Electronic Theses and Dissertations

Theses, Dissertations, and Major Papers

11-5-2020

Freshwater Beach Microbial Ecology, Community Dynamics and Adaptive Responses to Environmental Changes Using Metabarcoding and Transcriptomics

Abdolrazagh Hashemi Shahraki
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Freshwater Beach Microbial Ecology, Community Dynamics and Adaptive Responses to Environmental Changes Using Metabarcoding and Transcriptomics

By

Abdolrazagh Hashemi Shahraki

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Great Lakes Institute for Environmental Research

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

at the University of Windsor

Windsor, Ontario, Canada

2020

Freshwater Beach Microbial Ecology, Community Dynamics and Adaptive Responses to Environmental Changes Using Metabarcoding and Transcriptomics

by

Abdolrazagh Hashemi Shahraki

APPROVED BY:

_____
M. Coleman, External Examiner
University of Chicago

_____
T. Mao
MW Technologies Inc.

_____
T. Noel
Department of Integrative Biology

_____
K. Drouillard
Great Lakes Institute for Environmental Research

_____
D. Haffner
Great Lakes Institute for Environmental Research

_____
S. Chaganti, Co-Advisor
University of Michigan

_____
D. Heath, Co-Advisor
Great Lakes Institute for Environmental Research

October 6, 2020

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I.   Co-Authorship

I hereby declare that this dissertation incorporates materials that are the result of joint research, as follows:

Chapters 2-6 were completed under the supervision of my advisors, Dr. Daniel Heath and Dr. Subba Rao Chaganti. In all cases, the key ideas, primary contributions, data analysis and writing were carried out by me. The contribution of the co-authors was primarily through advice on the broad research ideas, review of results, participation in scientific discussion, literature review and subsequently in editing the presentation and thesis material.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis. I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

II.   Previous Publication

This dissertation includes two original papers that have been previously published in peer-reviewed journals, as follows:

Thesis Chapters	Authors	Title	Publication Status
2	Abdolrazagh Hashemi Shahraki, Subba Rao Chaganti, Daniel Heath	Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario	Published, Microbial Ecology
6	Abdolrazagh Hashemi Shahraki, Subba Rao Chaganti, Daniel Heath	Recreational water monitoring: Nanofluidic qRT-PCR chip for assessing beach water safety	Published, Environmental DNA

I certify that I have obtained written permission from the copyright owner(s) to include the above papers in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner (s) to include such material (s) in my thesis.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

The Laurentian Great Lakes (LGLs) represent the single most valuable natural resource on the North American continent and are a critical source of drinking water, important aquatic species habitat, water for the industrial sector and tourism/recreational activities as well as many other ecological services. LGLs ecosystems are changing rapidly due to climate change effects and are thus highly susceptible and responsive to any added anthropogenic stressors. The aquatic bacterial community affects critical ecosystem functions, such as nutrient cycling, water quality, recreational activities, etc., in ecosystems such as the LGLs, but perturbations can alter both the composition and functionality of the bacterial community. These changes can result in negative effects on whole ecosystem health with an associated loss in economic and social values.

Understanding the temporal and spatial variation in the composition, diversity and ultimately activity of the bacterial community is paramount in understanding overall ecosystem services providing by the bacterial community. Characterizing spatial and temporal variation (and the factors that contribute to it) can provide deeper insight into the processes and mechanisms operating in LGLs ecosystems, and ultimately improve our basic knowledge and ability to predict bacterial community composition, dynamics and function.

Enumeration of *Escherichia coli* as a bioindicator of human fecal contamination is widely used to quantify recreational water quality and safety. The inclusion of microbial source tracking (MST) as part of water quality monitoring along with E. coli and waterborne pathogens in a novel monitoring tool could help to determine the specific fecal source (e.g. human, dog, cattle, wildlife, etc.) and has great potential for accurate estimation of water-related health risks.

