funding support from the Government of Ontario (Canada-Ontario Agreement grant agreement #2607). Such support does not indicate endorsement by the Government of Ontario of the contents or conclusions of this contribution.

Authors' contributions

Conceptualization, H.E.S.; methodology H.E.S., F.S., and T.A.E; formal analysis and Experimentation, F.S., and K.L.T; writing—original draft preparation, F.S; writing—review and editing H.E.S., F.S., and T.A.E., supervision, H.E.S. All authors have read and agreed to the published version of the manuscript.

References:

1. USEPA (2009). National Epidemiologic and Environmental Assessment of Recreational Water Epidemiology Studies. EPA/600/R-10/168. Available at: http://www.epa.gov/neear/files/Report2009v5_508comp.pdf
2. Health Canada (2023). Canadian recreational water quality guidelines-Indicators of

fecal contamination: Overview Available at: https://www.canada.ca/en/health-canada/services/publications/healthy-living/recreational-water-quality-guidelines-indicators-fecal-contamination.html

3. Edge, Thomas A., and Stephen Hill. "Multiple lines of evidence to identify the sources of fecal pollution at a freshwater beach in Hamilton Harbour, Lake Ontario." Water research 41, no. 16 (2007): 3585-3594.
4. Edge, T. A., S. Hill, P. Seto, and J. Marsalek. "Library-dependent and library-independent microbial source tracking to identify spatial variation in faecal contamination sources along a Lake Ontario beach (Ontario, Canada)." Water Science and Technology 62, no. 3 (2010): 719-727.
5. USEPA. (2012). Recreational water quality criteria. (pp. 42): United States (US) Environmental Protection Agency (EPA). EPA-820-F-12–058. Available at: https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf. USEPA, 2009..
6. Ferguson, Donna M., John F. Griffith, Charles D. McGee, Stephen B. Weisberg, and Charles Hagedorn. "Comparison of *Enterococcus* species diversity in marine water and wastewater using Enterolert and EPA Method 1600." Journal of environmental and public health 2013 (2013).
7. Odonkor, Stephen T., and Joseph K. Ampofo. "*Escherichia coli* as an indicator of bacteriological quality of water: an overview." Microbiology research 4, no. 1 (2013): e2.

8. Brumfield, Kyle D., Joseph A. Cotruvo, Orin C. Shanks, Mano Sivaganesan, Jessica Hey, Nur A. Hasan, Anwar Huq, Rita R. Colwell, and Menu B. Leddy. "Metagenomic sequencing and quantitative real-time PCR for fecal pollution assessment in an urban watershed." Frontiers in water 3 (2021): 626849.
9. Ballesté, E., García-Aljaro, C. and Blanch, A.R., 2018. Assessment of the decay rates of microbial source tracking molecular markers and fecal indicator bacteria from different sources. Journal of Applied Microbiology, 125(6), pp.1938-1949.
10. Ahmed, W., Toze, S., Veal, C., Fisher, P., Zhang, Q., Zhu, Z., Staley, C. and Sadowsky, M.J., 2021. Comparative decay of culturable fecal indicator bacteria, microbial source tracking marker genes, and enteric pathogens in laboratory microcosms that mimic a sub-tropical environment. Science of The Total Environment, 751, p.141475.
11. Zimmer-Faust, A.G., Thulsiraj, V., Marambio-Jones, C., Cao, Y., Griffith, J.F., Holden, P.A. and Jay, J.A., 2017. Effect of freshwater sediment characteristics on the persistence of fecal indicator bacteria and genetic markers within a Southern California watershed. Water research, 119, pp.1-11.
12. Hannides, Angelos, Nicole Elko, Tiffany Roberts Briggs, Sung-Chan Kim, Annie Mercer, Kyeong Park, Brad Rosov, Ryan Searcy, and Michael Walther. "US beach water quality monitoring." Shore & Beach 89, no. 3 (2021): 26.
13. USEPA (2012). Recreational water quality criteria. (pp. 42): United States (US) Environmental Protection Agency (EPA). EPA-820-F-12-058. Available at:

https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf

14. Wade, T.J., Calderon, R.L., Brenner, K.P., Sams, E., Beach, M., Haugland, R., Wymer, L., Dufour, A.P., 2008. High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. Epidemiology, 375-383.
15. Wade, T.J., Sams, E., Brenner, K.P., Haugland, R., Chern, E., Beach, M., Wymer, L., Rankin, C.C., Love, D., Li, Q., 2010. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: a prospective cohort study. Environmental Health 9, 1-14.
16. Rochelle-Newall, Emma, Thi Mai Huong Nguyen, Thi Phuong Quynh Le, Oloth Sengtaheuanghoung, and Olivier Ribolzi. "A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions." Frontiers in microbiology 6 (2015): 308.
17. Islam, M.M., Hofstra, N. and Islam, M.A., 2017. The impact of environmental variables on faecal indicator bacteria in the Betna river basin, Bangladesh. Environmental Processes, 4, pp.319-332.
18. Harwood, Valerie J., Christopher Staley, Brian D. Badgley, Kim Borges, and Asja Korajkic. "Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes." FEMS microbiology reviews 38, no. 1 (2014): 1-40.
19. Nappier, Sharon P., Audrey Ichida, Kirsten Jaglo, Rich Haugland, and Kaedra R.

Jones. "Advancements in mitigating interference in quantitative polymerase chain reaction (qPCR) for microbial water quality monitoring." Science of the Total Environment 671 (2019): 732-740.

20. Cao, Yiping, Meredith R. Raith, and John F. Griffith. "Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment." Water research 70 (2015): 337-349.
21. Fresia, Pablo, Verónica Antelo, Cecilia Salazar, Matías Giménez, Bruno D'Alessandro, Ebrahim Afshinnekoo, Christopher Mason, Gastón H. Gonnet, and Gregorio Iraola. "Urban metagenomics uncover antibiotic resistance reservoirs in coastal beach and sewage waters." Microbiome 7, no. 1 (2019): 1-9.
22. Edge, Thomas A., Donald J. Baird, Guillaume Bilodeau, Nellie Gagné, Charles Greer, David Konkin, Glen Newton et al. "The Ecobiomics project: Advancing metagenomics assessment of soil health and freshwater quality in Canada." Science of The Total Environment 710 (2020): 135906.
23. Saleem, F., Edge, T.A. and Schellhorn, H.E.,. Validation of qPCR method for enterococci quantification at Toronto beaches: Application for rapid recreational water monitoring. Journal of Great Lakes Research, 48(3) (2022), pp.707-716.
24. Mohiuddin, Mahi M., Steven R. Botts, Athanasios Paschos, and Herb E. Schellhorn. "Temporal and spatial changes in bacterial diversity in mixed use watersheds of the Great Lakes region." Journal of Great Lakes Research 45, no. 1 (2019): 109-118.
25. APHA, 2012. Standard methods for the examination of water and wastewater.

American Public Health Association. American Water Works Association, Water Environment Federation.

26. Saleem, F., Schellhorn, H.E., Simhon, A. and Edge, T.A.,. Same-day *Enterococcus* qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. Canadian Journal of Public Health, (2023), pp.1-12.
27. USEPA. (2015). Method 1609.1: Enterococci in water by TaqMan® quantitative polymerase chain reaction (qPCR) with internal amplification control (IAC) assay. EPA-820-R-15–099. Available at: https://www.epa.gov/sites/default/files/2015-08/documents/method_1609-1-*Enterococcus*-iac_2015_3.pdf.
28. Andrews, Simon. "FastQC: a quality control tool for high throughput sequence data." (2010).
29. Buchfink, Benjamin, Chao Xie, and Daniel H. Huson. "Fast and sensitive protein alignment using DIAMOND." Nature methods 12, no. 1 (2015): 59-60.
30. Bağcı, Caner, Sina Beier, Anna Górska, and Daniel H. Huson. "Introduction to the analysis of environmental sequences: metagenomics with MEGAN." Evolutionary genomics: statistical and computational methods (2019): 591-604.
31. Bağcı, Caner, Sascha Patz, and Daniel H. Huson. "DIAMOND+ MEGAN: fast and easy taxonomic and functional analysis of short and long microbiome sequences." Current protocols 1, no. 3 (2021): e59.
32. Huson, D.H., Richter, D.C., Mitra, S., Auch, A.F. and Schuster, S.C.,. Methods for

comparative metagenomics. BMC bioinformatics, 10 (2009), pp.1-10.

33. Pereira, M.B., Wallroth, M., Jonsson, V. and Kristiansson, E.,. Comparison of normalization methods for the analysis of metagenomic gene abundance data. BMC genomics, 19 (2018), pp.1-17.
34. De Brito, Luciana Carla Neves, Janet Doolittle-Hall, Chun-Teh Lee, Kevin Moss, Wilson Bambirra Júnior, Warley Luciano Fonseca Tavares, Antônio Paulino Ribeiro Sobrinho, and Flávia Rocha Fonseca Teles. "The apical root canal system microbial communities determined by next-generation sequencing." Scientific Reports 10, no. 1 (2020): 10932.
35. Astudillo-García, Carmen, James J. Bell, Nicole S. Webster, Bettina Glasl, Jamaluddin Jompa, Jose M. Montoya, and Michael W. Taylor. "Evaluating the core microbiota in complex communities: a systematic investigation." Environmental microbiology 19, no. 4 (2017): 1450-1462.
36. Chernov, T. I., A. K. Tkhakakhova, and O. V. Kutovaya. "Assessment of diversity indices for the characterization of the soil prokaryotic community by metagenomic analysis." Eurasian Soil Science 48 (2015): 410-415.
37. Bae, Sungwoo, Colleen Lyons, and Nora Onstad. "A culture-dependent and metagenomic approach of household drinking water from the source to point of use in a developing country." Water research X 2 (2019): 100026.
38. Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics

assembly via succinct de Bruijn graph." Bioinformatics 31, no. 10 (2015): 1674-1676.

39. De Nies, Laura, Sara Lopes, Susheel Bhanu Busi, Valentina Galata, Anna Heintz-Buschart, Cedric Christian Laczny, Patrick May, and Paul Wilmes. "PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data." Microbiome 9 (2021): 1-14.
40. Parks, Donovan H., Gene W. Tyson, Philip Hugenholtz, and Robert G. Beiko. "STAMP: statistical analysis of taxonomic and functional profiles." Bioinformatics 30, no. 21 (2014): 3123-3124.
41. West, Robert M. "Best practice in statistics: Use the Welch t-test when testing the difference between two groups." Annals of clinical biochemistry 58, no. 4 (2021): 267-269.
42. Hu, W., Liang, J., Ju, F., Wang, Q., Liu, R., Bai, Y., Liu, H. and Qu, J.,. Metagenomics unravels differential microbiome composition and metabolic potential in rapid sand filters purifying surface water versus groundwater. Environmental Science & Technology, 54(8) (2020), pp.5197-5206.
43. Sun, F., Wang, C. and Yang, H.,. Physicochemical factors drive bacterial communities in an aquaculture environment. Frontiers in Environmental Science, 9 (2021), p.709541.
44. Miller, R.R., Montoya, V., Gardy, J.L., Patrick, D.M. and Tang, P.,. Metagenomics for pathogen detection in public health. Genome medicine, 5 (2013), pp.1-14.

45. Bae, S., Lyons, C. and Onstad, N.,. A culture-dependent and metagenomic approach of household drinking water from the source to point of use in a developing country. Water research X, 2 (2019), p.100026.

46. Lee, C.S., Kim, M., Lee, C., Yu, Z. and Lee, J.,. The microbiota of recreational freshwaters and the implications for environmental and public health. Frontiers in microbiology, 7, (2016), p.1826.

47. Hengy, M.H., Horton, D.J., Uzarski, D.G. and Learman, D.R.,. Microbial community diversity patterns are related to physical and chemical differences among temperate lakes near Beaver Island, MI. PeerJ, 5 (2017), p.e3937.

48. Nevers, Meredith B., Muruleedhara N. Byappanahalli, Cindy H. Nakatsu, Julie L. Kinzelman, Mantha S. Phanikumar, Dawn A. Shively, and Ashley M. Spoljaric. "Interaction of bacterial communities and indicators of water quality in shoreline sand, sediment, and water of Lake Michigan." Water research 178 (2020): 115671.

49. Rothenheber, Derek, and Stephen Jones. "Enterococcal concentrations in a coastal ecosystem are a function of fecal source input, environmental conditions, and environmental sources." Applied and Environmental Microbiology 84, no. 17 (2018): e01038-18.

50. Nakatsu, Cindy H., Muruleedhara N. Byappanahalli, and Meredith B. Nevers. "Bacterial community 16S rRNA gene sequencing characterizes riverine microbial impact on Lake Michigan." Frontiers in microbiology 10 (2019): 996.

51. Fisher, Jenny C., Ryan J. Newton, Deborah K. Dila, and Sandra L. McLellan. "Urban

microbial ecology of a freshwater estuary of Lake Michigan." Elementa 3 (2015): 000064.

52. Powers, Nicole C., Hailey R. Wallgren, Sandra Marbach, and Jeffrey W. Turner. "Relationship between rainfall, fecal pollution, antimicrobial resistance, and microbial diversity in an urbanized subtropical bay." Applied and environmental microbiology 86, no. 19 (2020): e01229-20.
53. Winters, A.D., Marsh, T.L., Brenden, T.O. and Faisal, M.,. Molecular characterization of bacterial communities associated with sediments in the Laurentian Great Lakes. Journal of Great Lakes Research, 40(3) (2014), pp.640-645.
54. Crevecoeur, Sophie, Thomas A. Edge, Linet Cynthia Watson, Susan B. Watson, Charles W. Greer, Jan JH Ciborowski, Ngan Diep et al. "Spatio-temporal connectivity of the aquatic microbiome associated with cyanobacterial blooms along a Great Lake riverine-lacustrine continuum." Frontiers in Microbiology 14 (2023): 1073753.
55. Paerl, H. W., and Otten, T. G. Harmful cyanobacterial blooms: causes, consequences, and controls. Microbial Ecology. 65 (2013), 995–1010.
56. Coffey, Rory, Michael J. Paul, Jen Stamp, Anna Hamilton, and Thomas Johnson. "A review of water quality responses to air temperature and precipitation changes 2: nutrients, algal blooms, sediment, pathogens." JAWRA Journal of the American Water Resources Association 55, no. 4 (2019): 844-868.
57. Halac, Silvana Raquel, Raquel del Valle Bazán, Nancy Beatriz Larrosa, Ana

Florencia Nadal, Ana Laura Ruibal-Conti, Maria Ines Rodriguez, Marcia Andrea Ruiz, and Abel Gerardo López. "First report on negative association between cyanobacteria and fecal indicator bacteria at San Roque reservoir (Argentina): impact of environmental factors." Journal of Freshwater Ecology 34, no. 1 (2019): 273-291.

58. Ransom-Jones, E., Jones, D.L., McCarthy, A.J. and McDonald, J.E., 2012. The Fibrobacteres: an important phylum of cellulose-degrading bacteria. Microbial ecology, 63, pp.267-281.
59. Jeong, J.Y., Park, H.D., Lee, K.H., Weon, H.Y. and Ka, J.O., 2011. Microbial community analysis and identification of alternative host-specific fecal indicators in fecal and river water samples using pyrosequencing. The Journal of Microbiology, 49, pp.585-594.
60. Muñoz-Zanzi, C., Mason, M.R., Encina, C., Astroza, A. and Romero, A., 2014. Leptospira contamination in household and environmental water in rural communities in southern Chile. International journal of environmental research and public health, 11(7), pp.6666-6680.
61. Fisher, J.C., Newton, R.J., Dila, D.K. and McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa, 3, p.000064.
62. Staley, Z., J. Grabuski, E. Sverko, and T.A. Edge. 2016. Comparison of microbial and chemical source tracking markers to identify fecal contamination sources in the Humber River and associated stormwater outfalls. Applied and Environmental Microbiology 82: 6357-6366.

63. Staley, Z.R., Chuong, J.D., Hill, S.J., Grabuski, J., Shokralla, S., Hajibabaei, M. and Edge, T.A., 2018. Fecal source tracking and eDNA profiling in an urban creek following an extreme rain event. Scientific reports, 8(1), p.14390.

64. Paulino, G.V.B., Félix, C.R., Silvan, C.G., Andersen, G.L. and Landell, M.F., 2020. Bacterial community and environmental factors associated to rivers runoff and their possible impacts on coral reef conservation. Marine pollution bulletin, 156, p.111233.

65. Carney, R.L., Labbate, M., Siboni, N., Tagg, K.A., Mitrovic, S.M. and Seymour, J.R., 2019. Urban beaches are environmental hotspots for antibiotic resistance following rainfall. Water Research, 167, p.115081.

66. Kraemer, Susanne A., Naila Barbosa da Costa, Anaiss oliva, Yannick Huot, and David A. Walsh. "A resistome survey across hundreds of freshwater bacterial communities reveals the impacts of veterinary and human antibiotics use." Frontiers in Microbiology 13 (2022): 995418.

67. Liguori, K., Keenum, I., Davis, B.C., Calarco, J., Milligan, E., Harwood, V.J. and Pruden, A., 2022. Antimicrobial resistance monitoring of water environments: a framework for standardized methods and quality control. Environmental science & technology, 56(13), pp.9149-9160.

Chapter 7. eDNA Metabarcoding-based Source Attribution of Fecal Indicator Bacteria Exceedances in Urban Freshwater Beaches, Sand and Rivers

This chapter demonstrates the application of DNA sequencing-based technology for the characterization of fecal contamination sources, including human sewage, wildlife, and domestic animals. Therefore, providing actionable insights for targeted risk management and intervention strategies. The successful application of eDNA metabarcoding and MST methods in this study highlights the potential for these techniques to enhance the accuracy and efficiency of water quality monitoring programs, which aligns with the aims and objectives of this thesis.

The text I present here is ***under review for publication****. The formatting has been changed from the journal version to be consistent throughout the thesis.*

Author Contributions

FS performed the experiments, analyzed the data, wrote the original manuscript, and revised it as required by the reviewers. EL, JLJ and KT helped with the experiments. HES and TAE supervised the study and reviewed the manuscript before submission to the journal.

Abstract:

Freshwater beach quality is routinely tested by measuring fecal indicator bacteria, which can assess water quality but cannot identify sources of fecal contamination. We compared eDNA metabarcoding and microbial source tracking (MST) digital PCR methods to identify fecal contamination sources in water and sand at four urban Lake Ontario beaches and two nearby river mouth locations. eDNA sequences matched mammal, bird, and fish taxa known in the study area. Human eDNA sequences were prominent in all water and sand samples such that they had less value for discriminating between sewage occurrence at sites. Mallard duck, muskrat, beaver, raccoon, gull, robin, chicken, red fox, and cow eDNA sequences were common across all locations. Dog, Canada goose, and swan eDNA sequences were more common in Toronto beach waters, suggesting localized sources. MST results were generally consistent with eDNA, such as finding the Gull4 DNA marker and the human mitochondrial DNA marker in most water and sand samples. Chicken, cow, and dog eDNA sequences and the human bacterial MST DNA marker often showed a higher probability of occurrence on Beach Action Value (BAV) exceedance days. The surprisingly widespread detection of chicken and cow eDNA sequences was likely from incompletely digested human food, raising caution for interpreting eDNA results related to food animals in sewage-contaminated urban settings. Combined use of MST and eDNA methods provided a more comprehensive characterization of potential fecal contamination sources, including diverse wildlife species at the human-animal One Health interface, that can guide targeted beach-specific water monitoring and risk management strategies.

Keywords: eDNA, metabarcoding, recreational waters, beach sands, beaches, rivers, fecal source tracking.

Introduction:

Fecal contamination is one of the main causes of the deterioration of water quality and is a concern for the long-term sustainability of recreational water ecosystems[1]. Traditional routine beach monitoring strategies rely on fecal indicator bacteria levels for water quality assessment. However, controlling the impact of fecal contamination on the sustainability of recreational waters requires comprehensive identification of fecal sources and such information cannot be obtained from fecal indicator bacteria levels alone[2, 3, 4]. Fecal source tracking to date has been mainly based on a microbial source tracking approach targeting microorganisms specific to a host organisms' gut, including humans, birds, and mammals[5,6,7]. Several microbial fecal source tracking markers, like the human *Bacteroides* HF183 marker, have been extensively tested[8]. However, there are still relatively few well-validated host-specific microbial DNA markers for animals, particularly for non-domestic animals. Mitochondrial sequences from human, mammal, and avian cells in environmental DNA (eDNA) offer another approach to expand the toolbox for detecting fecal contamination sources[9]. These eukaryotic host cells are continuously sloughed off in large numbers from the lining of human, mammal, and bird gastrointestinal tracts into fecal matter that is then shed into the environment[10].

DNA derived from mammalian or avian cells in water samples can provide information on the taxonomic diversity of potential fecal contamination sources in aquatic ecosystems[9]. Common methods for detecting human or animal host sequences in eDNA have employed species-specific primers and quantitative PCR[10] or digital PCR[11]. However, fecal

contamination in environmental waters can be from diverse animal sources, and such PCR-based approaches have typically targeted one marker at a time and are unable to provide a comprehensive diversity profile of potential fecal sources. In contrast, Next Generation Sequencing-based metabarcoding facilitates the comprehensive characterization of universal marker genes (such as mitochondrial 16S rRNA gene) in environmental DNA. This combination of mammal and avian eDNA metabarcoding, along with growing numbers of host-specific microbial source tracking DNA markers, offers the potential to comprehensively identify potential fecal contamination sources for recreational waters[12,13].

In this study, we compared eDNA metabarcoding and microbial source tracking DNA markers for more comprehensive profiling of potential fecal contamination sources in water and sand at four urban Lake Ontario beaches and two nearby river mouth locations. The questions we addressed were: 1) What are the potential fecal contamination sources for our four urban freshwater beaches? 2) How different are fecal contamination sources as determined by eDNA metabarcoding and microbial source tracking approaches? 3) What are the different fecal contamination sources between urban beach waters, beach sands, and rivers? And 4) Which fecal contamination sources are most associated with Beach Action Value (BAV) exceedances at beaches?

Materials and Methods:

Study Plan, Sample Collection and Fecal Indicator Bacteria Data:

We tested four freshwater beaches (two urban beaches in the City of Toronto and two urban beaches in the City of St Catherines on Lake Ontario) for beach sands at the St Catherines beaches and two rivers adjacent to the two Toronto beaches. Figure 1 indicates the geographical locations of each sampling location from beaches and rivers. There are municipal wastewater treatment plants that discharge in the vicinity of all four tested beach sites. For water sample collection from Toronto beaches and rivers, samples were collected three times a week for the 2021 summer season between June 01 and August 26 (total=309 samples). On each sampling day, eight samples were collected from Marie Curtis Park East Beach (3 samples: 30W, 30W replicate, and 32W), Sunnyside Beach (3 samples: 18W, 21W, and 21W replicate), Etobicoke Creek (1 sample), and Humber River (1 sample). Samples from St Catherines beaches were collected three times a week in the 2022 summer season between May 31 and September 06 (total=336 water and sand pore water samples). Eight samples were collected for each sampling day, including 2 water samples and 2 sand pore water samples at Lakeside Beach (LK3 and LK5 sampling sites) and 2 water and 2 sand pore water samples at Sunset Beach (SS3 and SS5 sampling sites). Sand pore water samples were obtained by digging to collect groundwater in the beach swash zone within about 1 meter of the lake. Samples were collected between 5:30 and 7 am for the 2021 sampling season (Toronto) and 8:00 and 9 am for the 2022 sampling season. Grab samples were collected in 500 mL sterile polyethylene terephthalate (PET) bottles and were delivered to the lab on ice

within 1 hour of sample collection. After delivery to the lab, samples were processed for *E. coli* by membrane filtration and *Enterococcus* by qPCR within a few hours. Fecal indicator bacteria data and categorizations of water samples in relation to Beach Action Values (BAV) were taken from our previously published work[14,15].

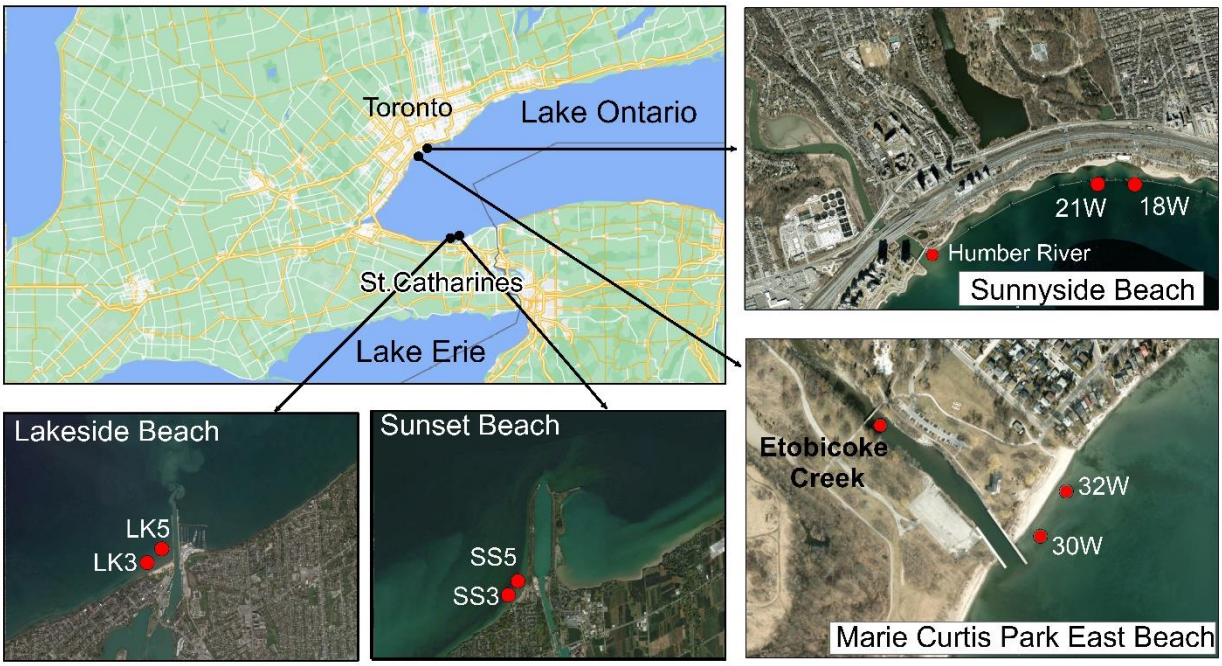


Figure 1. Geographical locations of Toronto and St Catherines (Niagara) sampling sites.

DNA Extraction and Sample Selection/Pooling for Metabarcoding:

For eDNA extraction, 300 mL of sample was filtered through a 0.22-micron nitrocellulose membrane filter (Millipore Corp., Bedford, MA). Filters were then processed for DNA extraction using the Norgen Soil Plus DNA Extraction kit (Norgen Biotek Corp., Canada) with minor modifications. We increased the bead beating time and changed the glass beads to zirconium beads to improve cell disruption. All the other steps were performed using the

manufacturer's protocol. For DNA sequencing (eDNA metabarcoding), 48 samples were selected from across beach water, beach sand, and river sampling locations (Supplementary Table 1). Sample selection first categorized samples as Beach Action Value Exceedance or non-exceedance samples based on fecal indicator levels (*E. coli* and *Enterococcus*). Then, for each beach location, 3 BAV-exceedance and 3 non-exceedance samples were selected, except for Lakeside Beach, for which we selected 2 samples each (due to only two BAV exceedance days). The 3 BAV-exceedance and 3 non-exceedance samples were selected for each river location according to the same sampling dates for their adjacent beaches. For beach sand locations at Lakeside and Sunset Beaches, the BAV-exceedance and non-exceedance samples were selected, corresponding to the same sampling dates selected for their adjacent beach water samples. DNA corresponding to sampling sites from a single location and for a single sampling day was pooled in equivalent volumes as a single sample for library preparation and DNA sequencing.

eDNA Metabarcoding Sequencing:

Mitochondrial 16S PCR was performed in two parts following Ragot and Villemur (2022)[13]: 1) Amplification of ~400 bp fragment with PCR cycles limited to 10 to attempt to reduce PCR amplification bias due to dominant taxa, and 2) Nested PCR using Illumina linker-attached primers and PCR product from first PCR with 35 cycles. For the first PCR, each PCR reaction (25 µL) comprised of 12.5 µL Hot Start PCR master mix (Thermoscientific Inc, USA), 1.0 µL of forward and reverse primers (10 µM), 2.0 µL DNA, and 8.5 µL of Nuclease-free water. The PCR protocol for the first PCR included initial denaturation at 95

°C for 10 min, 10 cycles of 95 °C for 30 sec, 58 °C for 1 min, and 72 °C for 40 sec, followed by final extension at 72 °C for 5 min. The reaction composition and PCR protocol remained the same for the nested PCR, except the PCR product from the first PCR was used as the DNA template, and the PCR protocol included touchdown PCR for annealing (69 to 59 °C for 10 cycles) to improve the annealing efficiency before the cycling steps. After amplicon PCR, amplicons were purified using Ampure XP magnetic beads (Beckman Coulter, California, USA). For index PCR (attachment of unique DNA barcodes to each sample's amplicons), we designed 7 P5 and 8 P7 primers, corresponding to 56 unique barcode combinations and ordered as Ultramers from IdtDNA (Coralville, USA). For the index PCR, each PCR reaction (25 µL) comprised of 12.5 µL Hot Start PCR master mix (Thermo Scientific Inc, USA), 2.0 µL of P5 and P7 indexing primers (5 µM), 5.0 µL of purified PCR product, and 6.5 µL of Nuclease-free water. The index PCR protocol included initial denaturation at 95 °C for 3 min, 8 cycles of 95 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 30 sec, followed by final extension at 72 °C for 5 min. Indexed PCR amplicons were purified using Ampure XP magnetic beads (Beckman Coulter, California, USA), followed by normalization to 4 nm for each sample and pooling. All the PCR reactions were performed on Bio-Rad CFX96 Touch Real-Time PCR (Bio-Rad Inc. USA). Paired-end sequencing (2x250, V2 kit) was performed on the Illumina Mi-Seq sequencing platform at Farncombe Sequencing Institute (McMaster University).

Digital PCR for Microbial Fecal Source Tracking Markers:

Assays for digital PCR (dPCR) used previously published primer and probe sets for the human HF183 marker[16], the seagull Gull4 marker[17], the human mitochondrial DNA marker[10], the Canada goose mitochondrial DNA marker[18], and the dog Dog3 marker[19]. The ruminant Rum2Bac DNA marker[20] and the swine Pig2Bac DNA marker[21] were run only on samples with cattle and pig sequences detected from eDNA analysis. Digital PCR reactions were conducted as a duplex assay for HF183 and Gull4 markers and for the human mitochondrial and goose mitochondrial markers, while other markers were run in a singleplex format. Comparative assay testing in singleplex and duplex formats has not indicated any interference between marker assays in each duplex. Each duplex dPCR reaction consisted of 1 µL nuclease-free water, 0.75 µL of each 900 nM forward and reverse primer, 0.75 µL of each 250 nM probe, 7.5 µL of QuantStudio™ 3D Digital PCR Master Mix v.2 (ThermoFisher) and 2.0 µL of extracted DNA template. Each singleplex dPCR reaction consisted of 1.95 µL nuclease-free water, 1.4 µL of each 900 nM forward and reverse primer, 0.75 µL of 250 nM probe, 7.5 µL of QuantStudio™ 3D Digital PCR Master Mix v.2 (ThermoFisher) and 2.0 µL of extracted DNA template. Reactions were loaded onto a 20,000 micro-well chip (QuantStudio™ 3D Digital PCR 20K Chip v2 with partition volume = 755 pL) using a QuantStudio™ 3D Chip Loader (ThermoFisher), and PCR assays were carried out using the ProFlex PCR System (ThermoFisher). Thermocycler settings were 96˚C for 10 min, followed by 40 cycles of 60˚C for 2 min and 98˚C for 30 s, then 60˚C for 2 min. Chips were then read using a QuantStudio™ 3D Digital PCR Instrument

(ThermoFisher). Results were analyzed using the QuantStudio™ AnalysisSuite™, which determined threshold fluorescence values for the ROX reference dye to identify the number of qualified PCR well partitions, as well as for the FAM and VIC dye signals to identify positive reactions for the DNA markers. The AnalysisSuite™ software applies a Poisson Plus modelling technique to determine concentrations of each target within the sample, and results were reported as DNA copies per 100 mL.

All no-template PCR controls were negative, indicating a lack of contamination to compromise dPCR assays. The number of wells analyzed on dPCR chips was typically between 16,000 to 18,000 wells. The sensitivity and specificity of the dPCR assays were outlined in Edge et al. (2021)[11], and 2-3 additional samples of gull feces, dog feces, cow feces, pig feces, sewage influent, as well as DNA standards prepared by the U.S. National Institute of Standards and Technology (SRM 2917, Plasmid DNA for Fecal Indicator Detection and Identification) were used as positive reference materials for dPCR runs. At least 4 positive wells (clearly cluster separated from negative wells) were set as a detection threshold for a DNA marker in dPCR assays. This threshold provided a clear basis for discriminating water sample dPCR results from all dPCR results for filter and DNA extraction blanks and no-template PCR control samples. The threshold was equivalent to a detection limit of about 14 DNA copies per 100ml.

Bioinformatics and Data Analysis:

For each sample, ~200,000 paired-end sequences were obtained following eDNA sequencing. Data quality was checked using FastQC[22], and sequences below the quality

threshold 30 were removed using the Fastp FastQ preprocessing tool[23]. After initial quality control, the APSCALE pipeline[24] for metabarcoding analysis was used for downstream bioinformatics. Bioinformatics pipeline steps included: 1) length filtration (Minimum length = 70, maximum length = 400 bp), Illumina barcodes removal and primer sequences trimming were performed using Cutadept[25], 2) Paired-end merging (Maxdiffpct = 25, maxdiffs = 199, and minovlen = 5), and Denoising (Alpha = 2, and minsize = 8) were performed using VSEARCH[26] to identify Exact Seq, and 3) LULU algorithm[27] was used for post-clustering curation to remove erroneous Exact Sequence Variants (ESVs). Exact Sequence Variants were taxonomically annotated using NCBI BLASTn (percentage identity >95% and alignment length/query coverage >80%) against the RefSeq nucleotide database. To normalize the datasets and avoid false positives, ESVs or taxonomic groups lower than 0.003% in abundance were considered absent and were removed from the datasets, as described previously for mt metabarcoding datasets[28]. The conditional probability of a taxa as a fecal contamination source was calculated using the following equation[29,30]:

$$P(H/T) = \frac{P(T/H)x\ Detection\ Frequency\ of\ a\ taxa}{P(T/H)\ x\ Detection\ Frequency + P(T/H')\ x\ P(H')}$$

Where P(H/T) is the conditional probability of a taxon as a potential fecal contamination source, P(T/H) is the Probability of Sequence Detection (BAV Exceedance), P(T/H') is the Probability of Sequence Detection (BAV Non-exceedance), and P(H') is the probability of not detecting a taxon in the water samples. Statistical comparison between groups was performed using Welch's t-test on GGPLOT and GGPUBR packages in R[31].

Results:

Quality Control Analyses:

An initial trial was conducted to test for the validity of mitochondrial 16S rRNA gene fragment amplification (Supplementary Figure 1). All the tested samples from rivers and beach sampling sites amplified the gene fragment in the 300-400 bp range, similar to the positive control (Salmon sperm DNA). Although a non-specific gene fragment at ~800 bp was detected, it was removed during the gene library purification steps using magnetic bead size selection. After the Next-generation sequencing, more than ~200,000 DNA reads (22.3 ± 2.2 $x10^4$) were generated for each sample (Supplementary Table 2). About 95-97% of the DNA reads passed the quality control parameters after primer trimming and removal of lower-quality (Quality score <30) sequences. In total, 369 Exact Sequence Variants (ESVs) were identified, out of which 200 were processed for sequence annotation after removing potentially erroneous ESVs using LULU filtering, which identifies errors by searching for co-occurrence patterns and sequence similarity measures[27].

Taxonomic Annotation and Characterization of eDNA Sequences:

Across all 48 samples (200 ESVs), mitochondrial DNA metabarcoding identified 61 taxa, including fish, birds, mammals and arthropods (Supplementary File 2). The fish, mammal and bird taxa identified were consistent with those known from our study area, and lists of species for each sampling site are provided in Supplementary Tables 3 and 4. In a few cases, we corrected taxonomic assignments to more generic identifications (European herring gull

= gull; Northern house martin = martin; Eurasian tree sparrow = sparrow). Supplementary Figure 2 shows the distribution of ESVs according to the mammal, bird, and fish categories in all samples. Mitochondrial DNA sequences associated with fish were relatively most abundant (55-78%), followed by mammals (20-43%) and birds (2-5%). While not the focus of this study, the predominant fish species detected by eDNA sequences in every sample was the round goby (*Neogobius melanostomus*), consistent with a previous study on the Great Lakes[32]. However, alewife (*Alosa pseudoharengus*), green sunfish (*Lepomis cyanellus*), longnose dace (*Rhinichthys cataractae*), central stoneroller (*Campostoma anomalum*), blacknosed dace (*Rhinichthys atratulus*), rock bass (*Ambloplites rupestris*), and bluntnose minnow (*Pimephales notatus*) were also widely prevalent. Beta diversity analysis was performed using organism types (fishes, mammals, and birds) to assess the divergence patterns among the sampling sites. Diversity analysis for fish and mammals sp. (Supplementary Figures 3 and 4) revealed two distinct but closely associated clusters. Samples from Toronto beaches and their adjacent rivers clustered together, while Niagara beach water and their adjacent beach sand samples aggregated into a separate cluster. Interestingly, compared to beach samples that grouped, samples from river sources were more broadly distributed, indicating a higher diversity of fish and mammal species sequences in rivers. However, for birds sp. (Supplementary Figure 5), samples from both beach locations (Toronto and Niagara) were closely associated in a single cluster.

eDNA Sequence Detection Frequency of Potential Fecal Contamination Sources

Mitochondrial DNA sequences resolved and taxonomically assigned to mammalian and bird fecal contamination sources are shown in Supplementary Figure 6. All 48 samples were positive for human ESVs, followed by *Castor canadensis* (Beaver: 97%), *Ondatra zibethicus* (Muskrat: 97%), *Anas platyrhynchos* (Mallard duck: 91%), *Larus* species (Gull: 87%), *Procyon lotor* (Raccoon: 87%), *Turdus migratorius* (Robin: 77%), *Gallus gallus* (Chicken: 73%), *Vulpes vulpes* (Fox: 50%), *Bos taurus* (Cow: 48%), *Canis lupis familiaris* (Dog: 33%), and *Branta canadensis* (Canada goose: 27%). Multiple ESVs were associated with most potential fecal sources, notably: human (7), squirrel (7), swan (7), robin (6), and 3 for beaver, dog, opossum, cow, red fox, Canada goose gull, and pigeon.

Supplementary Figure 7 shows human sequences were predominant in all the samples (87-96%), followed by beaver and muskrat (3-8%), mallard (2-5%), gull (2-5%), raccoon (1-5%), chicken (1-2%), and others (<1%). Figures 2 and 3 show the percentage detection of common mammalian and avian eDNA sequences across all beach and river study locations. Among the mammalian sequences, human (100%), muskrat/beaver (92-100%) and raccoon (75-90%) mt sequences showed the highest detection frequency across all four sample types. Dog, vole and cat mt sequences were more frequently detected in Toronto beach water and river samples than in Niagara beach water and adjacent sand samples. Among avian sequences, mallard (100%), gull (82-100%), robin (68-100%), and chicken (70-82%) showed the highest detection frequency across all sample types. Goose, swan, martin, and

pigeon mt sequences were more frequently detected for Toronto beaches and river samples than Niagara beach water and adjacent sand samples.

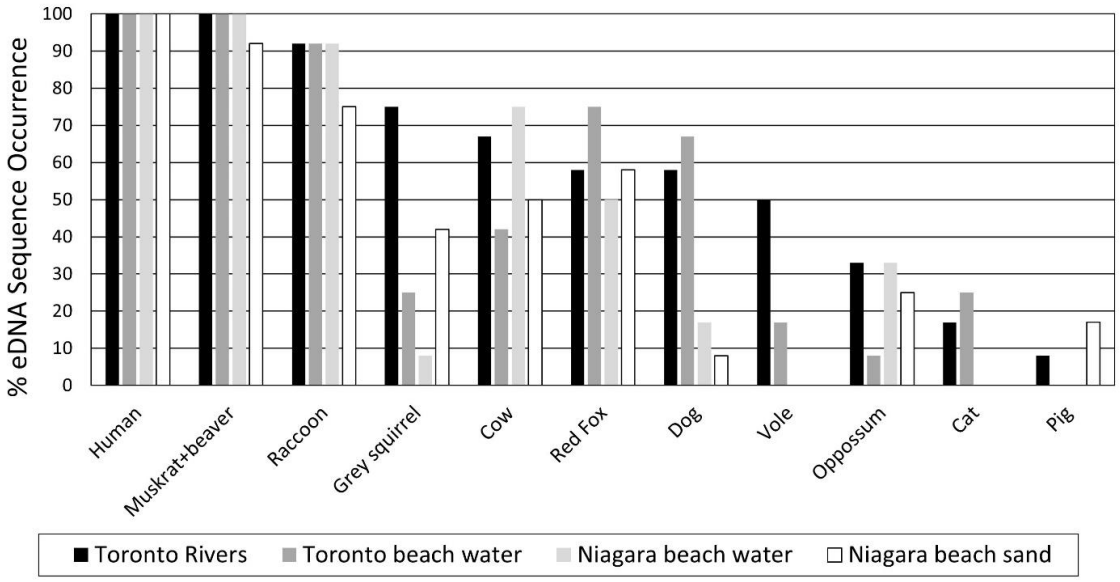


Figure 2. Detection percentage of most common mammalian eDNA sequences for beach, river, and sand samples.

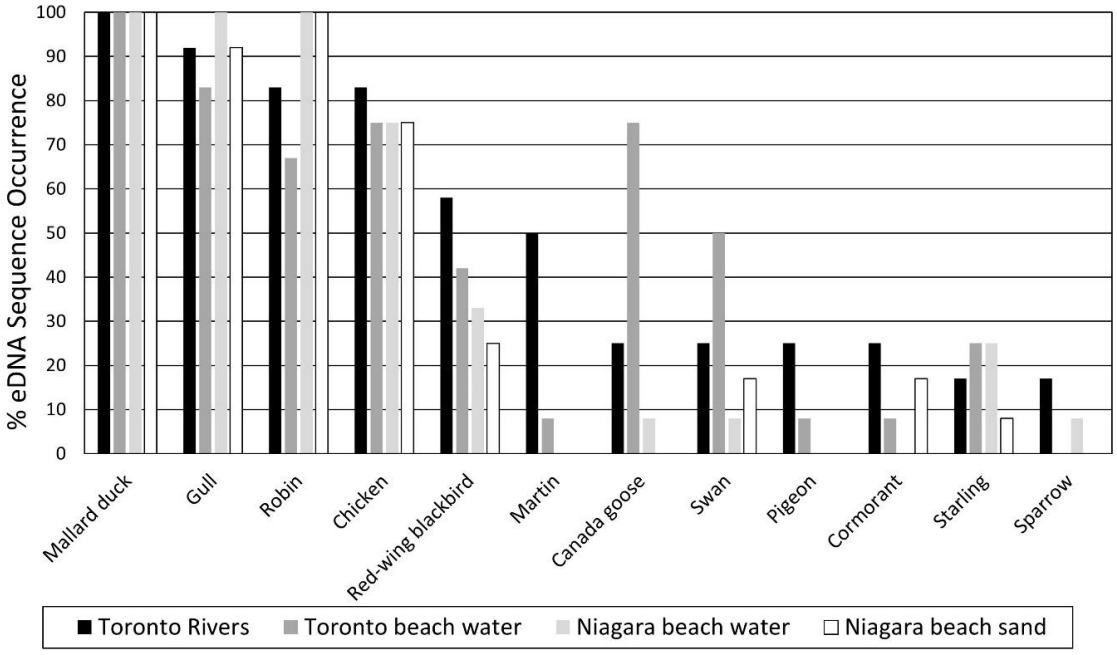


Figure 3. Detection percentage of most common avian eDNA sequences for beach, river, and sand samples.

eDNA Sequence Differential Abundance between the Sample Types/Locations

Several species were relatively more abundant (beaver (p = 4.7e-5), mallard (p = 1.0e-3), and chicken (p = 0.04) at Toronto locations (Supplementary Figures 8, 9 and 10). A comparison between Toronto beach and river samples identified that human and beaver mt sequences were relatively more abundant in river samples. Raccoon, gull, and chicken sequences were relatively more abundant in Toronto beach samples (Supplementary Figures 11, 12, 13, 14, and 15), though only chicken sequences were statistically significant (p = 0.004). No significant differences were observed between Niagara beach water and Niagara beach sand samples.

eDNA Occurrence Probability in Association to Fecal Indicator Beach Action Value Exceedances:

The occurrence of human, mammal, or bird eDNA sequences associated with BAV exceedance and non-exceedance conditions was calculated to estimate their potential as fecal contamination source indicators (Table 2). The probability of occurrence of humans and beavers as potential fecal contamination sources was 100% for all the sampling locations. Other highly probable fecal contamination sources included mallard, chicken, gull, raccoon, cow, muskrat, dog, Canada goose, fox, and robin. Comparison between the Toronto beach and adjacent rivers and Niagara beach water and adjacent beach sand identified similar patterns of eDNA sequence probability of occurrence. The probability of detecting Canada goose (67% for beaches and 40% for rivers) and dog (77% for beaches and 50% for rivers) eDNA sequences was higher for Toronto beaches than their adjacent rivers. Additionally, dog and Canada goose mt sequences were more frequently detected for Toronto beach and river samples than Niagara beach water and sand samples. The Canada goose mitochondrial eDNA sequence was not detected in any Niagara beach water or beach sand samples, while the dog eDNA sequence was only detected in a single sample. Comparison between Beach Action Value exceedance and non-exceedance beach water samples identified that beaver, muskrat, mallard and chicken mt sequences were relatively

more abundant in BAV exceedance samples (Supplementary Figures 16, 17, 18, and 19). However, only chicken ($p = 0.02$) were significant. Specifically, the higher abundance of

chicken mt sequences (p = 0.03) in BAV exceedance samples was mainly associated with Toronto beaches (Supplementary Figure 20).

Table 5. Probability of human, mammal, and bird species as potential fecal contamination sources for the beaches, rivers and sand.

Sampling Locations	Potential Fecal Sources	Probability of Sequence Detection (BAV Exceedance)	Probability of Sequence Detection (BAV Non-exceedance)	Conditional Probability as a Potential Fecal Contamination Source
Toronto Beaches	**Human**	100%	100%	100%
	Beaver	100%	100%	100%
	Mallard	100%	100%	100%
	Chicken	100%	50%	86%
	Gull	67%	100%	77%
	Racoon	100%	83%	93%
	Cow	33%	50%	32%
	Muskrat	83%	100%	90%
	Robin	50%	67%	52%
	Fox	67%	50%	65%
	Dog	83%	50%	77%
	Goose	67%	67%	67%
Toronto Rivers	**Human**	100%	100%	100%
	Beaver	100%	100%	100%
	Mallard	100%	100%	100%
	Chicken	100%	67%	88%
	Gull	83%	100%	90%
	Racoon	100%	83%	93%
	Cow	83%	33%	78%
	Muskrat	100%	100%	100%
	Robin	67%	100%	77%
	Fox	33%	50%	32%
	Dog	50%	50%	50%
	Goose	33%	17%	40%
	Human	100%	100%	100%
	Beaver	100%	100%	100%
	Mallard	83%	67%	79%
	Chicken	67%	50%	65%

	Gull	83%	83%	83%
Niagara	**Racoon**	83%	100%	90%
Beaches	**Cow**	50%	50%	50%
	Muskrat	100%	100%	100%
	Robin	83%	83%	83%
	Fox	67%	17%	74%
	Dog	17%	0%	17%*
	Goose	0%	17%	0%
	Human	100%	100%	100%
	Beaver	100%	83%	93%
Niagara	**Mallard**	100%	83%	93%
Beach Sand	**Chicken**	67%	50%	65%
	Gull	83%	100%	90%
	Racoon	67%	83%	71%
	Cow	67%	17%	74%
	Muskrat	100%	83%	93%
	Robin	67%	100%	77%
	Fox	50%	67%	51%
	Dog	17%	0%	17%*
	Goose		Not Detected	

*Only a single sample was positive for dog mt sequences. Conditional probability is normalized according to the probability of detection in response to BAV exceedances to avoid bias associated with the Bayesian theorem.

Digital PCR Detection Frequency of Potential Fecal Contamination Sources and Beach Action Value Exceedances

Four relatively common microbial source tracking markers for human, mammalian and avian fecal pollution sources were tested for the same 48 samples using a digital PCR method to compare with results from eDNA metabarcoding (Figure 2). Human mt (68-100%) and Gull4 (75-100%) markers were detected in the majority of the samples across our study locations, followed by human HF183 (9-82%) and Dog3 (22-42%). The frequency of the Gull4 marker was higher for beach water (100%) than for rivers (73%) and beach sand (82%) samples. Also, the probability of gulls as a potential fecal source was 100% for both Toronto and Niagara beach water (Table 3).