On the other hand, we still have an incomplete understanding of freshwater microbial ecology and community dynamics and their response to disturbance, particularly to human-related environmental stressors. We must be able to predict and track the sources of harmful bacterial outbreaks; this will require a clear understanding of the impact of environmental and anthropogenic stressors on microbial community diversity and function.

The research comprising this dissertation was designed to characterize broad to fine-scale temporal and spatial variation in freshwater bacterial community composition and gene transcription. Temporal variation in freshwater bacterial community composition (bi-hourly, monthly and seasonal variation) and gene transcription profile (seasonally) was significant while spatial variation was significant but limited in magnitude. A novel monitoring approach (nanofluidic TaqMan qRT-PCR) was designed and optimized for rapid and reliable monitoring of freshwater quality for waterborne pathogens, MST markers and *E. coli* as a bioindicator of fecal contamination. Finally, an experimental bacterial microcosm study was used to study the response of adapted (pre-exposure to different levels of nutrient stress) bacteria communities to very high nutrient stress. This experiment revealed that pre-exposure to a higher level of nutrient stress provides greater protection against community change than low levels when the bacterial community is challenged with a very high level of a stressor.

These cumulative insights into the temporal and spatial variation of the freshwater bacteria community composition and transcriptome, the development of a novel nanofluidic TaqMan qRT-PCR tool for detecting and quantifying harmful bacteria and our microcosm study outcomes provide baseline knowledge and tools which will be valuable for improving best management practices, monitoring and accurate prediction of changes in freshwater ecosystem function.

DEDICATION

I dedicate this dissertation to my wonderful wife, my kids, mom, dad, brothers and sister and all my family and friends for their continued love and support.

October 2020

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

the community of two sampling locations at the whole community level – minor differences exist and that is why both locations are presented in the heat map. The color code indicates the relative abundance of each OTU. The lines in the heatmap represent the relative abundance of each OTU across the BC of P2-Co, C-LND, C-HND, R-LND and R-HND of LP and SL. Taxa abbreviation: Acido; *Acidiobacteria*, Actin; *Actinobacteria*, Bacter; *Bacteroidetes*, Chlam; *Chlamydiae*, Chloro; *Chloroflexi*, Cyano; *Cyanobacteria*, Firm; *Firmicutes*, Plan; *Planctomycetes*, Proteo; *Proteobacteria* and Verru; *Verrucomicrobia*………………………………………………………………….

LIST OF APPENDICES

LIST OF ABBREVIATIONS/SYMBOLS

BC	Bacterial communities
BCC	Bacterial community composition
bp	Base-pair
CB	Cedar Beach
CND	Challenge nutrient dose
C-HND	Challenged of high nutrient dose
C-LND	Challenged of low nutrient dose
CH	Colchester Harbour Beach
cDNA	Complementary DNA
CPM	Count per million
cHABs	cyanobacterial harmful algal blooms
CT	cycle threshold
DOM	Dissolved organic matters
distLM	Distance-based linear model
dsDNA	double-stranded deoxyribonucleic acid
EPA	Environmental Protection Agency
ESV	Exact sequence variants
FDR	False discovery rate
FIB	Fecal indicator bacteria
PCo1	First principal coordinates
FC	Fold change
GTA	Gene transfer agents
GLMM	Generalised linear mixed model
GLM	Generalized liner model
HND	High nutrient dose
HB	Holiday Beach
HGT	Horizontal gene transfer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LP	Lakeview Park Beach
LGL	Laurentian Great Lakes
LDA	Linear discriminant analysis
LND	Low nutrient dose
ML	Maximum likelihood
MD	Mean difference
Tm	Melting temperature
MST	Microbial source tracking
MDS	Multi-dimensional scaling
NGS	Next-generation sequencing
N	Nitrogen
NTC	No-template control
OUT	Operational taxonomic unit

PERMANOVA	Permutational multivariate analysis of variance
P1-Co	Phase 1 control community
P2-Co	Phase 2 control community
P	Phosphate
PCoA	Principal coordinates analysis
PP	Point Pelee Beach
PRC	Principle response curve
Cq	Quantification cycle
QIIME	Quantitative Insights into Microbial Ecology
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative real time PCR
RC	Read count
R-LND	Relaxed of low nutrient dose
R-HND	Relaxed of high nutrient dose
RDP	Ribosomal database project
RIN	RNA integrity numbers
RNAseq	RNA sequencing
SP	Sand Point Beach
SL	Seaclift Beach
PCo2	Second principal coordinates
SSU rRNA	Small subunit ribosomal ribonucleic acid
SPRI	Solid Phase Reversible Immobilization
BLAST	The Basic Local Alignment Search Tool
PCo3	Third principal coordinates
TMM	Trimmed mean of M-values
UPGMA	Unweighted pair group method with arithmetic mean
VBNC	Viable but non-culturable state
WHO	World Health Organization

GENERAL INTRODUCTION

1.1 Introduction

The broad interest in microbial community ecology is a consequence of the postulate that interactions among microorganisms (microbe–microbe) are of essential importance for whole ecosystem dynamics and the evolutionary ecology of all organisms in the ecosystem. Microbes rarely live in isolation, but instead, coexist in complex ecologies with diverse symbiotic relationships (Saffo, 1992). The nature of the relationships between living organisms span a wide range, and include; win-win (mutualism), win-zero (commensalism), win-lose (parasitism, predation), zero-lose (amensalism), and lose-lose (competition) situations (Lidicker, 1979). The final outcome of these widespread interactions in microbial communities leads to either aggregations, avoidance or exclusions among microbial taxa (Berry and Widder, 2014; Konopka, 2009).

Microbiome diversity is typically described, at a basic level, in terms of within (i.e., alpha) and among communities (i.e., beta), and that compromises the bacterial community composition (BCC). All biologists who sample natural communities are plagued with the problem of how well a sample reflects a community's "true" diversity. New genetic techniques, mostly based on massively parallel (next-generation) sequencing (NGS), have revealed extensive microbial diversity that was previously undetected with culture-dependent methods and morphological identification, but exhaustive inventories of microbial communities still remain impractical (Hughes et al., 2001). Reliable estimates of microbial diversity would offer a means to address once intractable questions, such as 1) what processes control microbial diversity? 2) what temporal and spatial scales do microbial communities' function at? and 3) How do microbial communities affect ecosystem functioning? How are human beings affecting microbial communities? Several microbial studies have used alpha and beta diversity indices to characterize the diversity of the BCCs. Most microbial ecologists have borrowed diversity indices from plant and animal ecologists such as Shannon's and Simpson's indices of diversity (Ludwig et al., 1988), however, those indexes need a clear definition of species and an unambiguous identification of each individual, both

of which are difficult in bacteriology (Curtis et al., 2002). Still, others have proposed new diversity statistics specific to microbial samples (Watve and Gangal, 1996). Despite the interest, however, the value of those innovative microbial-specific tools has not yet been evaluated for BCs, while new potential approaches continue to be proposed but remain to be explored. However, specific but alpha (Chao 1 and Shannon) and beta (Bray-Curtis (dis)similarity) diversity indices are the most popular tools for measuring diversity within and among the microbial community samples (Leinster and Cobbold, 2012).

A growing body of studies suggests that BCC exhibits a wide range of temporal and spatial variation (Lear et al., 2014; Lymer et al., 2008). Abiotic factors (e.g., light (Hölker et al., 2015), temperature (Villaescusa et al., 2016) and nutrients (Lv et al., 2017)) as well as biotic factors (e.g., abundances of grazers (Wojciechowski et al., 2017) and viral lysis (Rodriguez-Brito et al., 2010)) change the nature of microbial interactions and ultimately the BCC on diel to seasonal cycles in freshwater ecosystems. How the bacterial communities (BCs) are shaped by these episodic or fluctuating biotic and abiotic factors is not fully understood, yet those BC changes ultimately impact the productivity and ecological services of freshwater BCCs. The BCC is closely related to their microscale physical and chemical environments (O'Donnell et al., 2007), and those factors are affected by human activities such as agriculture, urban development, industry, tourism (including aquatic recreational activities), among others (Nogales et al., 2011). In response to anthropogenic stressors, the microbial community may exhibit resistance (does not change), resilience (changes but recovers due to metabolic flexibility, tolerance, rapid growth rate and adaptive evolution), functional redundancy (microbial composition changes, but new taxa perform similar functions to those of the original community) (Allison and Martiny, 2008) and finally, community breakdown (microbial community and function change) (Muturi et al., 2017).