Interestingly, the human HF183 marker was only detected on BAV-exceedance beach days for Toronto beach sites. In contrast, HF183 was detected for both BAV exceedance and non-exceedance beach days for Niagara beach sites. Overall, the HF183 marker was detected more frequently for Niagara beach water (65%) than Toronto beach water (42%), however, the probability of detecting HF183 in association with BAV exceedances was higher for Toronto beach water (100%) than Niagara beach water (55%). Interestingly, the HF183 DNA marker only occurred in 8% of Niagara beach sand samples, in contrast to 100% sand detections of the human mt DNA marker and human eDNA sequences.

Although the overall detection frequency of the Dog3 marker was higher for Niagara beach water (42%) than Toronto beach water (25%), the probability of the Dog3 marker association with BAV exceedances was higher for Sunnyside Beach water (100%) than other beach

waters (Table 3). All the BAV exceedance samples positive for the Dog3 marker in Toronto were from Sunnyside Beach. Interestingly, the association of the Dog3 marker with BAV exceedances was higher for Toronto beach water than its adjacent rivers. While the Canada goose mt DNA marker was never detected by dPCR, Canada goose eDNA sequences were found in 67% of Toronto beach water samples, 25% of Toronto River samples, and 8% of Niagara beach water samples.

The ruminant fecal bacterial Rum2Bac DNA marker was tested on 16 of the cow eDNA +ve samples, and the pig fecal bacterial Pig2Bac DNA marker was tested on the 3 pig eDNA +ve samples. These two DNA markers for host-specific fecal bacteria were not detected in these samples. A comparison of the detection rate of food animal eDNA sequences (chicken, cow, pig) with human HF183 and mt dPCR markers and human eDNA sequences is provided in Supplementary Table 5. Chicken, cow, and pig eDNA sequences were almost always detected in Toronto River samples with the human HF183 DNA marker for sewage contamination. In contrast, these food-animal eDNA sequences were more commonly associated with human eDNA sequences and the human mt DNA marker than human-specific bacterial fecal marker (HF183) for Toronto and Niagara beach water and Niagara beach sand. In particular, chicken and cow eDNA sequences were always detected with human eDNA sequences and the human mt DNA marker in Niagara beach sand, but only rarely with the human HF183 DNA marker.

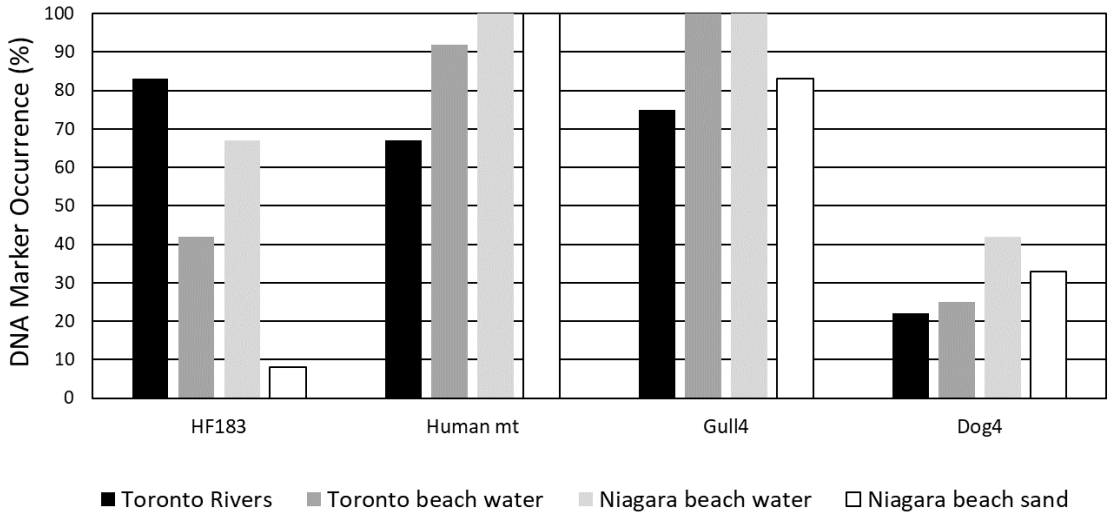


Figure 4. Detection percentage of Human, Gull, and Dog dPCR fecal source tracking DNA markers in beach, river and sand samples.

Table 6. Co-occurrence of microbial source tracking DNA markers from Beach, River, and Sand locations during BAV exceedance and non-exceedance days

Sampling Locations	Potential Fecal Contamination Sources	Probability of Sequence Detection (BAV Exceedance)	Probability of Sequence Detection (BAV Non-exceedance)	Conditional Probability as a Potential Fecal Contamination Source
Toronto Beaches	**HF183**	83%	0%	100%*
	Human Mt	83%	100%	90%
	Gull4	100%	100%	100%
	Dog3	50%	0%	100%*
Toronto Rivers	**HF183**	100%	67%	88%
	Human Mt	75%	60%	71%
	Gull4	83%	67%	79%
	Dog3	25%	20%	26%
Niagara Beaches	**HF183**	50%	83%	55%
	Human Mt	100%	100%	100%
	Gull4	100%	100%	100%
	Dog3	50%	33%	52%
Niagara Beach Sand	**HF183**	17%	0%	100%*
	Human Mt	100%	100%	100%
	Gull4	83%	83%	83%
	Dog3	17%	50%	14%

*Complete absence of sequence detection in BAV non-exceedance days compared to BAV exceedance days results in a 100% conditional probability of an organism as a potential fecal contamination source. We have provided all three probabilities for clarity of our message.

Discussion:

Fecal indicator bacteria, including *E. coli* and *Enterococcus*, commonly used for recreational water quality monitoring, cannot differentiate between fecal sources[33]. Most fecal source tracking studies have applied a microbial source tracking approach focused on a limited number of targets (source-specific microbial taxa). In this study, we applied eDNA metabarcoding and dPCR assays for microbial DNA markers as a more comprehensive approach to assess fecal source profiles for urban freshwater beaches along with their adjacent water sources (rivers/creeks) and beach sands.

Our study found eDNA from fishes, humans, beavers, muskrats, mallard ducks, and gulls was predominant across all our study locations. A study in two rural Quebec watersheds[30] found generally similar results, although that study detected black bear and moose eDNA, and had fewer dog, red fox, raccoon and gull eDNA detections than our urban area. Our results probably reflected the continuous eDNA shedding by specific animals living in close association with urban aquatic settings and the large human populations in our study area. Our study locations are large urban centers (Toronto and St Catherines), and they are all in proximity to municipal wastewater treatment plants. The relatively high numbers of human-associated sequences in all our samples were probably associated with discharges from nearby municipal wastewater treatment plants, combined sewer overflows, and sewage cross-connected stormwater outfalls. The regular occurrence of sewage contamination at our Toronto study locations has been well documented[34,35,36], and the results are in agreement

with previous studies finding higher levels of human mitochondrial DNA associated with sewage contamination[11,37,38,39,40].

A challenge with mitochondrial DNA is that it may not be exclusively derived from fecal sources. For example, in recreational water settings, the potential for bather shedding of human skin cells, saliva and hair, in addition to fecal matter, is well-established[12]. Human mitochondrial DNA has even been proposed as an indicator for detecting greywater from residential sources like showers or sinks that do not involve toilet flush water[41]. Human eDNA can be readily detected as a genetic by-catch in water, sand, and air environmental samples, even in areas with relatively low human habitation densities[42], which raises caution for interpreting human eDNA results from water samples as automatically inferring sewage contamination in some areas. In our urban sewage-impacted study areas, the high prevalence of human eDNA sequences in all water and sand samples precluded our ability to discriminate between sites that were differentially impacted by human fecal contamination. The microbial source tracking HF183 marker was a better indicator of human fecal contamination hotspots and contributions to BAV exceedances than the human mt DNA marker or human eDNA sequences.

An aspect requiring further examination is whether microbial source tracking DNA markers from anaerobic bacteria, and eukaryotic eDNA from different tissue compartments (e.g. feces, skin, hair, feathers, rotting carcass) can decay at significantly different rates and have different persistence and transport characteristics. Human mt DNA from sewage has been found to have decay rates similar to those of the microbial HF183 marker in freshwater

mesocosms[43,44]. However, it has been suggested that if the mitochondrial (mt) membrane remains intact, this could allow mt DNA to persist longer in the environment[45]. Some studies examining decay rates of eDNA in the environment have identified a slower decay rate of mtDNA than nuclear DNA[46,47,48], suggesting that mitochondrial eDNA markers may persist longer in the environment compared to bacterial or eukaryotic nuclear DNA markers. It is possible that a longer persistence and accumulation in beach sand may have contributed to human eDNA sequences and the human mitochondrial DNA marker occurring in 100% of our beach sand samples compared to only 8% for the microbial HF183 marker.

Similar to studies in Ontario and Quebec watersheds[12,30], our eDNA metabarcoding results detected a diverse range of animals likely impacting water quality. This was particularly the case for beavers and muskrats which may not be typically associated with urban recreational water settings. Detection of beavers and muskrats as a potential fecal contamination source for recreational waters could be very important for beach monitoring strategies as these animals are known to harbour protist pathogens like *Giardia* and *Cryptosporidium*, whose occurrence may be poorly correlated to conventional fecal indicator bacteria like *E. coli*[49,50]. An Ontario agricultural watershed study found muskrats to commonly occur at some sites, and they were also associated with an increased likelihood of occurrence of *Campylobacter* species in water samples[51]. While microbial source tracking *Bacteroides* DNA markers have been developed for beavers[52] and muskrats[53], they have not been well tested, particularly in urban settings.

Other urban mammal eDNA sequences commonly detected in our samples were for raccoons, red foxes, and dogs. Raccoons occur widely in our study area, and their feces can significantly contribute to fecal pollution in stormwater systems[54] and present *Salmonella* and other pathogen health risks[55]. Similar to our results, dogs have also been identified as locally important sources of fecal pollution at other beaches[56], which may guide the need for beach-specific remedial actions. Many of these urban wildlife species occur at the animal-human one Health interface and have the potential to contribute to the transmission of zoonotic pathogens or antimicrobial resistance to people in recreational water settings.

eDNA metabarcoding results also detected diverse potential avian fecal contamination sources, particularly Mallard ducks, gulls, and robins. Both eDNA sequencing and the Gull4 microbial source tracking DNA marker indicated the prominence of gull fecal contamination associated with beaches. The significance of gull fecal contamination at our sites is in agreement with previous studies conducted at Sunnyside Beach[34,57] and in riverine and urban coastal areas of Southern Ontario[58]. Another study on Lake Michigan beaches identified gulls as the predominant fecal contamination source, and intervention strategies to reduce gull occurrence significantly improved beach water quality[59].

Our findings also commonly detected Canada goose eDNA sequences at Toronto beaches, indicating their local importance, even though dPCR did not detect the Canada goose mitochondrial DNA marker. This was consistent with our Toronto River results, where human eDNA sequences were always prevalent, but dPCR less frequently detected the human mitochondrial DNA marker. More common detection of human and Canada goose

eDNA sequences may reflect the ability of eDNA sequence data to detect multiple DNA sequence variants (ESVs) for many fecal contamination sources[60]. The ability to detect a greater variability of DNA sequence variants may provide an advantage of eDNA metabarcoding over the specificity of PCR methods for detecting DNA markers.

An unexpected result from our eDNA metabarcoding was the common detection of chicken and cow (and some pig) sequences in all our urban study areas. These sequences could have been transported from farms, food processing facilities, or residential food waste further away from our study sites. However, it is unlikely such sources would have been so widespread to commonly impact all our study sites at two rivers and four different beaches in two different cities and for chicken eDNA sequences to be significantly associated with BAV exceedances. Chicken, cow, and pig eDNA sequences were always detected with human eDNA sequences, often with the human HF183 microbial source tracking marker, and in proximity to wastewater treatment plants and other known sewage sources in our study area. Potential concerns about human food DNA sequences carrying over into human feces have been previously identified[10,12,30,44,61]. A previous study on one of our Toronto watersheds[12] used a broader CO1-based eDNA metabarcoding approach and raised a similar concern based on detecting eDNA sequences for non-native food fish species like Tilapia (*Oreochromis* sp.) and Sea bass (*Serranidae* sp.) in water samples. The passage of food DNA through animal digestive tracts is increasingly investigated by metabarcoding diet analyses of wildlife scats, and a study[62] recently used eDNA metabarcoding to demonstrate the high occurrence of domestic dog DNA in red fox scats that was attributed to significant

coprophagia. In some situations, such as sewage-impacted aquatic ecosystems, it will be important to apply microbial source tracking DNA markers for common human food sources, including cattle, pig and chicken, alongside eDNA metabarcoding, to test for the possibility of eDNA sequences from undigested food waste in human feces.

eDNA metabarcoding identified a broader range of animal species likely impacting water quality in our study area than the more limited range of microbial source-tracking DNA markers available. However, we also identified limitations of eDNA metabarcoding that require further investigation. eDNA metabarcoding studies of fecal pollution should use microbial source tracking DNA markers in some situations as additional lines of evidence for fecal source attribution. Human eDNA PCR blocking techniques may prove useful for sewage-contaminated urban settings to better associate the occurrence of animal eDNA sequences with fecal contamination. Using fecal source tracking information from both eDNA metabarcoding and microbial source tracking DNA markers, combined with conventional fecal indicator bacteria, can help design targeted beach sustainability programs and risk management actions to control fecal contamination.

Conclusions:

1. eDNA metabarcoding of mammal and avian 16S rRNA gene sequences found human eDNA sequences to predominate in all urban beach water, sand, and river samples. PCR blocking for human eDNA sequences should be considered to avoid masking the detection of fecal contamination from less common animals in sewage-impacted areas.

2. Beaver, muskrat, gull, and mallard eDNA sequences were the most prominent animals detected across our study sites, indicating the potential for these wildlife species to contribute to fecal contamination and health risks at urban freshwater beaches.

3. Dogs and Canada goose eDNA sequences were most common at Sunnyside Beach, suggesting that these fecal contamination sources can be beach-specific and may require targeted remediation strategies.

4. Digital PCR assays for the bacterial Gull4 DNA marker and a human mitochondrial DNA marker were consistent with eDNA metabarcoding in detecting the widespread occurrence of these potential fecal sources.

5. Chicken and cow eDNA sequences were widely detected across all study sites but are suggested to be from sewage and incompletely digested human food origins.

6. eDNA metabarcoding can expand fecal source tracking capabilities for assessing diverse wildlife contributions to fecal pollution. However, microbial fecal source tracking markers for common food animals (e.g. chicken, cattle, and pigs) should be tested alongside eDNA metabarcoding analysis to address questions about the significance of coprophagy, non-fecal skin sources or occurrence of eDNA sequences from incompletely digested food in fecal pollution sources.

References:

1. Nevers, M.B., Byappanahalli, M.N., Shively, D., Buszka, P.M., Jackson, P.R. and Phanikumar, M.S., 2018. Identifying and eliminating sources of recreational water quality degradation along an urban coast. Journal of environmental quality, 47(5), pp.1042-1050.
2. Field, K.G. and Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water research, 41(16), pp.3517-3538.
3. Stachler, E., Akyon, B., de Carvalho, N.A., Ference, C. and Bibby, K., 2018. Correlation of crAssphage qPCR markers with culturable and molecular indicators of human fecal pollution in an impacted urban watershed. Environmental science & technology, 52(13), pp.7505-7512.
4. Saeidi, N., Gu, X., Tran, N.H., Goh, S.G., Kitajima, M., Kushmaro, A., Schmitz, B.W. and Gin, K.Y.H., 2018. Occurrence of traditional and alternative fecal indicators in tropical urban environments under different land use patterns. Applied and Environmental Microbiology, 84(14), pp.e00287-18.
5. Boehm, A.B., Van De Werfhorst, L.C., Griffith, J.F., Holden, P.A., Jay, J.A., Shanks, O.C., Wang, D. and Weisberg, S.B., 2013. Performance of forty-one microbial source tracking methods: a twenty-seven lab evaluation study. Water research, 47(18), pp.6812-6828.
6. Jardé, E., Jeanneau, L., Harrault, L., Quenot, E., Solecki, O., Petitjean, P., Lozach, S., Chevé, J. and Gourmelon, M., 2018. Application of a microbial source tracking

based on bacterial and chemical markers in headwater and coastal catchments. Science of The Total Environment, 610, pp.55-63.

7. Demeter, K., R. Linke, E. Ballesté, G. Reischer, R.E. Mayer, J. Vierheilig, C. Kolm, M.E. Stevenson, J. Derx, A.K.T. Kirsc hner, R. Sommer, O.C. Shanks, A.R. Blanch, J.B. Rose, W. Ahmed, and A.H. Farnleitner. 2023. Have genetic targets for faecal pollution diagnostics and source tracking revolutionized water quality analysis yet? FEMS Microbiology Reviews , 2023, 47 , 1–36. DOI: 10.1093/femsre/fuad028
8. Mayer, R.E., Reischer, G.H., Ixenmaier, S.K., Derx, J., Blaschke, A.P., Ebdon, J.E., Linke, R., Egle, L., Ahmed, W., Blanch, A.R. and Byamukama, D., 2018. Global distribution of human-associated fecal genetic markers in reference samples from six continents. Environmental science & technology, 52(9), pp.5076-5084.
9. Martellini, A., P. Payment, and R. Villemur. 2005. Use of eukaryotic mitochondrial DNA to differentiate human, bovine, porcine, and ovine sources in fecally contaminated surface water. Water Research 39: 541-548.
10. Caldwell, J.M., Raley, M.E. and Levine, J.F., 2007. Mitochondrial multiplex real-time PCR as a source tracking method in fecal-contaminated effluents. Environmental science & technology, 41(9), pp.3277-3283.
11. Edge, T.A., Boyd, R.J., Shum, P. and Thomas, J.L., 2021. Microbial source tracking to identify fecal sources contaminating the Toronto Harbour and Don River watershed in wet and dry weather. Journal of Great Lakes Research, 47(2), pp.366-377.

12. Staley, Z.R., Chuong, J.D., Hill, S.J., Grabuski, J., Shokralla, S., Hajibabaei, M. and Edge, T.A., 2018. Fecal source tracking and eDNA profiling in an urban creek following an extreme rain event. Scientific reports, 8(1), p.14390.
13. Ragot, R. and Villemur, R., 2022. eDNA profiling of mammals, birds, and fish of surface waters by mitochondrial metagenomics: application for source tracking of fecal contamination in surface waters. Environmental Monitoring and Assessment, 194, pp.1-13.
14. Saleem, F., Schellhorn, H.E., Simhon, A. and Edge, T.A., 2023a. Same-day *Enterococcus* qPCR results of recreational water quality at two Toronto beaches provide added public health protection and reduced beach days lost. Canadian Journal of Public Health, pp.1-12.
15. Saleem, F., Edge, T.A. and Schellhorn, H.E., 2023b. Application of Same-Day *Enterococcus* qPCR-Based Analyses for Quality Assessment of Shorelines (Water and Sand) at Recreational Beaches. Water, 15(13), p.2338.
16. Green, H.C., Haugland, R.A., Varma, M., Millen, H.T., Borchardt, M.A., Field, K.G., Walters, W.A., Knight, R., Sivaganesan, M., Kelty, C.A. and Shanks, O.C., 2014. Improved HF183 quantitative real-time PCR assay for characterization of human fecal pollution in ambient surface water samples. Applied and environmental microbiology, 80(10), pp.3086-3094.
17. Ryu, H., Griffith, J.F., Khan, I.U., Hill, S., Edge, T.A., Toledo-Hernandez, C., Gonzalez-Nieves, J. and Santo Domingo, J., 2012. Comparison of Gull feces-specific

assays targeting the 16S rRNA genes of Catellicoccus marimammalium and Streptococcus spp. Applied and Environmental Microbiology, 78(6), pp.1909-1916.

18. Caldwell, J.M. and Levine, J.F., 2009. Domestic wastewater influent profiling using mitochondrial real-time PCR for source tracking animal contamination. Journal of microbiological methods, 77(1), pp.17-22.
19. Green, H.C., White, K.M., Kelty, C.A. and Shanks, O.C., 2014. Development of rapid canine fecal source identification PCR-based assays. Environmental science & technology, 48(19), pp.11453-11461.
20. Mieszkin, S., Yala, J.F., Joubrel, R. and Gourmelon, M., 2010. Phylogenetic analysis of Bacteroidales 16S rRNA gene sequences from human and animal effluents and assessment of ruminant faecal pollution by real-time PCR. Journal of applied microbiology, 108(3), pp.974-984.
21. Mieszkin, S., Furet, J.P., Corthier, G. and Gourmelon, M., 2009. Estimation of pig fecal contamination in a river catchment by real-time PCR using two pig-specific Bacteroidales 16S rRNA genetic markers. Applied and environmental microbiology, 75(10), pp.3045-3054.
22. Brown, J., Pirrung, M. and McCue, L.A., 2017. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics, 33(19), pp.3137-3139.
23. Chen, S., Zhou, Y., Chen, Y. and Gu, J., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), pp.i884-i890.

24. Buchner, D., Macher, T.H. and Leese, F., 2022. APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. Bioinformatics, 38(20), pp.4817-4819.
25. Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal, 17(1), pp.10-12.
26. Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, p.e2584.
27. Frøslev, T.G., Kjøller, R., Bruun, H.H., Ejrnæs, R., Brunbjerg, A.K., Pietroni, C. and Hansen, A.J., 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature communications, 8(1), p.1188.
28. Drake, L.E., Cuff, J.P., Young, R.E., Marchbank, A., Chadwick, E.A. and Symondson, W.O., 2022. An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. Methods in Ecology and Evolution, 13(3), pp.694-710.
29. Waso, M., Khan, S., Khan, W., 2018a. Development and small-scale validation of a novel pigeon-associated mitochondrial DNA source tracking marker for the detection of fecal contamination in harvested rainwater. Sci. Total Environ. 615, 99–106.
30. Ragot, R., Lessard, F., Bélanger, A. and Villemur, R., 2023. Assessment of multiple fecal contamination sources in surface waters using environmental mitochondrial DNA metabarcoding. Science of The Total Environment, 898, p.165237.
31. Kassambara, A., 2020. Ggpubr: "ggplot2" based publication ready plots. R package version 0.4. 0, 438.

32. Andresa, K.J., D.M. Lodge, and J.Andrésa. 2023. Environmental DNA reveals the genetic diversity and population structure of an invasive species in the Laurentian Great Lakes. PNAS 2023 Vol. 120 No. 37 e2307345120. https://doi.org/10.1073/pnas.2307345120.
33. Tran, N.H., Gin, K.Y.H. and Ngo, H.H., 2015. Fecal pollution source tracking toolbox for identification, evaluation and characterization of fecal contamination in receiving urban surface waters and groundwater. Science of the Total Environment, 538, pp.38-57.
34. Staley, Z.R. and Edge, T.A., 2016. Comparative microbial source tracking methods for identification of fecal contamination sources at Sunnyside Beach in the Toronto region area of concern. Journal of Water and Health, 14(5), pp.839-850.
35. Staley, Z., J. Grabuski, E. Sverko, and T.A. Edge. 2016. Comparison of microbial and chemical source tracking markers to identify fecal contamination sources in the Humber River and associated stormwater outfalls. Applied and Environmental Microbiology 82: 6357-6366
36. Li, E., F. Saleem, T.A. Edge, and H.E. Schellhorn. 2024. Assessment of crAssphage as a human fecal source tracking marker in the lower Great Lakes. Science of the Total Environment 912 (2024) 168840.
37. Kapoor, V., Smith, C., Santo Domingo, J.W., Lu, T. and Wendell, D., 2013. Correlative assessment of fecal indicators using human mitochondrial DNA as a direct marker. Environmental science & technology, 47(18), pp.10485-10493.

38. Villemur, R., Imbeau, M., Vuong, M.N., Masson, L. and Payment, P., 2015. An environmental survey of surface waters using mitochondrial DNA from human, bovine and porcine origin as fecal source tracking markers. Water Research, 69, pp.143-153.
39. Hajj-Mohamad, M., M. Hachad, G. Deschamps, S. Sauvé, R. Villemur, M-A. Blais, M. Prévost, and S. Dorner. 2019. Fecal contamination of storm sewers: Evaluating wastewater micropollutants, human-specific Bacteroides 16S rRNA, and mitochondrial DNA genetic markers as alternative indicators of sewer cross connections. Science of the Total Environment 659 (2019) 548–560.
40. Tanvir Pasha, A.B.M., Hinojosa, J., Phan, D., Lopez, A. and Kapoor, V., 2020. Detection of human fecal pollution in environmental waters using human mitochondrial DNA and correlation with general and human-associated fecal genetic markers. Journal of water and health, 18(1), pp.8-18.
41. Zimmerman, B.D., Ashbolt, N.J., Garland, J.L., Keely, S. and Wendell, D., 2014. Human mitochondrial DNA and endogenous bacterial surrogates for risk assessment of graywater reuse. Environmental science & technology, 48(14), pp.7993-8002.
42. Whitmore, L., McCauley, M., Farrell, J.A., Stammnitz, M.R., Koda, S.A., Mashkour, N., Summers, V., Osborne, T., Whilde, J. and Duffy, D.J., 2023. Inadvertent human genomic bycatch and intentional capture raise beneficial applications and ethical concerns with environmental DNA. Nature Ecology & Evolution, 7(6), pp.873-888.
43. Ragot, R. and Villemur, R., 2022. Influence of temperature and water quality on the persistence of human mitochondrial DNA, human Hf183 Bacteroidales, fecal

coliforms, and enterococci in surface water in human fecal source tracking context. Science of The Total Environment, 838, p.156025.

44. Ren, W. and Feng, Y., 2023. Persistence of human-and cattle-associated Bacteroidales and mitochondrial DNA markers in freshwater mesocosms. Science of The Total Environment, 899, p.165742.
45. Harrison, J.B., Sunday, J.M. and Rogers, S.M., 2019. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B, 286(1915), p.20191409.
46. Foran, D.R., 2006. Relative degradation of nuclear and mitochondrial DNA: an experimental approach. Journal of forensic sciences, 51(4), pp.766-770.
47. Emmons, A.L., DeBruyn, J.M., Mundorff, A.Z., Cobaugh, K.L. and Cabana, G.S., 2017. The persistence of human DNA in soil following surface decomposition. Science & Justice, 57(5), pp.341-348.
48. Bylemans, J., Furlan, E.M., Gleeson, D.M., Hardy, C.M. and Duncan, R.P., 2018. Does size matter? An experimental evaluation of the relative abundance and decay rates of aquatic environmental DNA. Environmental Science & Technology, 52(11), pp.6408-6416.
49. Tsui,C.K.-M., R. Miller, M. Uyaguari-Diaz, P. Tang, C. Chauve, W. Hsiao, J. Isaac-Renton, and N. Prystajecky. 2018. Beaver Fever: Whole-Genome Characterization of Waterborne Outbreak and Sporadic Isolates To Study the Zoonotic Transmission of Giardiasis. mSphere 3:e00090-18.

50. Bitto, A., Aldras, A., 2009. Prevalence of Giardia and Cryptosporidium in muskrats in northeastern Pennsylvania and New Jersey. J. Environ. Health 71, 20–26.
51. Marti, R., Gannon, V.P., Jokinen, C., Lanthier, M., Lapen, D.R., Neumann, N.F., Ruecker, N.J., Scott, A., Wilkes, G., Zhang, Y. and Topp, E., 2013a. Quantitative multi-year elucidation of fecal sources of waterborne pathogen contamination in the South Nation River basin using Bacteroidales microbial source tracking markers. Water research, 47(7), pp.2315-2324.
52. Marti, R., Zhang, Y., Tien, Y.C., Lapen, D.R. and Topp, E., 2013b. Assessment of a new Bacteroidales marker targeting North American beaver (Castor canadensis) fecal pollution by real-time PCR. Journal of microbiological methods, 95(2), pp.201-206.
53. Marti, R., Y. Zhang, D.R. Lapen, and E. Topp. 2011. Development and validation of a microbial source tracking marker for the detection of fecal pollution by muskrats. Journal of Microbiological Methods 87 (1), 82e88.
54. Ram, J.L., B. Thompson, C. Turner, J.M. Nechvatal, H. Sheehan, and J. Bobrin. 2007. Identification of pets and raccoons as sources of bacterial contamination of urban storm sewers using a sequence-based bacterial source tracking method. Water Research 41(16):3605-14.
55. Bondo, K.J., D.L. Pearl, N. Janecko, P. Boerlin, R.J. Reid-Smith, J. Parmley, and C.M. Jardine. 2016. Impact of Season, Demographic and Environmental Factors on Salmonella Occurrence in Raccoons (Procyon lotor) from Swine Farms and Conservation Areas in Southern Ontario. PLoS ONE 11(9): e0161497.

56. Shrestha, A., Kelty, C.A., Sivaganesan, M., Shanks, O.C. and Dorevitch, S., 2020. Fecal pollution source characterization at non-point source impacted beaches under dry and wet weather conditions. Water Research, 182, p.116014.

57. Edge, T.A., Hill, S., Seto, P., and Marsalek, J., 2010. Library-dependent and library-independent microbial source tracking to identify spatial variation in faecal contamination sources along a Lake Ontario beach (Ontario, Canada). Water Science & Technology 62.3: 719-727.

58. Lu, J., H. Ryu, S. Hill, M. Schoen, N. Ashbolt, T.A. Edge, and J. Santo Domingo. 2011. Distribution and potential significance of a gull fecal marker in urban coastal and riverine areas of southern Ontario, Canada. Water Research 45: 3960-3968.

59. Converse, R.R., Kinzelman, J.L., Sams, E.A., Hudgens, E., Dufour, A.P., Ryu, H., Santo-Domingo, J.W., Kelty, C.A., Shanks, O.C., Siefring, S.D. and Haugland, R.A., 2012. Dramatic improvements in beach water quality following gull removal. Environmental science & technology, 46(18), pp.10206-10213.

60. Skelton, J., Cauvin, A. and Hunter, M.E., 2023. Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. Environmental DNA, 5(5), pp.1092-1104.

61. Kortbaoui, R., Locas, A., Imbeau, M., Payment, P. and Villemur, R., 2009. Universal mitochondrial PCR combined with species-specific dot-blot assay as a source-tracking method of human, bovine, Chicken, ovine, and porcine in fecal-contaminated surface water. Water research, 43(7), pp.2002-2010.

62. Waggershauser, C.N., Taberlet, P., Coissac, E., Kortland, K., Hambly, C. and Lambin, X., 2022. Interspecific coprophagia by wild red foxes: DNA metabarcoding reveals a potentially widespread form of commensalism among animals. Ecology and Evolution, 12(7), p.e9029.

Chapter 8. Concluding Remarks

As human activities and recreational use of freshwater beaches increase, the need for accurate and timely water quality monitoring has become paramount. Despite significant advancements in contemporary technologies such as PCR, DNA sequencing, electrochemical sensors, and DNA capture devices, integrating these innovations into public health responses is yet to be explored in detail. Over time, our collaborations with both public and private authorities, notably the Ministry of the Environment, Conservation and Parks (MECP) and Health Canada, have focused on advocating the viability of molecular water monitoring techniques as viable alternatives to traditional approaches (Saleem et al., 2024; Saleem et al., 2024b; Saleem et al., 2023; Saleem et al., 2023b; Saleem et al., 2022). Within the MacWater consortium (www.macwater.org), our efforts in the Great Lakes domain aim to encourage the adoption of advanced water quality monitoring technologies. This thesis is a testament to our ongoing commitment to validating and developing new molecular methodologies for enhancing the robustness and comprehensiveness of water monitoring strategies. The discussion below addresses the principal inquiries that instigated this research, illustrating how each segment of the thesis provides a conclusive rationale for adopting modern technologies to enhance routine monitoring of freshwater beaches.

6.1. Rapid Environmental Monitoring Strategies

Rapid environmental monitoring signifies a crucial advancement in environmental health science, balancing quick intervention with the demand for accuracy, highlighting a

commitment to proactive environmental management, and integrating innovative science into public health practices. By combining speed and precision, these methods can enhance our ability to ensure safer interactions between communities and natural water sources. As the first goal of my thesis, I wanted to explore a rapid alternative to conventional bacterial culturing, and we partnered with Public Health Authorities, including Toronto Public Health and Niagara Public Health, to explore the idea of rapid molecular methods for environmental monitoring. The qPCR-based strategy demonstrated the potential for timely and effective decision-making by the beach managers. Besides reducing the time required to detect harmful microbial contaminants, the rapid environmental assessment enables more responsive and informed decisions. After communicating our results in the form of manuscripts (Saleem et al., 2022; Saleem et al., 2023; Saleem et al., 2023b), public health and research authorities, including Kingston Public Health and New Brunswick Research and Productivity Council (RPC), appreciated our efforts as they tried to validate the qPCR-based strategy for their jurisdictions, which shows a real-life impact and practical application this work had on the decision-makers. Considering the economic and public-health-related consequences associated with erroneous or delayed environmental decision-making, implementing rapid technologies to augment routine quality monitoring programs is a step in the right direction.

6.2. Conserved Molecular Markers for Enhanced Accuracy in Environmental Monitoring

The exploration of conserved signature genes and proteins challenges traditional paradigms by emphasizing the specificity and reliability of conserved genetic elements over the broader, less discriminating universal markers. When used as a diagnostic tool, these highly conserved sequences can offer greater precision in detecting and monitoring microbial presence, facilitating more accurate and targeted environmental and clinical applications. By leveraging the inherent stability and specificity of conserved genes and proteins, we can achieve higher accuracy and efficacy in fields from microbial ecology to pathogen detection, ultimately advancing diagnostic applications. Through our proof of principle study on *E. coli*, we demonstrated that the applications of conserved molecular markers, in a broader sense, can also be used for other pathogens, including protists, viruses and Cyanobacteria, for rapid, more precise and robust diagnostic strategies.

6.3. Moving Beyond the Concept of Fecal Indicators

The narrow focus on fecal indicator bacteria neglects the complex and diverse microbial ecosystems that inhabit the environment, thereby overlooking uncommon microbial risk factors. Adopting more advanced and comprehensive technologies, including metagenomics for microbial assessment, allows for the detection of a broader spectrum of pathogens and harmful microorganisms, offering a more complete and accurate evaluation of potential health risks associated with changes in recreational water quality. For example, cyanobacterial blooms, cyanotoxins, and antibiotic-resistance genes are some microbial risk

factors that may not be accounted for if conventional fecal indicators are relied on. With the increase in anthropogenic activities, it has become necessary to evolve beyond traditional quality concepts for in-depth information to foster a more resilient and informed strategy for environmental health management.

6.4. Advancing Beyond Routine Quality Testing to Identify Primary Sources of Biological Environmental Contaminants

Fecal Source Tracking moves beyond the simplistic detection of fecal contamination, aiming instead to uncover the precise origins of such pollutants—whether they originate from human waste, agricultural runoffs, or wildlife. By identifying the root causes of environmental quality deterioration, fecal source tracking facilitates more nuanced and targeted interventions, enhancing our understanding of the underlying mechanisms driving changes in recreational water quality. However, fecal source tracking markers are only available for some mammalian and avian species and do not cover a broad spectrum of wildlife that can contribute to fecal contamination, failing to encompass the broader spectrum of wildlife that may contribute to fecal contamination. To address this limitation, we investigated the application of DNA sequencing information, which offers a comprehensive profile of potential fecal contamination sources. Integrating DNA sequencing into routine monitoring signifies a paradigm shift towards more precise, efficient, and informed management/sustainability of recreational water resources.

7. Future Directions

7.1. Advancements in DNA-Based Diagnostic Technologies

The rapid advancement of DNA-based diagnostic technologies has revolutionized the field of water quality monitoring. Our future goals include designing and developing a DNA capture device capable of simultaneously detecting and quantifying a diverse array of pathogens, including viruses, bacteria, and protists from environmental and biological samples. Additionally, we would like to explore the potential of long-read DNA sequencing and portable DNA sequencing technologies, including Oxford Nanopore, for on-site real-time environmental diagnostic applications.

We have partnered with environmental protection agencies, including the Ontario Ministry of Environment Conservation and Parks (MECP), Environment Canada, and local public health authorities, to promote and develop newer molecular methods. Moving forward, through our association with MacWater Diagnostics and our public and private partners, we aim to validate and establish newer diagnostic technologies, ensuring their real-life applications for communities and citizen scientists.

7.2. Practical Applications and Community Engagement

We emphasize the importance of engaging communities and citizen scientists in water quality monitoring to translate these advancements into practical applications. By validating and promoting advanced molecular diagnostic technologies, we aim to empower local communities to actively participate in maintaining the sustainability of recreational waters.

This community-driven approach enhances the effectiveness of monitoring programs and fosters a culture of environmental stewardship.

In summary, my work's future directions include the development and integration of advanced molecular strategies, collaboration with environmental authorities, and community engagement. These efforts collectively aim to ensure the quality and sustainability of recreational waters, addressing both current and emerging challenges in water quality monitoring.

Appendix

Cyanobacterial Algal Bloom Monitoring: Molecular Methods and Technologies for Freshwater Ecosystems

This chapter provides a comprehensive understanding of the state-of-the-art molecular tools available for monitoring Harmful Algal blooms in recreational waters, which is essential for formulating effective strategies for sustainability strategies. The integration of this review thereby strengthens the thesis's objective of developing innovative and effective molecular strategies for the quality monitoring and sustainability of recreational waters, ultimately contributing to the advancement of environmental monitoring and protection.

*The text I present here is a **peer-reviewed review article published** in the Microorganisms journal. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Jiang, J.L., Atrache, R., Paschos, A., Edge, T.A. and Schellhorn, H.E., 2023. Cyanobacterial algal bloom monitoring: molecular methods and technologies for freshwater ecosystems. Microorganisms, 11(4), p.851.*

Author Contributions

FS and JLJ performed the literature review, wrote the original manuscript, and revised it as required by the reviewers. RA and AP helped with the revisions. HES and TAE supervised and reviewed the manuscript for publication.

Abstract

Cyanobacteria (blue-green algae) can accumulate to form harmful algal blooms (HABs) on the surface of freshwater ecosystems under eutrophic conditions. Extensive HAB events can threaten local wildlife, public health, and the utilization of recreational waters. For the detection/quantification of cyanobacteria and cyanotoxins, both the United States Environmental Protection Agency (USEPA) and Health Canada increasingly indicate molecular methods can be useful. However, each molecular detection method has specific advantages and limitations for monitoring HABs in recreational water ecosystems. Rapidly developing modern technologies, including satellite imaging, biosensors, and machine learning/Artificial intelligence can be integrated with standard/conventional methods to overcome the limitations associated with traditional cyanobacterial detection methodology. We examine advances in cyanobacterial cell lysis methodology and conventional/modern molecular detection methods, including imaging techniques, polymerase chain reaction (PCR)/DNA sequencing, enzyme-linked immunosorbent assays (ELISA), mass spectrometry, remote sensing, and machine learning/AI-based prediction models. This review focuses specifically on methodologies likely to be employed for recreational water ecosystems, especially in the Great Lakes region of North America.

Keywords: cyanobacteria, harmful algal blooms, Great Lakes, cyanotoxins, microcystin, cyanobacteria lysis, molecular methods.

1.3.7. Introduction

Cyanobacteria (blue-green algae) are a diverse group of bacteria that, in comparison to other bacterial communities, can uniquely perform photosynthesis and modulate the environmental oxygen content [1,2]. Prolific growth under eutrophic conditions leads to the accumulation of cyanobacterial biomass and the formation of algal blooms in freshwater ecosystems [3]. Freshwater algal blooms mainly comprise one or more of *Aphanizomenon, Cylindrospermopsis, Dolichospermum, Microcystis, Nodularia, Planktothrix,* and *Trichodesmium,* which have regulatory impacts on the ecological processes of aquatic ecosystems [3–5]. For example, cyanobacteria species that uniquely utilize carbon dioxide and nitrogen-dependent metabolism are essential nitrogen-fixing organisms under anaerobic conditions [6,7]. However, bloom-infested freshwater lakes may harbour cyanotoxin-producing cyanobacterial species, including *Microcystis, Dolichospermum* (formerly *Anabaena*)*, Raphidiopsis* (formerly *Cylindrospermopsis*)*, and Planktothrix*, which can impair water quality [8–10].

Favourable environmental factors, including high temperature and availability of micronutrients, allow the cyanobacterial blooms to propagate, leading to hypoxic/anoxic conditions as blooms decay and consume oxygen [11,12]. Aside from the over-population/propagation, cyanobacterial blooms can render freshwater ecosystems unsuitable for drinking water and recreational uses by producing metabolites and metabolic byproducts with unpleasant olfactory and gustatory properties [13–16]. The predominant density of cyano-toxigenic species in water bodies leads to the formation of harmful algal blooms (HABs). *Microcystins*, *cylindrospermopsins*, anatoxins, and saxitoxins are the most common

cyanotoxins produced by harmful algal blooms, and doses as low as parts per billion can induce acute toxicity [17,18]. Toxigenic effects of HAB cyanotoxins can manifest directly through ingestion or indirectly through consuming contaminated food products, including fish, molluscs, and agricultural produce [19,20]. Therefore, raw freshwater sources, including lakes and rivers, are monitored seasonally for their cyanotoxin potential to avoid health or economical losses.

The North American Great Lakes, the largest freshwater system in the world, are important for drinking water, recreation, industry and agriculture. Due to shallow depths and temperate weather, these lakes, especially Lake Erie, face seasonal large-scale harmful algal bloom events [21]. Although Lake Erie experiences the most extensive bloom events among the Great Lakes, cyanoHABs and cyanotoxins are now also across all the Great Lakes [22–25]. Despite the geographical and economic significance of the great lakes, the dynamics of cyanobacterial bloom formation are poorly understood.

The water quality of the Great Lakes is monitored by both the United States Environmental Protection Agency (USEPA) and Environment and Climate Change Canada, as well as other Federal, State, and Provincial programs [26]. For cyanobacterial monitoring, USEPA and Health Canada identify commonly used methods based on microscopy, enzyme-linked immunosorbent assays (ELISA), protein phosphatase inhibition assays (PPIA), and high-performance liquid chromatography coupled with either mass spectrometry (LCMS) or ultraviolet/photodiode array detectors for cyanotoxin detection and quantification [27,28]. For microcystin detection in recreational waters, the USEPA recommends LCMS- and ELISA-based methods [29–31]. LCMS- and ELISA- based methods are highly sensitive to

microcystin detection, but ELISA-based methods cannot distinguish microcystin congeners [29–31]. The USEPA and Health Canada also indicating molecular biology methodology, including conventional polymerase chain reaction (PCR), quantitative real-time PCR (qPCR), and microarrays/DNA chips that can be useful for cyanobacteria monitoring [27,28]. However, standardized molecular biology protocols have yet to be established for harmful algal blooms or cyanotoxin detection. Alternative high-throughput techniques for assessing freshwater ecosystems, including atomic force microscopy (AFM) and biomonitoring, are currently available [32,33]. AFM can visualize microcystin strand formation in the presence of metal ions, and can also examine the nanomechanical/adhesion properties of algal cells [34,35]. Biomonitoring can determine the nutrient and metal composition in water samples, which directly impacts algal populations [36]. This review focuses on the applications, advantages and limitations associated with the molecular biological techniques for harmful algal bloom monitoring in recreational water ecosystems, including the Great Lakes.

This review will evaluate new molecular methods for harmful algal bloom monitoring at the Great Lakes ecosystems. Objectives of this comparative review are: 1) assessing cell lysis methods for extracting biomolecules from bloom-infested water samples, 2) evaluating molecular methods for cyanobacteria and cyanotoxin detection, and 3) addressing benefits and limitations associated with the molecular methods.

1.3.8. Methods for the Disruption and Lysis of Algal Bloom Cyanobacteria

Cyanobacterial density can be monitored by direct (cell mass) or indirect (intracellular biomolecules) methods [37]. However, USEPA guidelines for monitoring cyanotoxin require data for the total cyanotoxin concentration, including both extracellular and intracellular concentrations [27]. Complete lysis of cyanobacterial cells is necessary to obtain an accurate estimate of intracellular cyanotoxin concentrations. Unlike the cell walls of other bacteria, cyanobacterial cell walls have a much thicker and highly crosslinked peptidoglycan layer, presenting a challenging obstacle for molecular detection methods that require cell lysis [38–40]. This section evaluates the mechanisms, limitations, and benefits of lysis methods for cyanobacterial cells.

1.3.8.1. Chemical Cell Lysis

Chemical lysis employs detergents, enzymes, or organic solvents to disrupt the cell membrane. Detergents are the most common chemical agents to break down non-covalent interactions for cell lysis [41]. However, using a detergent-only lysis buffer may be insufficient; for *Microcystis aeruginosa*, this method only disrupts 37% of the total cell wall material (Table 1) [42]. For cyanobacterial cells, the addition of proteolytic enzymes (proteinase K and lysozyme) and reducing agents (dithiothreitol and β-mercaptoethanol) can further disrupt stabilizing covalent interactions and increase the lysis efficiency to near 100% (Table 1) [43]. Chemical lysis methods yield a high recovery of intact nucleic acids with lesser contamination from non-target biomolecules [43–45]. Evaluation of DNA extraction following chemical lysis with a chemical lysis buffer (Tris-Urea) provided yields from 230

to 479 µg/ml of intact DNA from species of *Rivularia*, *Dolichospermum, Synechocystis*, and *Synechococcus* [46].

Chemical lysis can provide a higher yield of the targeted biomolecule. However, it can also result in chemical contamination that can cause interference in the downstream analysis [47,48]. Detergents (e.g. Triton-X 100) or organic solvents (e.g. phenol) degrade or stop the DNA polymerase from binding to the DNA template during PCR [49]. Hence, the greater yield of DNA from enzyme-based methods may be negated by the potential inhibition of downstream diagnostic protocols, including PCR [49]. Additionally, contaminating phenol absorbs at 230 nm UV, which may lead to over-estimation or false positives [50].

1.3.8.2. Ultrasonic Cell Lysis

Ultrasonication uses rapidly changing sonic pressure to cause cavitation, which agitates and disrupts cellular membranes and cell walls [51,52]. For some filamentous strains, sonication can break down the filaments into smaller structures or single cells [53,54]. In the treatment of smaller samples, sonication can be achieved through an ultrasonic bath or probe, with the latter being more efficient in lysis [55] (bath sonication of *Microcystis aeruginosa* resulted in 73% lysis after 30 min (Table 1) [42], while probe sonication can yield 80% lysis after 5 min (Table 1) [56]).

An advantage of ultrasonic lysis is the elimination of potential chemical contaminants (associated with chemical lysis), so there is little interference of downstream diagnostic PCR assays by enzymatic inhibition. Although probe sonication increases the cell lysis efficiency with reduced sonication time, a fraction of cells can remain intact even with an increase in

sonication time, and over-exposure to sonication can cause nucleic acid fragmentation [46,56]. In addition, cellular debris and biomolecules remain in the lysate, which can interfere with subsequent analytical techniques [49]. For cyanobacterial cells, a longer processing time and subsequent purification steps may be required to obtain higher yield and purity [55].

1.3.8.3. Physical Cell Lysis

Physical cell lysis methods apply external forces, including bead beating, to rupture the cell membrane and cell wall. The types of physical lysis discussed can be divided into two categories: mechanical and cryogenic (Table 1). A commonly used method of mechanical lysis is bead beating, for which ceramic or glass beads are agitated together with the cell sample to achieve cell lysis [57]. Bead beating for 3 min with 0.5 mm glass beads can result in the lysis of 99% of *M. aeruginosa* cells (Table 1) [42]. However, for *Synechocystis*, only 50% of the total cells can be lysed after bead beating [43]. Cryogenic methods include freeze-thawing and lyophilization. During cycles of freezing in either dry ice or a freezer, cell swelling and ice formation on the cellular membrane and cell wall cause structural disruption when the samples are thawed [41]. Lyophilization is similar in principle in that cyanobacterial samples are frozen and then dehydrated using a freeze-dryer under a vacuum [42]. Lysis is ultimately achieved upon rehydration [42]. For *M. aeruginosa*, three freeze-thaw cycles (-70 °C for 10 min, then 37 °C for 5 min) resulted in the lysis of only 19% of total cells (Table 1) [42]. However, another trial demonstrated that almost no *M. aeruginosa* cells remained intact after one extensive freeze-thaw cycle (-20 °C for 12 h, then 25 °C for

2-4 h) (Table 1) [56]. On the other hand, lyophilization had very high efficiency, disrupting 92 and 98% of total cells after rehydration with deionized water and methanol, respectively (Table 1) [42].

Like ultrasonication, physical cell lysis methods also circumvent the issue of chemical contamination. Despite the inconsistencies in efficiency, bead beating remains a standard method for most cell lysis protocols [58–61]. DNA extraction from bead-beaten benthic cyanobacteria resulted in a concentration of 15 µg/ml [55], a lower recovery than the chemical lysis method. The efficiency of bead beating can differ widely depending on the morphology and the characteristics of the cyanobacterial cells (e.g. benthic species of cyanobacteria are challenging to disrupt due to their protective sheaths and mucilage [55]). For bead beating, the yield of the biomolecules can be increased by adjusting parameters, including beat size, bead count, and time [62,63]. Non-optimized parameters can lead to extensive DNA shearing, which may compromise the integrity of downstream analysis [46]. Compared to bead beating, cryogenic lysis, particularly lyophilization, provides high lysis efficiency approximately equal to that of chemical lysis with an optimized buffer cocktail [42,46]. Despite the higher lysis efficiency, lyophilization requires a long waiting time (upwards of days) for the freeze drying to be complete [42], thus limiting application for routine, rapid monitoring.