The BCC provides critical ecosystem services, such as nutrient cycling, water quality, recreational activities, etc., but sometimes perturbations can alter both the composition and functionality of the BCs and thus result in negative effects on whole ecosystem health with an associated loss in economic and social values (Allison and

Martiny, 2008). For example, significant loss of diversity and key specialized functions of microbial communities in response to long-term heavy metal stress has been documented (Singh et al., 2014). The Laurentian Great Lakes (LGLs) represent the single most valuable natural resource on the North American continent (McKenna Jr, 2019) and are a critical source of drinking water (Dodds et al., 2013). For example, Lake Erie serves as a source of drinking water (for ~11 million people), important aquatic species habitat, water for the industrial sector and tourism/recreational activities (Reutter, 2019). Degradation of freshwater quality also significantly reduces human recreational opportunities and results in economic loss (e.g., decreased tourism, reduced fishing activity, etc.) (Vörösmarty et al., 2010). In the US, for example, epidemiologic data indicate the risk for developing acute gastrointestinal illness symptoms is as high as 15 cases per 1000 swimmers which translates into $2.2-$3.7 billion economic burdens annually (DeFlorio-Barker et al., 2018). Recreational water quality and safety currently is monitored by enumeration of *Escherichia coli* as a bioindicator of human fecal contamination (Edberg et al., 2000), however, a growing body of literatures reports that *E. coli* is a poor health risk indicator for recreational waters (Jang et al., 2017; Odonkor and Ampofo, 2013). The inclusion of microbial source tracking (MST) as part of water quality monitoring could help to determine the specific fecal source (e.g. human, dog, cattle, wildlife, etc.) and has great potential for accurate estimation of water related health risks (Harwood et al., 2014).

We still have an incomplete understanding of freshwater microbial ecology and community dynamics and their response to disturbance, particularly to human-related environmental stressors. A better understanding of freshwater microbial ecology will allow us to identify potential biomarkers for monitoring and predicting the health status of aquatic ecosystems using rapid and novel molecular ecological technology. To better characterize freshwater microbial ecology, microbial co-occurrence (and exclusion), and microbial community dynamics must be assessed, not just tracking of harmful bacteria. We must be able to predict and track the sources of harmful bacterial outbreaks; this will require a clear understanding of the impact of environmental and anthropogenic stressors on microbial community diversity and function. To address

this critical need, five research projects (chapters 2-6) that focus on specific knowledge gaps make up my dissertation.

1.2 Research chapters

Chapter 2; Diel dynamics of freshwater bacterial communities at beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Short-term trends in freshwater BCC in large lake ecosystems are expected to be relatively stable due to the size and complexity of the habitat; however, this has not been systematically tested. We addressed this gap by short-term temporal (2-h intervals in a diel cycle) characterization of the BCC in Lakes Erie and St. Clair using metabarcoding and high-throughput sequencing of the 16S rRNA gene. Also, to provide a more accurate picture of freshwater safety and human health risk at fine-scale, we evaluated bi-hourly variation of *E. coli* levels using qPCR.

Chapter 3; Temporal and spatial variation in microbial community dynamics in large freshwater lake ecosystems: Laurentian Great Lakes Erie and St. Clair. Long-term trends in freshwater BCC and its dynamics are not yet well characterized, particularly in large lake ecosystems. We addressed this gap by temporally (15 months) and spatially (6 sampling locations) characterizing the BCC variation in Lakes Erie and St. Clair; two connected ecosystems in the LGLs using metabarcoding and high-throughput sequencing of the 16S rRNA gene.