1.3.8.4. Combinatorial Cell Lysis Methodologies

Lysis efficiency depends on the characteristics and morphology of the cells [64]. For cyanobacteria, employing several lysis methods in succession may be helpful to ensure

maximum lysis or to break down cells of highly resilient species. Combining bead beating with an optimized chemical lysis cocktail can achieve a 2-3-fold increase in its efficiency for *Synechocystis* and *Synechococcus spp*., allowing for maximal lysis of cyanobacterial cells [43]. Also, combining cryogenic and enzyme lysis methods increases DNA yield for filamentous *Arthrospira* species [65]. Additionally, including chemical lysis buffer, lysozyme, and proteinase K to bead beating can provide a 2-fold increase in extracted DNA compared to bead beating alone [55]. Complementarity of cell lysis methods with each other can increase lysis efficiency and DNA yield, as well as reduce the limitations associated with each method.

Table 1. Cyanobacterial cell lysis methods.

Lysis Method		**Lysis Efficiency**	**Advantages**	**Limitations**
Chemical	Detergent	37% [43]	High efficiency, high yield, low DNA degradation	Deposit contaminants that interfere with downstream assays
	Detergent-enzyme cocktail	100% [43]		
Ultrasonic	Bath sonication	73% [42]	Avoids chemical contaminants, increased purity of extracted biomolecule	Long processing time, incomplete lysis, DNA shearing
	Probe sonication	80% [56]		
Mechanical	Bead beating	50-99% [42,43]	Avoids chemical contaminants, increased purity of extracted biomolecule	DNA shearing (requires optimal bead beating parameters), inconsistent lysis efficiency depending on cell morphology

Cryogenic	Freeze-thaw	19-100% [42,56]	Avoids chemical contaminants, increased purity of extracted biomolecule	Inconsistent lysis efficiency
	Lyophilization	92-98% [42]	High efficiency, avoids chemical contaminants, increased purity of extracted biomolecule	Long waiting times can limit the use of rapid detection methods

1.3.9. Methods and Technologies for Cyanobacterial and Cyanotoxin Monitoring

Conventional strategies include biological assays and chromatographic technologies for determining cyanotoxin concentration, as well as quantitative enumeration of cyanobacteria cells. Recent advancements in molecular biology and computer science led to rapid improvements in PCR, DNA sequencing, microfluidics, and machine learning methods to probe and monitor cyanobacterial density/activity in recreational waters. This section will introduce methods for cyanobacteria/cyanotoxin monitoring, evaluate their current advantages and disadvantages, and evaluate each method's sensitivity or specificity in detecting harmful algal blooms. The focus is on the diagnostic molecules of microcystins and DNA, as they remain stable after extraction [66]. The sensitivities, benefits, and limitations of cyanotoxin quantification methods are summarized in Table 2, while those of cyanobacterial cell quantification methods are summarized in Table 3.

1.3.9.1. Conventional Methods and Techniques

The simplest protocol for cyanobacterial cell counting by microscopy is the Utermöhl method (for phytoplankton), in which sedimentation of cyanobacteria cells is performed before enumeration using an inverted microscope [67,68]. In addition to simplicity, microscopic methods can provide high specificity and allow for identification up to the genus and species levels [69,70]. However, microcystin levels are only weakly correlated with microscopic cell counts and may not fully account for the cyano-toxicity of the recreational waters [71]. Direct enumeration can also be time-consuming [70]; the Utermöhl method requires 24-48 h of sedimentation time, limiting its applicability to real-time, rapid monitoring [67,68]. Enumeration and identification of cyanobacteria via microscopy also require trained expertise in the field, and its reliability can significantly vary depending on the analyst's skill. Expertise limitation for microscopic cyanobacterial counting can be overcome by integrating computational models (e.g. PhytoNumb3rs [72]) or automated microscopy (e.g. FlowCam Cyano (Yokogawa Fluid Imaging Technologies, Maine, USA)) for cyanobacterial morphology. Atomic force microscopy can visualize even a single molecule in ultra-high resolution and can used along with conventional microscopic techniques to robustly examine the properties of both microcystin and cyanobacteria cells [32,34,35]. AFM enables assessments in liquid media that mimic intracellular environments, however, AFM requires sample immobilization [73,74]. Overall, quantitative enumeration can provide a suitable reference for more advanced monitoring technologies, but it can be somewhat inconsistent and timely when used by itself.

Enzyme-linked immunosorbent assays (ELISAs) can be used to quantify cyanotoxins using competitive binding between antibodies and the targeted toxins [75]. The current standard for microcystin detection is an assay that recognizes the β-amino acid, ADDA, (4E, 6E 3-amino-9-methoxy-2, 6, 8-trimethyl-10-phenyldeca-4, 6-dienoic acid) that is present in most of the penta- and heptapeptide congeners of microcystins and nodularins [76,77]. ADDA-based ELISA method is rapid (2-3h) and can be performed without pre-concentration on lake water samples. It is highly sensitive, being able to detect a minimum of 0.02-0.07 ng/mL of microcystins (Table 2), which is below the Health Canada guideline for microcystin-LR of 1.5 ng/mL [78]. Regarding commercial availability, ADDA-microcystin ELISA kits (Manuals available at: https://www.enzolifesciences.com/ALX-850-319/microcystins-adda-specific-elisa-kit/ and https://www.caymanchem.com/product/502000/microcystin-elisa-kit) report sensitivities ranging from 0.1 - 0.196 ng/mL (Enzo Life Sciences, Inc., New York, USA, Cayman Chemical, Michigan, USA) (Table 2). USEPA testing of the Abraxis ELISA Microcystins Strip Test (Gold Standard Diagnostics Horsham Inc. Warminster, USA) provided a minimum reporting limit of 0.30 ng/mL (Table 2) [31]. Commercially available kits can vary in terms of detection sensitivity/limit. Therefore, a comparative validation of ELISA kits should be performed to determine the optimal sensitivity for a recreation water ecosystem. One limitation of ELISA-based methods is their inability to distinguish between different congeners of microcystins and nodularins present in a sample [76,79]. Lack of specificity is reflected in common cross-reactions with microcystin degradation byproducts and metabolites, which can lead to overestimation/false-positive of cyano-toxicity [80–83].

Chromatography purifies compounds (including cyanotoxins) based on hydrophobic, hydrophilic, ionic or affinity-based mobility through a separation medium [84]. It can be coupled with a spectrophotometric detection method for quantification [18]. Liquid chromatography-mass spectrometry (e.g. liquid chromatography-electrospray ionization-high resolution mass spectrometry, ultra-high-performance liquid chromatography high-resolution mass spectrometry) can provide a sensitivity as low as 0.000004-0.02 ng/mL (Table 2) [85,86]. Aside from lower detection limits (high sensitivity), chromatography coupled with tandem mass spectrometry can differentiate between different cyanotoxin congeners [87]. As such, chromatography-coupled-mass spectrometry methods are highly sensitive and can resolve specific congeners of interest. Although, it is essential to note that chromatographic methods require specific congener standards for qualitative/quantitative assessment (pure quantity of targeted molecule). However, standards for only a few naturally-occurring microcystins are commercially available [80,88,89]. Additionally, chromatography requires highly trained personnel and high recurrent costs [90]. Overall, most conventional methods have limitations that can be overcome by integrating or coupling modern molecular biological or computational methods.

1.3.9.2. Polymerase Chain Reaction (PCR) and DNA Sequencing

Conventional polymerase chain reaction (PCR) uses amplification of organism-specific, genomic DNA sequences for qualitative (absence or presence) analysis, while real-time PCR or quantitative PCR incorporates DNA binding dye or probe to quantify the DNA target (cyanobacterial or cyanotoxin gene copy number) in purified DNA (e.g. DNA extracted from

an environmental sample including algal blooms) [91]. Newer digital PCR techniques are also available that partition purified DNA samples into thousands of oil droplets or wells on a chip to quantify the DNA target. The expression of actively transcribed cyanotoxin genes can be assessed using reverse transcription qPCR (RT-qPCR) to distinguish active from quiescent cells [92,93]. Universal gene target (16s rRNA) analysis, which has a 97% conserved region for cyanobacteria, is commonly used for qPCR studies [94–97]. One assessment reported a detection limit of twenty-five gene copies per reaction (Table 3) [98]. Although the 16s rRNA gene effectively quantifies cyano-toxigenic cyanobacteria, including *Microcystis* and *Planktothrix spp.* [77], it is not specific enough for all cyanobacteria and can result in over-estimation or false positives [99,100]. Attempts have been made to counter non-specificity by using cyanobacterial barcoding (integrating PCR amplification and next-generation DNA sequencing), which can allow the use of downstream bioinformatics to obtain a higher confidence level of cyanobacterial identification [101–103].

For cyanotoxin detection/quantification, the *mcyE* (microcystin gene cluster) assay detects all potential microcystin producers and the nodularin synthetase gene cluster [104]. The *mcyA* assay covers many, but not all, microcystin-producing strains [9], while the *mcyE* and *mcyG* assays are sensitive and specific to *Microcystis* and *Planktothrix* [105–107]. The detection limit for the *mcyE* assay was reported to be 3-63 gene copies per PCR reaction (Table 2) [108]. To increase the PCR detection specificity, conserved signature proteins (CSPs) and conserved signature indels (CSIs) that are highly specific to a particular clade of organisms can be used [109]. Thirty-nine Cyanobacteria-specific conserved signatures

proteins have been identified for cyanobacteria species [110]. These can be used for generating cyano-specific PCR/qPCR methods.

For application to the Great Lakes ecosystems, the USEPA and Health Canada have identified PCR and qPCR as useful methods for detecting cyanobacteria and cyanotoxins [27,28]. PCR methodology has been used for qualitative analysis of bloom-infested Lake Erie to determine that up to 42% of the total Cyanobacterial population is comprised of *Microcystis*, and the measured proportion of total *Microcystis* demonstrates a strong correlation with total microcystin concentrations [98,111]. qPCR and RT-qPCR have recently been demonstrated to be useful early warning tools of cyanobacteria blooms and cyanotoxin production in inland lakes in Ohio [77,112]. Characterization of the spatiotemporal variation associated with cyanobacterial blooms in Lake Erie has also been assessed using PCR, followed by high-throughput sequencing of the 16s rRNA gene [113]. These applications indicate the potential of using PCR/qPCR and DNA sequencing for cyanobacterial identification in the Great Lakes ecosystems.

Compared to traditional methods, including microscopic enumeration, qPCR can be more rapid and provide results within 3-4 h [77,90]. In addition, by incorporating the cyanotoxin gene quantification, the cyanotoxin potential of recreational waters, including the Great Lakes, can be determined [77]. A limitation of PCR is the potential to be inhibited by environmental contaminants, including cell debris, humic acids, detergents, and polysaccharides. These contaminants can degrade or sequester the DNA polymerase and nucleic acids, thus inhibiting PCR amplification [49]. Furthermore, DNA extraction efficiency can also be a limiting factor in PCR analysis; the efficiency of DNA recovery in

microbial identification can be approximately 30% or less due to the environmental matrix interferences and unique resilience of the cyanobacteria cell wall and sheath [91]. Additionally, when using specific target genes for PCR, unknown toxigenic species with environmental variations in the gene sequence may go undetected (under-representation of the counts) [114,115]. On the other hand, non-viable cells can get detected when using universal targets (over-representation of the data) [90]. While these are challenges for PCR approaches, they can also be addressed through rigorous quality control steps as demonstrated by the development of USEPA's standard *Enterococcus* qPCR method now being applied at Great Lakes beaches [116,117]. In conclusion, PCR-based techniques can efficiently and conveniently monitor cyanobacteria quantity and toxicity, and the associated limitations can be resolved by integrating other molecular methods.

1.3.9.3. Microfluidic and DNA Capture Devices

Microfluidic and DNA capture devices incorporate target-specific probes or fluorophores that can bind to the target molecule in the cellular/DNA extract and generate a signal corresponding to the amount of the targeted molecule [118,119]. A handheld microfluidic device capable of rapid and onsite detection and identification of microcystins and other cyanotoxins is available [120,121]. The device performs an automated ELISA assay as a disposable microfluidic cartridge with a detection limit of 0.16 ng/ml (Table 2), comparable to commercial ELISA kits for laboratory settings [120]. An antibody microarray chip called CYANOCHIP is available for cyanobacteria detection in soil and water samples; it has a detection limit of 100-1000 cells (Table 3) [122,123].

Based on DNA sequence detection, a microfluidic chip biosensor targeting the cyanopeptolin (a cyanotoxin produced by *Planktothrix* and *Microcystis spp.*) can provide a detection limit of 6×10^{-12} M of target DNA [124]. However, testing was only conducted with purified PCR products for this sensor. Similarly, a cantilever biosensor assay designed to recognize a conserved region of the 16s rRNA gene in *M. aeruginosa* can detect approximately 50 cells/mL in purified DNA and 500 cells/mL for river water samples (Table 3) [125]. DNA chip assays are also available for both the 16s rRNA and *mcyE* genes to monitor microcystin levels [126,127]. These have been tested for environmental samples, and the detection limit is as low as 1–5 fmol of DNA (Tables 2,3) [126].

A significant advantage of microfluidic and chip assays is their portability and usage of minimum sample volume [128]. The devices are also capable of providing rapid detection of the target molecules. For example, the CYANOchip takes 3 h, including incubation time to provide results [123]. For the microarray chip, multiple targets can be monitored and analyzed simultaneously using 17 antibodies [123], further enhancing its efficiency. These factors all increase the potential for onsite microfluidics and chip technology applications. However, testing on environmental complex samples can reduce the sensitivity due to the presence of background microorganisms [125]. A possible extension to microfluidic devices is the biomonitoring tool, which determines analytes impacting algal growth with high sensitivity and low sample volumes [36]. However, it is often difficult to directly apply the results of biomonitoring to make predictions on freshwater ecosystems [33,36]. Overall, microfluidic and DNA capture devices can provide portability, but

Table 2. Cyanotoxin (microcystin) detection and quantification methods.

Method	Sensitivity	Advantages	Limitations
Enzyme-linked immunosorbent assay (ELISA)	0.02-0.30 ng/mL [31,76]	Rapid, high sensitivity, the limit of detection within Health Canada guidelines	Low specificity, congener-independent, cross-reaction with cyanotoxin metabolites lead to false positives/overestimation
Liquid chromatography-mass spectrometry (LC/MS)	0.000004-0.02 ng/mL [85,86]	High sensitivity, the limit of detection within Health Canada guidelines, congener-specific	Few standards are commercially available, require highly trained personnel, high recurrent cost
Quantitative polymerase chain reaction (qPCR) targeting *mcyE*	3-63 gene copies per reaction [108]	Rapid allows both qualitative and quantitative analysis, allows assessment of cyanotoxin potential	Environmental contaminants can inhibit amplification, species with sequence variations may go undetected, detection of non-viable cells
Microfluidic device	0.16 ng/ml [120,121]	Rapid, portable and suitable for the field use, requires minimal sample volume	Reduced sensitivity, interference from sample background
DNA capture targeting *mcyE*	1–5 fmol of DNA [126]	Rapid, convenient, potential for field use, requires minimal sample volume	Reduced sensitivity, interference from sample background, Limited testing on complex environmental samples.

1.3.10. Future Directions

1.3.10.1. Remote Sensing/Satellite Imaging

Satellite imaging and remote sensing algorithms allow coverage of a larger geographical area than the methods based on water sample collection. Standard sensors for monitoring bloom dynamics include Landsat, Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectrometer (MERIS), and Ocean and Land Color Instrument (OLCI) due to their optimum temporal, spatial, and spectral resolutions [70,129].

A novel method is available to determine the magnitude of cyanoHABs across the United States using data from MERIS; the method extended to OLCI [130]. The current Cyanobacteria Index (CI), which uses MERIS and OLCI data, identifies cyanotoxin-producing cyanobacteria with 84% accuracy and a detection limit of 1 CI value (Table 3) [131,132]. Cyanotoxins cannot be directly measured using remote sensing as concentration can only be determined indirectly by detecting the surrogate pigments, including phycocyanin or chlorophyll a [133]. To appropriately measure these pigments, band ratio algorithms in the red and near-infrared spectral regions can be employed [129]. Evaluation of twenty-seven algorithms for cyanobacterial biovolume determination indicates that both chlorophyll a- and phycocyanin-based algorithms achieved high detection accuracy, with phycocyanin being slightly more sensitive than chlorophyll methods [134].

Remote sensing can especially be useful for inland water bodies [129], including the Great Lakes. The Cyanobacteria Assessment Network (CyAN) project by the USEPA aims to develop an early warning indicator system for cyanobacteria blooms in freshwater systems using satellite data records [135]. The project utilizes products for MERIS, OLCI, as well as MODIS. For this purpose, the potential for phytoplankton exposure at a Great Lakes beach was determined using two remote sensing methods: the MODIS ocean chlorophyll-a algorithm and the MERIS CI. As part of the EOLakeWatch (Earth Observation Lake Watch) program, Environment and Climate Change Canada monitors the water quality of lakes across Canada (including the Great Lakes) using satellite imagery [136]. Chlorophyll algorithms are applied to MERIS and OLCI data, from which algal bloom indices are derived [137].

There are specific challenges accompanying the application of remote sensing for bloom monitoring, a major one being the need to identify surrogate pigments. Variability in the correlation between chlorophyll a/phycocyanin and cyanotoxins can make it difficult to accurately infer water toxicity from satellite imagery, especially for microcystins [133]. Remote sensing is also limited to water systems that cover large surface areas, and it may be inaccessible for smaller, more secluded areas for recreation [70]. Additionally, differing cyanobacteria composition and optical properties in parts of Lake Erie can lead to misidentification by remote sensing algorithms [138]. Although remote sensing and satellite imaging provide a broader geographical range of detection, this methodology may be unable to distinguish between harmful algal blooms and other cyanobacterial biomass. Integrating remote sensing with cyanotoxin quantification can overcome the specificity limitation and provide more robust information.

1.3.10.2. Artificial Intelligence and Machine-Learning-Based Prediction Tools

Following the ever-increasing volume of available data on aquatic ecosystems, artificial intelligence and machine-learning algorithms can become valuable tools in making predictions on cyanobacterial bloom activities. There are multiple ways in which these algorithms can be applied to cyanobacteria monitoring, including cell imaging and water quality prediction deep learning algorithms [139–142]. Microscopic images of cyanobacteria samples can be processed for enumeration and species identification using imaging-based detection software. A fast regional convolutional neural network (R-CNN) was able to identify five major species of cyanobacteria (including *Microcystis spp.*) with average

precision values from 0.890 to 0.973 [141]. In the same study, a basic convolutional neural network (CNN) was able to quantify populations of 50-250 cyanobacterial cells accurately (Table 3). Another deep learning-based method for qualitative microscopic image processing using a convolution fusion network (CFN) outperformed classic CNN models in terms of accuracy with a prediction rate of 99.36% in classifying cyanobacteria cells [142]. A multi-objective hybrid evolutionary algorithm (MOHEA) can provide crucial threshold exceedances of local cyanobacteria outbreaks and forecast cyanobacteria activity seven days before bloom events [139]. Chlorophyll a, a compound necessary for photosynthesis, can be used to train machine-learning algorithms and predict the onset of cyanobacteria blooms. For example, an auto-regressive integrated moving average (ARIMA) model was developed to predict chlorophyll a level in Lake Taihu, China. This model demonstrates potential for use as a cyanobacterial bloom warning system [140].

Unlike conventional or molecular biology methods, models rooted in artificial intelligence and machine learning can solve more complex problems with a greater number of variables [143]. Once proper algorithms have been developed and implemented, it can reduce the need for onsite expertise in cyanobacteria monitoring [70]. However, these technologies also possess many challenges in their current state of development. For example, to reach an accurate and reliable prediction, it is necessary to integrate multiple algorithms [144]. The dataset size needed is also an issue, as most collected data (70-80%) is used as a training set [145]. Only the remaining 20-30% is used as test data to measure prediction accuracy, creating a need for an immensely more extensive and diverse data set to observe all the possible environmental patterns [145]. The AI/machine learning models produced under

such strict guidelines may not be applicable across geospatial differences or in environments with drastically different physical and chemical differences. These limitations must first be addressed for the widespread adoption of AI and machine learning technologies to be feasible.

Table 3. Cyanobacteria detection and quantification methods.

Method	Sensitivity	Advantages	Limitations
Microscopic enumeration	Not applicable	Simplicity, identification up to the genus and species levels	Time-consuming, requires trained personnel, and accuracy is dependent on the skill of the analyst
Quantitative polymerase chain reaction (qPCR) targeting 16s rRNA	25 gene copies per reaction [98]	Rapid allows both qualitative and quantitative analysis	Non-specificity of a target gene, environmental contaminants inhibit amplification, detection of unviable cells
Antibody microarray chip (CYANOCHIP)	100 cells [122,123]	Rapid, convenient, potential for field use, requires minimal sample volume	Reduced sensitivity, interference from sample background
Biosensor assay	50-500 cells/mL [125]	Rapid, convenient, potential for field use, requires minimal sample volume	Reduced sensitivity, interference from sample background
DNA capture device targeting 16s rRNA	1–5 fmol of DNA [126]	Rapid, convenient, potential for field use, requires minimal sample volume	Reduced sensitivity, interference from sample background
Remote sensing	1 CI value [132]	Extensive coverage of the geographical area, useful for inland bodies of water	Variability in the correlation between surrogate pigment and toxicity, inaccessible for smaller areas, and differing cyanobacterial

			composition leads to misinterpretation
Artificial intelligence (convolutional neural network)	50 cells [141]	Capable of complex analysis, reducing the need for onsite expertise	Need to integrate multiple algorithms for high reliability, requires extensive and diverse datasets, not applicable across geospatial differences

1.3.11. Conclusions

Cyanobacterial harmful algal blooms threaten the conservation of essential freshwater ecosystems, including the North American Great Lakes. Conventional/standard molecular methods for cyanobacterial detection are available. However, specific limitations are associated with each method.

1. Cyanobacterial cells resist disruption methods, making quantitative recovery of potential diagnostic molecules difficult. However, optimizing protocol parameters and combining multiple lysis methods can lead to complete disruption.

2. ELISA-based toxin detection methods (including microfluidic devices) alone cannot resolve or quantify all microcystin congeners. Chromatography-mass spectrometry methods can unambiguously identify microcystins and other cyanotoxins. However, the obtained information is difficult to incorporate into a public health response due to the lack of commercially available standards.

3. DNA diagnostic methods (PCR, DNA capture devices) targeting the 16s rRNA gene, while useful for other bacteria, is of limited value for cyanobacteria due to insufficient specificity. This obstacle can be countered by metabarcoding or targeting a toxin gene like the *Mcy* (microcystin synthetase) gene cluster.

4. Bloom monitoring may be aided by novel technologies, particularly in quickly establishing spatiotemporal characteristics of specific events. These technologies can augment traditional characterization methods in producing a public health response. However, while ongoing, standardization of common tools for this ancillary information still needs to be completed.

5. Newer/modern technologies, including satellite imaging, biosensors and machine learning/artificial intelligence methods, can be integrated with the convention-al/standard molecular methods to overcome the problems associated with cyano-bacterial detection in recreational water ecosystems, including the Great Lakes.

Author Contributions: Literature survey J.L.J., and F.S.; validation, H.E.S; writing—original draft preparation, J.L.J., and F.S; writing—review and editing J.L.J., F.S., T.A.E., H.E.S; supervision, H.E.S.

Funding: This research was funded by NSERC Alliance Award, grant number 554507-20.

Acknowledgements: We would like to thank all members of the Schellhorn lab for their contributions to this research.

Conflicts of Interest: The authors declare no conflict of interest.

1.4. References

1. Bekker, A.; Holland, H.D.; Wang, P.-L.; Rumble, D.; Stein, H.J.; Hannah, J.L.; Coetzee, L.L.; Beukes, N.J. Dating the Rise of Atmospheric Oxygen. Nature 2004, 427, 117–120, doi:10.1038/nature02260.

2. Sánchez-Baracaldo, P.; Cardona, T. On the Origin of Oxygenic Photosynthesis and Cyanobacteria. New Phytol. 2020, 225, 1440–1446, doi:10.1111/nph.16249.

3. Rastogi, R.P.; Madamwar, D.; Incharoensakdi, A. Bloom Dynamics of Cyanobacteria and Their Toxins: Environmental Health Impacts and Mitigation Strategies. Front. Microbiol. 2015, 6, doi:10.3389/fmicb.2015.01254.

4. Huisman, J.; Codd, G.A.; Paerl, H.W.; Ibelings, B.W.; Verspagen, J.M.H.; Visser, P.M. Cyanobacterial Blooms. Nat. Rev. Microbiol. 2018, 16, 471–483, doi:10.1038/s41579-018-0040-

5. Hoiczyk, E.; Hansel, A. Cyanobacterial Cell Walls: News from an Unusual Prokaryotic Envelope. J. Bacteriol. 2000, 182, 1191–1199, doi:10.1128/JB.182.5.1191-1199.2000.

6. Bothe, H.; Schmitz, O.; Yates, M.G.; Newton, W.E. Nitrogen Fixation and Hydrogen Metabolism in Cyanobacteria. Microbiol. Mol. Biol. Rev. MMBR 2010, 74, 529–551, doi:10.1128/MMBR.00033-10.

7. Paerl, H. The Cyanobacterial Nitrogen Fixation Paradox in Natural Waters. F1000Research 2017, 6, 244, doi:10.12688/f1000research.10603.1.

8. Van Dolah, F.M. Marine Algal Toxins: Origins, Health Effects, and Their Increased Occurrence. Environ. Health Perspect. 2000, 108 Suppl 1, 133–141, doi:10.1289/ehp.00108s1133.

9. Hisbergues, M.; Christiansen, G.; Rouhiainen, L.; Sivonen, K.; Börner, T. PCR-Based Identification of Microcystin-Producing Genotypes of Different Cyanobacterial Genera. Arch. Microbiol. 2003, 180, 402–410, doi:10.1007/s00203-003-0605-9.

10. Plaas, H.E.; Paerl, H.W. Toxic Cyanobacteria: A Growing Threat to Water and Air Quality. Environ. Sci. Technol. 2021, 55, 44–64, doi:10.1021/acs.est.0c06653.

11. Rabalais, N.N.; Díaz, R.J.; Levin, L.A.; Turner, R.E.; Gilbert, D.; Zhang, J. Dynamics and Distribution of Natural and Human-Caused Hypoxia. Biogeosci-ences 2010, 7, 585–619, doi:10.5194/bg-7-585-2010.

12. Paerl, H.W. Controlling Harmful Cyanobacterial Blooms in a Climatically More Extreme World: Management Options and Research Needs. J. Plankton Res. 2017, 39, 763–771, doi:10.1093/plankt/fbx042.

13. Izaguirre, G.; Taylor, W.D. A Guide to Geosmin- and MIB-Producing Cyanobacteria in the United States. Water Sci. Technol. J. Int. Assoc. Water Pollut. Res. 2004, 49, 19–24.

14. Jüttner, F.; Watson, S.B. Biochemical and Ecological Control of Geosmin and 2-Methylisoborneol in Source Waters. Appl. Environ. Microbiol. 2007, 73, 4395–4406, doi:10.1128/AEM.02250-06.

15. Wang, Z.; Song, G.; Li, Y.; Yu, G.; Hou, X.; Gan, Z.; Li, R. The Diversity, Origin, and Evolutionary Analysis of Geosmin Synthase Gene in Cyanobacteria. Sci. Total Environ. 2019, 689, 789–796, doi:10.1016/j.scitotenv.2019.06.468.

16. Watson, S.B.; Ridal, J.; Boyer, G.L. Taste and Odour and Cyanobacterial Toxins: Impairment, Prediction, and Management in the Great Lakes. Can. J. Fish. Aquat. Sci. 2008, 65, 1779–1796, doi:10.1139/F08-084.

17. Miller, T.R.; Beversdorf, L.J.; Weirich, C.A.; Bartlett, S.L. Cyanobacterial Toxins of the Laurentian Great Lakes, Their Toxicological Effects, and Numeri-cal Limits in Drinking Water. Mar. Drugs 2017, 15, 160, doi:10.3390/md15060160.

18. Svirčev, Z.; Lalić, D.; Bojadžija Savić, G.; Tokodi, N.; Drobac Backović, D.; Chen, L.; Meriluoto, J.; Codd, G.A. Global Geographical and Historical Over-view of Cyanotoxin Distribution and Cyanobacterial Poisonings. Arch. Toxicol. 2019, 93, 2429–2481, doi:10.1007/s00204-019-02524-4.

19. Ferrão-Filho, A. da S.; Kozlowsky-Suzuki, B. Cyanotoxins: Bioaccumulation and Effects on Aquatic Animals. Mar. Drugs 2011, 9, 2729–2772, doi:10.3390/md9122729.

20. Shi, L.; Du, X.; Liu, H.; Chen, X.; Ma, Y.; Wang, R.; Tian, Z.; Zhang, S.; Guo, H.; Zhang, H. Update on the Adverse Effects of Microcystins on the Liver. Environ. Res. 2021, 195, 110890, doi:10.1016/j.envres.2021.110890.

21. Steffen, M.M.; Belisle, B.S.; Watson, S.B.; Boyer, G.L.; Wilhelm, S.W. Status, Causes and Controls of Cyanobacterial Blooms in Lake Erie. J. Gt. Lakes Res. 2014, 40, 215–225, doi:10.1016/j.jglr.2013.12.012.

22. Makarewicz, J.C.; Boyer, G.L.; Lewis, T.W.; Guenther, W.; Atkinson, J.; Arnold, M. Spatial and Temporal Distribution of the Cyanotoxin Microcystin-LR in the Lake Ontario Ecosystem: Coastal Embayments, Rivers, Nearshore and Offshore, and Upland Lakes. J. Gt. Lakes Res. 2009, 35, 83–89, doi:10.1016/j.jglr.2008.11.010.

23. Almuhtaram, H.; Cui, Y.; Zamyadi, A.; Hofmann, R. Cyanotoxins and Cyanobacteria Cell Accumulations in Drinking Water Treatment Plants with a Low Risk of Bloom Formation at the Source. Toxins 2018, 10, 430, doi:10.3390/toxins10110430.

24. Bartlett, S.L.; Brunner, S.L.; Klump, J.V.; Houghton, E.M.; Miller, T.R. Spatial Analysis of Toxic or Otherwise Bioactive Cyanobacterial Peptides in Green Bay, Lake Michigan. J. Gt. Lakes Res. 2018, 44, 924–933, doi:10.1016/j.jglr.2018.08.016.

25. Sterner, R.W.; Reinl, K.L.; Lafrancois, B.M.; Brovold, S.; Miller, T.R. A First Assessment of Cyanobacterial Blooms in Oligotrophic Lake Superior. Limnol. Oceanogr. 2020, 65, 2984–2998, doi:10.1002/lno.11569.

26. United States Environmental Protection Agency; Environment and Climate Change Canada 2022 Progress Report of the Parties. Pursuant to the 2012 Canada-United States Great Lakes Water Quality Agreement. Cat. No.: En161-25E-PDF. EPA 905R22003. Available at: https://binational.net/wp-content/uploads/2022/07/2022-Progress-Report-of-the-Parties.pdf

27. United States Environmental Protection Agency Determination of Cyanotoxins in Drinking and Ambient Freshwaters Available online: https://www.epa.gov/cyanohabs/determination-cyanotoxins-drinking-and-ambient-freshwaters (accessed on 18 February 2023).

28. Health Canada Guidelines for Canadian Recreational Water Quality - Cyanobacteria and Their Toxins Available online: https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidance-canadian-recreational-water-quality-cyanobacteria-toxins.html (ac-cessed on 18 February 2023).

29. Shoemaker, J.A.; Tettenhorst, D.R.; de la Cruz, A. Method 544. Determination of Microcystins and Nodularins in Drinking Water by Solid Phase Extrac-tion and Liquid Chromatography/Tandem Mass Spectrometry (LC/MS/MS) 2015. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=306953&simpleSearch=1&searchAll=544

30. Shoemaker, J.A.; Tettenhorst, D.R.; de la Cruz, A. Single Laboratory Validated Method for Determination of Microcystins and Nodularin in Ambient Freshwaters by Solid

Phase Extraction and Liquid Chromatography/ Tandem Mass Spectrometry (LC/MS/MS) 2017. Available at: https://www.epa.gov/sites/default/files/2017-11/documents/microcystin_method_for_ambient_water_nov_2017.pdf

31. Zaffiro, A.; Rosenblum, L.; Wendelken, S.C. Method 546: Determination of Total Microcystins and Nodularins in Drinking Water and Ambient Water by Adda Enzyme-Linked Immunosorbent Assay 2016. Available at: https://www.epa.gov/sites/default/files/2016-09/documents/method-546-determination-total-microcystins-nodularins-drinking-water-ambient-water-adda-enzyme-linked-immunosorbent-assay.pdf

32. Marcuello, C. Present and Future Opportunities in the Use of Atomic Force Microscopy to Address the Physico-Chemical Properties of Aquatic Ecosys-tems at the Nanoscale Level. Int. Aquat. Res. 2022, 14, 231–240, doi:10.22034/iar.2022.1965012.1317.

33. Zhou, Q.; Zhang, J.; Fu, J.; Shi, J.; Jiang, G. Biomonitoring: An Appealing Tool for Assessment of Metal Pollution in the Aquatic Ecosystem. Anal. Chim. Acta 2008, 606, 135–150, doi:10.1016/j.aca.2007.11.018.

34. Ceballos-Laita, L.; Marcuello, C.; Lostao, A.; Calvo-Begueria, L.; Velazquez-Campoy, A.; Bes, M.T.; Fillat, M.F.; Peleato, M.-L. Microcystin-LR Binds Iron, and Iron Promotes Self-Assembly. Environ. Sci. Technol. 2017, 51, 4841–4850, doi:10.1021/acs.est.6b05939.

35. Pillet, F.; Dague, E.; Pečar Ilić, J.; Ružić, I.; Rols, M.-P.; Ivošević DeNardis, N. Changes in Nanomechanical Properties and Adhesion Dynamics of Algal Cells during Their Growth. Bioelectrochemistry Amst. Neth. 2019, 127, 154–162, doi:10.1016/j.bioelechem.2019.02.011.

36. Respondek, Z.; Jerz, D.; Świsłowski, P.; Rajfur, M. Active Biomonitoring of Heavy Metal Concentrations in Aquatic Environment Using Mosses and Algae. Water 2022, 14, 3335, doi:10.3390/w14203335.

37. Jin, C.; Mesquita, M.M.F.; Deglint, J.L.; Emelko, M.B.; Wong, A. Quantification of Cyanobacterial Cells via a Novel Imaging-Driven Technique with an Integrated Fluorescence Signature. Sci. Rep. 2018, 8, 9055, doi:10.1038/s41598-018-27406-0.

38. Billi, D.; Grilli Caiola, M.; Paolozzi, L.; Ghelardini, P. A Method for DNA Extraction from the Desert Cyanobacterium Chroococcidiopsis and Its Appli-cation to Identification of FtsZ. Appl. Environ. Microbiol. 1998, 64, 4053–4056, doi:10.1128/AEM.64.10.4053-4056.1998.

39. Fiore, M.F.; Moon, D.H.; Tsai, S.M.; Lee, H.; Trevors, J.T. Miniprep DNA Isolation from Unicellular and Filamentous Cyanobacteria. J. Microbiol. Methods 2000, 39, 159–169, doi:10.1016/S0167-7012(99)00110-4.

40. Zhang, J.-Y.; Lin, G.-M.; Xing, W.-Y.; Zhang, C.-C. Diversity of Growth Patterns Probed in Live Cyanobacterial Cells Using a Fluorescent Analog of a Peptidoglycan Precursor. Front. Microbiol. 2018, 9, doi:10.3389/fmicb.2018.00791.

41. Shehadul Islam, M.; Aryasomayajula, A.; Selvaganapathy, P.R. A Review on Macroscale and Microscale Cell Lysis Methods. Micromachines 2017, 8, 83, doi:10.3390/mi8030083.

42. Kim, I.S.; Nguyen, G.H.; Kim, S.; Lee, J.; Yu, H.-W. Evaluation of Methods for Cyanobacterial Cell Lysis and Toxin (Microcystin-LR) Extraction Using Chromatographic and Mass Spectrometric Analyses. Environ. Eng. Res. 2009, 14, 250–254.

43. Mehta, K.K.; Evitt, N.H.; Swartz, J.R. Chemical Lysis of Cyanobacteria. J. Biol. Eng. 2015, 9, 10, doi:10.1186/s13036-015-0007-y.

44. Bhardwaj, B.; Agrawal, A.; Ledwani, L.; Vaiphei, S.T.; Jain, S. An Efficient Method for DNA Extraction from Cyanobacteria Isolated from Hypersaline and Marine Environments. J. Phycol. 2019, 55, 733–737, doi:10.1111/jpy.12847.

45. Kim Tiam, S.; Comte, K.; Dalle, C.; Duval, C.; Pancrace, C.; Gugger, M.; Marie, B.; Yéprémian, C.; Bernard, C. Development of a New Extraction Method Based on High-Intensity Ultra-Sonication to Study RNA Regulation of the Filamentous Cyanobacteria Planktothrix. PloS One 2019, 14, e0222029, doi:10.1371/journal.pone.0222029.

46. Singh, S.P.; Rastogi, R.P.; Häder, D.-P.; Sinha, R.P. An Improved Method for Genomic DNA Extraction from Cyanobacteria. World J. Microbiol. Biotechnol. 2011, 27, 1225–1230, doi:10.1007/s11274-010-0571-8.

47. Xiao, Z.; Yang, G.; Yan, D.; Li, S.; Chen, Z.; Li, W.; Wu, Y.; Chen, H. Effects of Diverse Materials-Based Methods on DNA Extraction for Clostridium Difficle from Stool Samples. Mater. Express 2019, 9, 509–516, doi:10.1166/mex.2019.1520.

48. Emaus, M.N.; Anderson, J.L. Simultaneous Cell Lysis and DNA Extraction from Whole Blood Using Magnetic Ionic Liquids. Anal. Bioanal. Chem. 2020, 412, 8039–8049, doi:10.1007/s00216-020-02941-w.

49. Schrader, C.; Schielke, A.; Ellerbroek, L.; Johne, R. PCR Inhibitors – Occurrence, Properties and Removal. J. Appl. Microbiol. 2012, 113, 1014–1026, doi:10.1111/j.1365-2672.2012.05384.x.

50. Koshy, L.; Anju, A.L.; Harikrishnan, S.; Kutty, V.R.; Jissa, V.T.; Kurikesu, I.; Jayachandran, P.; Jayakumaran Nair, A.; Gangaprasad, A.; Nair, G.M.; et al. Evaluating Genomic DNA Extraction Methods from Human Whole Blood Using Endpoint and Real-Time PCR Assays. Mol. Biol. Rep. 2017, 44, 97–108, doi:10.1007/s11033-016-4085-9.

51. Brown, R.B.; Audet, J. Current Techniques for Single-Cell Lysis. J. R. Soc. Interface 2008, 5, S131–S138, doi:10.1098/rsif.2008.0009.focus.

52. Pandur, Ž.; Dular, M.; Kostanjšek, R.; Stopar, D. Bacterial Cell Wall Material Properties Determine *E. coli* Resistance to Sonolysis. Ultrason. Sonochem. 2022, 83, 105919, doi:10.1016/j.ultsonch.2022.105919.

53. Wu, X.; Zarka, A.; Boussiba, S. A Simplified Protocol for Preparing DNA from Filamentous Cyanobacteria. Plant Mol. Biol. Report. 2000, 18, 385–392, doi:10.1007/BF02825067.

54. Rajasekhar, P.; Fan, L.; Nguyen, T.; Roddick, F.A. A Review of the Use of Sonication to Control Cyanobacterial Blooms. Water Res. 2012, 46, 4319–4329, doi:10.1016/j.watres.2012.05.054.

55. Gaget, V.; Keulen, A.; Lau, M.; Monis, P.; Brookes, J.D. DNA Extraction from Benthic Cyanobacteria: Comparative Assessment and Optimization. J. Appl. Microbiol. 2017, 122, 294–304, doi:10.1111/jam.13332.

56. Greenstein, K.E.; Zamyadi, A.; Wert, E.C. Comparative Assessment of Physical and Chemical Cyanobacteria Cell Lysis Methods for Total Microcys-tin-LR Analysis. Toxins 2021, 13, 596, doi:10.3390/toxins13090596.

57. Danaeifar, M. New Horizons in Developing Cell Lysis Methods: A Review. Biotechnol. Bioeng. 2022, 119, 3007–3021, doi:10.1002/bit.28198.

58. Dahlgren, K.K.; Gates, C.; Lee, T.; Cameron, J.C. Proximity-Based Proteomics Reveals the Thylakoid Lumen Proteome in the Cyanobacterium Syn-echococcus Sp. PCC 7002. Photosynth. Res. 2021, 147, 177–195, doi:10.1007/s11120-020-00806-y.

59. Shinde, S.; Singapuri, S.; Jiang, Z.; Long, B.; Wilcox, D.; Klatt, C.; Jones, J.A.; Yuan, J.S.; Wang, X. Thermodynamics Contributes to High Limonene Productivity in Cyanobacteria. Metab. Eng. Commun. 2022, 14, e00193, doi:10.1016/j.mec.2022.e00193.

60. De Porcellinis, A.; Frigaard, N.-U.; Sakuragi, Y. Determination of the Glycogen Content in Cyanobacteria. J. Vis. Exp. JoVE 2017, 56068, doi:10.3791/56068.

61. Hood, R.D.; Higgins, S.A.; Flamholz, A.; Nichols, R.J.; Savage, D.F. The Stringent Response Regulates Adaptation to Darkness in the Cyanobacterium Synechococcus Elongatus. Proc. Natl. Acad. Sci. 2016, 113, E4867–E4876, doi:10.1073/pnas.1524915113.

62. Bürgmann, H.; Pesaro, M.; Widmer, F.; Zeyer, J. A Strategy for Optimizing Quality and Quantity of DNA Extracted from Soil. J. Microbiol. Methods 2001, 45, 7–20, doi:10.1016/S0167-7012(01)00213-5.

63. Scharf, S.; Bartels, A.; Kondakci, M.; Pfeffer, K.; Henrich, B.; Haas, R. Introduction of a Bead Beating Step Improves Fungal DNA Extraction from Selected Patient Specimens. Int. J. Med. Microbiol. 2020, 310, 151443, doi:10.1016/j.ijmm.2020.151443.

64. D'Hondt, E.; Martín-Juárez, J.; Bolado, S.; Kasperoviciene, J.; Koreiviene, J.; Sulcius, S.; Elst, K.; Bastiaens, L. 6 - Cell Disruption Technologies. In Microal-gae-Based Biofuels and Bioproducts; Gonzalez-Fernandez, C., Muñoz, R., Eds.; Woodhead Publishing Series in Energy; Woodhead Publishing, 2017; pp. 133–154 ISBN 978-0-08-101023-5.

65. Morin, N.; Vallaeys, T.; Hendrickx, L.; Natalie, L.; Wilmotte, A. An Efficient DNA Isolation Protocol for Filamentous Cyanobacteria of the Genus Ar-throspira. J. Microbiol. Methods 2010, 80, 148–154, doi:10.1016/j.mimet.2009.11.012.

66. Wu, X.; Hou, L.; Lin, X.; Xie, Z. Chapter 12 - Application of Novel Nanomaterials for Chemo- and Biosensing of Algal Toxins in Shellfish and Water. In Novel Nanomaterials

for Biomedical, Environmental and Energy Applications; Wang, X., Chen, X., Eds.; Micro and Nano Technologies; Elsevier, 2019; pp. 353–414 ISBN 978-0-12-814497-8.

67. Verschoor, M.J.; Powe, C.R.; McQuay, E.; Schiff, S.L.; Venkiteswaran, J.J.; Li, J.; Molot, L.A. Internal Iron Loading and Warm Temperatures Are Precondi-tions for Cyanobacterial Dominance in Embayments along Georgian Bay, Great Lakes. Can. J. Fish. Aquat. Sci. 2017, 74, 1439–1453, doi:10.1139/cjfas-2016-0377.

68. McKay, R.M.; Frenken, T.; Diep, N.; Cody, W.R.; Crevecoeur, S.; Dove, A.; Drouillard, K.G.; Ortiz, X.; Wintermute, J.; Zastepa, A. Bloom Announcement: An Early Autumn Cyanobacterial Bloom Co-Dominated by Aphanizomenon Flos-Aquae and Planktothrix Agardhii in an Agriculturally-Influenced Great Lakes Tributary (Thames River, Ontario, Canada). Data Brief 2020, 30, 105585, doi:10.1016/j.dib.2020.105585.

69. Bukowska, A.; Bielczyńska, A.; Karnkowska, A.; Chróst, R.J.; Jasser, I. Molecular (PCR-DGGE) versus Morphological Approach: Analysis of Taxonomic Composition of Potentially Toxic Cyanobacteria in Freshwater Lakes. Aquat. Biosyst. 2014, 10, 2, doi:10.1186/2046-9063-10-2.

70. Almuhtaram, H.; Kibuye, F.A.; Ajjampur, S.; Glover, C.M.; Hofmann, R.; Gaget, V.; Owen, C.; Wert, E.C.; Zamyadi, A. State of Knowledge on Early Warning Tools for Cyanobacteria Detection. Ecol. Indic. 2021, 133, 108442, doi:10.1016/j.ecolind.2021.108442.

71. Chen, K.; Allen, J.; Lu, J. Community Structures of Phytoplankton with Emphasis on Toxic Cyanobacteria in an Ohio Inland Lake during Bloom Season. J. Water Resour. Prot. 2017, 9, 1–29, doi:10.4236/jwarp.2017.911083.

72. Vadrucci, M.R.; Roselli, L.; Castelluccia, D.; Di Festa, T.; Donadei, D.; Florio, M.; Longo, E.; D'Arpa, S.; Maci, F.; Ranieri, S.; et al. PhytoNumb3rs: An Easy-to-Use Computer Toolkit for Counting Microalgae by the Utermöhl Method. Ecol. Inform. 2018, 46, 147–155, doi:10.1016/j.ecoinf.2018.06.007.

73. Doktycz, M.J.; Sullivan, C.J.; Hoyt, P.R.; Pelletier, D.A.; Wu, S.; Allison, D.P. AFM Imaging of Bacteria in Liquid Media Immobilized on Gelatin Coated Mica Surfaces. Ultramicroscopy 2003, 97, 209–216, doi:10.1016/S0304-3991(03)00045-7.

74. Guan, D.; Barraud, C.; Charlaix, E.; Tong, P. Noncontact Viscoelastic Measurement of Polymer Thin Films in a Liquid Medium Using Long-Needle Atomic Force Microscopy. Langmuir 2017, 33, 1385–1390, doi:10.1021/acs.langmuir.6b04066.

75. Qian, S.S.; Chaffin, J.D.; DuFour, M.R.; Sherman, J.J.; Golnick, P.C.; Collier, C.D.; Nummer, S.A.; Margida, M.G. Quantifying and Reducing Uncertainty in Estimated Microcystin Concentrations from the ELISA Method. Environ. Sci. Technol. 2015, 49, 14221–14229, doi:10.1021/acs.est.5b03029.

76. Fischer, W.J.; Garthwaite, I.; Miles, C.O.; Ross, K.M.; Aggen, J.B.; Chamberlin, A.R.; Towers, N.R.; Dietrich, D.R. Congener-Independent Immunoassay for Microcystins and Nodularins. Environ. Sci. Technol. 2001, 35, 4849–4856, doi:10.1021/es011182f.

77. Lu, J.; Struewing, I.; Wymer, L.; Tettenhorst, D.R.; Shoemaker, J.; Allen, J. Use of QPCR and RT-QPCR for Monitoring Variations of Microcystin Producers and as an Early Warning System to Predict Toxin Production in an Ohio Inland Lake. Water Res. 2020, 170, 115262, doi:10.1016/j.watres.2019.115262.

78. Health Canada Guidelines for Canadian Drinking Water Quality: Guideline Technical Document – Cyanobacterial Toxins Available online: https://www.canada.ca/en/health-canada/services/publications/healthy-living/guidelines-canadian-drinking-water-quality-guideline-technical-document-cyanobacterial-toxins-document.html (accessed on 18 February 2023).

79. Douglas Greene, S.B.; LeFevre, G.H.; Markfort, C.D. Improving the Spatial and Temporal Monitoring of Cyanotoxins in Iowa Lakes Using a Multiscale and Multi-Modal Monitoring Approach. Sci. Total Environ. 2021, 760, 143327, doi:10.1016/j.scitotenv.2020.143327.

80. Mountfort, Douglas.O.; Holland, P.; Sprosen, J. Method for Detecting Classes of Microcystins by Combination of Protein Phosphatase Inhibition Assay and ELISA: Comparison with LC-MS. Toxicon 2005, 45, 199–206, doi:10.1016/j.toxicon.2004.10.008.

81. He, X.; Stanford, B.D.; Adams, C.; Rosenfeldt, E.J.; Wert, E.C. Varied Influence of Microcystin Structural Difference on ELISA Cross-Reactivity and Chlo-rination Efficiency of Congener Mixtures. Water Res. 2017, 126, 515–523, doi:10.1016/j.watres.2017.09.037.

82. Aranda-Rodriguez, R.; Jin, Z. Evaluation of Field Test Kits to Detect Microcystins: 2010 Study; Health Canada, 2011;

83. Birbeck, J.A.; Westrick, J.A.; O'Neill, G.M.; Spies, B.; Szlag, D.C. Comparative Analysis of Microcystin Prevalence in Michigan Lakes by Online Concen-tration LC/MS/MS and ELISA. Toxins 2019, 11, 13, doi:10.3390/toxins11010013.

84. Coskun, O. Separation Techniques: Chromatography. North. Clin. Istanb. 2016, 3, 156–160, doi:10.14744/nci.2016.32757.

85. Flores, C.; Caixach, J. An Integrated Strategy for Rapid and Accurate Determination of Free and Cell-Bound Microcystins and Related Peptides in Natu-ral Blooms by Liquid Chromatography–Electrospray-High Resolution Mass Spectrometry and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight/Time-of-Flight Mass Spectrometry Using Both Positive and Negative Ionization Modes. J. Chromatogr. A 2015, 1407, 76–89, doi:10.1016/j.chroma.2015.06.022.

86. Filatova, D.; Núñez, O.; Farré, M. Ultra-Trace Analysis of Cyanotoxins by Liquid Chromatography Coupled to High-Resolution Mass Spectrometry. Toxins 2020, 12, 247, doi:10.3390/toxins12040247.

87. Greer, B.; McNamee, S.E.; Boots, B.; Cimarelli, L.; Guillebault, D.; Helmi, K.; Marcheggiani, S.; Panaiotov, S.; Breitenbach, U.; Akçaalan, R.; et al. A Vali-dated UPLC-MS/MS Method for the Surveillance of Ten Aquatic Biotoxins in European Brackish and Freshwater Systems. Harmful Algae 2016, 55, 31–40, doi:10.1016/j.hal.2016.01.006.

88. Lawrence, J.F.; Niedzwiadek, B.; Menard, C.; Lau, B.P.Y.; Lewis, D.; Kuper-Goodman, T.; Carbone, S.; Holmes, C. Comparison of Liquid Chromatog-raphy/Mass Spectrometry, ELISA, and Phosphatase Assay for the Determination of Microcystins in Blue-Green Algae Products. J. AOAC Int. 2001, 84, 1035–1044, doi:10.1093/jaoac/84.4.1035.

89. Kumar, P.; Rautela, A.; Kesari, V.; Szlag, D.; Westrick, J.; Kumar, S. Recent Developments in the Methods of Quantitative Analysis of Microcystins. J. Biochem. Mol. Toxicol. 2020, 34, e22582, doi:10.1002/jbt.22582.

90. Gaget, V.; Lau, M.; Sendall, B.; Froscio, S.; Humpage, A.R. Cyanotoxins: Which Detection Technique for an Optimum Risk Assessment? Water Res. 2017, 118, 227–238, doi:10.1016/j.watres.2017.04.025.

91. Kralik, P.; Ricchi, M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. Front. Microbiol. 2017, 8, doi:10.3389/fmicb.2017.00108.

92. Pinto, F.; Pacheco, C.C.; Ferreira, D.; Moradas-Ferreira, P.; Tamagnini, P. Selection of Suitable Reference Genes for RT-QPCR Analyses in Cyanobacteria. PloS One 2012, 7, e34983, doi:10.1371/journal.pone.0034983.

93. Barón-Sola, Á.; Campo, F.F. del; Sanz-Alférez, S. Dynamics of Cylindrospermopsin Production and Toxin Gene Expression in Aphanizomenon Ovalisporum. Adv. Microbiol. 2016, 6, 381–390, doi:10.4236/aim.2016.65037.

94. Chiu, Y.-T.; Chen, Y.-H.; Wang, T.-S.; Yen, H.-K.; Lin, T.-F. A QPCR-Based Tool to Diagnose the Presence of Harmful Cyanobacteria and Cyanotoxins in Drinking Water Sources. Int. J. Environ. Res. Public. Health 2017, 14, 547, doi:10.3390/ijerph14050547.

95. Oh, K.-H.; Jeong, D.-H.; Shin, S.-H.; Cho, Y.-C. Simultaneous Quantification of Cyanobacteria and Microcystis Spp. Using Real-Time PCR. J. Microbiol. Biotechnol. 2012, 22, 248–255, doi:10.4014/jmb.1109.09051.

96. Te, S.H.; Chen, E.Y.; Gin, K.Y.-H. Comparison of Quantitative PCR and Droplet Digital PCR Multiplex Assays for Two Genera of Bloom-Forming Cya-nobacteria, Cylindrospermopsis and Microcystis. Appl. Environ. Microbiol. 2015, 81, 5203–5211, doi:10.1128/AEM.00931-15.

97. Luo, X.; Li, J.; Chang, T.; He, H.; Zhao, Y.; Yang, X.; Zhao, Y.; Xu, Y. Stable Reference Gene Selection for RT-QPCR Analysis in Synechococcus Elongatus PCC 7942 under Abiotic Stresses. BioMed Res. Int. 2019, 2019, e7630601, doi:10.1155/2019/7630601.

98. Rinta-Kanto, J.M.; Ouellette, A.J.A.; Boyer, G.L.; Twiss, M.R.; Bridgeman, T.B.; Wilhelm, S.W. Quantification of Toxic Microcystis Spp. during the 2003 and 2004 Blooms in Western Lake Erie Using Quantitative Real-Time PCR. Environ. Sci. Technol. 2005, 39, 4198–4205, doi:10.1021/es048249u.

99. Khomutovska, N.; Sandzewicz, M.; Łach, Ł.; Suska-Malawska, M.; Chmielewska, M.; Mazur-Marzec, H.; Cegłowska, M.; Niyatbekov, T.; Wood, S.A.; Puddick, J.; et al.

Limited Microcystin, Anatoxin and Cylindrospermopsin Production by Cyanobacteria from Microbial Mats in Cold Deserts. Toxins 2020, 12, 244, doi:10.3390/toxins12040244.

100. Zupančič, M.; Kogovšek, P.; Šter, T.; Remec Rekar, Š.; Cerasino, L.; Baebler, Š.; Krivograd Klemenčič, A.; Eleršek, T. Potentially Toxic Planktic and Ben-thic Cyanobacteria in Slovenian Freshwater Bodies: Detection by Quantitative PCR. Toxins 2021, 13, 133, doi:10.3390/toxins13020133.

101. Li, X.; Huo, S.; Zhang, J.; Ma, C.; Xiao, Z.; Zhang, H.; Xi, B.; Xia, X. Metabarcoding Reveals a More Complex Cyanobacterial Community than Morpho-logical Identification. Ecol. Indic. 2019, 107, 105653, doi:10.1016/j.ecolind.2019.105653.

102. Casero, M.C.; Velázquez, D.; Medina-Cobo, M.; Quesada, A.; Cirés, S. Unmasking the Identity of Toxigenic Cyanobacteria Driving a Multi-Toxin Bloom by High-Throughput Sequencing of Cyanotoxins Genes and 16S RRNA Metabarcoding. Sci. Total Environ. 2019, 665, 367–378, doi:10.1016/j.scitotenv.2019.02.083.

103. MacKeigan, P.W.; Garner, R.E.; Monchamp, M.-È.; Walsh, D.A.; Onana, V.E.; Kraemer, S.A.; Pick, F.R.; Beisner, B.E.; Agbeti, M.D.; da Costa, N.B.; et al. Comparing Microscopy and DNA Metabarcoding Techniques for Identifying Cyanobacteria Assemblages across Hundreds of Lakes. Harmful Algae 2022, 113, 102187, doi:10.1016/j.hal.2022.102187.

104. Jungblut, A.-D.; Neilan, B.A. Molecular Identification and Evolution of the Cyclic Peptide Hepatotoxins, Microcystin and Nodularin, Synthetase Genes in Three Orders of Cyanobacteria. Arch. Microbiol. 2006, 185, 107–114, doi:10.1007/s00203-005-0073-5.

105. Tillett, D.; Parker, D.L.; Neilan, B.A. Detection of Toxigenicity by a Probe for the Microcystin Synthetase A Gene (McyA) of the Cyanobacterial Genus Microcystis: Comparison of Toxicities with 16S RRNA and Phycocyanin Operon (Phycocyanin Intergenic Spacer) Phylogenies. Appl. Environ. Microbiol. 2001, 67, 2810–2818, doi:10.1128/AEM.67.6.2810-2818.2001.

106. Furukawa, K.; Noda, N.; Tsuneda, S.; Saito, T.; Itayama, T.; Inamori, Y. Highly Sensitive Real-Time PCR Assay for Quantification of Toxic Cyanobacteria Based on Microcystin Synthetase A Gene. J. Biosci. Bioeng. 2006, 102, 90–96, doi:10.1263/jbb.102.90.

107. Ngwa, F.; Madramootoo, C.; Jabaji, S. Monitoring Toxigenic Microcystis Strains in the Missisquoi Bay, Quebec, by PCR Targeting Multiple Toxic Gene Loci. Environ. Toxicol. 2014, 29, 440–451, doi:10.1002/tox.21770.

108. Ngwa, F.F.; Madramootoo, C.A.; Jabaji, S. Development and Application of a Multiplex QPCR Technique to Detect Multiple Microcystin-Producing Cyanobacterial Genera in a Canadian Freshwater Lake. J. Appl. Phycol. 2014, 26, 1675–1687, doi:10.1007/s10811-013-0199-9.

109. Gupta, R.S. Impact of Genomics on Clarifying the Evolutionary Relationships amongst Mycobacteria: Identification of Molecular Signatures Specific for the Tuberculosis-Complex of Bacteria with Potential Applications for Novel Diagnostics and Therapeutics. High-Throughput 2018, 7, 31, doi:10.3390/ht7040031.

110. Gupta, R.S.; Mathews, D.W. Signature Proteins for the Major Clades of Cyanobacteria. BMC Evol. Biol. 2010, 10, 24, doi:10.1186/1471-2148-10-24.

111. Rinta-Kanto, J.M.; Konopko, E.A.; DeBruyn, J.M.; Bourbonniere, R.A.; Boyer, G.L.; Wilhelm, S.W. Lake Erie Microcystis: Relationship between Microcys-tin Production, Dynamics of Genotypes and Environmental Parameters in a Large Lake. Harmful Algae 2009, 8, 665–673, doi:10.1016/j.hal.2008.12.004.

112. Duan, X.; Zhang, C.; Struewing, I.; Li, X.; Allen, J.; Lu, J. Cyanotoxin-Encoding Genes as Powerful Predictors of Cyanotoxin Production during Harmful Cyanobacterial Blooms in an Inland Freshwater Lake: Evaluating a Novel Early-Warning System. Sci. Total Environ. 2022, 830, 154568, doi:10.1016/j.scitotenv.2022.154568.

113. Crevecoeur, S.; Edge, T.A.; Watson, L.C.; Watson, S.B.; Greer, C.W.; Ciborowski, J.J.H.; Diep, N.; Dove, A.; Drouillard, K.G.; Frenken, T.; et al. Spa-tio-Temporal Connectivity of the Aquatic Microbiome Associated with Cyanobacterial Blooms along a Great Lake Riverine-Lacustrine Continuum. Front. Microbiol. 2023, 14, doi:10.3389/fmicb.2023.1073753.

114. Eiler, A.; Gonzalez-Rey, C.; Allen, S.; Bertilsson, S. Growth Response of Vibrio Cholerae and Other Vibrio Spp. to Cyanobacterial Dissolved Organic Matter and Temperature in Brackish Water. FEMS Microbiol. Ecol. 2007, 60, 411–418, doi:10.1111/j.1574-6941.2007.00303.x.

115. Gaget, V.; Hobson, P.; Keulen, A.; Newton, K.; Monis, P.; Humpage, A.R.; Weyrich, L.S.; Brookes, J.D. Toolbox for the Sampling and Monitoring of Benthic Cyanobacteria. Water Res. 2020, 169, 115222, doi:10.1016/j.watres.2019.115222.

116. Dorevitch, S.; Shrestha, A.; DeFlorio-Barker, S.; Breitenbach, C.; Heimler, I. Monitoring Urban Beaches with QPCR vs. Culture Measures of Fecal Indi-cator Bacteria: Implications for Public Notification. Environ. Health 2017, 16, 45, doi:10.1186/s12940-017-0256-y.

117. Saleem, F.; Edge, T.A.; Schellhorn, H.E. Validation of QPCR Method for Enterococci Quantification at Toronto Beaches: Application for Rapid Recrea-tional Water Monitoring. J. Gt. Lakes Res. 2022, 48, 707–716, doi:10.1016/j.jglr.2022.02.008.

118. Knob, R.; Hanson, R.L.; Tateoka, O.B.; Wood, R.L.; Guerrero-Arguero, I.; Robison, R.A.; Pitt, W.G.; Woolley, A.T. Sequence-Specific Sepsis-Related DNA Capture and Fluorescent Labeling in Monoliths Prepared by Single-Step Photopolymerization in Microfluidic Devices. J. Chromatogr. A 2018, 1562, 12–18, doi:10.1016/j.chroma.2018.05.042.

119. Wei, Y.-J.; Zhao, Y.-N.; Zhang, X.; Wei, X.; Chen, M.-L.; Chen, X.-W. Biochemical Analysis Based on Optical Detection Integrated Microfluidic Chip. TrAC Trends Anal. Chem. 2023, 158, 116865, doi:10.1016/j.trac.2022.116865.

120. Jiao, H. Handheld Microfluidic Device for Cyanobacteria Toxin Detection and Monitoring | Phase I; United States Environmental Protection Agency, 2013; Available at: https://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.abstractDetail/abstract_id/10003/report/F

121. Jiao, H. Handheld Microfluidic Device for Cyanobacteria Toxin Detection and Monitoring | Phase II; United States Environmental Protection Agency, 2017; Available at: https://cfpub.epa.gov/ncer_abstracts/index.cfm/fuseaction/display.abstractDetail/abstract_id/10416

122. Blanco, Y.; Quesada, A.; Gallardo-Carreño, I.; Aguirre, J.; Parro, V. CYANOCHIP: An Antibody Microarray for High-Taxonomical-Resolution Cyano-bacterial Monitoring. Environ. Sci. Technol. 2015, 49, 1611–1620, doi:10.1021/es5051106.

123. Blanco, Y.; Moreno-Paz, M.; Parro, V. Experimental Protocol for Detecting Cyanobacteria in Liquid and Solid Samples with an Antibody Microarray Chip. J. Vis. Exp. JoVE 2017, 54994, doi:10.3791/54994.

124. Ölcer, Z.; Esen, E.; Ersoy, A.; Budak, S.; Sever Kaya, D.; Yağmur Gök, M.; Barut, S.; Üstek, D.; Uludag, Y. Microfluidics and Nanoparticles Based Am-perometric Biosensor

for the Detection of Cyanobacteria (Planktothrix Agardhii NIVA-CYA 116) DNA. Biosens. Bioelectron. 2015, 70, 426–432, doi:10.1016/j.bios.2015.03.052.

125. Johnson, B.N.; Mutharasan, R. A Cantilever Biosensor-Based Assay for Toxin-Producing Cyanobacteria Microcystis Aeruginosa Using 16S RRNA. Envi-ron. Sci. Technol. 2013, 47, 12333–12341, doi:10.1021/es402925k.

126. Rantala, A.; Rizzi, E.; Castiglioni, B.; De Bellis, G.; Sivonen, K. Identification of Hepatotoxin-Producing Cyanobacteria by DNA-Chip. Environ. Microbiol. 2008, 10, 653–664, doi:10.1111/j.1462-2920.2007.01488.x.

127. Sipari, H.; Rantala-Ylinen, A.; Jokela, J.; Oksanen, I.; Sivonen, K. Development of a Chip Assay and Quantitative PCR for Detecting Microcystin Synthe-tase E Gene Expression. Appl. Environ. Microbiol. 2010, 76, 3797–3805, doi:10.1128/AEM.00452-10.

128. Khan, M.; Mao, S.; Li, W.; Lin, J.-M. Microfluidic Devices in the Fast-Growing Domain of Single-Cell Analysis. Chem. – Eur. J. 2018, 24, 15398–15420, doi:10.1002/chem.201800305.

129. Shi, K.; Zhang, Y.; Qin, B.; Zhou, B. Remote Sensing of Cyanobacterial Blooms in Inland Waters: Present Knowledge and Future Challenges. Sci. Bull. 2019, 64, 1540–1556, doi:10.1016/j.scib.2019.07.002.

130. Mishra, S.; Stumpf, R.P.; Schaeffer, B.A.; Werdell, P.J.; Loftin, K.A.; Meredith, A. Measurement of Cyanobacterial Bloom Magnitude Using Satellite Re-mote Sensing. Sci. Rep. 2019, 9, 18310, doi:10.1038/s41598-019-54453-y.

131. Mishra, S.; Stumpf, R.P.; Schaeffer, B.; Werdell, P.J.; Loftin, K.A.; Meredith, A. Evaluation of a Satellite-Based Cyanobacteria Bloom Detection Algorithm Using Field-Measured Microcystin Data. Sci. Total Environ. 2021, 774, 145462, doi:10.1016/j.scitotenv.2021.145462.

132. Coffer, M.M.; Schaeffer, B.A.; Darling, J.A.; Urquhart, E.A.; Salls, W.B. Quantifying National and Regional Cyanobacterial Occurrence in US Lakes Using Satellite Remote Sensing. Ecol. Indic. 2020, 111, 105976, doi:10.1016/j.ecolind.2019.105976.

133. Stumpf, R.P.; Davis, T.W.; Wynne, T.T.; Graham, J.L.; Loftin, K.A.; Johengen, T.H.; Gossiaux, D.; Palladino, D.; Burtner, A. Challenges for Mapping Cy-anotoxin Patterns from Remote Sensing of Cyanobacteria. Harmful Algae 2016, 54, 160–173, doi:10.1016/j.hal.2016.01.005.

134. Beck, R.; Xu, M.; Zhan, S.; Liu, H.; Johansen, R.A.; Tong, S.; Yang, B.; Shu, S.; Wu, Q.; Wang, S.; et al. Comparison of Satellite Reflectance Algorithms for Estimating Phycocyanin Values and Cyanobacterial Total Biovolume in a Temperate Reservoir Using Coincident Hyperspectral Aircraft Imagery and Dense Coincident Surface Observations. Remote Sens. 2017, 9, 538, doi:10.3390/rs9060538.

135. United States Environmental Protection Agency Cyanobacteria Assessment Network (CyAN) Available online: https://www.epa.gov/water-research/cyanobacteria-assessment-network-cyan (accessed on 24 February 2023).

136. Environment and Climate Change Canada EOLakeWatch: Satellite Earth Observations for Lake Monitoring Available online: https://www.canada.ca/en/environment-climate-change/services/water-overview/satellite-earth-observations-lake-monitoring.html (accessed on 26 February 2023).

137. Environment and Climate Change Canada Remote Sensing of Algal Blooms Available online: https://www.canada.ca/en/environment-climate-change/services/water-overview/satellite-earth-observations-lake-monitoring/remote-sensing-algal-blooms.html (accessed on 26 February 2023).

138. Binding, C.E.; Zastepa, A.; Zeng, C. The Impact of Phytoplankton Community Composition on Optical Properties and Satellite Observations of the 2017 Western Lake Erie Algal Bloom. J. Gt. Lakes Res. 2019, 45, 573–586, doi:10.1016/j.jglr.2018.11.015.

139. Cao, H.; Recknagel, F.; Bartkow, M. Spatially-Explicit Forecasting of Cyanobacteria Assemblages in Freshwater Lakes by Multi-Objective Hybrid Evo-lutionary Algorithms. Ecol. Model. 2016, 342, 97–112, doi:10.1016/j.ecolmodel.2016.09.024.

140. Chen, Q.; Guan, T.; Yun, L.; Li, R.; Recknagel, F. Online Forecasting Chlorophyll a Concentrations by an Auto-Regressive Integrated Moving Average Model: Feasibilities and Potentials. Harmful Algae 2015, 43, 58–65, doi:10.1016/j.hal.2015.01.002.

141. Baek, S.-S.; Pyo, J.; Pachepsky, Y.; Park, Y.; Ligaray, M.; Ahn, C.-Y.; Kim, Y.-H.; Ahn Chun, J.; Hwa Cho, K. Identification and Enumeration of Cyanobac-teria Species Using

a Deep Neural Network. Ecol. Indic. 2020, 115, 106395, doi:10.1016/j.ecolind.2020.106395.

142. Gaur, A.; Pant, G.; Jalal, A.S. Computer-Aided Cyanobacterial Harmful Algae Blooms (CyanoHABs) Studies Based on Fused Artificial Intelligence (AI) Models. Algal Res. 2022, 67, 102842, doi:10.1016/j.algal.2022.102842.

143. Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A Review of the Application of Machine Learning in Water Quality Evaluation. Eco-Environ. Health 2022, 1, 107–116, doi:10.1016/j.eehl.2022.06.001.

144. Xu, T.; Coco, G.; Neale, M. A Predictive Model of Recreational Water Quality Based on Adaptive Synthetic Sampling Algorithms and Machine Learning. Water Res. 2020, 177, 115788, doi:10.1016/j.watres.2020.115788.

145. Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. Water 2022, 14, 1384, doi:10.3390/w14091384.

Characterization of Taxonomic and Functional Dynamics Associated with Harmful Algal Bloom Formation in Recreational Water Ecosystems

This chapter exemplifies the type of cutting-edge research necessary for advancing our understanding and management of Harmful Algal Blooms (HABs) in recreational waters. By evaluating the applicability and effectiveness of various molecular techniques, including metabarcoding, metagenomics, qPCR, and ELISA-based methods, this study provides a comprehensive analysis of HABs dynamics in the Great Lakes region.

*The text I present here is a **peer-reviewed manuscript published** in the Toxins journal. The formatting has been changed from the journal version to be consistent throughout the thesis.*

***Saleem, F**., Atrache, R., Jiang, J.L., Tran, K.L., Li, E., Paschos, A., Edge, T.A. and Schellhorn, H.E., 2024. Characterization of Taxonomic and Functional Dynamics Associated with Harmful Algal Bloom Formation in Recreational Water Ecosystems. Toxins, 16(6), p.263.*

Author Contributions

FS performed the experiments, analyzed the data, wrote the original manuscript, and revised it as required by the reviewers. EL, JLJ, RA, AP and KT helped with the experiments. HES and TAE supervised the study and reviewed the manuscript before submission to the journal.

Abstract

Harmful algal bloom (HAB) formation leads to the eutrophication of water ecosystems and may render recreational lakes unsuitable for human use. We evaluated the applicability and comparison of metabarcoding, metagenomics, qPCR, and ELISA-based methods for cyanobacteria/cyanotoxin detection in bloom and non-bloom sites for the Great Lakes region. DNA sequencing-based methods robustly identified differences between bloom and non-bloom samples (e.g., the relative prominence of *Anabaena* and *Planktothrix*). Shotgun sequencing strategies also identified the enrichment of metabolic genes typical of cyanobacteria in bloom samples, though toxin genes were not detected, suggesting deeper sequencing or PCR methods may be needed to detect low-abundance toxin genes. PCR and ELISA indicated microcystin levels and microcystin gene copies were significantly more abundant in bloom sites. However, not all bloom samples were positive for microcystin, possibly due to bloom development by non-toxin-producing species. Additionally, microcystin levels were significantly correlated (positively) with microcystin gene copy number but not with total cyanobacterial 16S gene copies. In summary, next-generation sequencing-based methods can identify specific taxonomic and functional targets, which can be used for absolute quantification methods (qPCR and ELISA) to augment conventional water monitoring strategies.

Keywords: harmful algal blooms; metagenomics; microcystin; next-generation sequencing; microbiology

1.4.1. Introduction

Harmful algal bloom (HAB) formation is a natural event caused by the mass increase in phytoplankton proliferation in water ecosystems and is mainly due to Cyanobacteria, Diatoms and Dinoflagellates [1]. Among the major constituents of algal blooms, Cyanobacteria (blue-green algae) are of particular concern for the sustainability of freshwater ecosystems. Under favorable environmental conditions, Cyanobacteria can mass proliferate, thus leading to the deterioration of recreational water quality [2,3]. Several cyanobacterial species can produce cyanotoxins, leading to HABs, rendering the water ecosystems toxigenic and unusable for recreational purposes [4]. Environmental monitoring agencies monitor recreational waters for the formation/persistence of HABs and, depending on the detection methods/water quality thresholds, post warnings at those water bodies used for recreational purposes [5]. Closure of water bodies can last for months and ultimately result in recreational and economic losses [6]. However, bloom formation by itself may not necessarily be indicative of toxigenic Cyanobacteria [7], leading to unnecessary postings. Therefore, it is imperative to have robust HAB detection methods to augment existing water monitoring strategies.

Conventional detection methods are based on microscopy for cyanobacterial cell identification and cell counting [8], polymerase chain reaction for cyanobacterial/cyanotoxin gene copies detection/quantification [9], liquid chromatography/mass spectrometry (LC/MS) and enzyme-linked immunosorbent/chromatography assays for direct assessment of cyanotoxins [10]. Microcystin, the most commonly found cyanotoxin worldwide [11], is

a primary focus for detection and monitoring. The United States Environmental Protection Agency (USEPA) has developed a protocol for the determination of total microcystin/nodularins (MCs/NODs) in ambient waters using ADDA-ELISA [12], which targets a conserved part of microcystin. Additionally, the microcystin synthetase (*mcyE*) assay for qPCR can detect all microcystin congeners and the nodularin synthetase gene clusters [13]. In the environment, microcystins have more than 279 variants [14], encoded by the *mcy* gene cluster. PCR-based approaches can be useful in detecting/quantifying cyanotoxin genes and transcripts [15]. There is a strong positive correlation between microcystin gene transcripts and microcystin levels, suggesting PCR-based indicators as potential diagnostic measures [16]. Specifically, a positive correlation between PCR-based measures and microcystin levels has been observed in environmental waters from the US [16,17] and China [18]. However, the relationship between toxins and environmental factors, including internal loading [19] and temperature [20], is well established and has been validated for cyanotoxin prediction models for water ecosystems [19]. Specifically, internal nutrient loading and temperature can be primary drivers of weekly variations in toxin levels, while longer time frame changes in toxin levels are mainly associated with external nutrient loading [21].

Although environmental monitoring agencies widely use the conventional methods, each of the methods suffers from specific limitations, including an inability to detect unculturable Cyanobacteria [22], microscopy requires extensive knowledge of phytoplankton structures and does not determine toxin production [23], the inability of PCR-based methods to

distinguish between viable and non-viable cells [24], and potential cross-reactivity (false positive) using ELISA methods [25]. Additionally, conventional methods provide limited information regarding target-specific Cyanobacteria and cyanotoxins. However, cyanobacterial species/cyanotoxin levels differ from one environment to another [26], and limited information obtained from the conventional methods may not provide a comprehensive assessment for water quality monitoring.

Advancements in next-generation sequencing technologies have resulted in new analytical methodologies, such as metagenomics, that can provide a comprehensive view of microbial/cyanobacterial dynamics in recreational waters [27]. Amplicon sequencing/16S metabarcoding can provide only taxonomic information, while whole genome/shotgun sequencing can provide both taxonomic and functional changes in cyanobacterial communities [28]. Thus, implementing comprehensive information obtained from metagenomics with conventional methods can provide region-specific cyanobacterial trends, which can augment the water management/monitoring programs/strategies. However, the potential of metagenomics as an analytical technique still needs to be assessed and validated. In collaboration with the Ministry of the Environment, Conservation, and Parks (MECP), the current study evaluates the applicability and limitations of metagenomics and conventional ELISA and qPCR-based methods for cyanobacterial/algal bloom monitoring in the Great Lakes region. The questions addressed in this study are as follows: (1) Can metagenomics-based methods unambiguously identify Cyanobacteria species from complex recreational water samples? (2) Are there identifiable taxonomic and metabolic (functional

genes) changes associated with bloom-enriched and non-bloom complex water samples? (3) Is there any correlation between MC/NOD concentrations and cyanobacterial 16S rRNA gene copies, as well as the correlation between *mcyE* and MC/NOD, and between *mcyE* and cyanobacterial 16S rRNA gene copies?

1.4.2. Results

A total of 108 lake water samples were collected and categorized into bloom and non-bloom samples using microscopy. All the samples were processed for 16S metabarcoding, and then a subset of samples (n = 24) showing ≥10% cyanobacterial abundance was used for shotgun DNA sequencing, microcystin ELISA, and cyanotoxin gene qPCR.

Quality Analyses

The data quality or variability in qualitative parameters can impact results from DNA sequencing, qPCR, and ELISA methods. Before the data analysis, we assessed each dataset for quality analysis. Supplementary Table S1 demonstrates the quality control analytics for cyanobacteria 16S and cyanotoxin qPCR analysis. Standard curve values used to determine gene copy counts in qPCR assays are also listed in Supplementary Table S1. The total cyanobacterial 16S rRNA assay, as well as *mcyE*/*ndaF*, *CyrA*, and *SxtA* assays, demonstrated high linearity between 10 and 100,000 gene copies/reaction, with coefficients of determination (R2) ranging between 0.9996 and 0.9999. Intercept and slope values for generated standard curves ranged between 38.5 and 39.14 and between −3.25 and −3.46, respectively. The calculated efficiencies of each standard curve ranged from 94.432% to 101.17%.

More than 90% of DNA sequencing reads passed the quality control criteria described in the methods section.

Supplementary Table S2 details the quality control analytics for ELISA analysis. The seven LFBs' mean recovery was 117% ± 13%, demonstrating robust recovery. For the confirmation of the MRL (0.4 ng/mL), the upper prediction interval of results (PIR) limit was 115%, and the lower PIR limit was 54%. The five LRBs recovered less than 14% of the MRL to demonstrate an acceptable system background. For the analysis batch, the calibration curve had the coefficients a = 1.4, b = 0.73, c = 1.8, and d = 0.17, and a correlation coefficient (r2) of 0.99. Five standards had a coefficient of variation (CV) of less than 10%, and one standard had a CV of less than 12%. The four LRB wells recovered less than 5.8% of the MRL. The LFBs had recoveries between 112 and 140%. The low-CV had a recovery of 52%, and the QCS had a recovery of 85%. The LFSM/LFSMD had mean recoveries of 69–70%, CVs of 0.77–7.1%, and a relative difference of 1.0%. Lastly, the sample duplicates had CVs between 0.090 and 12%.

Taxonomic Composition Overview of Bloom and Non-Bloom Freshwater Lake Water Samples (Metabarcoding)

All the samples (blooms and non-blooms) were first evaluated for overall taxonomic diversity using 16S metabarcoding (Figure 1). For non-bloom samples, Proteobacteria (32–85%) was the most dominant bacterial group, followed by Actinobacteria (4–35%), Bacteroidetes (3–30%), and Cyanobacteria (0.6–25%). Similar to non-bloom samples, Proteobacteria (19–75%) were the most abundant phyla in bloom samples, followed by Bacteroidetes (11–60%), Verrucomicrobia (5–60%), Cyanobacteria (2–60%),

Actinobacteria (3–35%), and Planctomycetes (2–30%). Bloom samples also showed a higher prevalence of Planctomycetes and Verrucomicrobia than non-bloom samples.

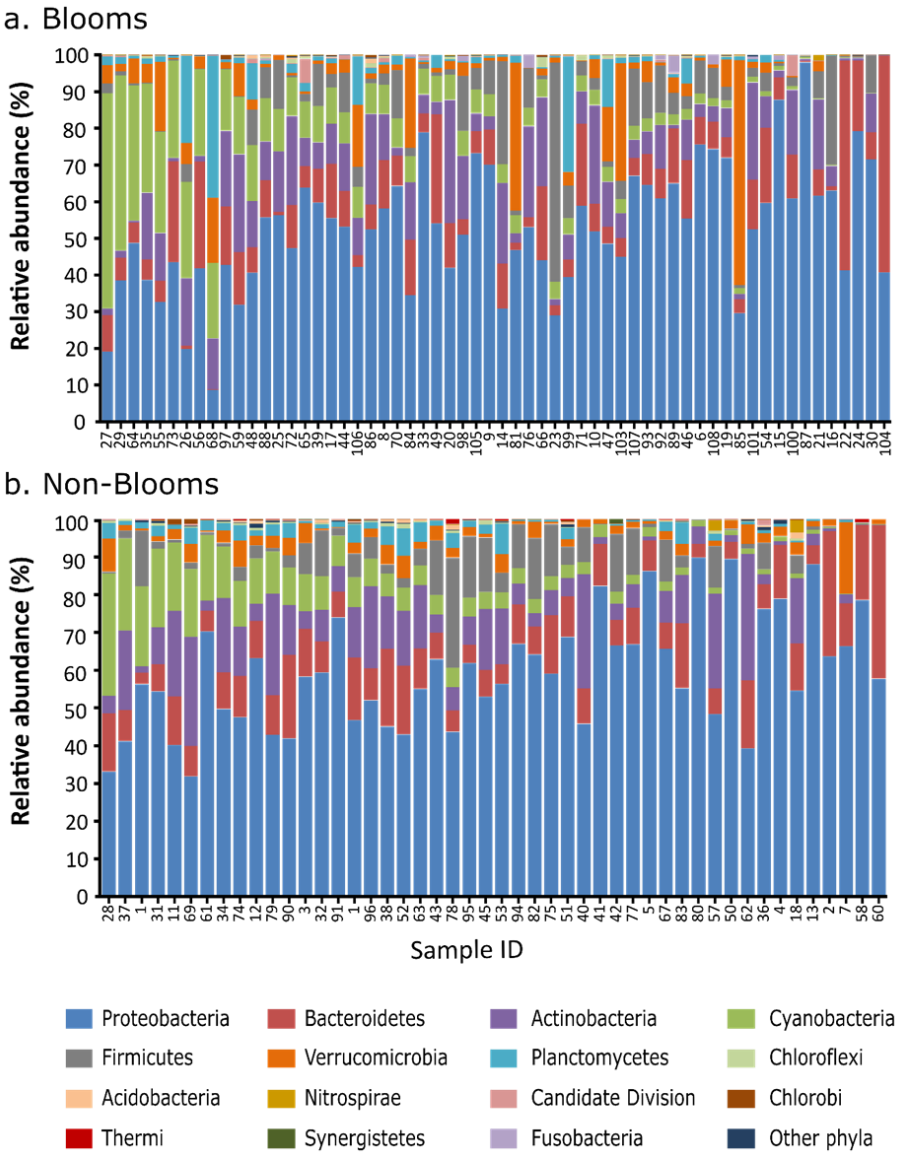


Figure 1. Taxonomic composition on phyla level for (**a**) bloom and (**b**) non-bloom samples using 16S metabarcoding.

Differential Taxonomic and Functional Composition of Bloom and Non-Bloom Freshwater Samples (Shotgun Metagenomics)

Twenty-four samples (12 bloom and 12 non-bloom) were analyzed using shotgun metagenomics to characterize taxonomic and functional dynamics associated with bloom and non-bloom samples. At the phylum level (Figure 2), Proteobacteria (20–75%) was most dominant for all non-bloom samples, while Cyanobacteria (10–90%) was most abundant in seven (60%) bloom samples. Similarly, compared to non-bloom samples, cyanobacterial abundance was significantly higher (p-value = 1.2×10^{-9} in bloom samples (Figures 3 and S1). Additionally, we observed an inversely proportional relationship between Bacteroidetes and Cyanobacteria for bloom samples. Cyanobacterial sequences were resolved to lower taxonomic levels to observe genus-level differences between bloom and non-bloom samples (Supplementary Figure S2). At the genus level, *Anabaena*, *Planktothrix*, *Microcystis*, and *Synechococcus* were abundant in both bloom and non-bloom samples (Supplementary Figure S2). Non-bloom samples were taxonomically (cyanobacterial) diverse, while bloom samples were dominated mainly by a single cyanobacterial genus (Supplementary Figure S2).

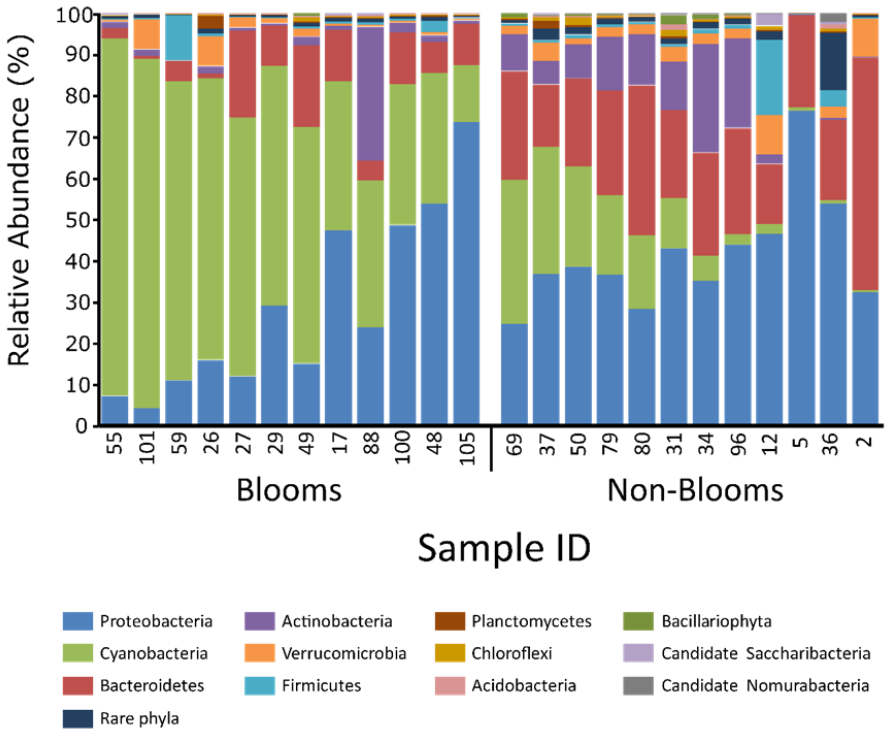


Figure 2. Taxonomic composition on phyla level for bloom and non-bloom samples using shotgun sequencing.

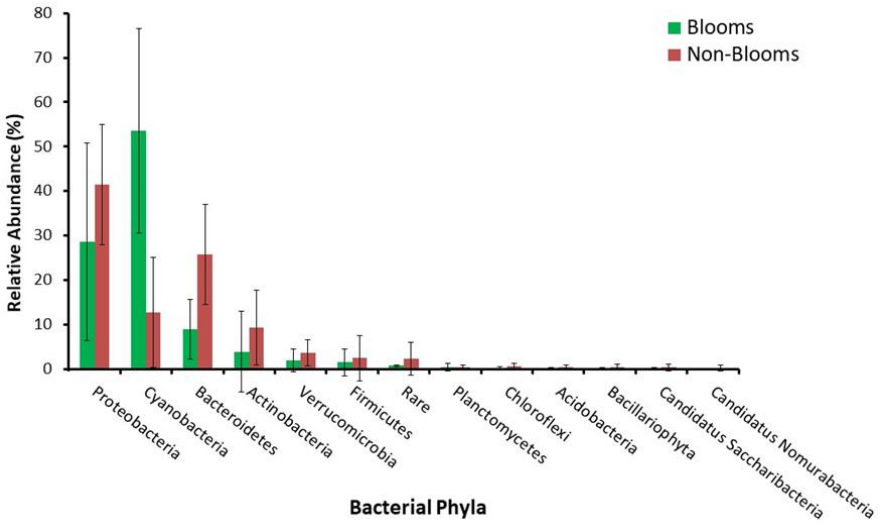


Figure 3. Differential abundance of bacterial phyla between bloom and non-bloom samples using shotgun sequencing.

Overall differential abundance profiling identified that *Anabaena* (p-value = 3.3×10^{16}) and *Planktothrix* (p-value = 9.1×10^{-10}) were significantly more abundant in bloom samples than non-bloom samples (Figures 4 and S3). The diversity analysis between the bloom and non-bloom samples on phylum and genus levels agglomerated most of the samples in close proximity to each other (Figure 5). However, diversity matrices using functional gene annotations showed differentiation between bloom and non-bloom samples.

Differential gene abundance profiling was performed to characterize specific functional differences between bloom and non-bloom samples (Supplementary Figure S4). Bloom samples demonstrated comparatively higher metabolic potential than non-bloom samples. Genes associated with cell division, cell wall, vitamins, co-factors and pigment synthesis, nucleic acid metabolism, cyanobacterial electron transport system, cyanobacterial heterocyst

formation, and β-carboxysomes were differentially abundant in bloom samples compared to non-bloom samples. Additionally, genes associated with nitrogen and phosphorus metabolism were also more abundant in bloom samples (Supplementary Table S3).

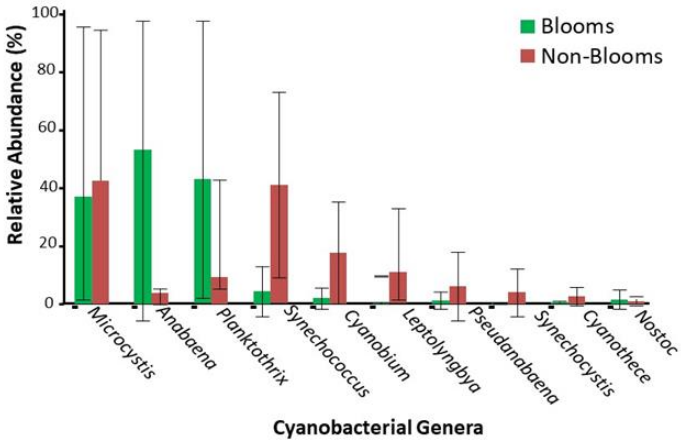


Figure 4. Differential abundance of bacterial genera between bloom and non-bloom samples using shotgun sequencing.

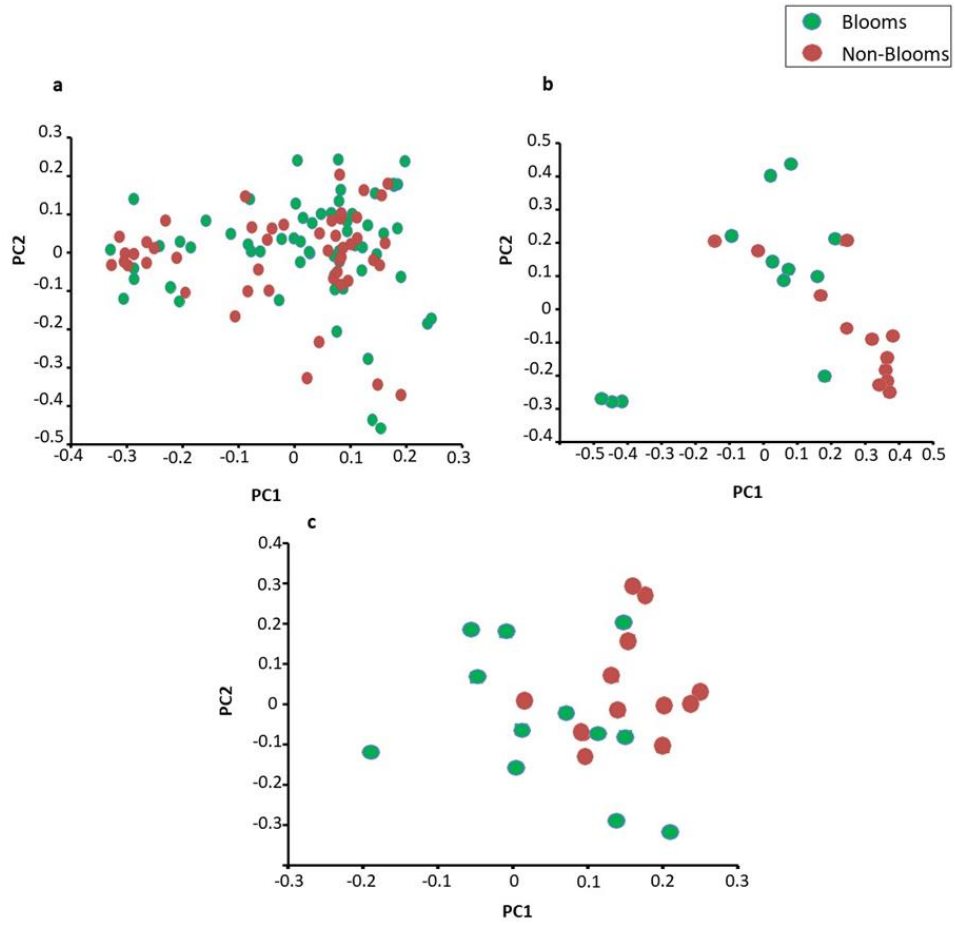


Figure 5. Beta diversity (PcoA) plots using weighted unifrac distance matrix on (**a**) order level (16S metabarcoding), (**b**) genus level (shotgun sequencing), and (**c**) functional (SEED subsystems level 2).

Trends of Microcystin/Nodularin Levels and Cyanotoxin Gene Copies in Bloom and Non-bloom Recreation Water Samples

Table 1 shows the microcystin/nodularin concentrations, 16S rRNA gene copies, and *mcyE* gene copies of bloom and non-bloom samples. Three out of six (50%) bloom samples had

MC/NOD levels (207, 73, and 3092 ng/mL, respectively) above the Health Canada recreational water guideline of 10 ng/mL, while two out of ten (20%) non-bloom samples (50, 69) had MC/NOD levels above the MRL, at 0.80 and 0.61 ng/mL, respectively. However, MC/NOD levels in non-bloom samples remained below the recreational water quality threshold.

Table 1. The microcystin/nodularin concentrations (ng/mL), total cyanobacterial 16S rRNA gene copies, *mcyE* gene copies, and *SxtA* gene copies of bloom and non-bloom samples. Only MC/NOD values above the MRL (0.40 ng/mL) were reported. Only gene copy number values higher than the lower limit of their standard curve range were reported. -—Not detected. [a]—Gene copies/5 ng of DNA.

Sample ID	Bloom Presence	Microcystins Nodularin (ng/mL)	16S rRNA [aPl]	*mcyE* [a]	*SxtA* [a]
26	Bloom	-	32,550	-	-
55		-	6013	11	-
59		210	10,039,974	534	86
100		73	23,499	15	-
101		-	6,430,707	15	-
105		3100	18,744	78	-
2	Non-bloom	-	102	-	-
5		-	3851	-	-
12		-	104,903	-	-
31		-	16,194	-	-
34		-	170	-	-
37		-	107,731	-	-
50		0.80	266,112	453	-
69		0.61	145,307	92	-
79		-	43,904	33	-
96		-	61,219	67	-

Both bloom and non-bloom samples were positive for cyanobacterial 16S sequences (Table 1). Additionally, five out of six bloom samples (83%) showed the presence of microcystin gene copies, while only four out of 10 non-bloom samples (40%) showed the presence of microcystin gene copies. Furthermore, only one bloom sample demonstrated the presence of saxitoxin genes, and all the tested samples were negative for the cylindrospermopsin gene. The mean gene copies of total cyanobacterial 16S rRNA and *mcyE* per 5 ng of DNA showed no significant differences between bloom and non-bloom samples (p-values = 0.19 and 0.83, respectively). Interestingly, all the bloom samples were positive for microcystin-producing genera, but detectable microcystin levels were only observed for samples predominated by *Anabaena* and *Microcystis*. Correlation analysis was performed to observe the relationship between microcystin levels and cyanobacterial 16S/microcystin gene copies. Total MCs/NODs and *mcyE* gene copies were moderately (positively) correlated, with R = 0.52 and p-value = 0.038. Cyanobacterial 16S rRNA and *mcyE* gene copies were also moderately correlated, with R = 0.62 and p-value = 0.010. The correlation between total MCs/NODs and cyanobacterial 16S rRNA gene copies was non-significant (R = 0.44, p = 0.092).

1.4.3. Discussion

Cyanobacterial blooms are a global concern for the sustainability of water ecosystems, including the Great Lakes. Harmful algal blooms (HABs) can synthesize cyanotoxins that lead to the eutrophication of recreational waters [29,30]. Thus, continuous monitoring of water sources is necessary to avoid economic and health risks associated with the enrichment of Harmful algal blooms. Bloom development may not necessarily indicate toxigenicity, as not all cyanobacterial communities can produce cyanotoxins [7]. Additionally, cyanotoxin

gene expression can be modulated by environmental factors, including nitrogen, phosphorus, and temperature [31], and conventional methods for cyanobacterial identification may not relate to cyanotoxin production or fully gauge the public health risks. Molecular techniques, including DNA sequencing, qualitative/quantitative PCR, and ELISA, can provide broad or targeted information that can circumvent the limitations associated with conventional methods. In this study, we are examining the potential of molecular technologies to augment traditional water monitoring strategies.

All the water samples (n = 108) were first analyzed using 16S metabarcoding to obtain a taxonomic overview of bloom-enriched and non-bloom samples. Verrucomicrobia and Planctomycetes sequences revealed a positive association with cyanobacterial abundance, which substantiates the studies that have identified these microbial taxa in eutrophic environments [32,33]. The association of Verrocomicrobia and Planctomycetes with cyanobacterial blooms can be due to a nutrient-rich environment, and both of these taxa have the potential to utilize polysaccharides synthesized by cyanobacterial communities [34,35]. Although 16S metabarcoding identified some taxonomic differences between the bloom and non-bloom samples, Proteobacteria sequences dominated most of the samples, and there was little difference in terms of cyanobacterial relative abundance. The inability of amplicon-based sequencing to differentiate between the bloom and non-bloom samples can be due to PCR bias that favors the over-amplification of dominant microbial taxa, including Proteobacteria [36,37]. Compared to 16S metabarcoding, shotgun metagenomics more efficiently characterized a greater abundance of cyanobacterial sequences in bloom-enriched

samples, substantiating concerns that Cyanobacteria are the causative agents of many bloom formations in the Great Lakes region [38,39].

Additionally, shotgun sequencing was used to characterize the enrichment of specific cyanotoxin-producing genera in bloom samples, further validating the potential of using metagenomics sequencing-based methods to obtain more robust taxonomic profiles for bloom-specific response strategies. β-carboxysomes genes were differentially more abundant in bloom-enriched samples, and these genes are found predominantly in freshwater Cyanobacteria [40], which suggests Cyanobacteria to be the primary metabolically active component of microbial communities in the bloom-enriched water ecosystems. Although multiple metabolic pathways characteristic of cyanobacteria were differentially abundant in bloom-enriched samples, cyanotoxin genes were not detected in our datasets. Despite cyanobacterial DNA sequences being abundant, cyanotoxin genes may not represent a dominant proportion of total DNA [41], which hinders the detection of these genes without a higher depth of DNA sequencing [42]. For real-time detection purposes, qPCR or ELISA-based methods would better evaluate the toxigenic potential of bloom-enriched complex water samples. Currently, in Canada, there are no standard definitions for bloom designation at the federal or provincial level. While the initial assessments for bloom designation in this study were based on an initial microscopic assessment, more objective, quantitative criteria would be desirable that include comprehensive toxin identification and assessment. This is challenging because of the large number of microcystin congeners as well as the potential presence of other types of toxins (e.g., saxitoxin and anatoxin) that are not routinely monitored.

We observed only moderate positive correlations between cyanobacterial 16S rRNA and *mcyE* gene copies and between total MCs/NODs and *mcyE* gene copies, which is in agreement with previous studies from Canada [43], South Africa [44], and the USA [15,45], showing positive correlations between microcystins and *mcyE* gene copies. The correlation between total MCs/NODs and cyanobacterial 16S rRNA gene copies was non-significant (supported by shotgun sequencing), possibly due to bloom enrichment by non-toxic cyanobacterial communities [46]. Therefore, total cyanobacterial PCR markers may not be suitable for water quality assessment (testing cyanotoxin production) for freshwater ecosystems [41,47,48]. Additionally, samples with lower total cyanobacterial gene copies can still show high levels of cyanotoxins, which may be due to the proliferation of toxin-producing Cyanobacteria [49].

The filtration of LFBs using nylon syringe filters significantly reduced the MC/NOD recovery, likely due to the protein-binding properties of nylon filters [50]. USEPA method 546 for ELISA-based microcystin analysis recommends glass fiber filters [12], but other studies have reported success using regenerated cellulose and PES filters [10,45]. We did not observe any differences in microcystin recovery with the change in freezing temperature (−20 and −80 °C), which is in agreement with studies using different freezing temperatures for microcystin analysis [10,12,51]. Using crude DNA for qPCR resulted in variability of threshold cycle (Ct) values between the replicates, possibly due to PCR inhibition by cellular material [52,53]. However, crude DNA identified higher gene copies than purified DNA from the same samples, possibly because of DNA loss during purification cycles [54]. Therefore, DNA purification can reduce PCR inhibition but may underrepresent the

cyanotoxin gene copies. We did not observe cyanotoxin genes in metagenomics datasets (including bloom samples), which may be due to the requirement of higher sequencing depth for cyanotoxicity detection using DNA sequencing-based methods. Although conventional methods, including PCR and ELISA for cyanobacterial identification, can provide results within a few hours, they are limited in terms of information obtained and can be laborious for practical purposes when targeting a diverse cyanobacterial population. In comparison, next-generation DNA sequencing-based methodologies, including metagenomics, can provide a broader and robust overview of complex cyanobacterial communities, which can be difficult to differentiate visually or require multiple genetic markers for identification [55]. Additionally, there can be site/region-specific proliferation of particular cyanobacteria [56], which may not be identified using conventional molecular markers. However, DNA sequencing-based methods can provide a comprehensive taxonomic profile, which can be used for developing more specific and region-specific monitoring strategies. Additionally, bloom formation may not necessarily indicate the proliferation of toxin-producing cyanobacterial communities, thus, DNA-sequencing methods should be supplemented with real-time detection methods, including ELISA and qPCR, for more robust identification of cyanobacterial diversity and cyanotoxicity. This study provides a comprehensive comparative overview of molecular methods for Cyanobacteria/cyanotoxin detection in freshwater ecosystems and can improve conventional water monitoring strategies.

1.4.4. Conclusions

Our study concludes the following that in freshwater ecosystems:

1. Non-bloom samples showed higher cyanobacterial taxonomic diversity compared to bloom samples, suggesting the dominance of specific Cyanobacteria genera during bloom formation.

2. Anabaena and *Planktothrix* were predominantly abundant in bloom samples than in non-bloom samples, possibly due to these species being common HAB bloom-forming algae in the Great Lakes region. Additionally, detectable microcystin levels were only observed for samples with a higher relative abundance of *Anabaena* and *Planktothrix*, suggesting their role in cyanotoxicity.

3. Compared to non-bloom samples, bloom samples showed significant cyanobacterial metabolic gene diversity, including nitrogen and phosphorus genes, which may represent a higher nutrient uptake and processing due to bloom enrichment.

4. Microcystin/nodularin levels positively correlated with *mcyE* gene copy numbers, indicating that microcystin gene copies can robustly estimate microcystin production in freshwater ecosystems.

5. Cyanobacterial 16S rRNA gene copy numbers showed a significant positive correlation with *mcyE* gene copies but not with microcystin/nodularin levels. Total cyanobacterial abundance may not indicate cyanotoxin production in freshwater ecosystems.

1.4.5. Materials and Methods

Study Design and Sample Collection

A total of 108 lake water samples were collected across the Great Lakes Region (Canada) by the MECP during the 2014 and 2015 sampling seasons (May to September). Figure 6 shows the locations of sampling sites for this study. Of these, 101 were collected from freshwater lakes, three from rivers, three from ponds, and one from a creek. The samples were categorized as bloom-enriched (n = 60) or non-blooms (n = 48) using light microscopy, and these same terminologies are employed throughout this manuscript. The samples were collected in 500 mL sterile polyethylene terephthalate (PET) bottles and stored at −80 °C. Samples were delivered to the McMaster lab in 2016 and stored at −80 °C until further use for DNA sequencing ELISA and qPCR testing.

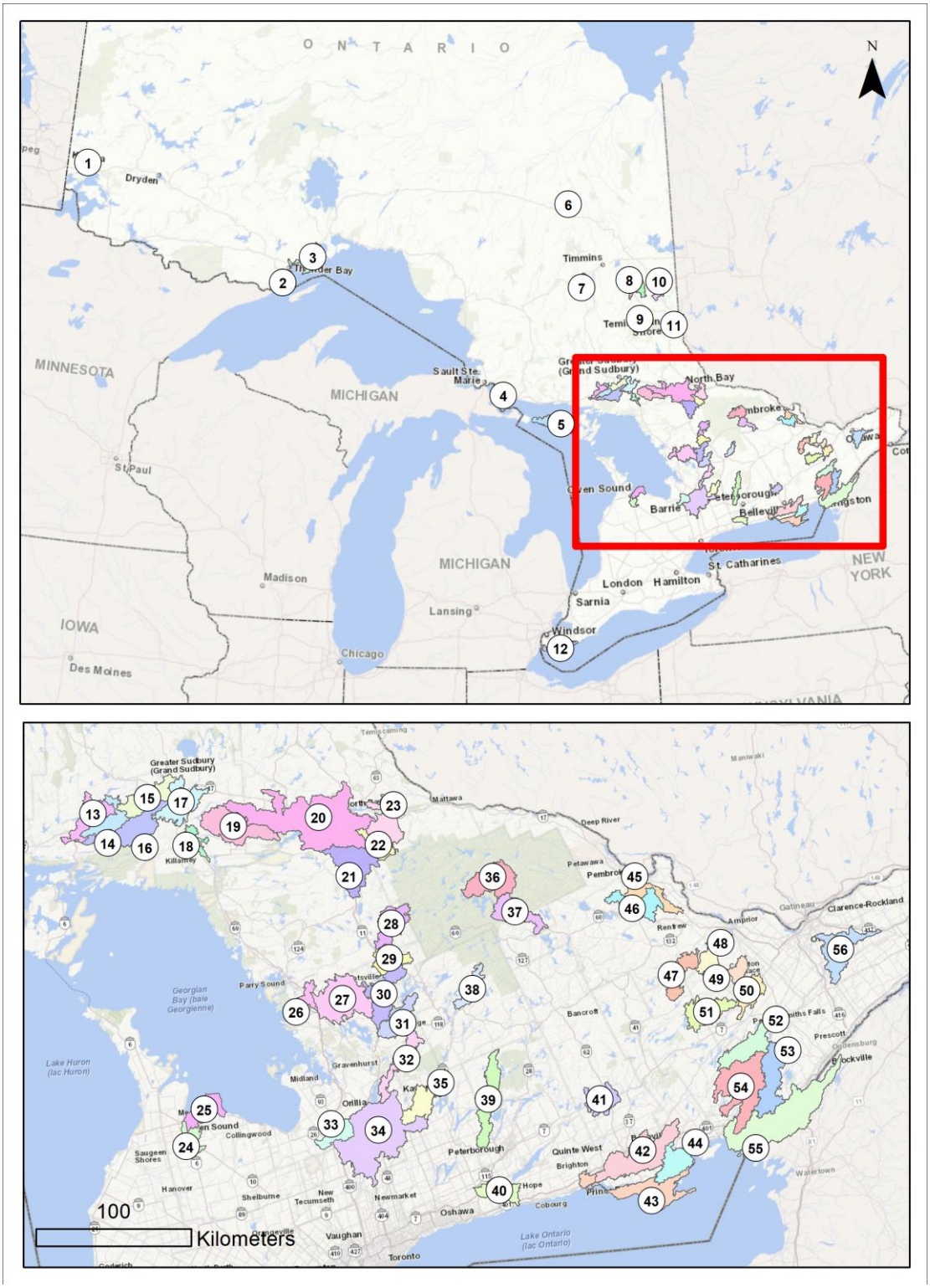

ONTARIO
N
MINNESOTA
MICHIGAN
IOWA
Dryden
Thunder Bay
Timmins
Sault Ste. Marie
North Bay
Owen Sound
Barrie
Peterborough
Bellevill
St Catharines
London
Hamilton
Sarnia
Windsor
St Paul
Madison
Lansing
Chicago
Des Moines
NEW YORK
Greater Sudbury (Grand Sudbury)
Mattawa
Deep River
Pembroke
Renfrew
Arnprior
Gatineau
Clarence-Rockland
Prescott
Brockville
Parry Sound
Georgian Bay (baie Georgienne)
Lake Huron (lac Huron)
Bancroft
Midland
Orillia
Gravenhurst
Collingwood
Saugeen Shores
Hanover
Shelburne
New Tecumseth
Newmarket
Oshawa
Vaughan
Toronto
Lake Ontario (lac Ontario)
Quinte West
Brighton
Cobourg
Hope
Watertown
100
Kilometers

Figure 6. Sampling locations employed in this study. (Contains information licensed under the Open Government License—Ontario).

DNA Extraction

Water samples were centrifuged at 10,000× *g* for 15 min. The supernatant was discarded, and the DNA was extracted from the cell pellet using the Norgen Soil Plus Extraction Kit (Norgen Biotek, *Thorold, ON,* Canada). Additionally, 10 µg/mL of lysozyme and 200 mM of β-mercaptoethanol were added to maximize the cyanobacterial cell lysis [57]. DNA was eluted in a final volume of 150 µL (TE Buffer) and stored at −80 °C.

DNA Sequencing Library Preparation

DNA extracted from all the samples (n = 108) was used for the Amplicon Sequencing (16S metabarcoding), while a smaller subset (n = 24, 12 bloom and 12 non-bloom) of samples was used for shotgun sequencing. For 16S metabarcoding library preparation, a 16S sequencing protocol (support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html ; accessed on 1 June 2023) was followed. Briefly, the V3–V4 region (550 bp) of the bacterial 16S rRNA gene was amplified using DreamTaq Hotstart Mastermix (Thermo Scientific, Waltham, MA, USA). The PCR reaction mix was composed of 12.5 µL of master mix, 2.5 µL (1.0 µmole) of Forward Primer, 2.5 µL (1.0 µmole) of Reverse Primer, 5.0 µL of DNA, and 2.5 µL nuclease-free water. The amplicon PCR protocol included initial denaturation at 95 °C for 3 min, 35 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, followed by

final extension at 72 °C for 5 min. Amplicon PCR products were purified using AmPure XP beads (Beckman Coulter Inc., Brea, CA, USA). Purified PCR products were subjected to index PCR using a Nextera XT index kit (Illumina Inc., *Hayward,* CA, USA). The second round of purification was performed using AmPure XP beads, and purified DNA libraries were pooled in equimolar concentrations after DNA quantification using QUBIT fluorometer (Thermo Scientific, MA, USA). All the PCR reactions were performed on a CFX-96 Touch PCR Detection System (BioRad, Santa Rosa, CA, USA). The shotgun sequencing library was prepared using Nextera XT DNA Sample Prep Kit (Illumina Inc., CA, USA). DNA sequencing was performed on MiSeq and HiSeq DNA sequencing platforms for 16S metabarcoding and shotgun sequencing at the Farncombe Sequencing facility (McMaster University, Hamilton, ON, Canada).

Metagenomics Data Analysis

Subsequently, 16S metabarcoding (Amplicon Sequencing) reads were processed using Quantitative Insights into Microbial Ecology (QIIME) using default parameters [58,59]. First, DNA sequence quality was estimated using FastQC [60], followed by adapter removal using Trimmomatic [61]. Sequencing reads were filtered to remove sequences with quality thresholds and lengths lower than 25 and 60, respectively. Quality-filtered reads were clustered into operational taxonomic units (OTUs) at 97% sequence similarity and assigned to their respective taxa using the RDP Classifier [62] against the Greengenes Database [63]. For shotgun sequencing, quality-filtered reads were aligned against the NCBI RefSeq database using the DIAMOND aligner [64], followed by taxonomic and functional

annotation using Metagenome Analyzer (MEGAN) with an e-value cutoff of 1.0×10^{-5} [65,66]. All the statistical analysis was performed on R using the VEGAN [67] and DESeq2 [68] library packages.

Total Microcystins/Nodularins Determination by ELISA

ADDA-ELISA 96-well kits for microcystins/nodularins (MCs/NODs) were obtained from Gold Standard Diagnostics (Davis, CA, USA). Sample preparation was conducted according to of USEPA Method 546 [12]. Samples were lysed by three freeze–thaw cycles (freeze at −80 °C and thaw at 35 °C in a water bath) and filtered using 0.45 um PES membrane syringe filters (Cytiva, Marlborough, MA, USA). The immunosorbent assay was then conducted according to the manufacturer's instructions (available at https://www.goldstandarddiagnostics.com/pub/media/productattachments/files/m/i/microcystins_nodularins-adda-elisa-user-guide-520011.pdf; accessed on 1 June 2023). The kit utilizes an indirect ELISA, with wells that are pre-coated with an analog of microcystins conjugated to a protein. In total, 50 µL of the standards, controls, and samples were first loaded into the wells in duplicates. This step was followed by adding 50 µL of the primary antibody solution. After 90 min of incubation, the wells were washed 3X with 1X Wash Buffer. Next, 100 µL of the secondary antibody–HRP conjugate solution was added to the wells, followed by 30 min incubation and three washes. A total of 100 µL of the color solution was then added to the wells, followed by 20 min of incubation away from sunlight and the addition of 50 µL of the stop solution. The absorbance of each well was determined at 450 nm with a BioTek Synergy H1 Multimode Microplate Reader (Agilent, Santa Clara,

CA, USA) within 15 min of adding the stop solution. Samples that exceeded the maximum standard concentration were diluted with the sample diluent/LRB provided in the kit.

ELISA Quality Controls

Microcystin-LR (MC-LR) standards (10 ug/mL in methanol) were obtained from Gold Standard Diagnostics (Davis, CA, USA) [69]. The MC-LR standard was diluted to a concentration of 100 ng/mL and stored at −20 °C to be used for spiking the Laboratory Fortified Blank (LFB), Laboratory Fortified Sample Matrix (LFSM), and Laboratory Fortified Sample Matrix Duplicate (LFSMD). In the analysis batch, the following quality control elements were included: one quality control sample (QCS, 0.75 ng/mL), one LRB, two LFBs, one LFSM, and one LFSMD [12]. To confirm the assay response using primary calibration standards, a QCS provided along with the kit was assayed with each plate. The LRB was obtained from a reagent water source separate from the sample diluent supplied in the kit, and it was aliquoted into four wells throughout the plate. The LFB and LFSM/LFSMD were fortified with 1.0 ng/mL of MC-LR, and each was loaded into one well. For the initial demonstration of capability (IDC), seven LFBs were fortified with 0.5 ug/mL of MC-LR to test precision and accuracy. The LFBs were then lysed, filtered, and assayed. Five LRBs were lysed, filtered, and assayed in the same analysis batch as the LFBs, including a low-range calibration verification (Low-CV) control to demonstrate an acceptable system background. To confirm the minimum reporting limit (MRL), seven LFBs were fortified at the proposed MRL concentration (0.40 ng/mL). The LFBs were then lysed,

filtered, and assayed in an analysis batch containing one Low-CV, one QCS, and two LRBs. The threshold values for the IDC results are shown in Supplementary Table S2.

During MRL confirmation, the seven LFBs (0.40 ng/mL) were frozen at −80 °C. However, an additional three LFBs at the same MC-LR concentration were frozen at −20 °C to test if changes in freeze temperature led to changes in absorbance values/MC concentration. Welch's *t*-test was performed subsequently to determine if there was a significant difference in the mean absorbances of the two sets of LFBs produced by the two freeze temperatures. Initial MRL confirmation tests were also filtered with 0.45 um nylon membrane syringe filters, which were later switched out for 0.45 um polyethersulfone (PES) membrane syringe filters (Cytiva, Marlborough, MA, USA). The absorbance values of the LFBs were compared using Welch's t-test to determine if there was a significant difference in the mean recovery between the sets filtered using the two syringe filter membranes.

Cyanobacteria and Cyanotoxin Gene qPCR Analysis and Quality Control

Cyanobacterial and cyanotoxin gene copies were measured using qPCR analysis using CyanoDTec Total Cyanobacteria and Toxin Kit (Phytoxigene™, *Akron,* OH, USA) as per manufacturer's instructions (available at https://static1.squarespace.com/static/531043b0e4b013842a3999f0/t/5d788d085bd75417004e0916/1568181527263/CyanoDTec+Procedure+Ver9.pdf; accessed on 1 June 2023). The cyanobacterial assay was used to measure 16S rRNA gene copy numbers, while cyanotoxin assays to measure microcystin/nodularin, saxitoxin, and cylindrospermopsin gene copy numbers. All qPCR reactions were run on the CFX96 Touch Real-Time PCR Detection System (BioRad Inc., CA, USA), and results were processed and analyzed using Bio-Rad

CFX Maestro Version 4.1 software. Standard curves for the cyanobacterial and cyanotoxins were generated using CyanoDTec CyanoNAS (Phytoxigene™, OH, USA), followed by the generation of composite five-point standard curves for each assay with a range of 10–100,000 gene copies. Each qPCR reaction contained 20 µL of CyanoDTec master mix containing dNTPs, DNA polymerase, and primer–probe solution and 5 µL of extracted and purified DNA from samples. qPCR reactions were subject to initial denaturation at 95 °C for two minutes, followed by 40 cycles of 95 °C for 15 s, followed by 60 °C for 45 s. Gene copy numbers were calculated using the slope–intercept equation generated within the composite standard curve. A sample was confirmed to have 16S or toxin gene copies if the determined gene copy number was above the lower limit of the standard curve range and were only accepted as valid results if the internal amplification control threshold-cycle (Ct) value did not differ by more than 1.5 when compared to non-template controls.

qPCR reactions were run alongside non-template controls to monitor for inhibition or potential false positives. An internal amplification control (IAC) uses non-specific DNA and primers independent of cyanobacterial reactions as a gauge for whether samples are being inhibited by external sources (i.e., contaminants). As a challenge test, we tested crude and purified DNA from homogenous *Microcystis aeruginosa* culture before analyzing the samples for qPCR.

ELISA and qPCR Data Analysis

A four-parameter logistic calibration curve was generated using the absorbance (450 nm) of 6 MC-LR standards at concentrations of 0, 0.15, 0.40, 1.0, 2.0, and 5.0 ng/mL (provided by kit). Equation (1) was fitted to the absorbance values of the standards through regression

analysis using the Solver add-in in Excel. This calibration curve was used to determine the MC/NOD concentration of the sample duplicates. The mean concentration of the two duplicates was then taken as the final concentration of MCs/NODs in each sample.

$$y = ((a - d))/(1 + (x/c)^b) + d \quad (1)$$

Equation (1): The four-parameter logistic calibration curve for ELISA analysis. As described in USEPA Method 546, y is the absorbance (450 nm), and x is the MC/NOD concentration. The coefficients, a, b, c, d, are calculated using logistic regression analysis.

Using the standard curve equation generated for the 16s rRNA and cyanotoxin gene copy assays, initial gene copy numbers are determined for each reaction using Equation (2).

$$g = 10\text{^}((-(w - m))/s) \quad (2)$$

Equation (2): Conversion of observed Ct value to initial gene copy counts, where g is the initial gene copy count, w is the observed Ct value, and m and s are the y-intercept and slope of the generated standard curve, respectively.

The data for total MCs/NODs and cyanobacterial 16S rRNA/*mcyE* gene copies were tested for normality using Shapiro–Wilk's ($p < 0.05$ for all three datasets), followed by Spearman correlation analysis (http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r) accessed on 1 June 2023.

Supplementary Materials: The following supporting information can be downloaded at www.mdpi.com/xxx/s1, Figure S1: Differentially Abundant Phyla. The positive log2FoldChange indicates enrichment at bloom sites, and the negative log2Fold Change

shows enrichment at non-bloom sites. The threshold-adjusted *p*-value for significance is 0.05; Figure S2: Relative Abundance of cyanobacterial genera for (a) Bloom and (b) Non-bloom samples using Shotgun Sequencing; Figure S3: Differentially Abundant Cyanobacterial Genera. The positive log2FoldChange indicates enrichment of taxa at bloom sites. The associated *p*-values for each cyanobacterial genera are indicated. The threshold-adjusted *p*-value for significance is 0.05; Figure S4: Differential Abundance of Functional Analysis. The SEED subsystems database was used at level 2 for the differential abundance of functions. Positive log2foldchange indicates differential abundance at bloom sites and negative log2foldchange at non-bloom sites. The threshold of significance was an adjusted *p*-value of 0.0001; Table S1: Quality control analytics for DNA Sequencing and qPCR analysis; Table S2: Quality control analytics for ELISA analysis, including control elements from the Initial Demonstration of Capability and the Analysis Batch; Table S3: Differential Abundance of Nitrogen and Phosphorus-associated Functions. The adjusted *p*-value cut-off is 0.05. Positive L2FC indicates enrichment at bloom sites.

Author Contributions: validation, F.S. and R.A.; methodology, F.S., R.A., H.E.S. and A.P.; investigation, R.A., J.L.J. and K.L.T.; writing—original draft preparation, F.S.; writing—review and editing F.S., T.A.E., K.L.T., J.L.J., E.L., and H.E.S.; supervision, H.E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ontario Ministry of the Environment, Conservation and Parks grant number 2607.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to sponsor privacy restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

1.4.6. References

1. Zohdi, E.; Abbaspour, M. Harmful algal blooms (red tide): A review of causes, impacts and approaches to monitoring and prediction. Int. J. Environ. Sci. Technol. 2019, 16, 1789–1806.

2. Huisman, J.; Codd, G.A.; Paerl, H.W.; Ibelings, B.W.; Verspagen, J.M; Visser, P.M. Cyanobacterial blooms. Nat. Rev. Microbiol. 2018, 16, 471–483.

3. Zhang, W.; Liu, J.; Xiao, Y.; Zhang, Y.; Yu, Y.; Zheng, Z.; Liu, Y.; Li, Q. The Impact of Cyanobacteria Blooms on the Aquatic Environment and Human Health. Toxins 2022, 14, 658. https://doi.org/10.3390/toxins14100658.

4. Pound, H.L.; Martin, R.M.; Sheik, C.S.; Steffen, M.M.; Newell, S.E.; Dick, G.J.; McKay, R.M.L.; Bullerjahn, G.S.; Wilhelm, S.W. Environmental Studies of Cyanobacterial Harmful Algal Blooms Should Include Interactions with the Dynamic Microbiome. Environ. Sci. Technol. 2021, 55, 12776–12779. https://doi.org/10.1021/acs.est.1c04207.

5. Davis, T.W.; Stumpf, R.; Bullerjahn, G.S.; McKay, R.M.L.; Chaffin, J.D.; Bridgeman, T.B.; Winslow, C. Science meets policy: A framework for determining

impairment designation criteria for large waterbodies affected by cyanobacterial harmful algal blooms. Harmful Algae 2019, 81, 59–64. https://doi.org/10.1016/j.hal.2018.11.016.

6. Smith, R.B.; Bass, B.; Sawyer, D.; Depew, D.; Watson, S.B. Estimating the economic costs of algal blooms in the Canadian Lake Erie Basin. Harmful Algae 2019, 87, 101624. https://doi.org/10.1016/j.hal.2019.101624.

7. Chaffin, J.D.; Mishra, S.; Kane, D.D.; Bade, D.L.; Stanislawczyk, K.; Slodysko, K.N.; Jones, K.W.; Parker, E.M.; Fox, E.L. Cy-anobacterial blooms in the central basin of Lake Erie: Potentials for cyanotoxins and environmental drivers. J. Great Lakes Res. 2019, 45, 277–289. https://doi.org/10.1016/j.jglr.2018.12.006.

8. Catherine, A.; Maloufi, S.; Congestri, R.; Viaggiu, E; Pilkaityte, R. Cyanobacterial samples: Preservation, enumeration, and biovolume measurements. In Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; pp. 313–330.

9. Fortin, N.; Aranda-Rodriguez, R.; Jing, H.; Pick, F.; Bird, D.; Greer, C.W. Detection of Microcystin-Producing Cyanobacteria in Missisquoi Bay, Quebec, Canada, Using Quantitative PCR. Appl. Environ. Microbiol. 2010, 76, 5105–5112. https://doi.org/10.1128/aem.00183-10.

10. Prescott, M.D.; Jia, A.; Guo, Y.C.; Di Giovanni, G.D.; Li, W.L.; Garcia, E.A.; Vanderford, B.J.; Eaton, A.D. Interlaboratory comparisons of cyanotoxin analysis by ELISA and LC–MS/MS. AWWA Water Sci. 2023, 5, e1342. https://doi.org/10.1002/aws2.1342.

11. Svirčev, Z.; Lalić, D.; Savić, G.B.; Tokodi, N.; Backović, D.D.; Chen, L.; Meriluoto, J.; Codd, G.A. Global geographical and historical overview of cyanotoxin distribution and cyanobacterial poisonings. Arch. Toxicol. 2019, 93, 2429–2481. https://doi.org/10.1007/s00204-019-02524-4.

12. Zaffiro, A.; Rosenblum, L.; Wendelken, S.C. Method 546: Determination of Total Microcystins and Nodularins in Drinking Water and Ambient Water by Adda Enzyme-Linked Immunosorbent Assay; United States Environmental Protection Agency: Cincinnati, OH, USA, 2016. Available online: https://www.epa.gov/sites/default/files/2016-09/documents/method-546-determination-total-microcystins-nodularins-drinking-water-ambient-water-adda-enzyme-linked-immunosorbent-assay.pdf (accessed 1 June 2023).

13. Jungblut, A.-D.; Neilan, B.A. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nod-ularin, synthetase genes in three orders of cyanobacteria. Arch. Microbiol. 2006, 185, 107–114. https://doi.org/10.1007/s00203-005-0073-5.

14. Abdallah, M.F.; Van Hassel, W.H.; Andjelkovic, M.; Wilmotte, A.; Rajkovic, A. Cyanotoxins and food contamination in de-veloping countries: Review of their types, toxicity, analysis, occurrence and mitigation strategies. Toxins 2021, 13, 786.

15. Lu, J.R.; Struewing, I.; Wymer, L.; Tettenhorst, D.R.; Shoemaker, J.; Allen, J. Use of qPCR and RT-qPCR for monitoring variations of microcystin producers and as an early

warning system to predict toxin production in an Ohio inland lake. Water Res. 2020, 170, 115262. https://doi.org/10.1016/j.watres.2019.115262.

16. Pacheco, A.B.F.; Guedes, I.A.; Azevedo, S. Is qPCR a reliable indicator of cyanotoxin risk in freshwater? Toxins 2016, 8, 172.

17. Duan, X.; Zhang, C.; Struewing, I.; Li, X.; Allen, J.; Lu, J. Cyanotoxin-encoding genes as powerful predictors of cyanotoxin production during harmful cyanobacterial blooms in an inland freshwater lake: Evaluating a novel early-warning system. Sci. Total. Environ. 2022, 830, 154568. https://doi.org/10.1016/j.scitotenv.2022.154568.

18. Li, D.; Gu, A.Z.; He, M. Quantification and genetic diversity of total and microcystin-producing Microcystis during blooming season in Tai and Yang-cheng lakes, China. J. Appl. Microbiol. 2014, 116, 1482–1494. https://doi.org/10.1111/jam.12456.

19. Elhabashy, A.; Li, J.; Sokolova, E. Water quality modeling of a eutrophic drinking water source: Impact of future climate on Cyanobacterial blooms. Ecol. Model. 2023, 477, 110275. https://doi.org/10.1016/j.ecolmodel.2023.110275.

20. Caen, A.; Mathias, J.-D.; Latour, D. How do seasonal temperature variations influence interplay between toxic and non-toxic cyanobacterial blooms? Evidence from modeling and experimental data. Harmful Algae 2024, 134, 102606. https://doi.org/10.1016/j.hal.2024.102606.

21. Jacquemin, S.J.; Doll, J.C.; Johnson, L.T.; Newell, S.E. Exploring long-term trends in microcystin toxin values associated with persistent harmful algal blooms in Grand Lake St Marys. Harmful Algae 2023, 122, 102374. https://doi.org/10.1016/j.hal.2023.102374.

22. Crnkovic, C.M.; May, D.S.; Orjala, J. The impact of culture conditions on growth and metabolomic profiles of freshwater cy-anobacteria. J. Appl. Phycol. 2018, 30, 375–384. https://doi.org/10.1007/s10811-017-1275-3.

23. Xiao, X.; Sogge, H.; Lagesen, K.; Tooming-Klunderud, A.; Jakobsen, K.S.; Rohrlack, T. Use of High Throughput Sequencing and Light Microscopy Show Contrasting Results in a Study of Phytoplankton Occurrence in a Freshwater Environment. PLoS ONE 2014, 9, e106510. https://doi.org/10.1371/journal.pone.0106510.

24. Wood, S.A.; Pochon, X.; Ming, W.; von Ammon, U.; Woods, C.; Carter, M.; Smith, M.; Inglis, G.; Zaiko, A. Considerations for incorporating real-time PCR assays into routine marine biosecurity surveillance programmes: A case study targeting the Med-iterranean fanworm (Sabella spallanzanii) and club tunicate (Styela clava). Genome 2019, 62, 137–146. https://doi.org/10.1139/gen-2018-0021.

25. Preece, E.P.; Moore, B.C.; Swanson, M.E.; Hardy, F.J. Identifying best methods for routine ELISA detection of microcystin in seafood. Environ. Monit. Assess. 2015, 187, 1–10. https://doi.org/10.1007/s10661-014-4255-y.

26. Chorus, I.; Fastner, J.; Welker, M. Cyanobacteria and Cyanotoxins in a Changing Environment: Concepts, Controversies, Challenges. Water 2021, 13, 2463. https://doi.org/10.3390/w13182463.

27. Affe, H.M.D.J.; Rigonato, J.; Nunes, J.M.D.C; Menezes, M. Metagenomic analysis of cyanobacteria in an oligotrophic tropical estuary, South Atlantic. Front. Microbiol. 2018, 9, 1393.

28. Saleem, F.; Jiang, J.L.; Atrache, R.; Paschos, A.; Edge, T.A; Schellhorn, H.E. Cyanobacterial Algal Bloom Monitoring: Molecular Methods and Technologies for Freshwater Ecosystems. Microorganisms 2023, 11, 851.

29. Codd, G.A.; Testai, E.; Funari, E.; Svirčev, Z. Cyanobacteria, cyanotoxins, and human health. In Water Treatment for Purification from Cyanobacteria and Cyanotoxins; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2020; pp. 37–68.

30. Zhou, B.; Cai, X.; Wang, S.; Yang, X. Analysis of the causes of cyanobacteria bloom: A review. J. Resour. Ecol. 2020, 11, 405–413.

31. Bashir, F.; Bashir, A.; Bouaïcha, N.; Chen, L.; Codd, G.A.; Neilan, B.; Xu, W.L.; Ziko, L.; Rajput, V.D.; Minkina, T; et al. Cy-anotoxins, biosynthetic gene clusters, and factors modulating cyanotoxin biosynthesis. World J. Microbiol. Biotechnol. 2023, 39, 241.

32. Haukka, K.; Kolmonen, E.; Hyder, R.; Hietala, J.; Vakkilainen, K.; Kairesalo, T.; Haario, H.; Sivonen, K. Effect of Nutrient Loading on Bacterioplankton Community

Composition in Lake Mesocosms. Microb. Ecol. 2006, 51, 137–146. https://doi.org/10.1007/s00248-005-0049-7.

33. Louati, I.; Pascault, N.; Debroas, D.; Bernard, C.; Humbert, J.-F.; Leloup, J. Structural Diversity of Bacterial Communities As-sociated with Bloom-Forming Freshwater Cyanobacteria Differs According to the Cyanobacterial Genus. PLoS ONE 2015, 10, e0140614. https://doi.org/10.1371/journal.pone.0140614.

34. Pizzetti, I.; Fuchs, B.M.; Gerdts, G.; Wichels, A.; Wiltshire, K.H.; Amann, R. Temporal variability of coastal Planctomycetes clades at Kabeltonne station, North Sea. Appl. Environ. Microbiol. 2011, 77, 5009–5017.

35. Cai, H.Y.; Yan, Z.S.; Wang, A.J.; Krumholz, L.R.; Jiang, H.L. Analysis of the attached microbial community on mucilaginous cyanobacterial aggregates in the eutrophic Lake Taihu reveals the importance of Planctomycetes. Microb. Ecol. 2013, 66, 73–83.

36. He, S.; Stevens, S.L.R.; Chan, L.-K.; Bertilsson, S.; del Rio, T.G.; Tringe, S.G.; Malmstrom, R.R.; McMahon, K.D. Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. mSphere 2017, 2, e00277-17. https://doi.org/10.1128/msphere.00277-17.

37. McLaren, M.R.; Willis, A.D; Callahan, B.J. Consistent and correctable bias in metagenomic sequencing experiments. Elife 2019, 8, e46923.

38. Drengenes, C.; Eagan, T.M.L.; Haaland, I.; Wiker, H.G.; Nielsen, R. Exploring protocol bias in airway microbiome studies: One versus two PCR steps and 16S rRNA gene

region V3 V4 versus V4. BMC Genom. 2021, 22, 1–15. https://doi.org/10.1186/s12864-020-07252-z.

39. Valadez-Cano, C.; Hawkes, K.; Calvaruso, R.; Reyes-Prieto, A; Lawrence, J. Amplicon-based and metagenomic approaches provide insights into toxigenic potential in understudied Atlantic Canadian lakes. Facets 2022, 7, 194–214.

40. MacKeigan, P.W.; Garner, R.E.; Monchamp, M.E.; Walsh, D.A.; Onana, V.E.; Kraemer, S.A.; Pick, F.R.; Beisner, B.E.; Agbeti, M.D.; da Costa, N.B; et al. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assem-blages across hundreds of lakes. Harmful Algae 2022, 113, 102187.

41. Rae, B.D.; Long, B.M.; Whitehead, L.F.; Förster, B.; Badger, M.R.; Price, G.D. Cyanobacterial Carboxysomes: Microcompart-ments that Facilitate CO2 Fixation. J. Mol. Microbiol. Biotechnol. 2013, 23, 300–307. https://doi.org/10.1159/000351342.

42. Graham, J.L.; Dubrovsky, N.M.; Foster, G.M.; King, L.R.; Loftin, K.A.; Rosen, B.H.; Stelzer, E.A. Cyanotoxin occurrence in large rivers of the United States. Inland Waters 2020, 10, 109–117. https://doi.org/10.1080/20442041.2019.1700749.

43. Zaheer, R.; Noyes, N.; Polo, R.O.; Cook, S.R.; Marinier, E.; Van Domselaar, G.; Belk, K.E.; Morley, P.S.; McAllister, T.A. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci. Rep. 2018, 8, 5890. https://doi.org/10.1038/s41598-018-24280-8.

44. Conradie, K.R.; Barnard, S. The dynamics of toxic Microcystis strains and microcystin production in two hypertrofic South African reservoirs. Harmful Algae 2012, 20, 1–10. https://doi.org/10.1016/j.hal.2012.03.006.

45. Ngwa, F.F.; Madramootoo, C.A.; Jabaji, S. Comparison of cyanobacterial microcystin synthetase (mcy) E gene transcript levels, mcy E gene copies, and biomass as indicators of microcystin risk under laboratory and field conditions. Microbiologyopen 2014, 3, 411–425. https://doi.org/10.1002/mbo3.173.

46. Padovan, A.; Kennedy, K.; Gibb, K. A microcystin synthesis mcyE/ndaF gene assay enables early detection of microcystin production in a tropical wastewater pond. Harmful Algae 2023, 127, 102476. https://doi.org/10.1016/j.hal.2023.102476.

47. Francy, D.S.; Brady, A.M.G.; Ecker, C.D.; Graham, J.L.; Stelzer, E.A.; Struffolino, P.; Dwyer, D.F.; Loftin, K.A. Estimating microcystin levels at recreational sites in western Lake Erie and Ohio. Harmful Algae 2016, 58, 23–34. https://doi.org/10.1016/j.hal.2016.07.003.

48. Baxa, D.V.; Kurobe, T.; Ger, K.A.; Lehman, P.W.; Teh, S.J. Estimating the abundance of toxic Microcystis in the San Francisco Estuary using quantitative real-time PCR. Harmful Algae 2010, 9, 342–349. https://doi.org/10.1016/j.hal.2010.01.001.

49. Beversdorf, L.J.; Chaston, S.D.; Miller, T.R.; McMahon, K.D. Microcystin mcyA and mcyE Gene Abundances Are Not Appro-priate Indicators of Microcystin Concentrations in Lakes. PLoS ONE 2015, 10, e0125353. https://doi.org/10.1371/journal.pone.0125353.

50. D'angelo, E. Development and evaluation of a sensitive, Diffusive Gradients in Thin-Films (DGT) method for determining microcystin-LR concentrations in freshwater and seawater. Harmful Algae 2019, 89, 101668. https://doi.org/10.1016/j.hal.2019.101668.

51. Li, Y.; Whitaker, J.S.; Skibbe, S.A.; Piaskowski, M.A. Microwave cell lysis for ADDA-ELISA analysis of total microcystins and nodularins. AWWA Water Sci. 2021, 3. https://doi.org/10.1002/aws2.1215.

52. Shanks, O.C.; Sivaganesan, M.; Peed, L.; Kelty, C.A.; Blackwood, A.D.; Greene, M.R.; Noble, R.T.; Bushon, R.N.; Stelzer, E.A.; Kinzelman, J.; et al. Interlaboratory Comparison of Real-Time PCR Protocols for Quantification of General Fecal Indicator Bacteria. Environ. Sci. Technol. 2012, 46, 945–953. https://doi.org/10.1021/es2031455.

53. Rinrid, K.; Wonnapinij, P.; Vuttipongchaikij, S.; Shutsrirung, A.; Suttangkakul, A. Assessing DNA extraction methods for metagenomic analysis from crop soil in Thailand. Agric. Nat. Resour. 2022, 56, 4.

54. Saenz, J.S.; Roldan, F.; Junca, H.; Arbeli, Z. Effect of the extraction and purification of soil DNA and pooling of PCR amplifi-cation products on the description of bacterial and archaeal communities. J. Appl. Microbiol. 2019, 126, 1454–1467. https://doi.org/10.1111/jam.14231.

55. Almuhtaram, H.; Kibuye, F.A.; Ajjampur, S.; Glover, C.M.; Hofmann, R.; Gaget, V.; Owen, C.; Wert, E.C.; Zamyadi, A. State of knowledge on early warning tools for

cyanobacteria detection. Ecol. Indic. 2021, 133, 108442. https://doi.org/10.1016/j.ecolind.2021.108442.

56. Ribeiro, K.F.; Duarte, L.; Crossetti, L.O. Everything is not everywhere: A tale on the biogeography of cyanobacteria. Hydrobi-ologia 2018, 820, 23–48. https://doi.org/10.1007/s10750-018-3669-x.

57. Mehta, K.K.; Evitt, N.H.; Swartz, J.R. Chemical lysis of cyanobacteria. J. Biol. Eng. 2015, 9, 10.

58. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Gonzalez Peña, A.; Goodrich, J.K.; Gordon, J.I.; et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 2010, 7, 335–336.

59. Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science using QIIME 2. Nat. Biotechnol. 2019, 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9.

60. Wingett, S.W; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research 2018, 7, 1338.

61. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 2014, 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

62. Lan, Y.; Wang, Q.; Cole, J.R.; Rosen, G.L. Using the RDP Classifier to Predict Taxonomic Novelty and Reduce the Search Space for Finding Novel Organisms. PLoS ONE 2012, 7, e32491. https://doi.org/10.1371/journal.pone.0032491.

63. McDonald, D.; Price, M.N.; Goodrich, J.; Nawrocki, E.P.; DeSantis, T.Z.; Probst, A.; Andersen, G.L.; Knight, R.; Hugenholtz, P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012, 6, 610–618. https://doi.org/10.1038/ismej.2011.139.

64. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 2015, 12, 59–60. https://doi.org/10.1038/nmeth.3176.

65. Huson, D.H.; Mitra, S. Introduction to the analysis of environmental sequences: Metagenomics with megan. In Evolutionary Genomics: Statistical and Computational Methods; Springer: Berlin/Heidelberg, Germany, 2012; Volume 2, pp. 415–429.

66. Bağcı, C.; Beier, S.; Górska, A.; Huson, D.H. Introduction to the Analysis of Environmental Sequences: Metagenomics with MEGAN. In Evolutionary Genomics: Statistical and Computational Methods; Springer: Berlin/Heidelberg, Germany, 2019; pp. 591–604.

67. Dixon, P. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 2003, 14, 927–930.

68. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014, 15, 550.

69. Aranda-Rodriguez, R.; Jin, Z.; Harvie, J.; Cabecinha, A. Evaluation of three field test kits to detect microcystins from a public health perspective. Harmful Algae 2015, 42, 34–42. https://doi.org/10.1016/j.hal.2015.01.001.

Additional File for Chapter 2

ESM Table S1. Geographical coordinates of sampling sites.

Sampling Location	Sampling Site	Latitude	Longitude
Marie Curtis East Beach	30W	43.585610	-79.540054
	32W	43.585110	-79.540560
Sunnyside Beach	18W	43.636612	-79.452670
	21W	43.637110	-79.457530
Etobicoke Creek	Etobicoke Creek	43.586781	-79.54373
Humber River	Humber River	43.632082	-79.47099

ESM Table S2. Time taken by each step of sample processing in USEPA 1609.1 method for Enterococci quantification in recreational water ecosystems.

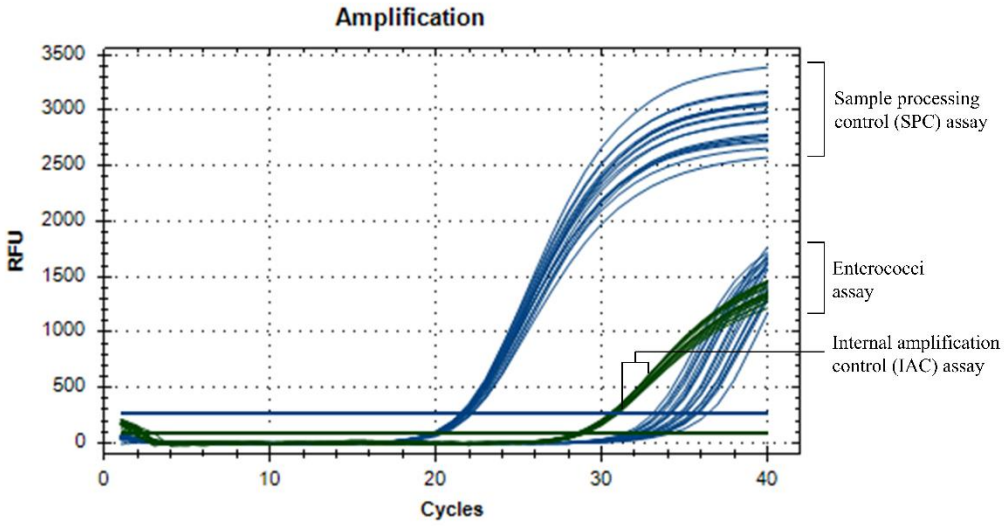
Sample Processing Step	Time Required (minutes)
Filtration	30-40 min
DNA Extraction	15-20 min
qPCR Plate Preparation	15-20 min
qPCR Plate Setup	5-10 min
qPCR run and Results Reporting	130-150 min
Total Time	185-210 min

ESM Table S3. Consumables Cost Comparison Between *E. coli* Enumeration by Culture and USEPA 1609.1 qPCR-Based Method for Recreational Water Analysis.

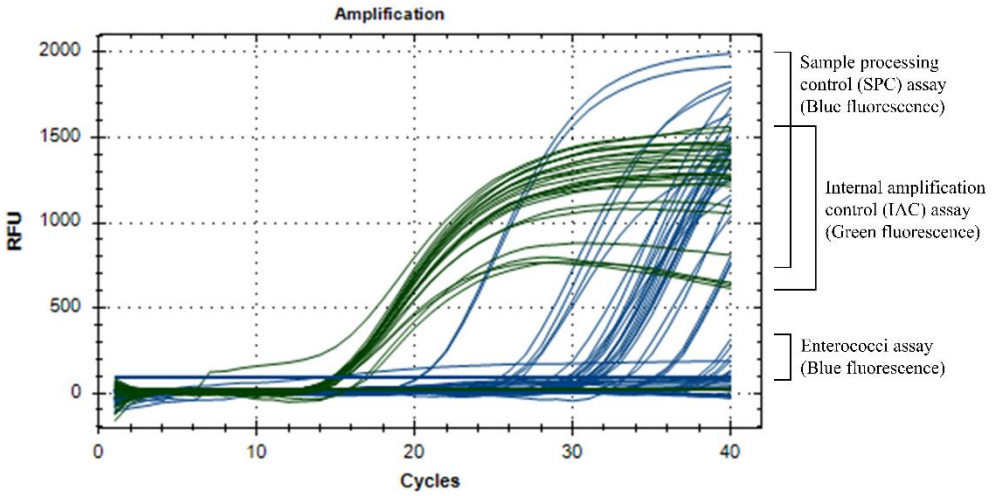
Consumable	Amount/Volume Required per Sample/Reaction	Manufacturer/Catalog Number	Cost (CAD)
USEPA 1609.1 qPCR-Based Method			
TaqMan™ Environmental Master Mix 2.0	25 µL for duplicate reactions	Fischer Scientific/4396838	3.93
0.4µ polycarbonate membrane filter	1 filter	Fischer Scientific/HTTP04700	1.43
TAQMAN Probe enterococci assay	80nM	Integrated DNA Technologies	0.63
TAQMAN Probe Sample Processing Control assay	80nM	Integrated DNA Technologies	0.63
TAQMAN Probe Internal Amplification Control assay	80nM	Integrated DNA Technologies	0.63
Enterococci/Internal Amplification Control Forward Primer	1µM	Integrated DNA Technologies	0.29
Enterococci/Internal Amplification Control Reverse Primer	1µM	Integrated DNA Technologies	0.29
Sample Processing Control Forward Primer	1µM	Integrated DNA Technologies	0.27
Sample Processing Control Forward Primer	1µM	Integrated DNA Technologies	0.27
2 mL preloaded with Sigma G1277 glass beads	1	Generite	2.24
200µL Filter Micropipette Tips	2	Fischer Scientific	0.24
20µL Filter Micropipette Tips	2	Fischer Scientific	0.25

Approximate Total Cost per Sample			~11.1 CAD
***E. coli* Enumeration Culture Method (x2 for Dilution Series Analysis)**			
Differential Coliform Agar	12mL	Oxoid/CM1038	0.52
47 mm Petri Dish	2	Fischer Scientific/09720500	0.62
47 mm gridded cellulose membrane filters	2	Fischer Scientific/14555597	2.94
Approximate Total Cost Per Sample			~4.08 CAD

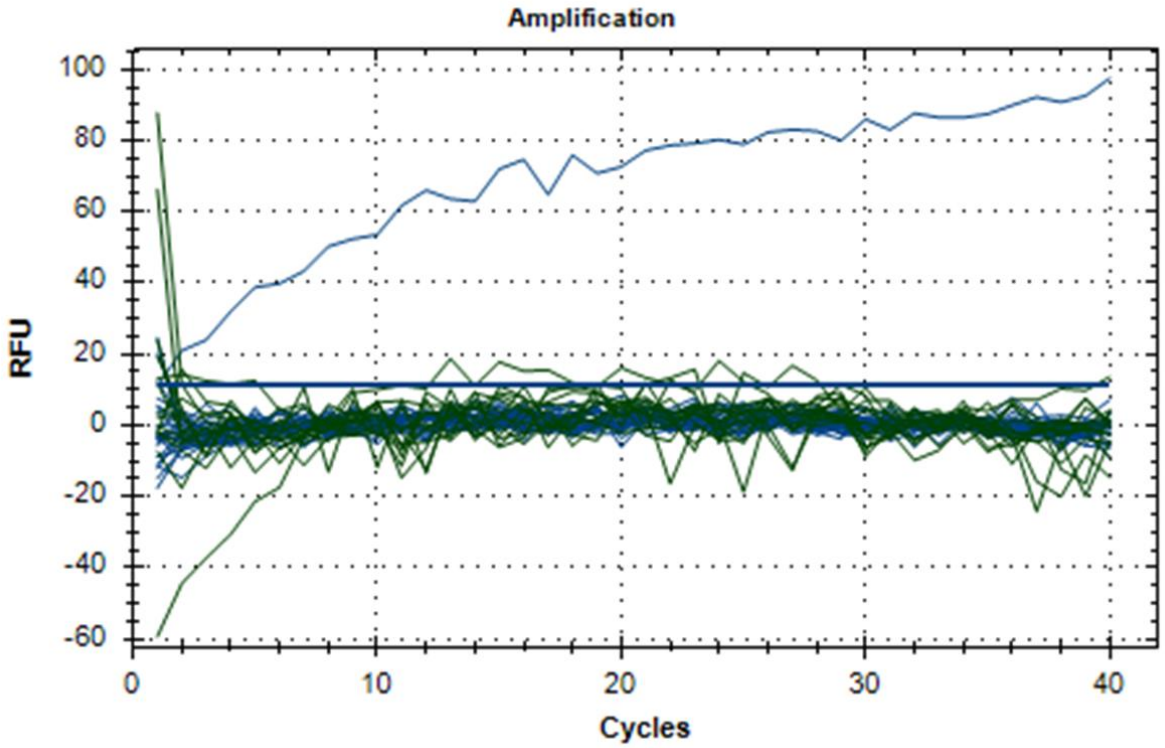
***Cost comparison does not include the equipment required for analysis, the technician time for analyses or costs of sampling and transport.**



ESM Figure S1. qPCR amplification plot with good DNA recovery and no inhibition for USEPA 1609.1 method



ESM Figure S2. qPCR amplification plot with poor DNA recovery for USEPA 1609.1 method.



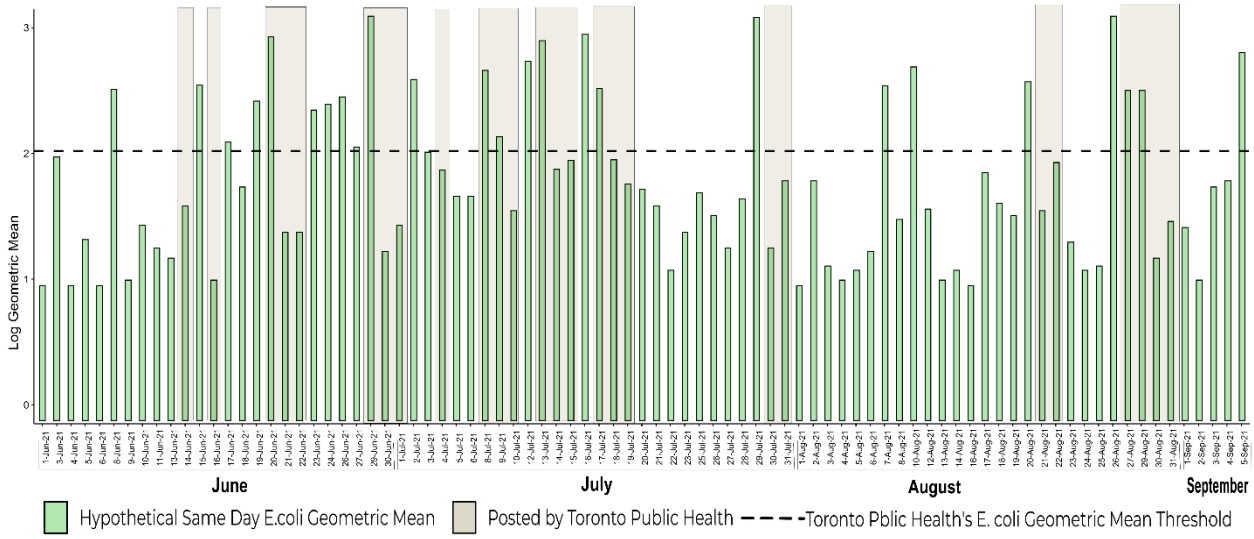
ESM Figure S3. qPCR amplification plot with complete inhibition for USEPA 1609.1 method.

Additional File for Chapter 3

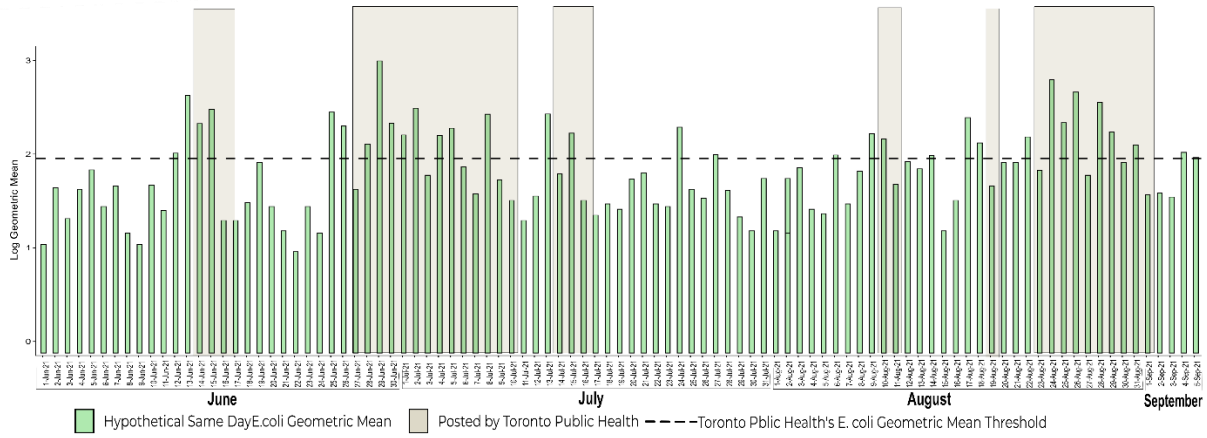
Supplementary Information

Supplementary Table 1. Geographical coordinates of sampling sites for both Marie Curtis Park East and Sunnyside Beaches

Beach	Sampling Site (Transect)	Latitude	Longitude
Marie Curtis Park East Beach	30W	43.585610	-79.540054
	32W	43.585110	-79.540560
Sunnyside Beach	18W	43.636612	-79.452670
	21W	43.637110	-79.457530



Supplementary Figure 1. Difference in Beach Postings for Marie Curtis Park East Beach according to Toronto Public Health's data for summer 2021, if hypothetically, *E. coli* culturing results were available on the same day of sample collection.

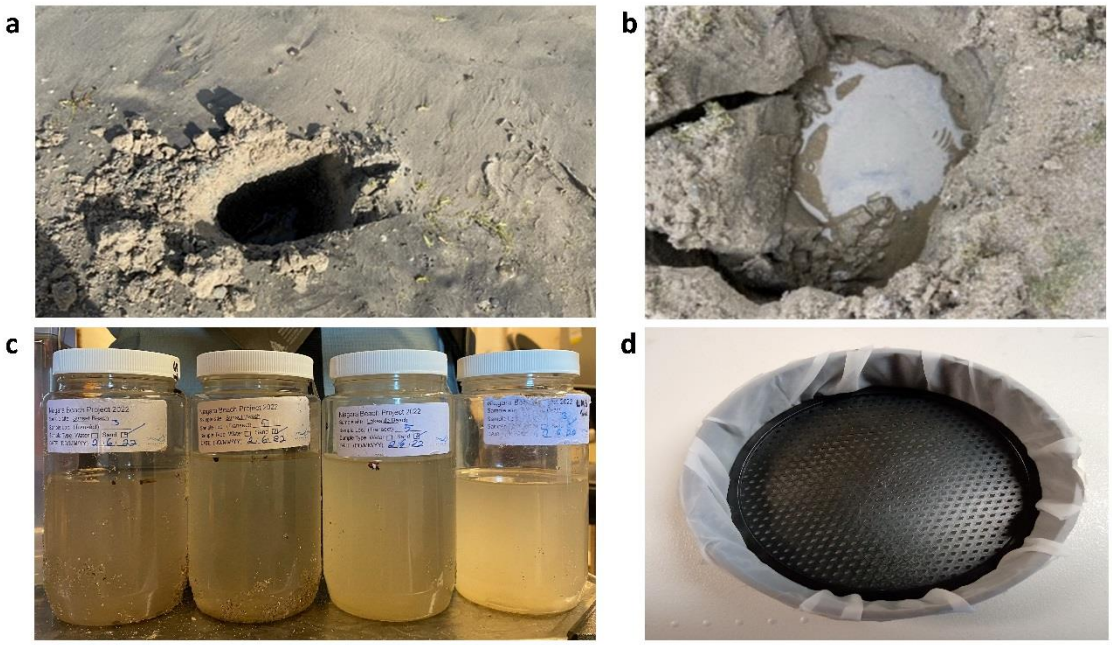


Supplementary Figure 2. Difference in Beach Postings for Sunnyside Beach according to Toronto Public Health's data for summer 2021, if hypothetically, *E. coli* culturing results were available on the same day of sample collection.

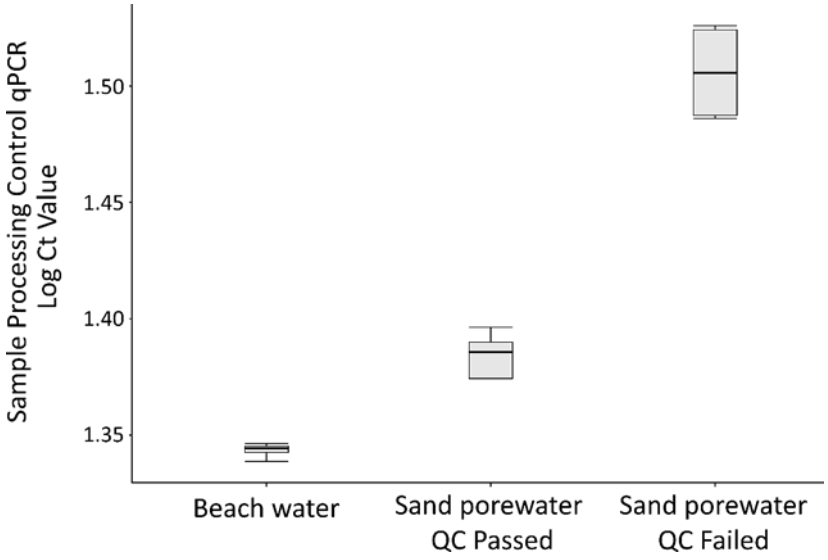
Additional File for Chapter 4

Sampling Site	Coordinates
Sunset Beach 3 (SS3)	43.2250201N, -79.2217576E
Sunset Beach 5 (SS5)	43.2260731N, -79.2208840E
Lakeside Beach 3 (LK3)	43.2045321N, -79.2662505E
Lakeside Beach 5 (LK5)	43.2050625N, -79.2651227E

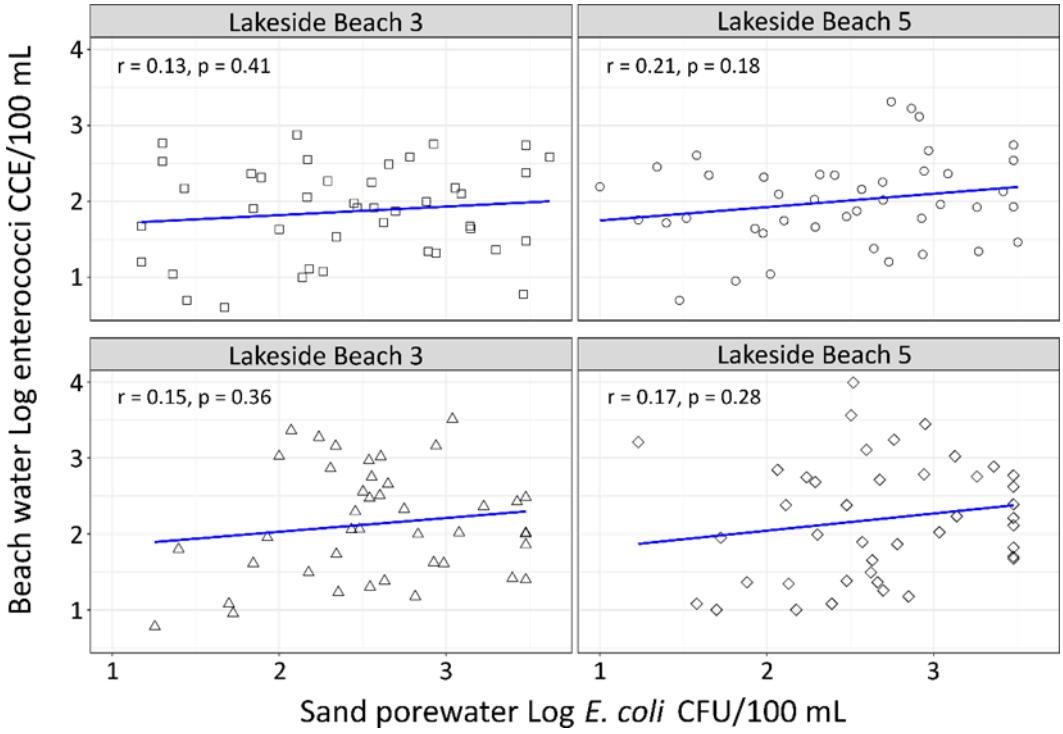
Supplementary Table 1. Geographical coordinates for sampling sites from Lakeside and Sunset beaches.



Supplementary Figure 1. Collection of sand-porewater samples from foreshore sand regions of sampling sites. a) Hole for sand-porewater collection, b) Seeped groundwater inside sand-porewater collection hole, c) Sand-porewater collected in 500mL sterile PET bottles and d) Prefiltration assembly with 50µ Nylon filter enclosed in plastic mesh for sand-porewater samples.

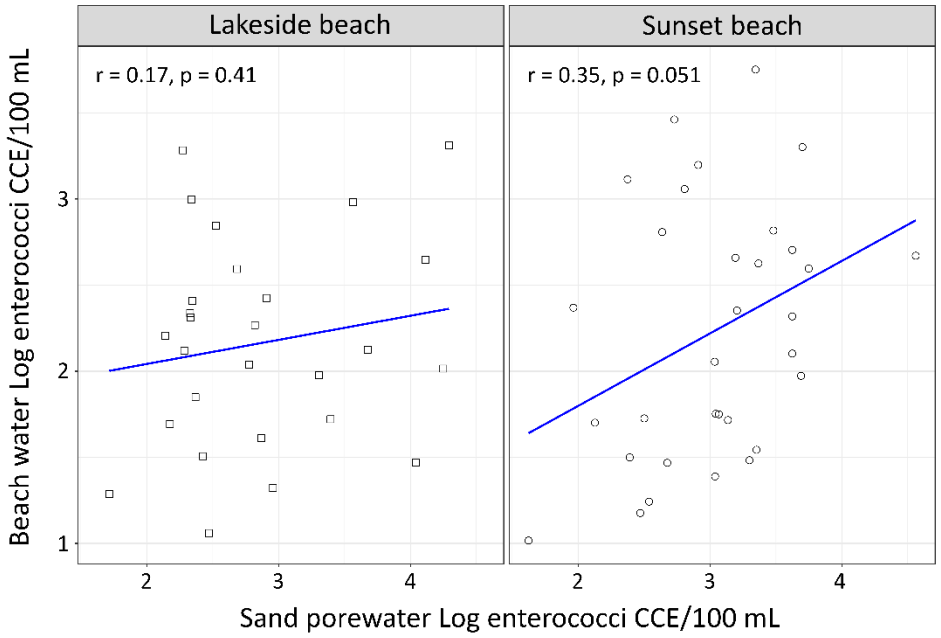


Supplementary Figure 2. Comparison between sample processing control (SPC; Salmon sperm DNA) Ct values from beach water, QC-passed sand-porewater and QC-failed sand-porewater samples.

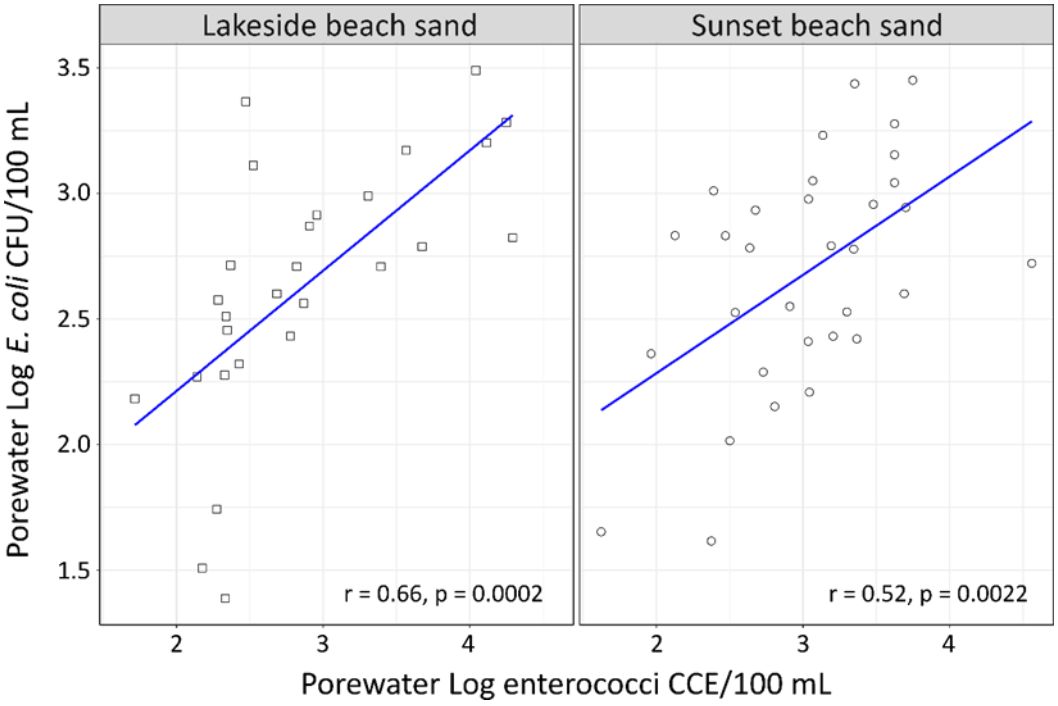


Supplementary Figure 3. Correlation between sand-porewater Log *E. coli* CFU/100mL and beach water enterococci calibrator cell equivalents (CCE/100mL) for Lakeside and Sunset

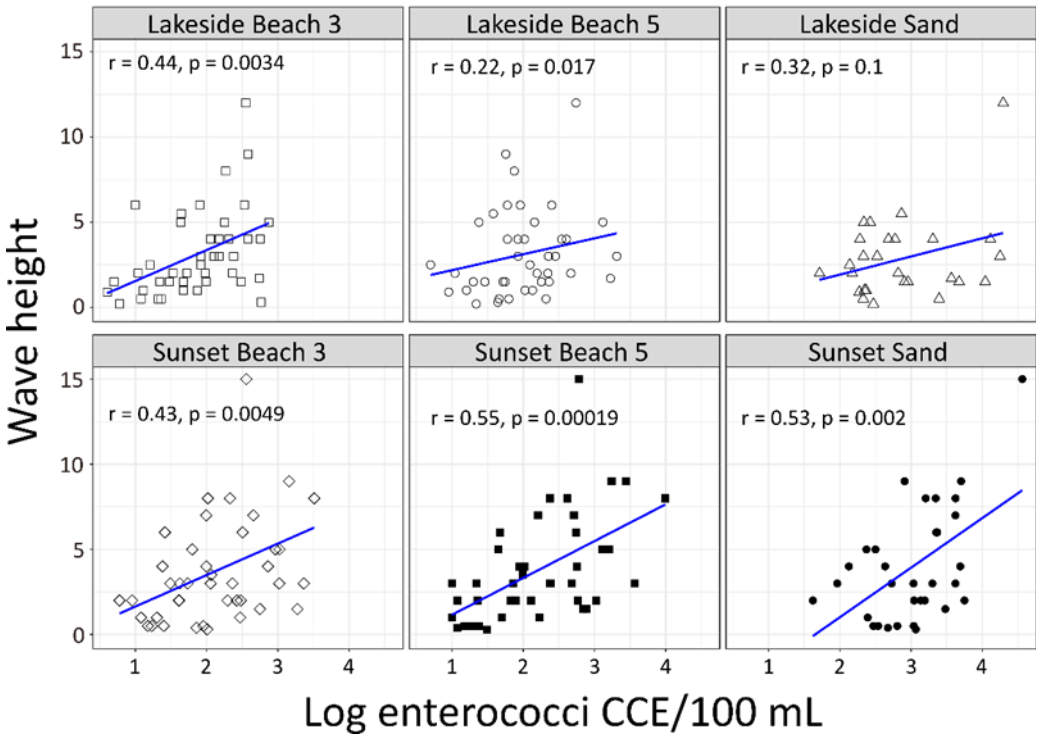
Beaches. Correlation analysis was performed by using Pearson Correlation at a 95% confidence interval.



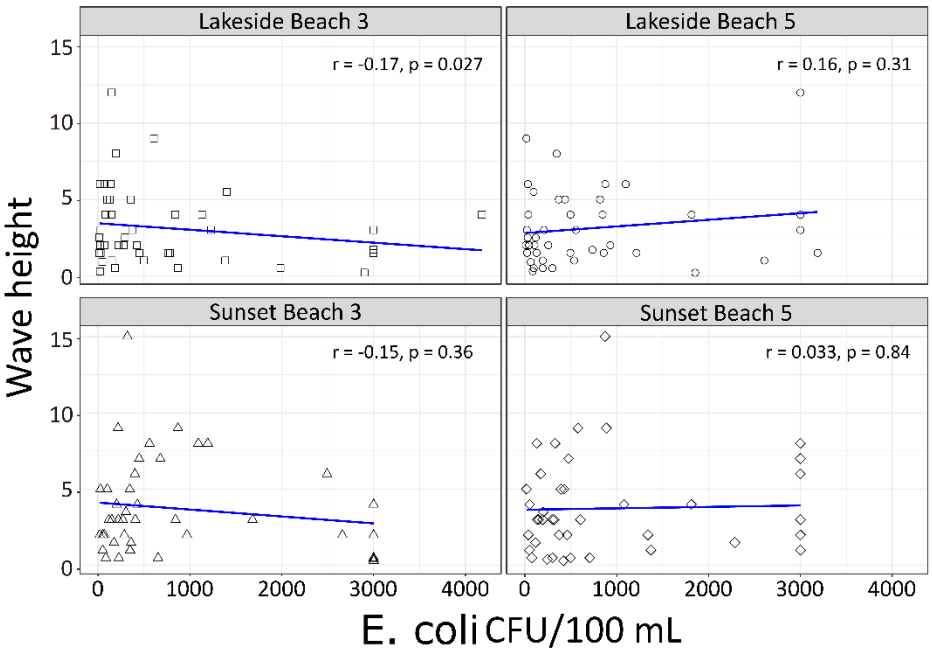
Supplementary Figure 4. Correlation between enterococci calibrator cell equivalents (CCE/100mL) of beach water and sand-porewater from Lakeside and Sunset Beaches. Correlation analysis was performed by using Pearson Correlation at a 95% confidence interval.



Supplementary Figure 5. Correlation between sand-porewater Log *E. coli* CFU/100mL and enterococci calibrator cell equivalents (CCE/100mL) for Lakeside and Sunset Beaches. Correlation analysis was performed by using Pearson Correlation at a 95% confidence interval.



Supplementary Figure 6. Correlation between wave height and Log enterococci calibrator cell equivalents (CCE/100mL) for sand-porewater and beach water from Lakeside and Sunset Beaches. Correlation analysis was performed by using Pearson Correlation at a 95% confidence interval. Wave height data were obtained from Niagara Public Health for analysis.



Supplementary Figure 7. Correlation between Wave height and Sand pore water Log *E. coli* CFU/100mL for Lakeside and Sunset beaches. Correlation analysis was performed by using Pearson Correlation at 95% confidence interval. Wave height data was obtained from Niagara Public Health for analysis.

Supplementary Table 2. Beach posting status in comparison to *E. coli* culturing beach water threshold exceedances for Sand-porewater samples from Lakeside and Sunset Beaches.

Lakeside Beach Sand-porewater		
Beach Status	***E. coli* Exceedance**	**No. of Days (Percentage)**
Open	**Yes**	25 (60%)
Open	No	14 (33%)
Posted	No	1 (2%)
Posted	Yes	2 (5%)
Sunset Beach Sand-porewater		
Open	**Yes**	22 (53%)
Open	No	7 (17%)
Posted	No	3 (7%)
Posted	Yes	10 (23%)

Additional Files for Chapter 5

Supplementary Table 1. Primers and Probes targeting *E. coli/Shigella*-specific CSPs for Conventional PCR and Sanger Sequencing.

CSP/Primer Name	Primer Sequence (5'-3')	PCR Product Size (bp)
YahL-F-Sanger	AGCTCCGCACAATAATTTGATG	756
YahL-R-Sanger	CTGTCACCTAATTCCTGGACTC	
YdjO-F-Sanger	GCCTGGCTTTCGACTCTTT	610
YdjO-R-Sanger	GAATGTGCCTGTAGCGAGAA	
YjfZ-F-Sanger	GCCATTAAGCAATGTCCTTCAG	772
YjfZ-R-Sanger	AAGAGATGACGGTTGCAGAG	

	Sequence (5'->3')
Forward primer	ACAGACGCGCCCATTAAGC
Reverse primer	CGTCCAGAACAGAGAGCAATAA

Products on target templates

>NZ_CP055055.1 Shigella dysenteriae strain SWHEFF_49 chromosome, complete genome

```
product length = 112
Forward primer  1        ACAGACGCGCCCATTAAGC  19
Template        1027580  ...................  1027598

Reverse primer  1        CGTCCAGAACAGAGAGCAATAA  22
Template        1027691  ......................  1027670
```

>NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome

```
product length = 112
Forward primer  1       ACAGACGCGCCCATTAAGC  19
Template        401461  ...................  401479

Reverse primer  1       CGTCCAGAACAGAGAGCAATAA  22
Template        401572  ......................  401551
```

>NZ_NIYS01000111.1 Shigella boydii strain ESBL-W3-2 NODE_57_length_30368_cov_63.70

```
product length = 112
Forward primer  1     ACAGACGCGCCCATTAAGC  19
Template        9580  ...................  9562

Reverse primer  1     CGTCCAGAACAGAGAGCAATAA  22
Template        9469  ......................  9490
```

>NZ_CP055292.1 Shigella sonnei strain SE6-1 chromosome, complete genome

```
product length = 112
Forward primer  1       ACAGACGCGCCCATTAAGC  19
Template        885612  ...................  885594

Reverse primer  1       CGTCCAGAACAGAGAGCAATAA  22
Template        885501  ......................  885522
```

>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome

```
product length = 112
Forward primer  1       ACAGACGCGCCCATTAAGC  19
Template        344764  ...................  344782

Reverse primer  1       CGTCCAGAACAGAGAGCAATAA  22
Template        344875  ......................  344854
```

Supplementary Figure 1. In-silico PCR for *YaHL* primer set using NCBI RefSeq Genome Database.

	Sequence (5'->3')
Forward primer	TTCTCGCTACAGGCACATTC
Reverse primer	GGCGATGCATACTGACTCAT

Products on target templates

>NZ_CP055055.1 Shigella dysenteriae strain SWHEFF_49 chromosome, complete genome

```
product length = 99
Forward primer  1        TTCTCGCTACAGGCACATTC  20
Template        2594763  ....................  2594744

Reverse primer  1        GGCGATGCATACTGACTCAT  20
Template        2594665  ....................  2594684
```

>NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome

```
product length = 98
Forward primer  1        TTCTCGCTACAGGCACATTC  20
Template        2412340  ....................  2412321

Reverse primer  1        GGCGATGCATACTGACTCAT  20
Template        2412243  ....................  2412262
```

>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome

```
product length = 99
Forward primer  1        TTCTCGCTACAGGCACATTC  20
Template        1812490  ....................  1812471

Reverse primer  1        GGCGATGCATACTGACTCAT  20
Template        1812392  ....................  1812411
```

>NZ_CP055292.1 Shigella sonnei strain SE6-1 chromosome, complete genome

```
product length = 99
Forward primer  1        TTCTCGCTACAGGCACATTC  20
Template        4076145  .....A..............  4076164

Reverse primer  1        GGCGATGCATACTGACTCAT  20
Template        4076243  ....................  4076224
```

>NZ_NIYS01000126.1 Shigella boydii strain ESBL-W3-2 NODE_65_length_22488_cov_34.93

```
product length = 99
Forward primer  1    TTCTCGCTACAGGCACATTC  20
Template        183  .....A..............  202

Reverse primer  1    GGCGATGCATACTGACTCAT  20
Template        281  ....................  262
```

Supplementary Figure 2. In-silico PCR for *YdjO* primer set using NCBI RefSeq Genome database.

	Sequence (5'->3')
Forward primer	CAACAGGACGTATGCTCTATCG
Reverse primer	GCCGTAAACCTTCTGCTAACTC

Products on target templates

>NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome

```
product length = 114
Forward primer  1        CAACAGGACGTATGCTCTATCG  22
Template        5410998  ......................  5410977

Reverse primer  1        GCCGTAAACCTTCTGCTAACTC  22
Template        5410885  ......................  5410906
```

>NZ_NIYS01000040.1 Shigella boydii strain ESBL-W3-2 NODE_12_length_97529_cov_43.691

```
product length = 114
Forward primer  1      CAACAGGACGTATGCTCTATCG  22
Template        88002  ......................  88023

Reverse primer  1      GCCGTAAACCTTCTGCTAACTC  22
Template        88115  ......................  88094
```

>NC_000913.3 Escherichia coli str. K-12 substr. MG1655, complete genome

```
product length = 114
Forward primer  1        CAACAGGACGTATGCTCTATCG  22
Template        4556049  ......................  4556028

Reverse primer  1        GCCGTAAACCTTCTGCTAACTC  22
Template        4555936  ......................  4555957
```

>NC_004337.2 Shigella flexneri 2a str. 301 chromosome, complete genome

```
product length = 114
Forward primer  1        CAACAGGACGTATGCTCTATCG  22
Template        4368385  ......................  4368406

Reverse primer  1        GCCGTAAACCTTCTGCTAACTC  22
Template        4368498  ......................  4368477
```

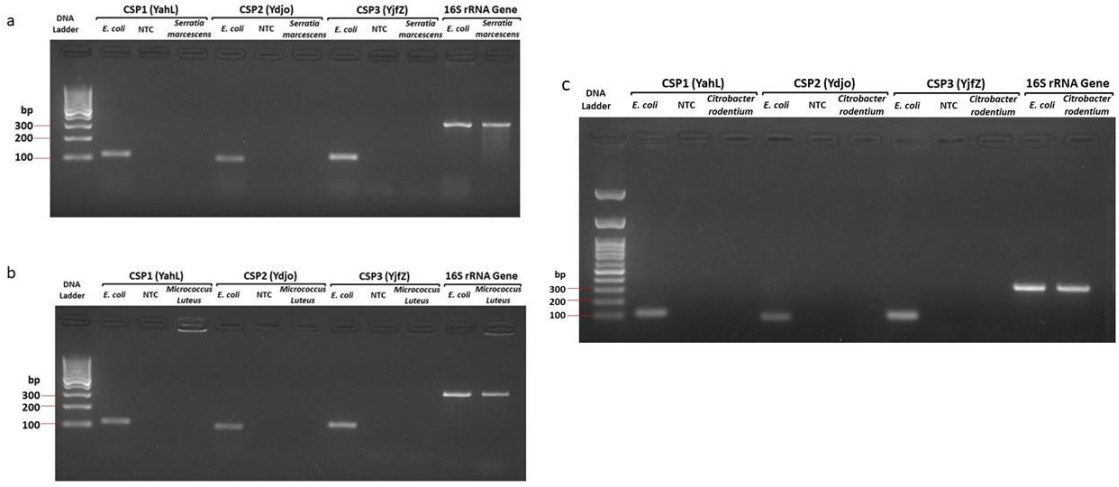
>NZ_CP055055.1 Shigella dysenteriae strain SWHEFF_49 chromosome, complete genome

```
product length = 114
Forward primer  1       CAACAGGACGTATGCTCTATCG  22
Template        516676  ......................  516655

Reverse primer  1       GCCGTAAACCTTCTGCTAACTC  22
Template        516563  ..................T..  516584
```

Supplementary Figure 3. In-silico PCR for *YjfZ* primer set using NCBI RefSeq Genome database.

Supplementary Table 2. Summarized description of in-silico PCR using NCBI nr database for primer sets for *YahL*, *YdjO* and *YjfZ E. coli*/*Shigella*-specific Conserved Signature Proteins/genes.

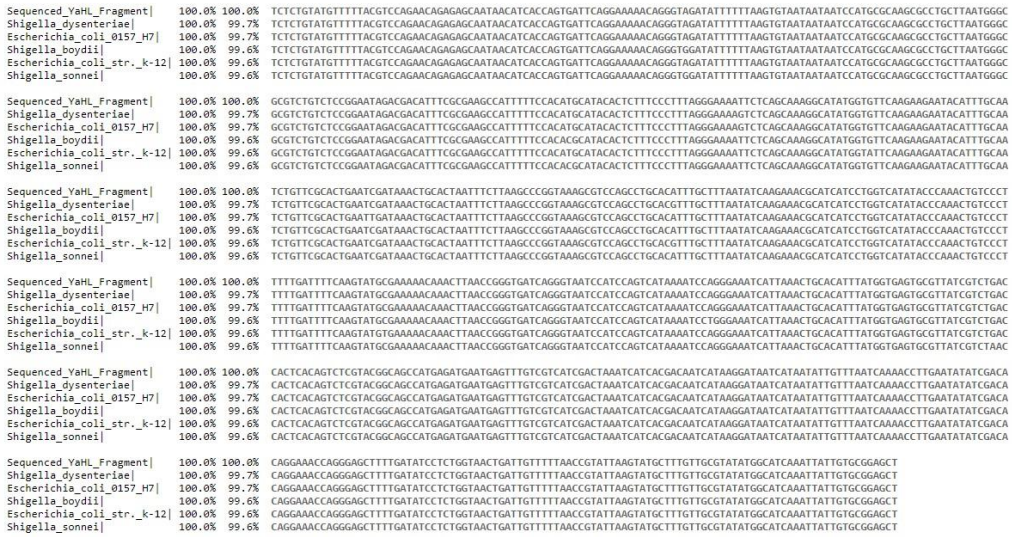
Conserved Signature Protein/Genes	PCR Product Size (bp)	In-silico *E. coli* Matches/Hits	In-silico *Shigella* Matches/Hits	In-silico non-*E. coli* and Non-*Shigella* Matches/Hits	*Eschericia*/*Shigella* Species Matched
YahL	112	950	51	0	*Eschericia coli, Shigella dysenteriae, Shigella sonnei, Shigella flexneri*
YdjO	99	919	82	0	
YjfZ	114	949	57	0	



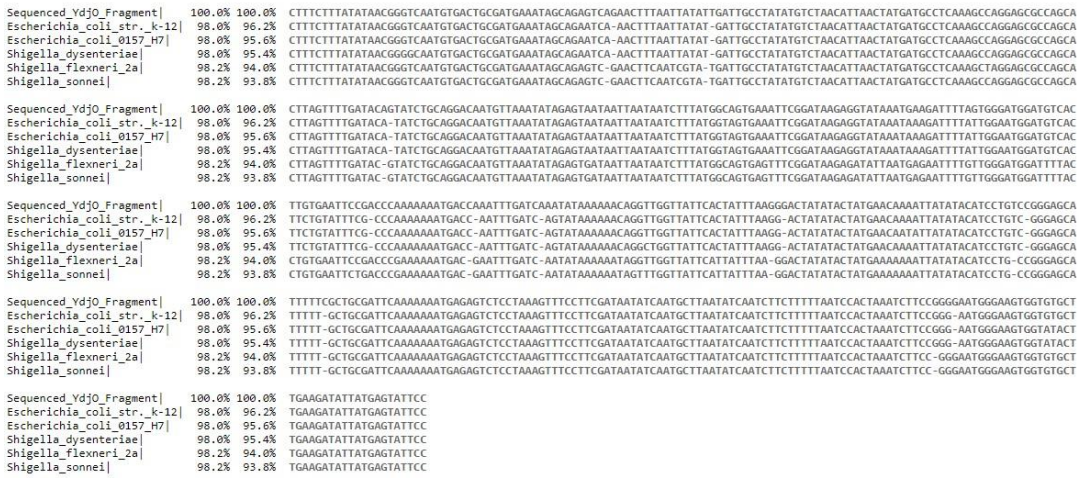
Supplementary Figure 4. Experimental validation of *E. coli*/*Shigella*-specific Conserved Signature Proteins/Genes based primer sets against (a) *Serratia marcescens* (in-group negative control), (b) *Micrococcus luteus* (out-group negative control) and (c) *Citrobacter rodentium* (in-group negative control).

Supplementary Table 3. Mean Threshold Cycle (Cq) values for *YahL, YdjO* and *YjfZ* qPCR assays using *E. coli* DNA (Positive Control), Wastewater DNA (Positive Control), Negative controls and Non-template control.

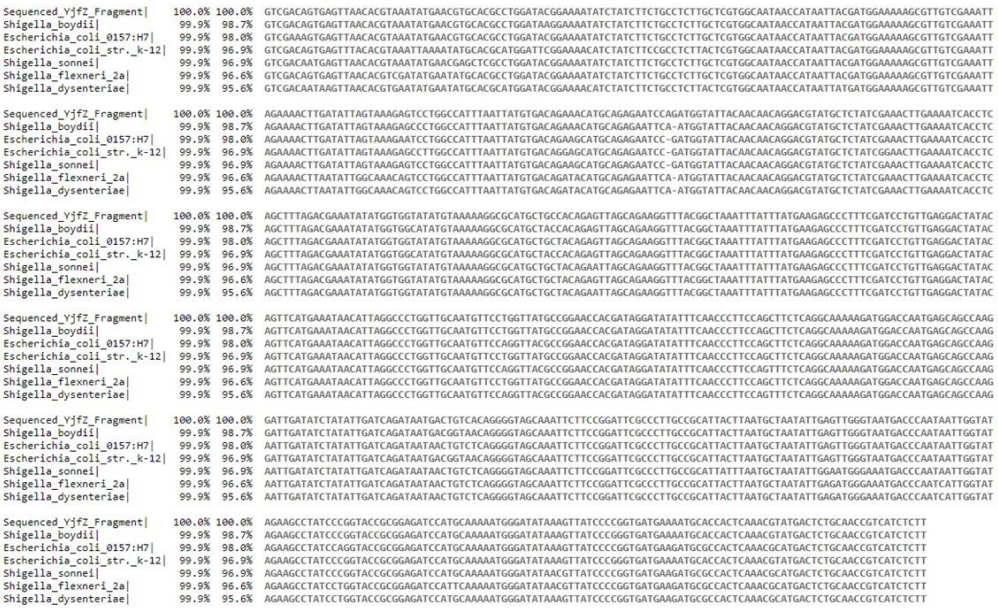
	qPCR Assay	Mean Cq Value
Wastewater DNA	*YahL*	29.2
	YdjO	30.0
	YjfZ	32.3
E. coli K12 Genomic DNA	*YahL*	18.8
	YdjO	17.2
	YjfZ	16.1
Non-Template/Negative Controls	All three qPCR assays	No amplification/fluorescence Signal



Supplementary Figure 5. Multiple Sequence Alignment of Sequenced YahL CSP DNA fragment (from wastewater DNA) against NCBI RefSeq Reference Genome Database. From left to right, Lane 1: Taxonomic matches, Lane 2: Query (Sequenced Fragment) coverage, Lane 3: Percentage Identity, and Lane 4: Sequence alignment.



Supplementary Figure 6. Multiple Sequence Alignment of Sequenced *YdjO* CSP DNA fragment (from wastewater DNA) against NCBI RefSeq Reference Genome Database. From left to right, Lane 1: Taxonomic matches, Lane 2: Query (Sequenced Fragment) coverage, Lane 3: Percentage Identity, and Lane 4: Sequence alignment.



Supplementary Figure 7. Multiple Sequence Alignment of Sequenced YjfZ CSP DNA fragment (from wastewater DNA) against NCBI RefSeq Reference Genome Database. From left to right, Lane 1: Taxonomic matches, Lane 2: Query (Sequenced Fragment) coverage, Lane 3: Percentage Identity, and Lane 4: Sequence alignment.

Supplementary Table 4. Correlation between three qPCR assays using Pathogenic and Environmental *E. coli* isolates (n = 18).

Conserved Signature Proteins		Correlation Coefficient	P-value
YahL	*YdjO*	0.92	2.79E-04
YahL	*YjfZ*	0.74	9.40E-04
YdjO	*YjfZ*	0.73	9.40E-04

Supplementary Table 5. Detection efficiency of three qPCR assays for recreational water samples (n = 30).

qPCR Assay	Number of Positive Samples	Detection Efficiency (%)
YahL	30	100
YdjO	28	93
YjfZ	29	96

	Sequence (5'->3')
Forward primer	TGATTGGCAAAATCTGGCCG
Reverse primer	GAAATCGCCCAAATCGCCAT

Products on target templates

>NZ_CP055055.1 Shigella dysenteriae strain SWHEFF_49 chromosome, complete genome

```
product length = 211
Forward primer  1        TGATTGGCAAAATCTGGCCG  20
Template        1295212  ....................  1295231

Reverse primer  1        GAAATCGCCCAAATCGCCAT  20
Template        1295422  ....................  1295403
```

>NZ_JAINCF010000042.1 Escherichia whittamii strain C2-3 c_000000000042, whole genome shotgun sequence

```
product length = 211
Forward primer  1      TGATTGGCAAAATCTGGCCG  20
Template        27011  ....................  27030

Reverse primer  1      GAAATCGCCCAAATCGCCAT  20
Template        27221  ....................  27202
```

>NZ_CACSXJ020000003.1 Escherichia marmotae strain H1-003-0086-C-F, whole genome shotgun sequence

```
product length = 211
Forward primer  1       TGATTGGCAAAATCTGGCCG  20
Template        152559  ....................  152540

Reverse primer  1       GAAATCGCCCAAATCGCCAT  20
Template        152349  ....................  152368
```

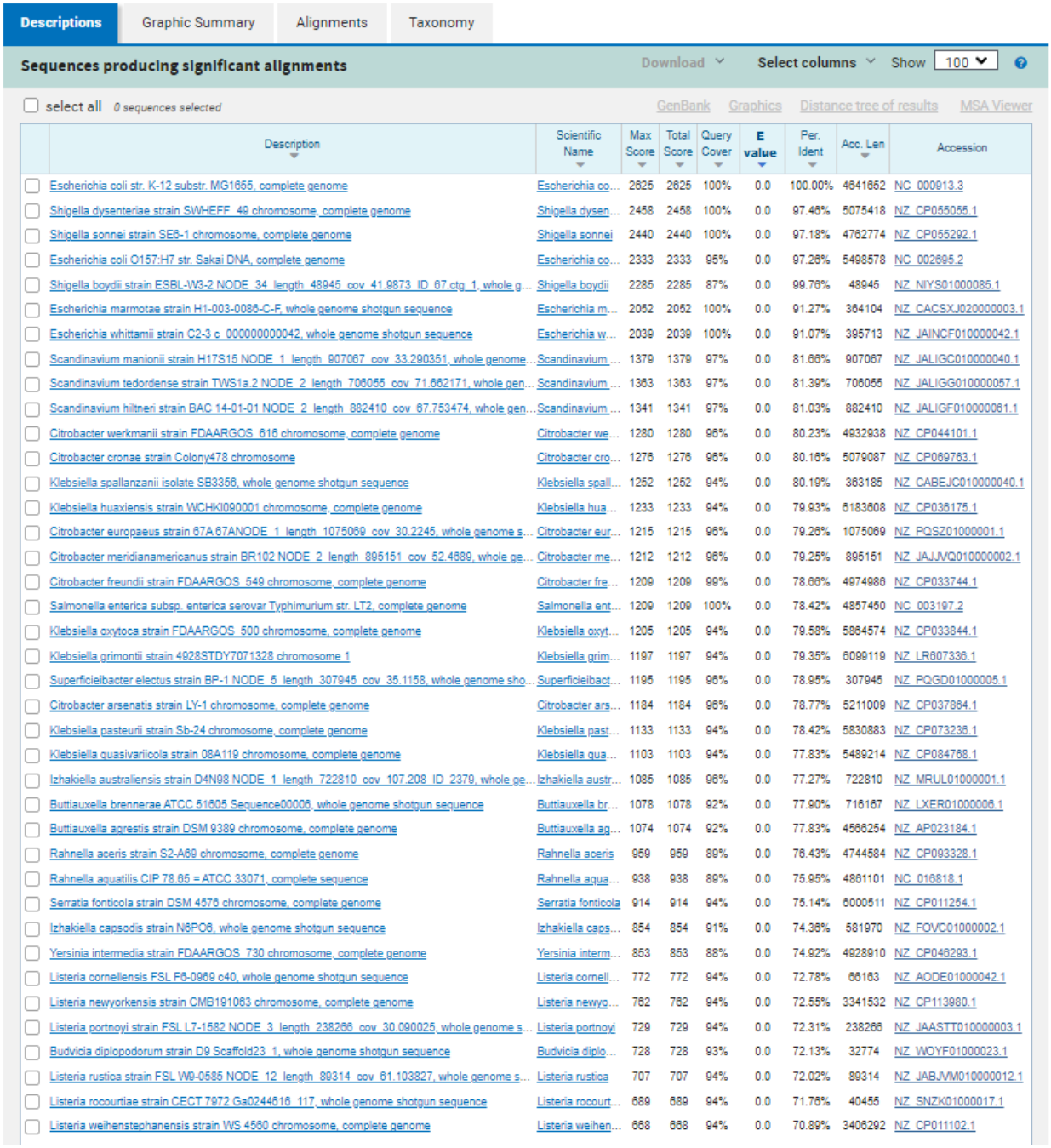
>NZ_CP055292.1 Shigella sonnei strain SE6-1 chromosome, complete genome

```
product length = 211
Forward primer  1       TGATTGGCAAAATCTGGCCG  20
Template        699641  ....................  699622

Reverse primer  1       GAAATCGCCCAAATCGCCAT  20
Template        699431  ....................  699450
```

>NC_002695.2 Escherichia coli O157:H7 str. Sakai DNA, complete genome

```
product length = 211
Forward primer  1       TGATTGGCAAAATCTGGCCG  20
Template        630628  ....................  630647

Reverse primer  1       GAAATCGCCCAAATCGCCAT  20
```

Supplementary Figure 8. In-silico PCR results for *ybbW* primer pair described in Walker et al., 2017) for *E. coli-specific* qPCR assay.



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all *0 sequences selected* GenBank Graphics Distance tree of results MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Escherichia coli str. K-12 substr. MG1655, complete genome	Escherichia co...	2625	2625	100%	0.0	100.00%	4641652	NC_000913.3
Shigella dysenteriae strain SWHEFF_49 chromosome, complete genome	Shigella dysen...	2458	2458	100%	0.0	97.46%	5075418	NZ_CP055055.1
Shigella sonnei strain SE6-1 chromosome, complete genome	Shigella sonnei	2440	2440	100%	0.0	97.18%	4762774	NZ_CP055292.1
Escherichia coli O157:H7 str. Sakai DNA, complete genome	Escherichia co...	2333	2333	95%	0.0	97.26%	5498578	NC_002695.2
Shigella boydii strain ESBL-W3-2 NODE_34_length_48945_cov_41.9873_ID_67.ctg_1, whole g...	Shigella boydii	2285	2285	87%	0.0	99.76%	48945	NZ_NIYS01000085.1
Escherichia marmotae strain H1-003-0086-C-F, whole genome shotgun sequence	Escherichia m...	2052	2052	100%	0.0	91.27%	364104	NZ_CACSXJ020000003.1
Escherichia whittamii strain C2-3 c_000000000042, whole genome shotgun sequence	Escherichia w...	2039	2039	100%	0.0	91.07%	395713	NZ_JAINCF010000042.1
Scandinavium manionii strain H17S15 NODE_1_length_907067_cov_33.290351, whole genome...	Scandinavium ...	1379	1379	97%	0.0	81.66%	907067	NZ_JALIGC010000040.1
Scandinavium tedordense strain TWS1a.2 NODE_2_length_706055_cov_71.662171, whole gen...	Scandinavium ...	1363	1363	97%	0.0	81.39%	706055	NZ_JALIGG010000057.1
Scandinavium hiltneri strain BAC 14-01-01 NODE_2_length_882410_cov_67.753474, whole gen...	Scandinavium ...	1341	1341	97%	0.0	81.03%	882410	NZ_JALIGF010000061.1
Citrobacter werkmanii strain FDAARGOS_616 chromosome, complete genome	Citrobacter we...	1280	1280	96%	0.0	80.23%	4932938	NZ_CP044101.1
Citrobacter cronae strain Colony478 chromosome	Citrobacter cro...	1276	1276	96%	0.0	80.16%	5079087	NZ_CP069763.1
Klebsiella spallanzanii isolate SB3356, whole genome shotgun sequence	Klebsiella spall...	1252	1252	94%	0.0	80.19%	363185	NZ_CABEJC010000040.1
Klebsiella huaxiensis strain WCHKI090001 chromosome, complete genome	Klebsiella hua...	1233	1233	94%	0.0	79.93%	6183608	NZ_CP036175.1
Citrobacter europaeus strain 67A 67ANODE_1_length_1075069_cov_30.2245, whole genome s...	Citrobacter eur...	1215	1215	96%	0.0	79.26%	1075069	NZ_PQSZ01000001.1
Citrobacter meridianamericanus strain BR102 NODE_2_length_895151_cov_52.4689, whole ge...	Citrobacter me...	1212	1212	96%	0.0	79.25%	895151	NZ_JAJJVQ010000002.1
Citrobacter freundii strain FDAARGOS_549 chromosome, complete genome	Citrobacter fre...	1209	1209	99%	0.0	78.66%	4974986	NZ_CP033744.1
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2, complete genome	Salmonella ent...	1209	1209	100%	0.0	78.42%	4857450	NC_003197.2
Klebsiella oxytoca strain FDAARGOS_500 chromosome, complete genome	Klebsiella oxyt...	1205	1205	94%	0.0	79.58%	5864574	NZ_CP033844.1
Klebsiella grimontii strain 4928STDY7071328 chromosome 1	Klebsiella grim...	1197	1197	94%	0.0	79.35%	6099119	NZ_LR607336.1
Superficieibacter electus strain BP-1 NODE_5_length_307945_cov_35.1158, whole genome sho...	Superficieibact...	1195	1195	96%	0.0	78.95%	307945	NZ_PQGD01000005.1
Citrobacter arsenatis strain LY-1 chromosome, complete genome	Citrobacter ars...	1184	1184	96%	0.0	78.77%	5211009	NZ_CP037864.1
Klebsiella pasteurii strain Sb-24 chromosome, complete genome	Klebsiella past...	1133	1133	94%	0.0	78.42%	5830883	NZ_CP073236.1
Klebsiella quasivariicola strain 08A119 chromosome, complete genome	Klebsiella qua...	1103	1103	94%	0.0	77.83%	5489214	NZ_CP084768.1
Izhakiella australiensis strain D4N98 NODE_1_length_722810_cov_107.208_ID_2379, whole ge...	Izhakiella austr...	1085	1085	96%	0.0	77.27%	722810	NZ_MRUL01000001.1
Buttiauxella brennerae ATCC 51605 Sequence00006, whole genome shotgun sequence	Buttiauxella br...	1078	1078	92%	0.0	77.90%	716167	NZ_LXER01000006.1
Buttiauxella agrestis strain DSM 9389 chromosome, complete genome	Buttiauxella ag...	1074	1074	92%	0.0	77.83%	4566254	NZ_AP023184.1
Rahnella aceris strain S2-A69 chromosome, complete genome	Rahnella aceris	959	959	89%	0.0	76.43%	4744584	NZ_CP093328.1
Rahnella aquatilis CIP 78.65 = ATCC 33071, complete sequence	Rahnella aqua...	938	938	89%	0.0	75.95%	4861101	NC_016818.1
Serratia fonticola strain DSM 4576 chromosome, complete genome	Serratia fonticola	914	914	94%	0.0	75.14%	5000511	NZ_CP011254.1
Izhakiella capsodis strain N6PO6, whole genome shotgun sequence	Izhakiella caps...	854	854	91%	0.0	74.38%	581970	NZ_FOVC01000002.1
Yersinia intermedia strain FDAARGOS_730 chromosome, complete genome	Yersinia interm...	853	853	88%	0.0	74.92%	4928910	NZ_CP046293.1
Listeria cornellensis FSL F6-0969 c40, whole genome shotgun sequence	Listeria cornell...	772	772	94%	0.0	72.78%	66163	NZ_AODE01000042.1
Listeria newyorkensis strain CMB191063 chromosome, complete genome	Listeria newyo...	762	762	94%	0.0	72.55%	3341532	NZ_CP113980.1
Listeria portnoyi strain FSL L7-1582 NODE_3_length_238266_cov_30.090025, whole genome s...	Listeria portnoyi	729	729	94%	0.0	72.31%	238266	NZ_JAASTT010000003.1
Budvicia diplopodorum strain D9 Scaffold23_1, whole genome shotgun sequence	Budvicia diplo...	728	728	93%	0.0	72.13%	32774	NZ_WOYF01000023.1
Listeria rustica strain FSL W9-0585 NODE_12_length_89314_cov_61.103827, whole genome s...	Listeria rustica	707	707	94%	0.0	72.02%	89314	NZ_JABJVM010000012.1
Listeria rocourtiae strain CECT 7972 Ga0244616_117, whole genome shotgun sequence	Listeria rocourt...	689	689	94%	0.0	71.76%	40455	NZ_SNZK01000017.1
Listeria weihenstephanensis strain WS 4560 chromosome, complete genome	Listeria weihen...	668	668	94%	0.0	70.89%	3406292	NZ_CP011102.1

Supplementary Figure 9. BLASTn results for *ybbW* gene (described by Walker et al., 2017) against the NCBI RefSeq Representative genomes.

Additional Files for Chapter 6

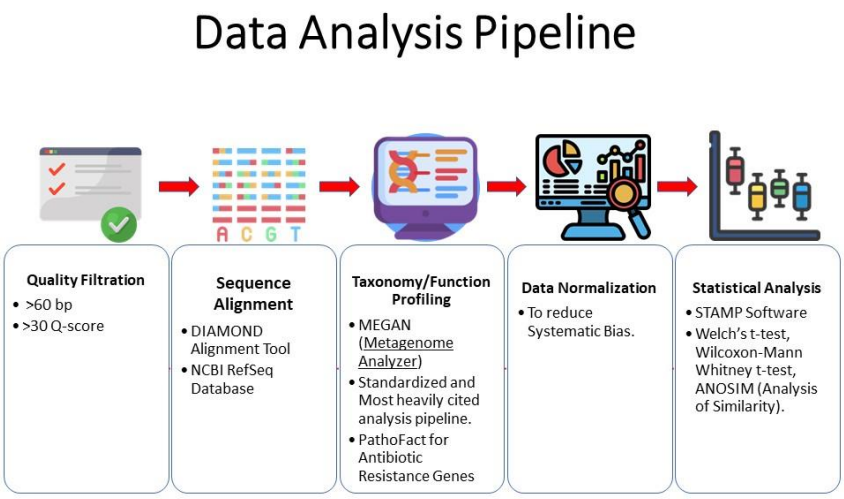
Supplementary Table 1. Distribution of samples for shotgun sequencing.

Sampling Location	Number of Beach Action Value Exceedance Day Samples	Number of Beach Action Value Non-Exceedance Day Samples
Marie Curtis Park East Beach	9	9
Sunnyside Beach	3	3
Etobicoke Creek	18	
Humber River	6	

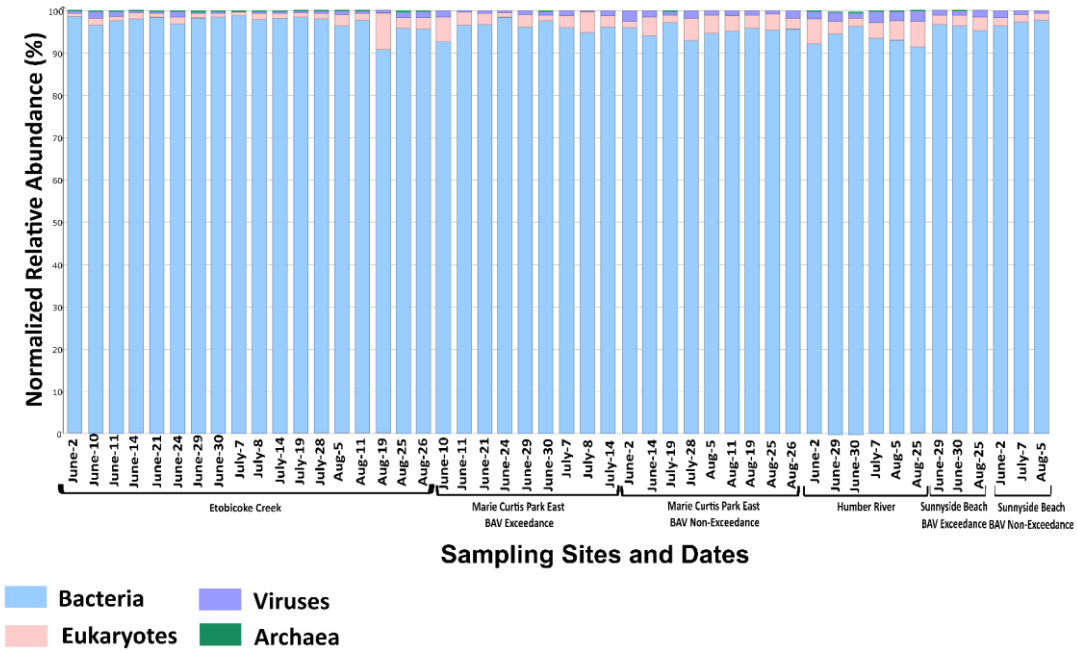
Supplementary Table 2. Metagenomics and qPCR Quality Analytics.

Data Analytics	Number of Sequences (Mean ± Standard Deviation)	Sampling Locations
Before Quality Filtration	12,925,145 ± 3,216,845	Marie Curtis Park East Beach and Etobicoke Creek
After Quality Filtration	10,694,704 ± 2,788,732	
Before Quality Filtration	13,739,200 ± 3,523,743	Sunnyside Beach and Humber River
After Quality Filtration	11,479,959 ± 3,204,589	
qPCR Standard Curve Quality Parameters		
Quality Parameter	**Value**	**Assay Type**
Calibration curve (R^2)	0.992	*Enterococcus* (LsrRNA, 23S rRNA)
Slope	-3.25	
Intercept	38.66	
Efficiency (%)	100	
Calibration curve (R^2)	0.9998	Microcystin (mcyE)
Slope	-3.37	
Intercept	38.51	
Efficiency (%)	97.84	
Calibration curve (R^2)	0.9996	Saxitoxin (sxtA)
Slope	-3.33	
Intercept	38.61	
Efficiency (%)	99	
Calibration curve (R^2)	0.9996	Cylindrospermopsin (cyrA)
Slope	-3.29	
Intercept	38.92	

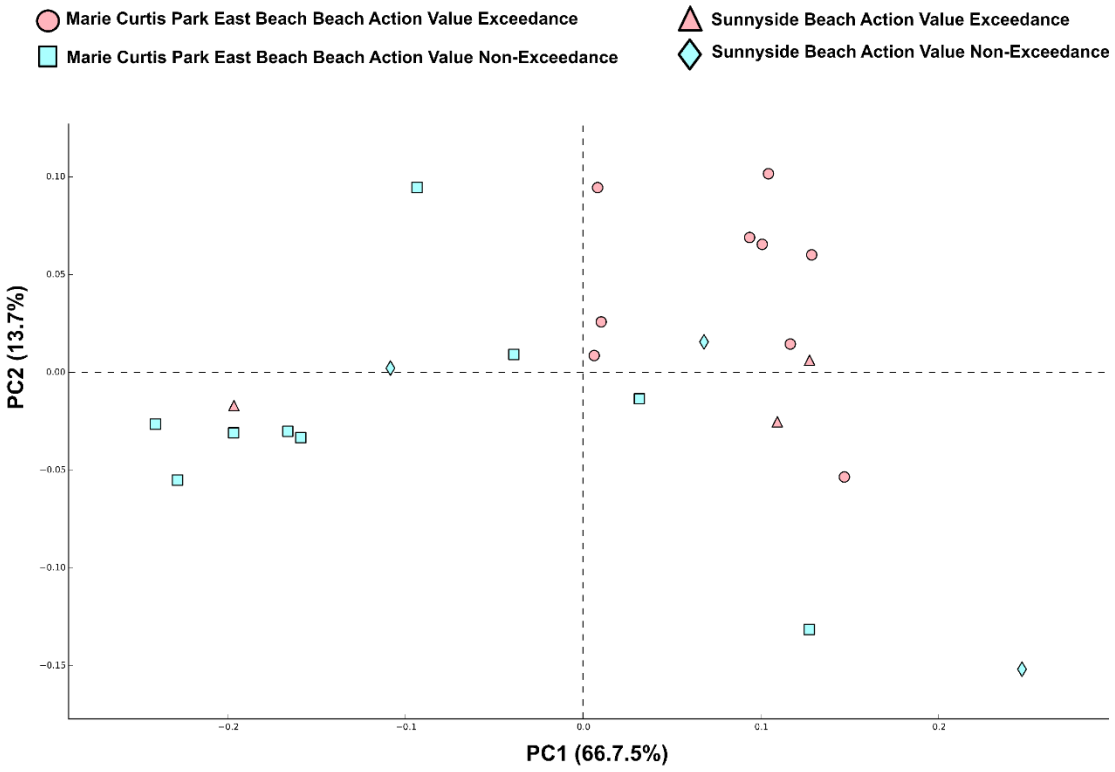
Efficiency (%)	101	Total Cyanobacteria (16S rRNA)
Calibration curve (R^2)	0.9999	
Slope	-3.46	
Intercept	39.13	
Efficiency (%)	94	



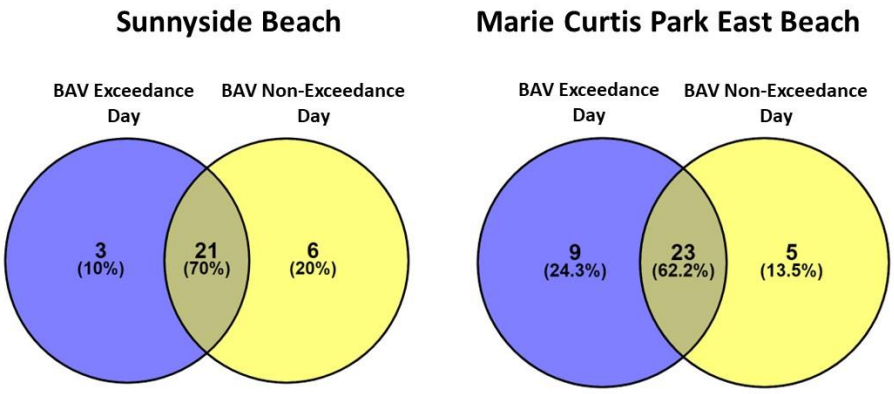
Supplementary Figure 1. Flowchart describing data filtration and analysis.



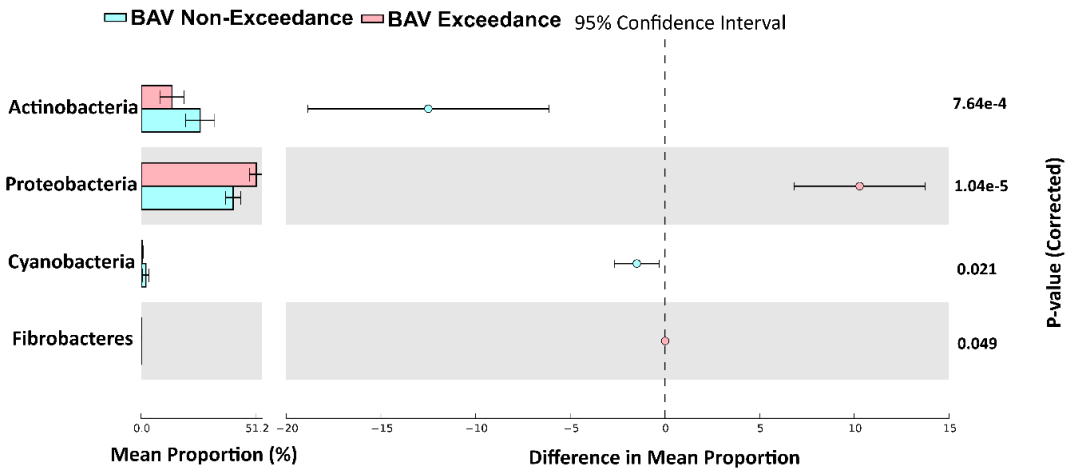
Supplementary Figure 2. Relative percentage taxonomic abundance of samples from Marie Curtis Park East Beach, Sunnyside Beach, Etobicoke Creek, and Humber River on Kingdom Level.



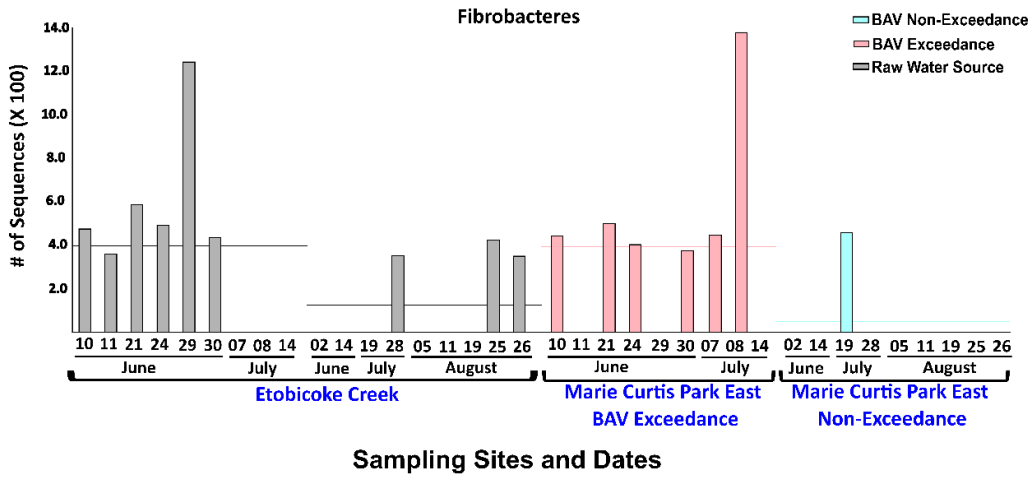
Supplementary Figure 3. Differential abundance Principal Component Analysis plots for Beach Action Value Exceedance and Non-Exceedance samples from Marie Curtis Park East and Sunnyside Beaches.



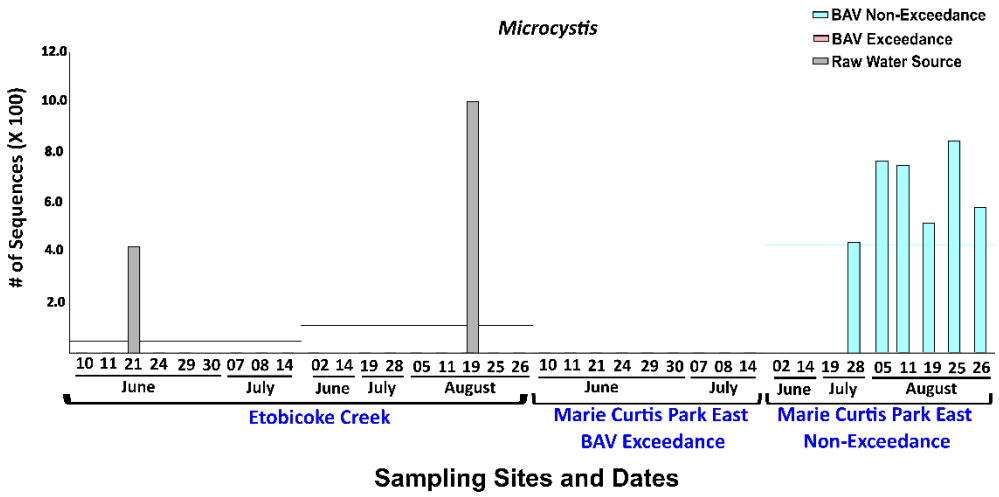
Supplementary Figure 4. Venn Diagram Analysis of Shared/Unique bacterial genera in Core Microbiome between Beach Action Value (BAV) Exceedance Beach Days and Non-Exceedance Beach Days from Marie Curtis Park East and Sunnyside Beaches.



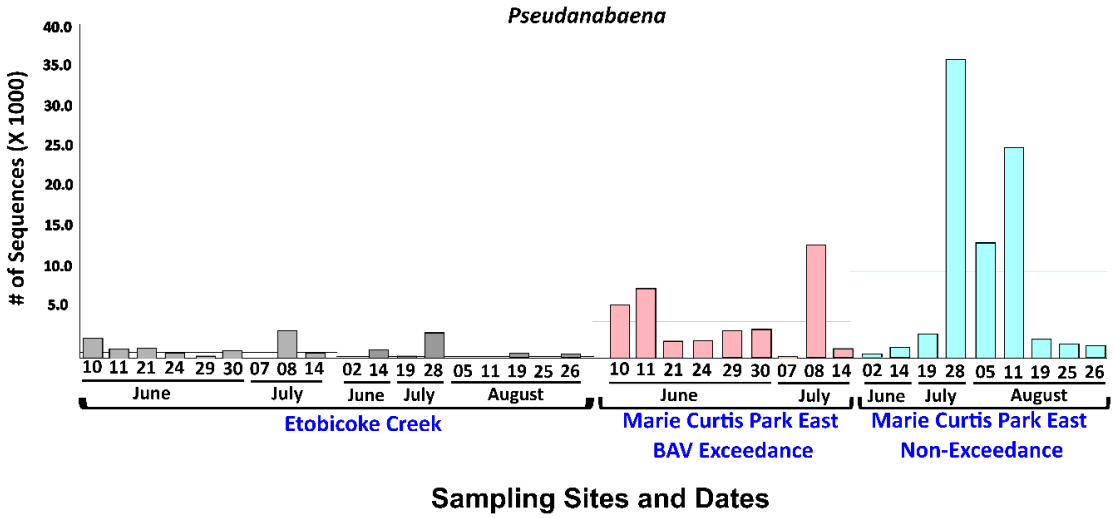
Supplementary Figure 5. Differential abundance extended error bar plot on phylum level for Beach Action Value Exceedance and Non-Exceedance samples from Marie Curtis Park East Beach.



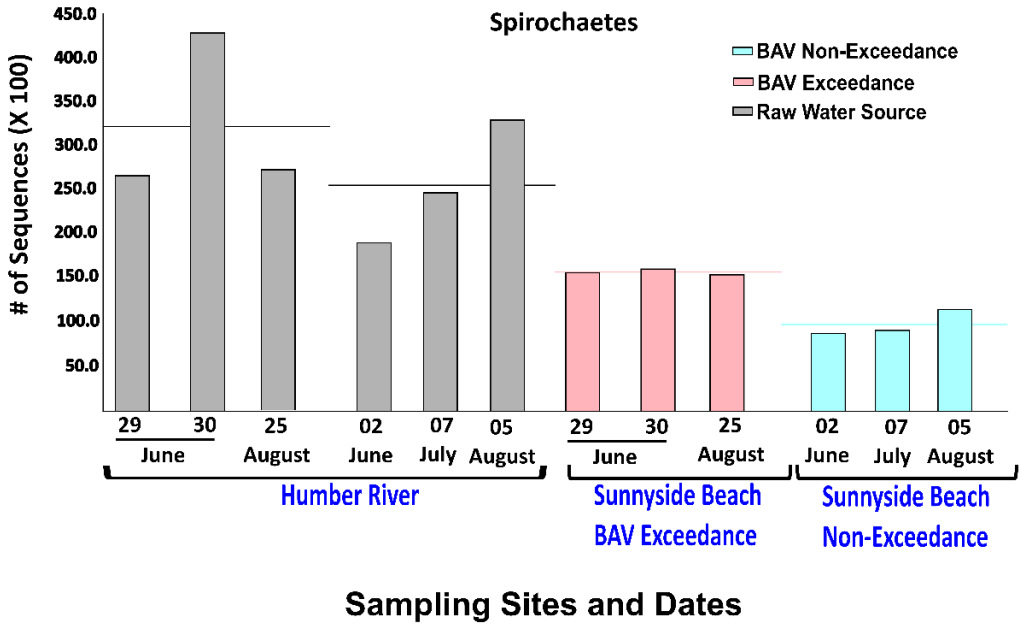
Supplementary Figure 6. Bar plot of Fibrobacteres abundance for Marie Curtis Park East Beach and Etobicoke Creek samples. The horizontal line represents the average number of normalized sequences for each group.



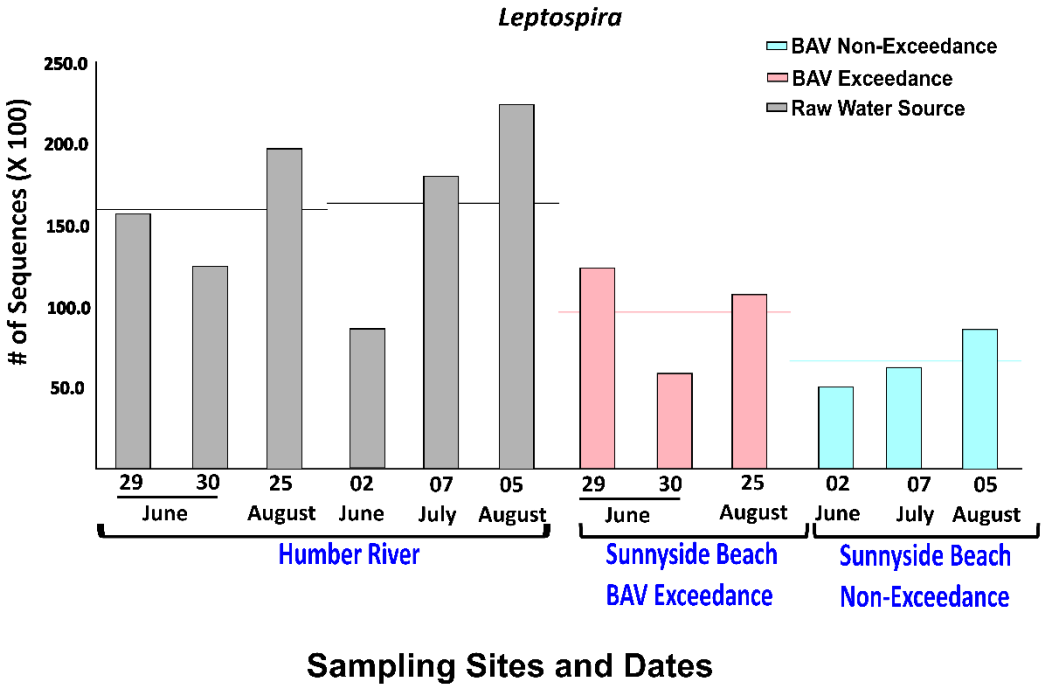
Supplementary Figure 7. Bar plot of *Microcystis* abundance for Marie Curtis Park East Beach and Etobicoke Creek samples. The horizontal line represents the average number of normalized sequences for each group.



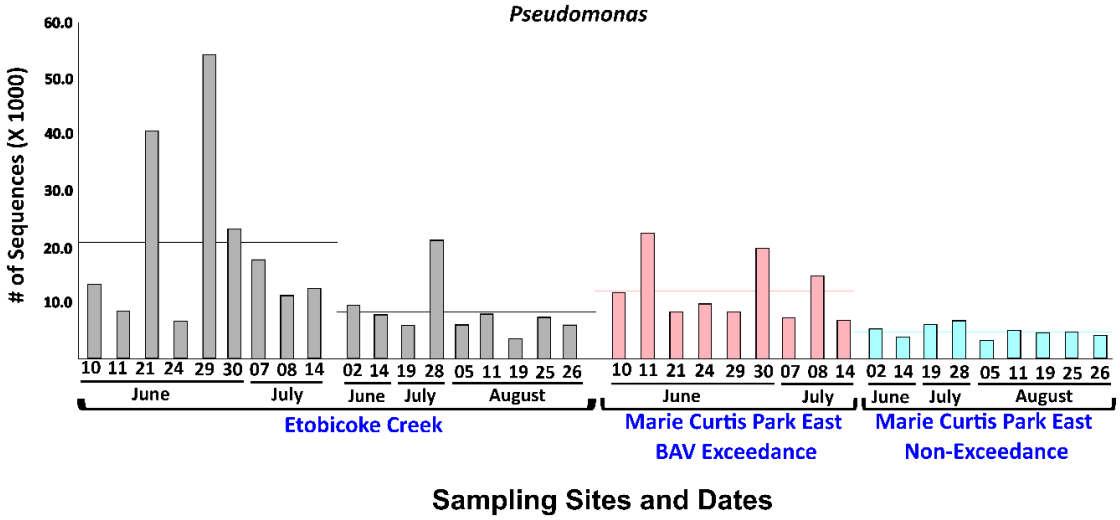
Supplementary Figure 8. Bar plot of *Pseudanabaena* abundance for Marie Curtis Park East Beach and Etobicoke Creek samples. The horizontal line represents the average number of normalized sequences for each group.



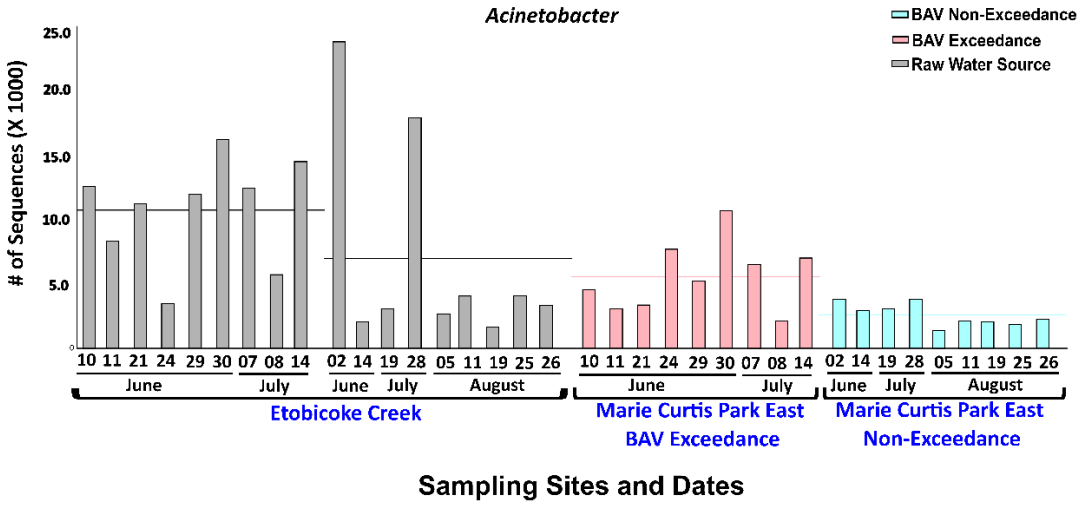
Supplementary Figure 9. Bar plot of Spirochaetes abundance for samples from Sunnyside Beach and Humber River. The horizontal line represents the average number of normalized sequences for each group.



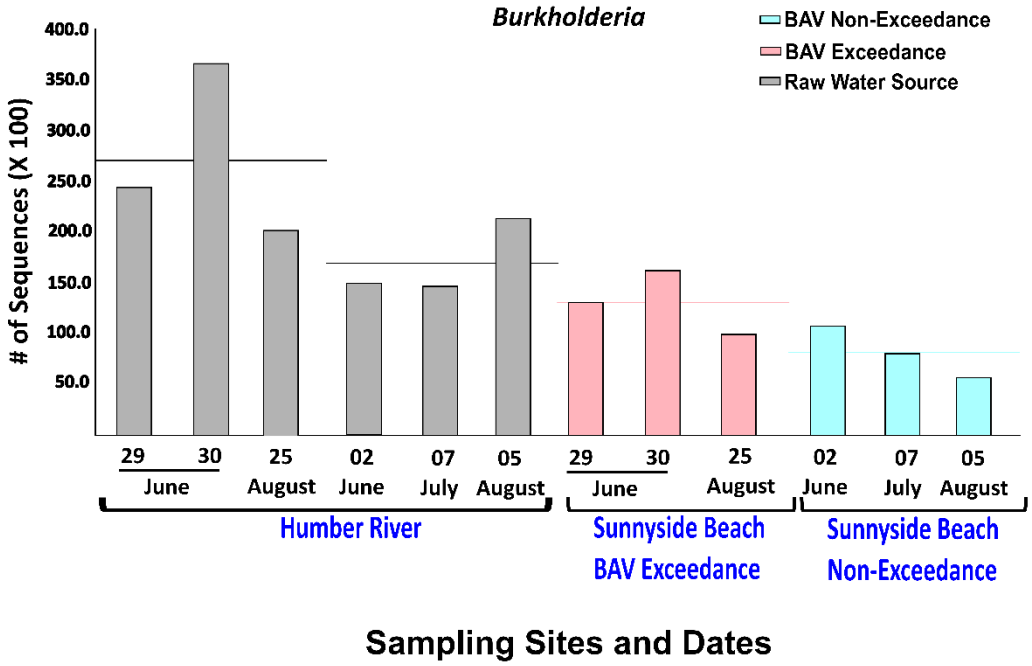
Supplementary Figure 10. Bar plot of *Leptospira* abundance for samples from Sunnyside Beach and Humber River. The horizontal line represents the average number of normalized sequences for each group.



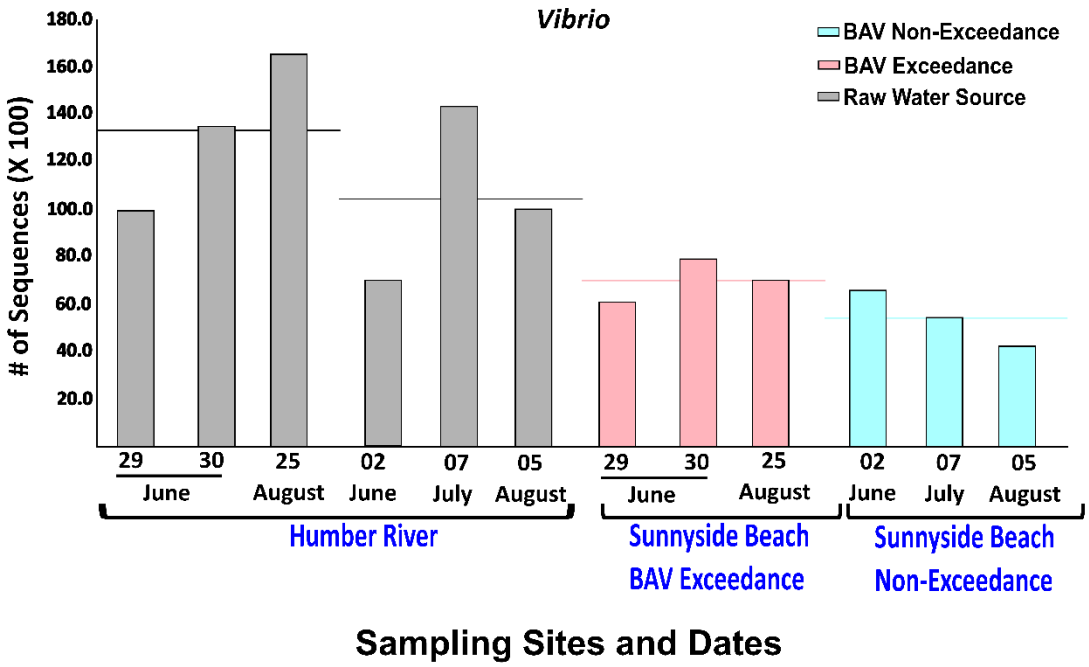
Supplementary Figure 11. Bar plot of *Pseudomonas* abundance for Marie Curtis Park East Beach and Etobicoke Creek samples. The horizontal line represents the average number of normalized sequences for each group.



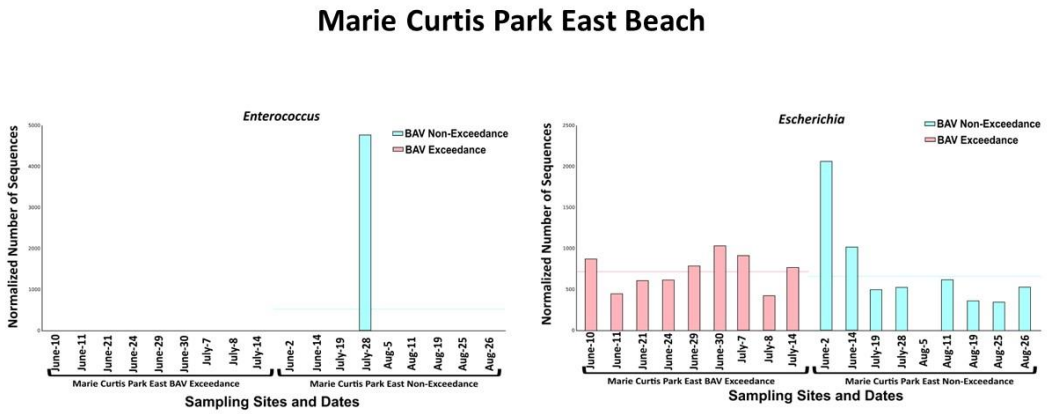
Supplementary Figure 12. Bar plot of *Acinetobacter* abundance for samples from Marie Curtis Park East Beach and Etobicoke Creek. The horizontal line represents the average number of normalized sequences for each group.



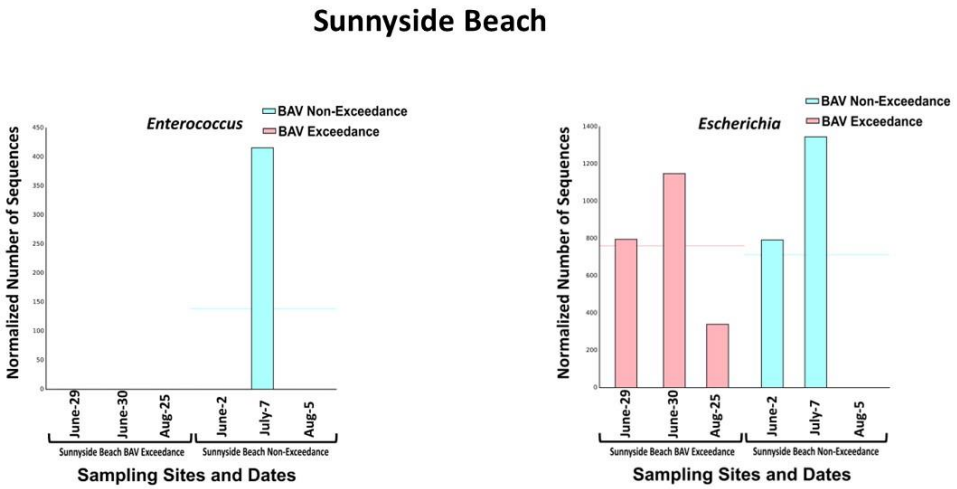
Supplementary Figure 13. The bar plot of *Burkholderia* abundance for Sunnyside Beach and Humber River samples. The horizontal line represents the average number of normalized sequences for each group.



Supplementary Figure 14. Bar plot of *Vibrio* abundance for samples from Sunnyside Beach and Humber River. The horizontal line represents the average number of normalized sequences for each group.



Supplementary Figure 15. Bar plots of *Escherichia* and *Enterococcus* abundance for Marie Curtis Park East Beach samples. The horizontal line represents the average number of normalized sequences for each group.



Supplementary Figure 16. Bar plots of *Escherichia* and *Enterococcus* abundance for samples from Sunnyside Beach. The horizontal line represents the average number of normalized sequences for each group.

Supplementary Table 3. Correlation analysis between Faecal indicator densities and Antibiotic Resistance Genes for Marie Curtis Park East Beach, Sunnyside Beach, Etobicoke Creek and Humber River.

Faecal Indicator	Antibiotic Resistance	Correlation Coefficient	P-value	Sampling Site
E. coli	Aminoglycoside Resistance	0.4	0.09	Marie Curtis Park East Beach
	Beta-lactam Resistance	0.1	0.4	
	Multidrug Resistance	0.2	0.3	
	Tetracycline Resistance	0.1	0.6	
	Macrolides-Lincosamides Resistance	0.09	0.7	
Enterococcus	Aminoglycoside Resistance	0.3	0.2	Marie Curtis Park East Beach
	Beta-lactam Resistance	0.08	0.7	
	Multidrug Resistance	0.1	0.4	
	Tetracycline Resistance	0.1	0.6	
	Macrolides-Lincosamides Resistance	0.03	0.8	
E. coli	Aminoglycoside Resistance	0.2	0.2	Etobicoke Creek
	Beta-lactam Resistance	0.1	0.5	
	Multidrug Resistance	0.1	0.6	
	Tetracycline Resistance	0.2	0.3	
	Macrolides-Lincosamides Resistance	0.3	0.1	
	Aminoglycoside Resistance	0.3	0.1	

	Beta-lactam Resistance	0.2	0.2	
Enterococcus	Multidrug Resistance	0.1	0.6	Etobicoke Creek
	Tetracycline Resistance	0.2	0.2	
	Macrolides-Lincosamides Resistance	0.3	0.1	
	Aminoglycoside Resistance	0.02	1	
	Beta-lactam Resistance	0.3	0.5	
E. coli	Multidrug Resistance	0.02	1	Sunnyside Beach
	Tetracycline Resistance	0.6	0.1	
	Macrolides-Lincosamides Resistance	0.05	0.9	
	Aminoglycoside Resistance	0.4	0.4	
	Beta-lactam Resistance	0.5	0.2	
Enterococcus	Multidrug Resistance	0.5	0.2	Sunnyside Beach
	Tetracycline Resistance	0.6	0.2	
	Macrolides-Lincosamides Resistance	0.6	0.1	
	Aminoglycoside Resistance	0.4	0.3	
	Beta-lactam Resistance	0.4	0.3	
E. coli	Multidrug Resistance	0.2	0.6	Humber River
	Tetracycline Resistance	0.5	0.2	
	Macrolides-Lincosamides Resistance	0.08	0.9	

Enterococcus	Aminoglycoside Resistance	0.7	0.1	Humber River
	Beta-lactam Resistance	0.8	0.06	
	Multidrug Resistance	-0.8	0.03	
	Tetracycline Resistance	-0.8	0.03	
	Macrolides-Lincosamides Resistance	0.4	0.3	

Additional Files for Chapter 7

Supplementary Table 1. Description of samples processed for eDNA metabarcoding.

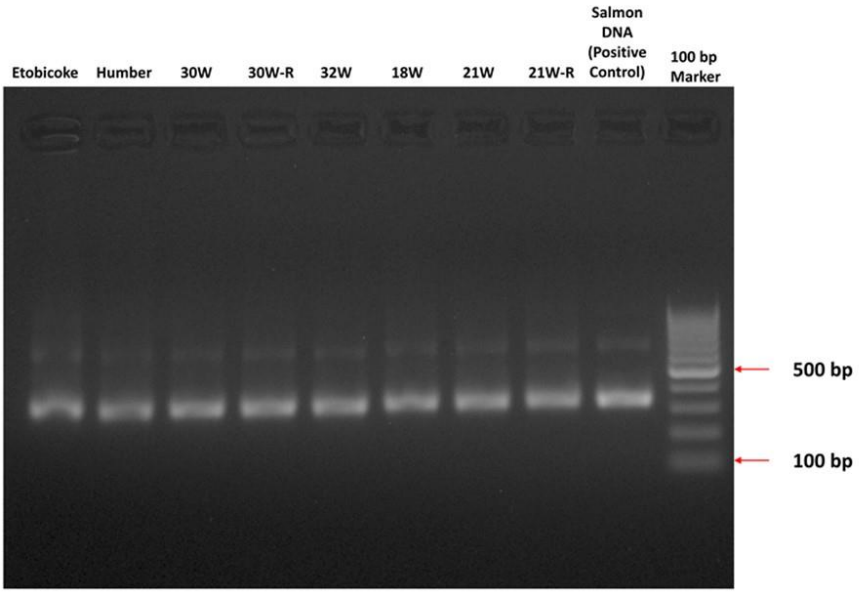
Sample Name	Sample Description	Sample Type	Sampling Location
MC-Ex-June10	Marie Curtis BAV Exceedance Day	Beach Water	Toronto
MC-Ex-June30	Marie Curtis BAV Exceedance Day	Beach Water	Toronto
MC-Ex-July7	Marie Curtis BAV Exceedance Day	Beach Water	Toronto
MC-Non-ex-June2	Marie Curtis BAV Non-Exceedance Day	Beach Water	Toronto
MC-Non-ex-July19	Marie Curtis BAV Non-Exceedance Day	Beach Water	Toronto
MC-Non-ex-August25	Marie Curtis BAV Non-Exceedance Day	Beach Water	Toronto
SS-Ex-June29	Sunnyside BAV Exceedance Day	Beach Water	Toronto
SS-Ex-June30	Sunnyside BAV Exceedance Day	Beach Water	Toronto
SS-Ex-August25	Sunnyside BAV Exceedance Day	Beach Water	Toronto
SS-Non-ex-June2	Sunnyside BAV Non-Exceedance Day	Beach Water	Toronto
SS-Non-ex-July7	Sunnyside BAV Non-Exceedance Day	Beach Water	Toronto
SS-Non-ex-August5	Sunnyside BAV Non-Exceedance Day	Beach Water	Toronto
Hum-Ex-June29	Humber River BAV Exceedance Day	Source Water	Toronto
Hum-Ex-June30	Humber River BAV Exceedance Day	Source Water	Toronto
Hum-Ex-August25	Humber River BAV Exceedance Day	Source Water	Toronto
Hum-Non-ex-June2	Humber River BAV Non-Exceedance Day	Source Water	Toronto
Hum-Non-ex-July7	Humber River BAV Non-Exceedance Day	Source Water	Toronto

Hum-Non-ex-August5	Humber River BAV Non-Exceedance Day	Source Water	Toronto
ETOB-Ex-June10	Etobicoke Creek BAV Exceedance Day	Source Water	Toronto
ETOB-Ex-June30	Etobicoke Creek BAV Exceedance Day	Source Water	Toronto
ETOB-Ex-July7	Etobicoke Creek BAV Exceedance Day	Source Water	Toronto
ETOB-Non-ex-June2	Etobicoke Creek BAV Non-Exceedance Day	Source Water	Toronto
ETOB-Non-ex-July28	Etobicoke Creek BAV Non-Exceedance Day	Source Water	Toronto
ETOB-Non-ex-August26	Etobicoke Creek BAV Non-Exceedance Day	Source Water	Toronto
LK-Ex-August4	Lakeside Beach BAV Exceedance Day	Beach Water	Niagara
LK-Ex-August9	Lakeside Beach BAV Exceedance Day	Beach Water	Niagara
Sun-Ex-August11	Sunset Beach BAV Exceedance Day	Beach Water	Niagara
LK-Non-ex-August10	Lakeside Beach BAV Non-Exceedance Day	Beach Water	Niagara
LK-Non-ex-August17	Lakeside Beach BAV Non-Exceedance Day	Beach Water	Niagara
Sun-Non-ex-August30	Sunset Beach BAV Non-Exceedance Day	Beach Water	Niagara
Sun-Ex-June7	Sunset Beach BAV Exceedance Day	Beach Water	Niagara
Sun-Ex-July21	Sunset Beach BAV Exceedance Day	Beach Water	Niagara
Sun-Ex-August9	Sunset Beach BAV Exceedance Day	Beach Water	Niagara
Sun-Non-Ex-June8	Sunset Beach BAV Non-Exceedance Day	Beach Water	Niagara
Sun-Non-Ex-July27	Sunset Beach BAV Non-Exceedance Day	Beach Water	Niagara
Sun-Non-Ex-August25	Sunset Beach BAV Non-Exceedance Day	Beach Water	Niagara
LK-Sand-Ex-June23	Lakeside Sand	Shoreline Sand-porewater	Niagara

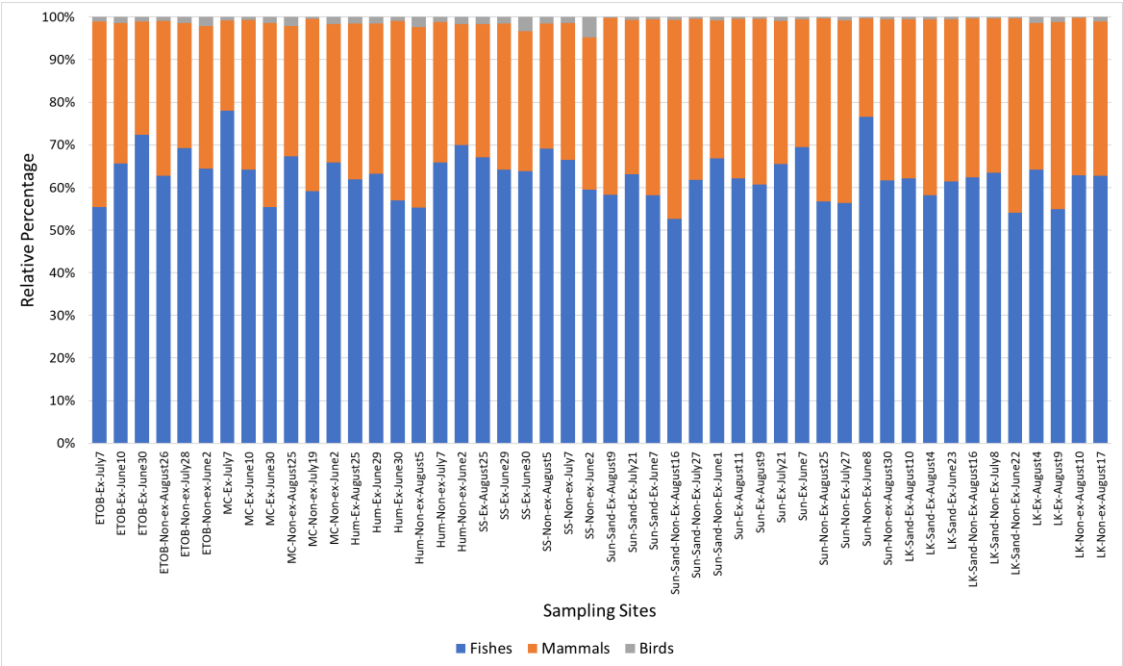
LK-Sand-Ex-August4	Lakeside Sand	Shoreline Sand-porewater	Niagara
LK-Sand-Ex-August10	Lakeside Sand	Shoreline Sand-porewater	Niagara
LK-Sand-June22	Lakeside Sand	Shoreline Sand-porewater	Niagara
LK-Sand-July8	Lakeside Sand	Shoreline Sand-porewater	Niagara
LK-Sand-August17	Lakeside Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Ex-June7	Sunset Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Ex-July21	Sunset Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Ex-August9	Sunset Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Non-Ex-June1	Sunset Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Non-Ex-July27	Sunset Sand	Shoreline Sand-porewater	Niagara
Sun-Sand-Non-Ex-August16	Sunset Sand	Shoreline Sand-porewater	Niagara

Supplementary Table 2. Quality control analytics before and after each eDNA analysis pipeline filtration step (n = 48).

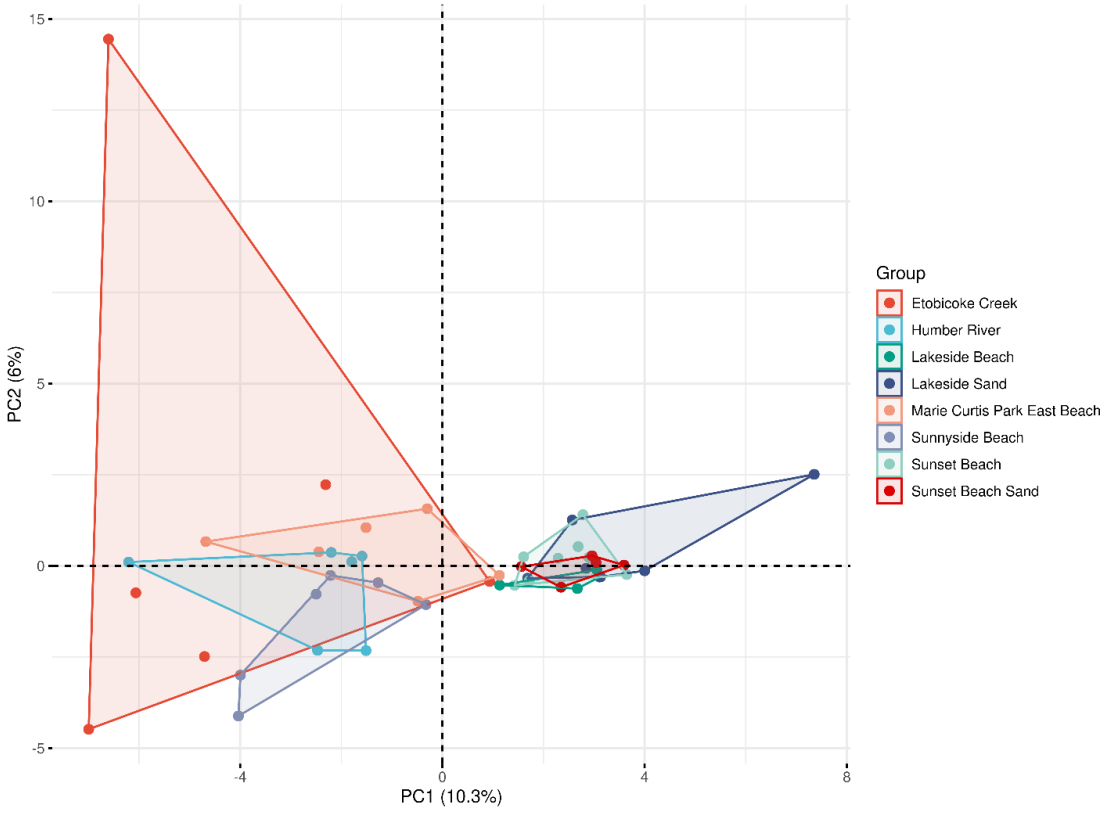
Data Quality Step	Process Summary	
Paired-end Merging	Before Merging	222805 ± 224
	After Merging	219605 ± 222
Trimming	Before Trimming	219605 ± 222
	After Trimming	219551 ± 224
Quality Filtering	Processed Sequences	219551 ± 224
	Passed Sequences	215993 ± 222
Dereplication	Processed Sequences	215993 ± 222
	Unique Sequences	40653 ± 732
Exact Sequence Variants (ESVs)	Before Filtration	369
	After Filtration	200



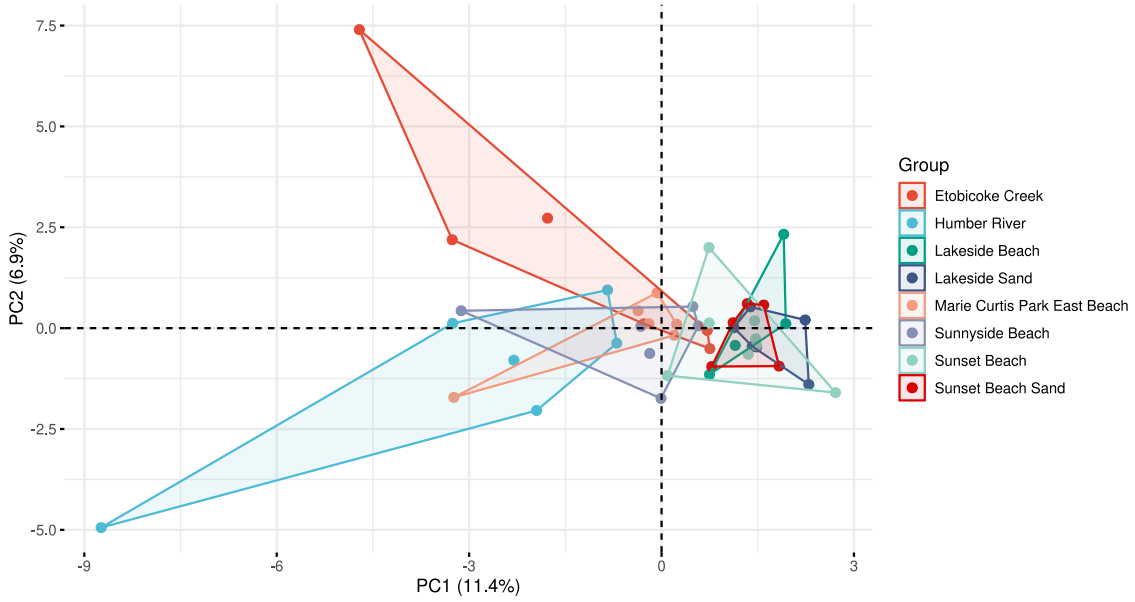
Supplementary Figure 1. Amplicon PCR for mitochondrial 16S rRNA gene.



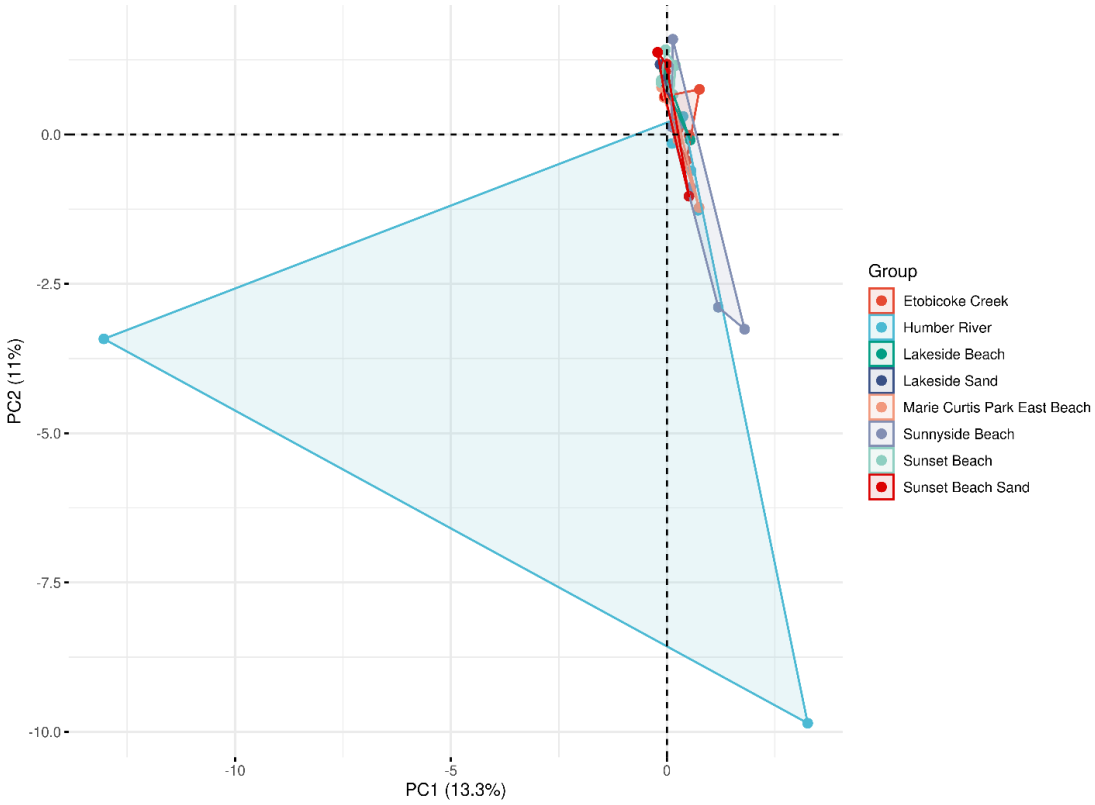
Supplementary Figure 2. Relative abundance of Exact Sequence Variants associated with Fishes, Mammals and Birds in all samples (n = 48).



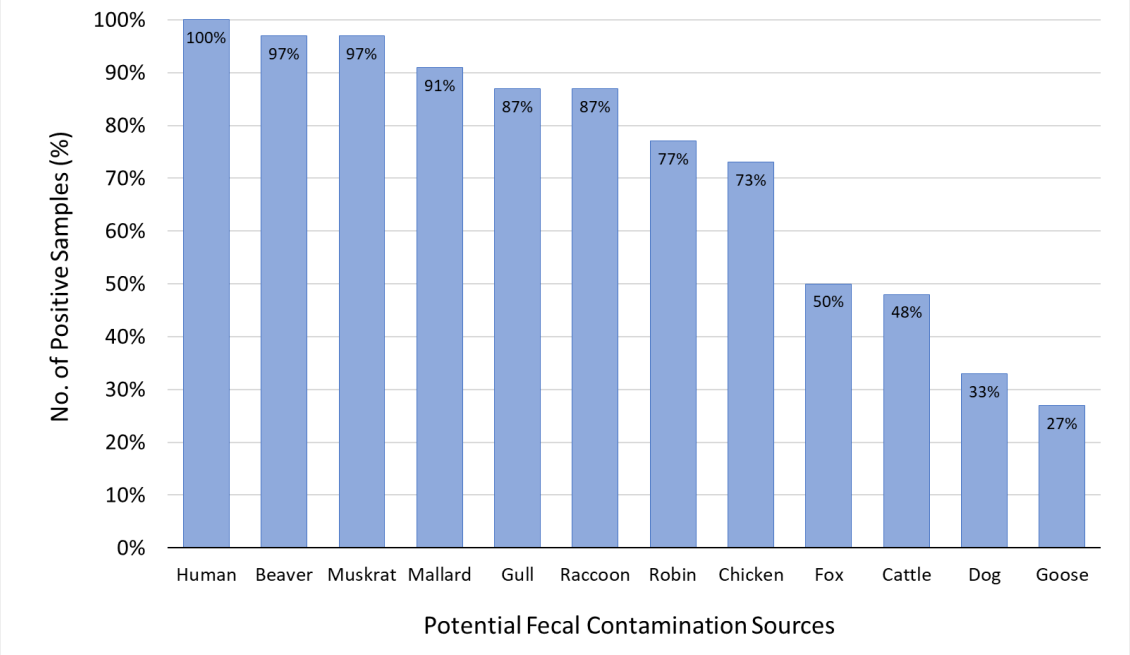
Supplementary Figure 3. Principle Component Analysis (Beta Diversity) of Fish species for Toronto and Niagara beaches, rivers and sand samples.



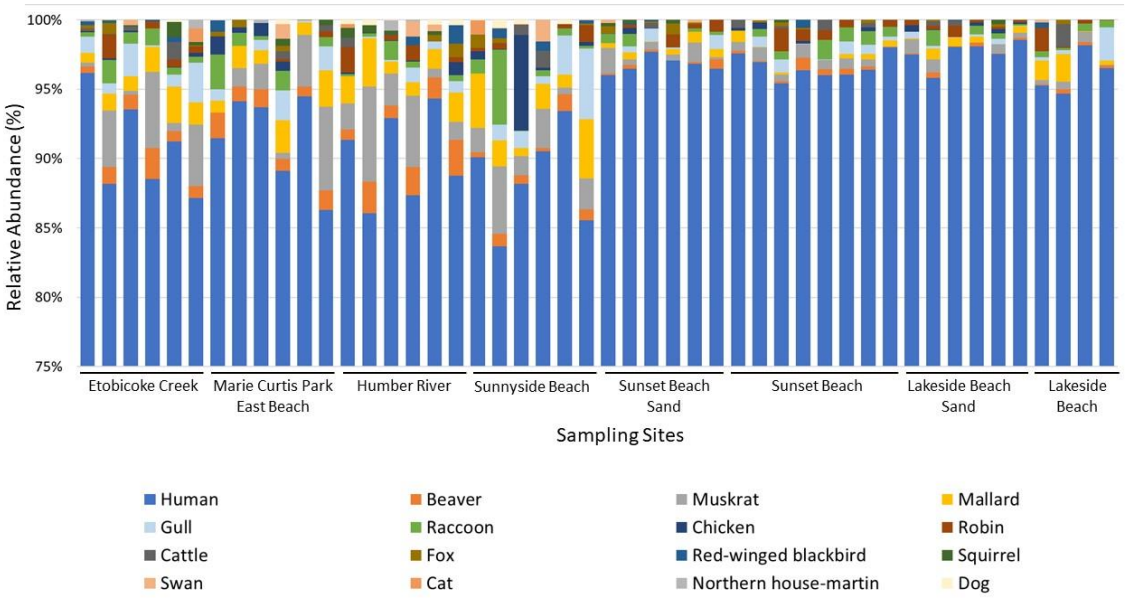
Supplementary Figure 4. Principle Component Analysis (Beta Diversity) of Mammal species for Toronto and Niagara beaches, rivers and sand samples.



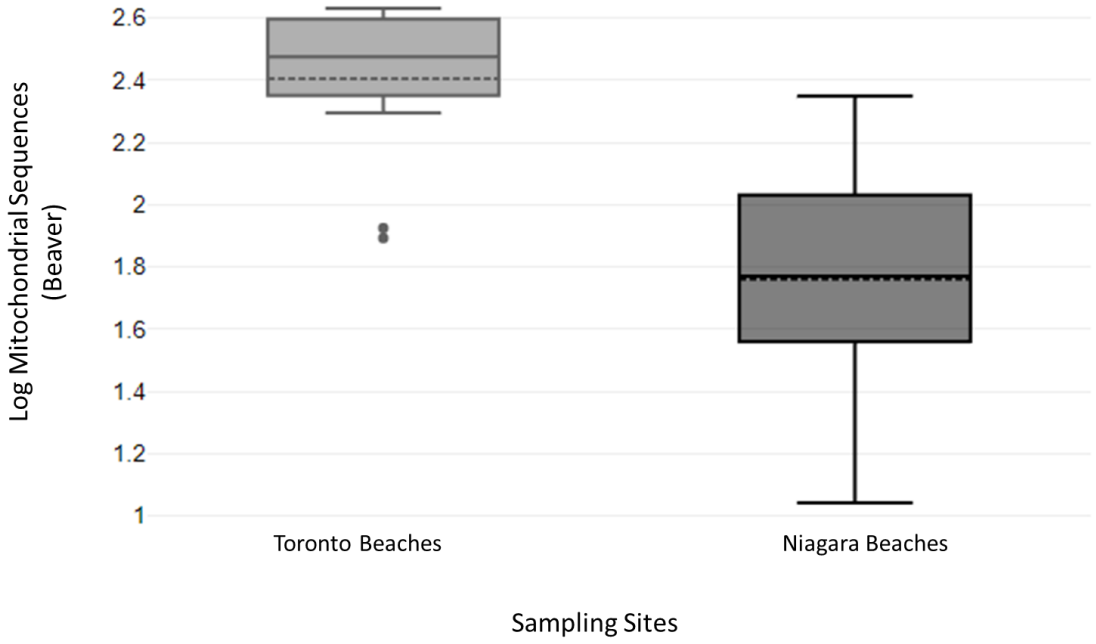
Supplementary Figure 5. Principle Component Analysis (Beta Diversity) of Bird species for Toronto and Niagara beaches, rivers and sand samples.



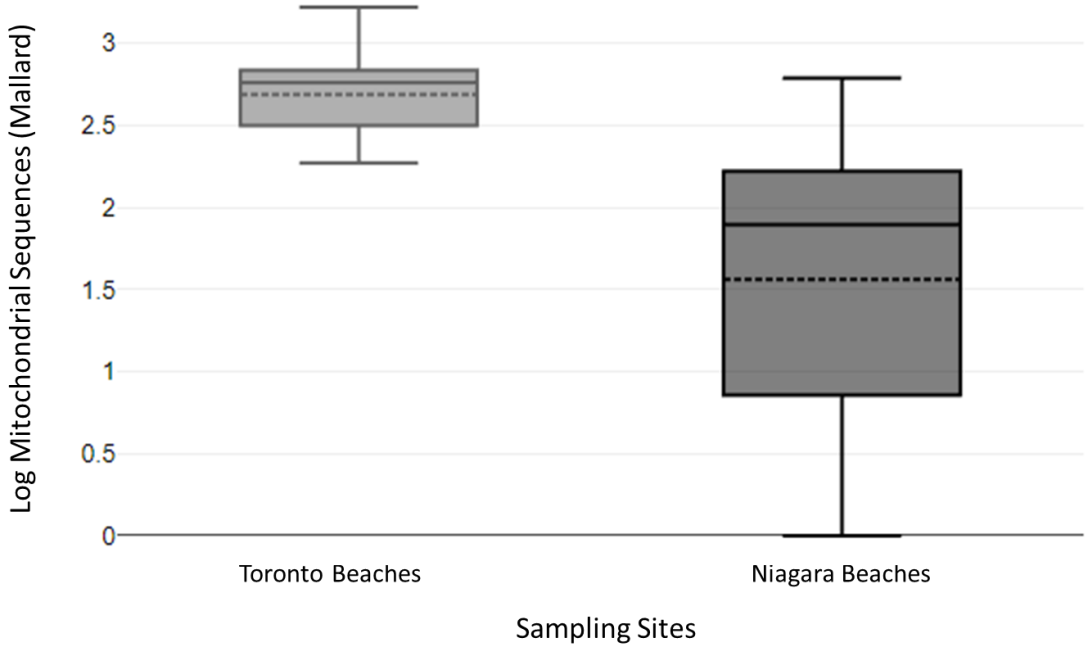
Supplementary Figure 6. Detection percentage (Number of positive samples) of common fecal contamination sources in the beach, river, and sand samples (n = 48).



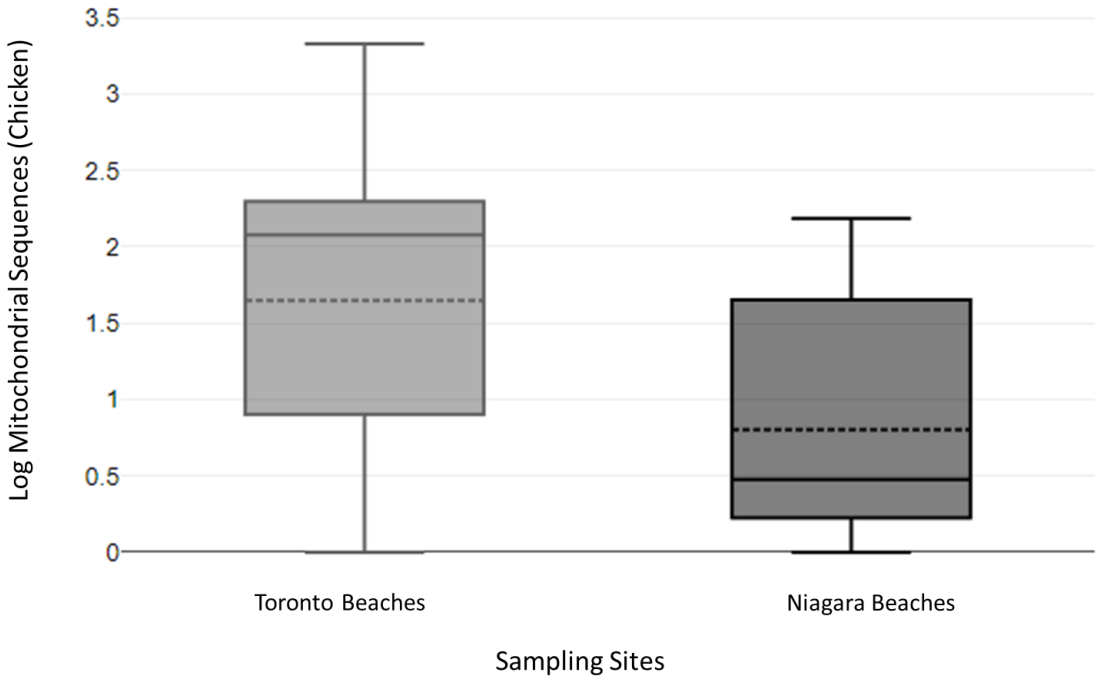
Supplementary Figure 7. Relative abundance of potential mammalian and bird fecal contamination sources for beaches, rivers and sand samples.



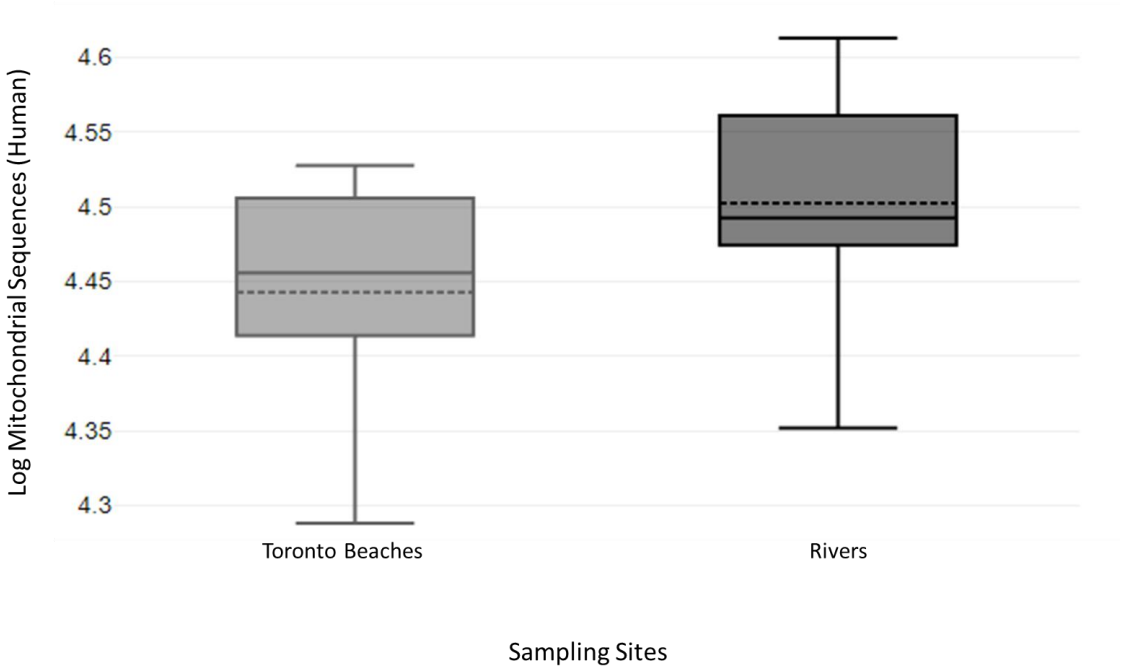
Supplementary Figure 8. Comparative abundance of Beaver mitochondrial sequences between Toronto and Niagara beaches.



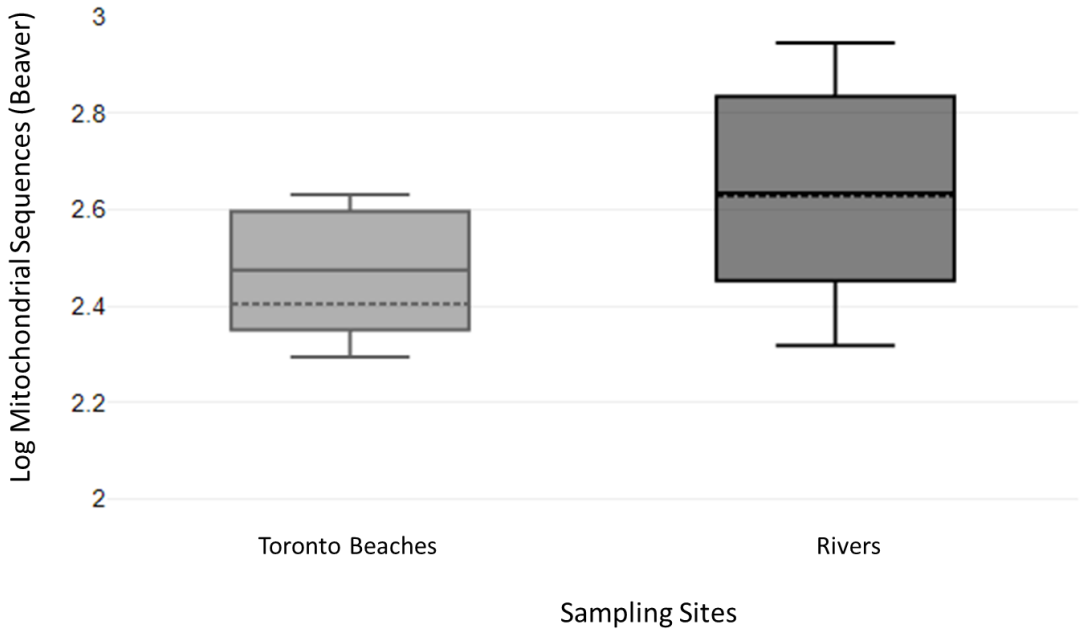
Supplementary Figure 9. Comparative abundance of Mallard mitochondrial sequences between Toronto and Niagara beaches.



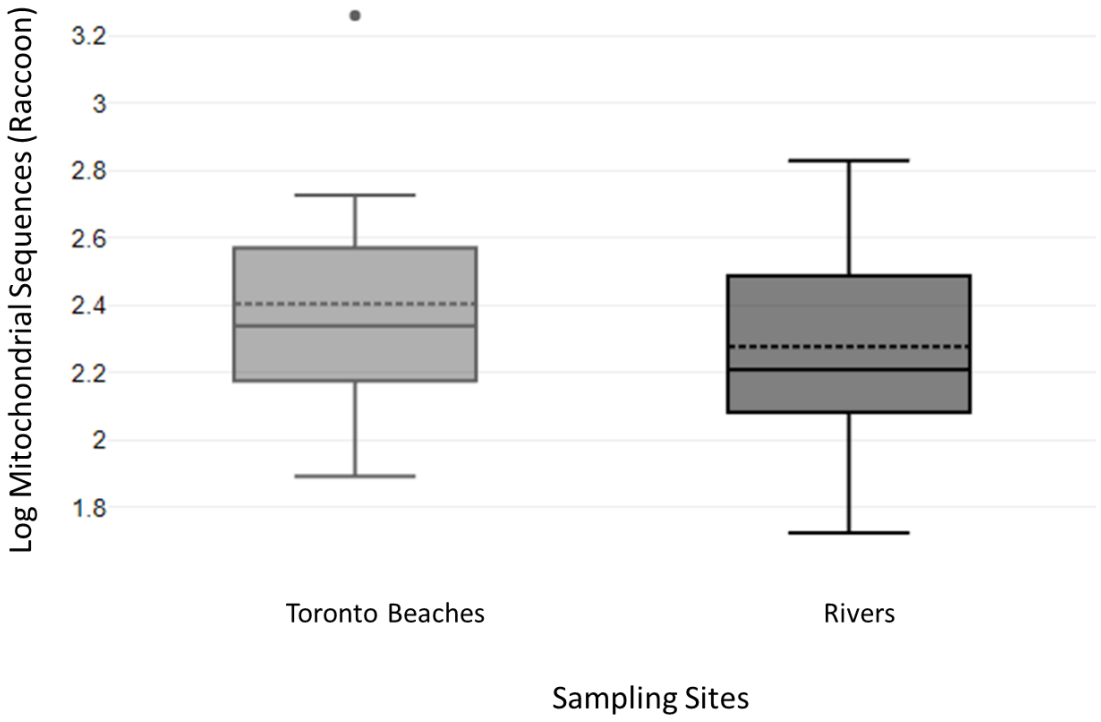
Supplementary Figure 10. Comparative abundance of Chicken mitochondrial sequences between Toronto and Niagara beaches.



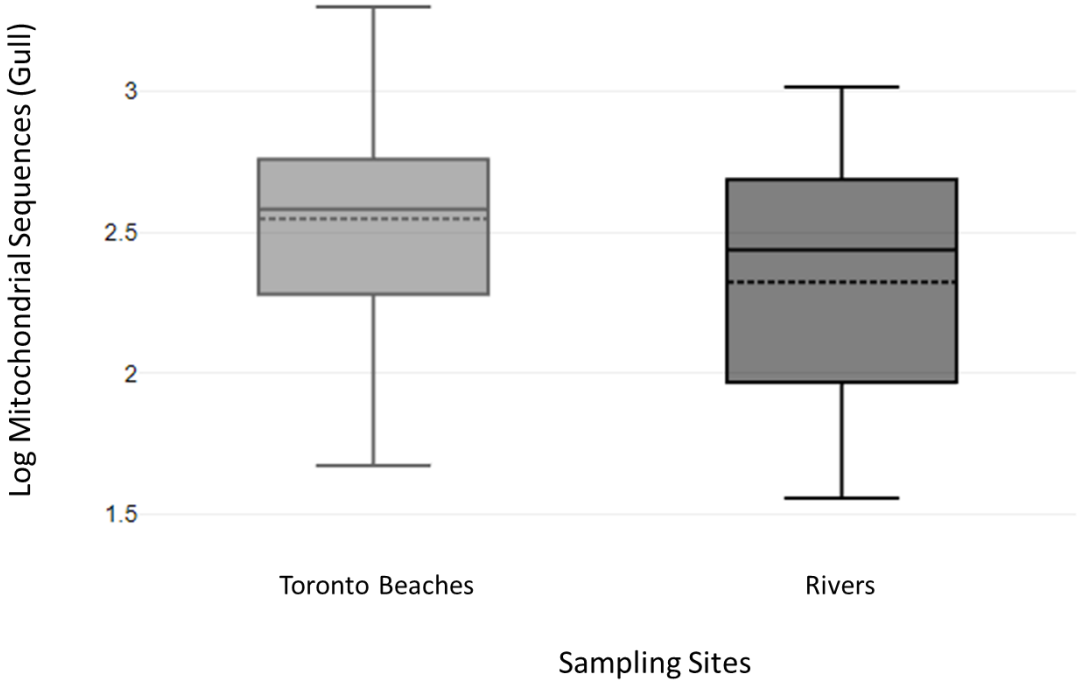
Supplementary Figure 11. Comparative abundance of Human mitochondrial sequences between Toronto beaches and river samples.



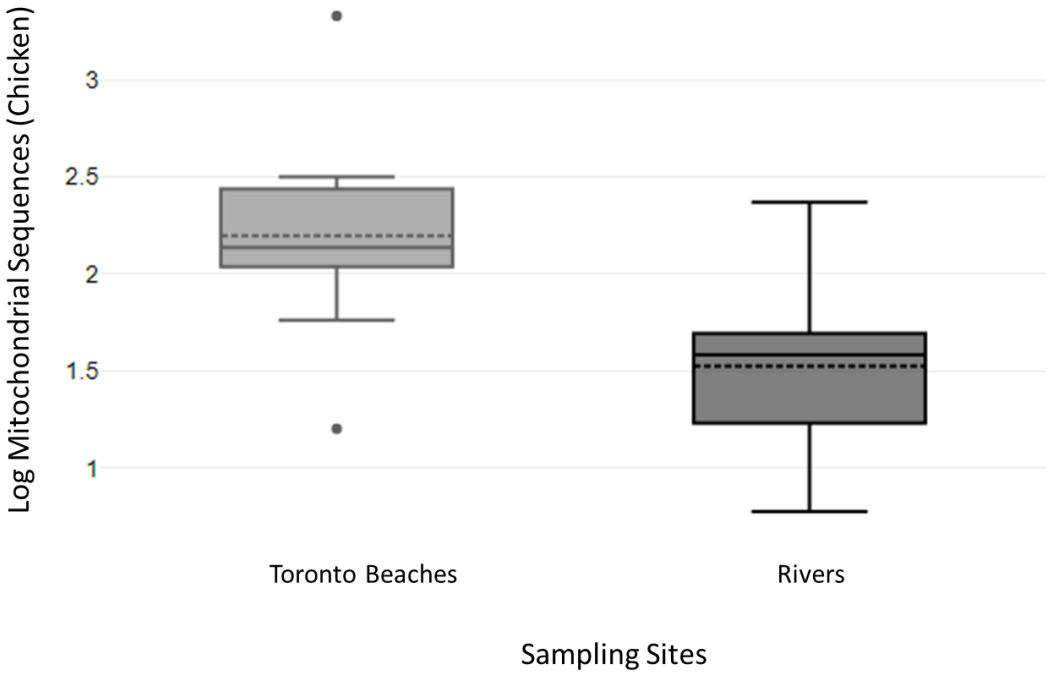
Supplementary Figure 12. Comparative abundance of Beaver mitochondrial sequences between Toronto beaches and river samples.



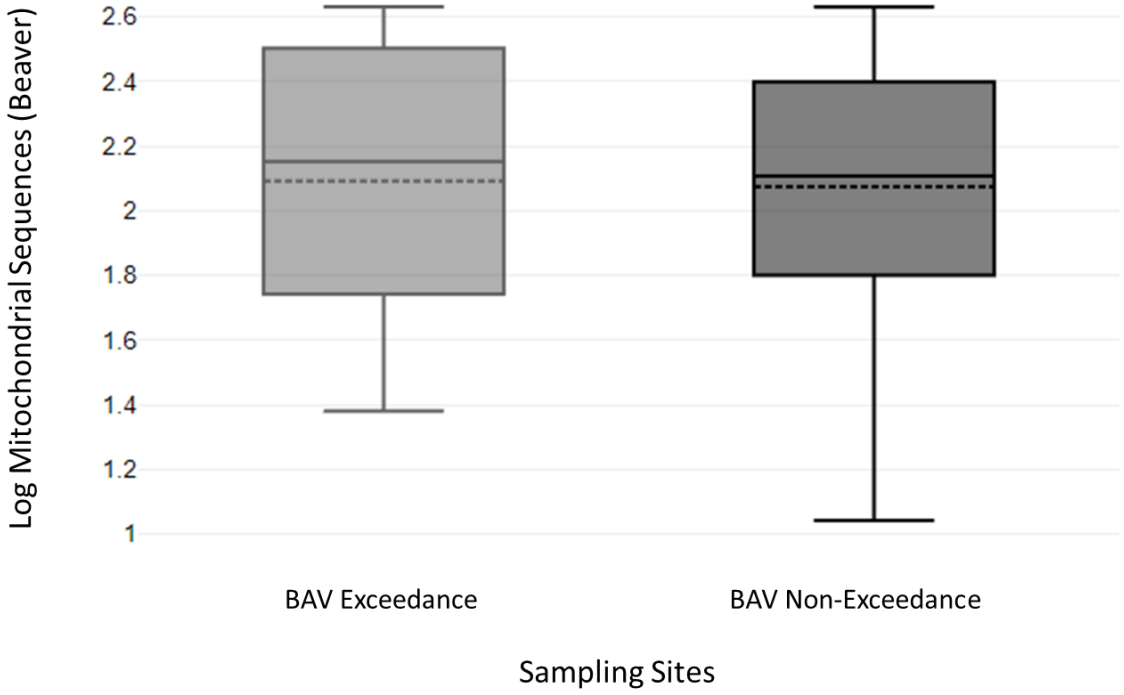
Supplementary Figure 13. Comparative abundance of Raccoon mitochondrial sequences between Toronto beaches and river samples.



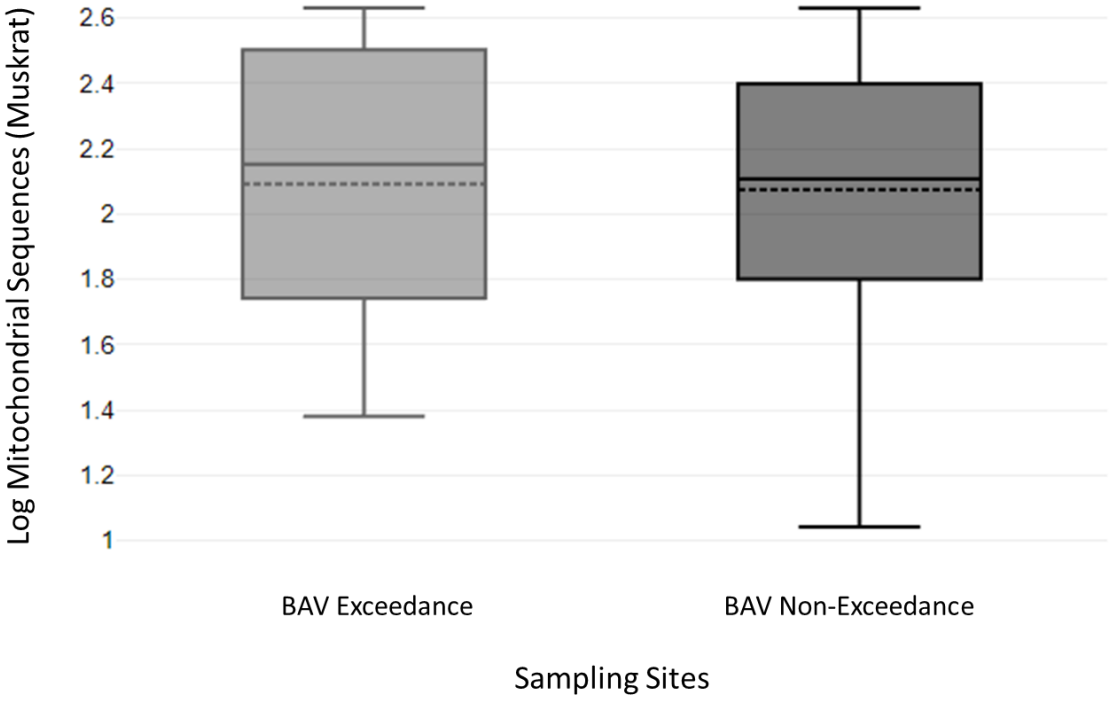
Supplementary Figure 14. Comparative abundance of Gull mitochondrial sequences between Toronto beaches and river samples.



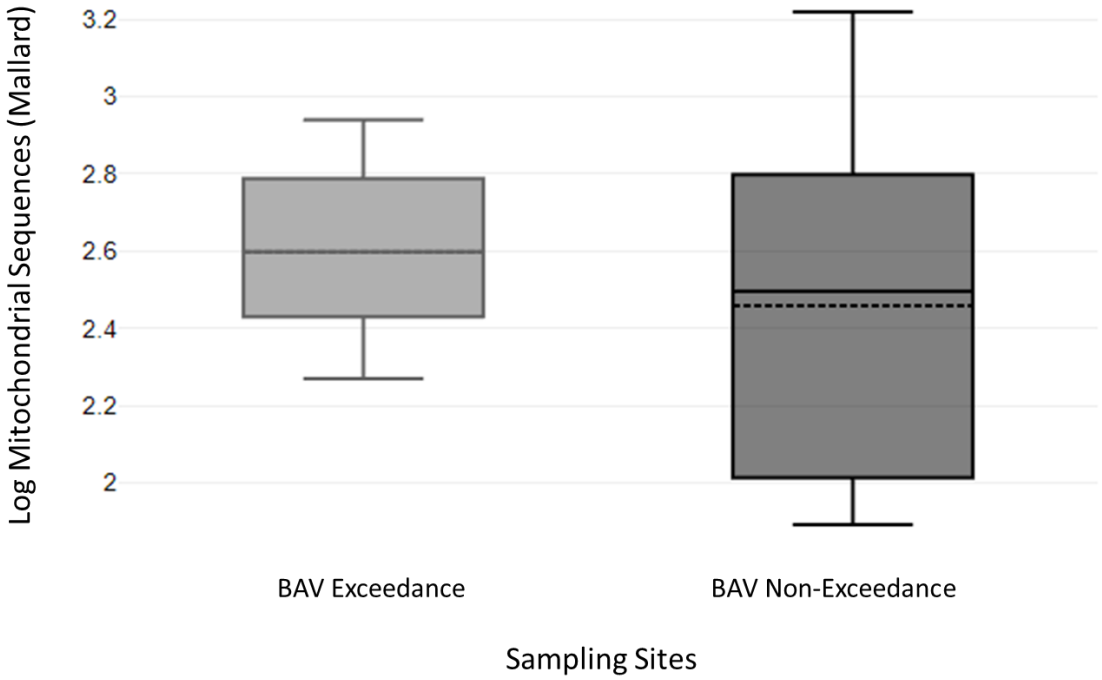
Supplementary Figure 15. Comparative abundance of Chicken mitochondrial sequences between Toronto beaches and river samples.



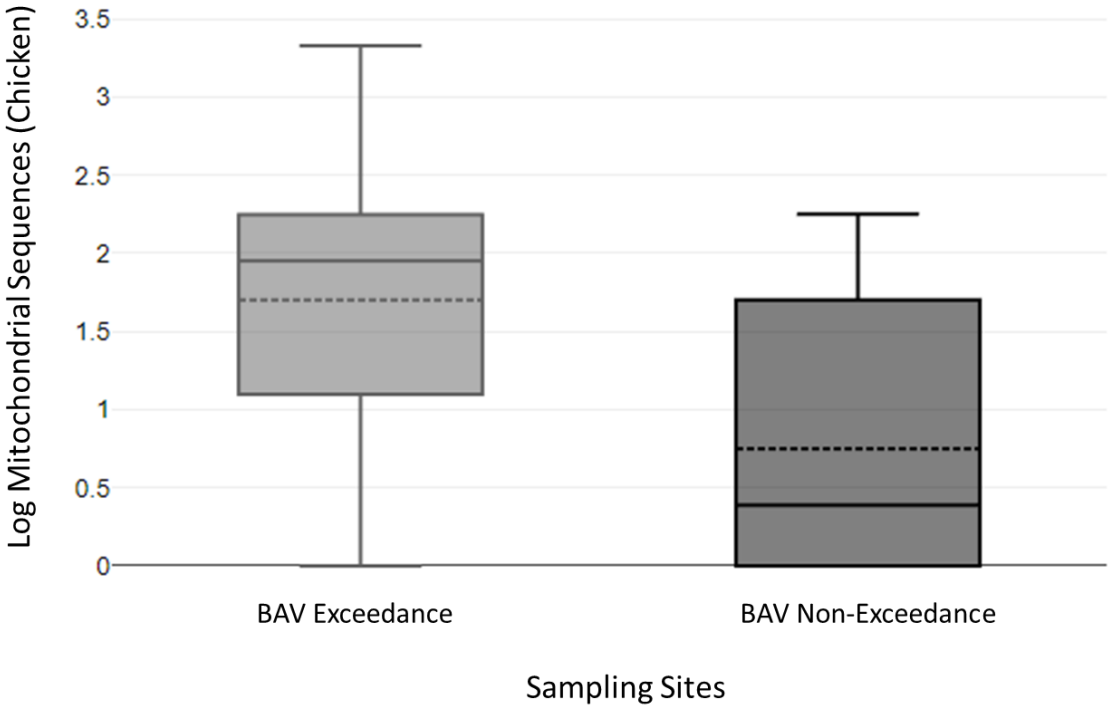
Supplementary Figure 16. Comparative abundance of Beaver mitochondrial sequences between Beach Action Value Exceedance and Non-exceedance Beach water samples.



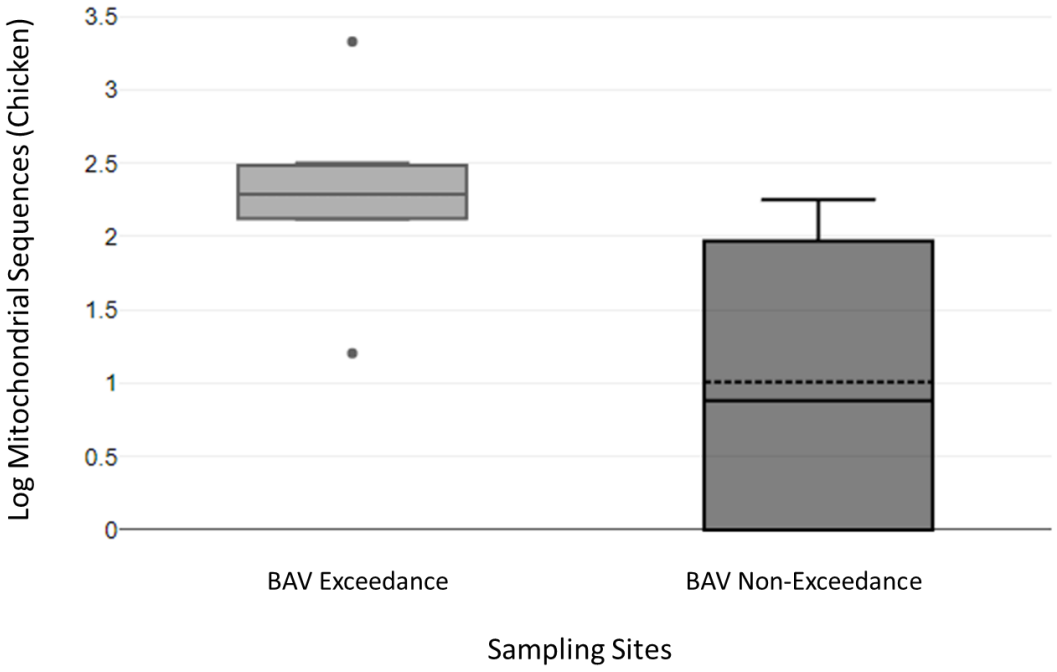
Supplementary Figure 17. Comparative abundance of Muskrat mitochondrial sequences between Beach Action Value Exceedance and Non-exceedance Beach water samples.



Supplementary Figure 18. Comparative abundance of Mallard mitochondrial sequences between Beach Action Value Exceedance and Non-exceedance Beach water samples.



Supplementary Figure 19. Comparative abundance of Chicken mitochondrial sequences between Beach Action Value Exceedance and Non-exceedance Beach water samples.



Supplementary Figure 20. Comparative abundance of Chicken mitochondrial sequences between Beach Action Value Exceedance and Non-exceedance shoreline water samples from Toronto beaches.

Supplementary Table 5. Comparison of the frequency of occurrence of food animal eDNA sequences (chicken, cow, pig) with human HF183 and mt dPCR markers and human eDNA sequences.

		HF183 % +ve	**HuMt % +ve**	**Human eDNA % +ve**
Toronto beaches n=12	Human eDNA +ve n=12	42	92	100
	Chicken eDNA +ve n=9	56	89	100
	Cow eDNA +ve n= 5	40	100	100
Niagara beaches n=12	Human eDNA +ve n=12	67	100	100
	Chicken eDNA +ve n=9	56	100	100
	Cow eDNA +ve n= 9	78	100	100
Niagara sand n=12	Human eDNA +ve n=12	8	100	100
	Chicken eDNA +ve n=9	11	100	100
	Cow eDNA +ve n= 6	17	100	100
	Pig eDNA +ve n=2	50	100	100
Toronto rivers n=12	Human eDNA +ve n=12	83	67*	100
	Chicken eDNA +ve n=10	90	71*	100
	Cow eDNA +ve n=8	100	50*	100
	Pig eDNA +ve n=1	100	100	100
Total samples	Human eDNA +ve n=48	50	91*	100
	Chicken eDNA +ve n=37	59	91*	100
	Cow eDNA +ve n= 28	64	88*	100

	Pig eDNA +ve n=3	67	100	100

* n=9 rather than 12 due to insufficient amount of DNA for PCR.