Chapter 4; Functional and virulence genes variation of bacterial communities in large freshwater lake ecosystems. Little is known regarding the seasonal and spatial variation in the ecological activity of the freshwater BC, yet metatranscriptomics provides a powerful new approach to characterize gene transcription variation at the community level. We addressed this gap using metatrascriptomic analysis of the freshwater BC transcription profiles at three public beaches in Lake Erie (2 beaches) and Lake St. Clair (1 beach), over winter, summer and fall (2019).

Chapter 5; Nutrient stress drives adaptive changes in freshwater bacteria communities: Response to secondary stress mediated by horizontal gene flow and reduced community change. Little is known about the potential effects of aquatic stressors, such as nutrient overloading, on the adaptive response of the BC, including such processes such as biodiversity change, community composition shifts and

horizontal gene transfer (HGT). Here we designed a two-phase study with an adaptive phase (pre-stress freshwater BC to low and high levels of nutrients at moderate levels) and a challenge phase (challenge the adapted freshwater BC to very high levels of nutrients) and studied the BC shift and diversity loss in both phases (and in the control community) by metabarcoding and high-throughput sequencing of the 16S rRNA gene. We also measured the prevalence of HGT genes using qPCR and markers for class I integrons (Int1 gene) and Gene Transfer Agents (GTA) in the adapted and challenged communities to explore the role of HGT in adaptation against a different level of stressors.

Chapter 6; Recreational water monitoring: nanofluidic qRT-PCR chip for assessing beach water safety. For effective real-time monitoring of recreational water and subsequent risk management, a robust, high-throughput quantitative tool that targets fecal indicator bacteria (FIBs), MST markers and waterborne bacterial pathogens is needed. We developed a panel of novel quantitative real-time PCR (qRT-PCR) assays and printed and tested them on the OpenArray platform to detect and quantify 2 FIBs, 7 MST markers and 15 bacterial pathogens in recreational freshwater.

The outcome of my thesis will increase our ability to quantify and predict changes in freshwater BCs over time and in response to ecological perturbations. My research addresses temporal and spatial BCC variation (chapters 2 and 3) as well as their functional variation (chapter 4). I also address the possibility of microbial "community evolution", or the adaptive response of freshwater BCC to human-derived stressors (chapter 5). To improve our ability to rapidly quantify risks associated with harmful freshwater bacteria, and to track their source (and hence improve predictability), I designed, optimized and validated a nanofluidic TaqMan OpenArray chip (chapter 6). Throughout my thesis I employed a variety of advanced molecular genetic techniques, my thesis chapters describe their use; however, below I provide an overview of these techniques for readers who may not be familiar with them.

1.3　References

Allison, S.D., Martiny, J.B., 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105(Supplement 1) 11512-11519.

Berry, D., Widder, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology 5 219.

Curtis, T.P., Sloan, W.T., Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences 99(16) 10494-10499.

DeFlorio-Barker, S., Wing, C., Jones, R.M., Dorevitch, S., 2018. Estimate of incidence and cost of recreational waterborne illness on United States surface waters. Environmental Health 17(1) 3.

Dodds, W.K., Perkin, J.S., Gerken, J.E., 2013. Human impact on freshwater ecosystem services: a global perspective. Environmental Science & Technology 47(16) 9061-9068.

Edberg, S., Rice, E., Karlin, R., Allen, M., 2000. Escherichia coli: the best biological drinking water indicator for public health protection. Journal of Applied Microbiology 88(S1) 106S-116S.

Harwood, V.J., Staley, C., Badgley, B.D., Borges, K., 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) 1-40.

Hölker, F., Wurzbacher, C., Weißenborn, C., Monaghan, M.T., Holzhauer, S.I., Premke, K., 2015. Microbial diversity and community respiration in freshwater sediments influenced by artificial light at night. Philosophical Transactions of the Royal Society B: Biological Sciences 370(1667) 20140130.

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., Bohannan, B.J., 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Applied and environmental microbiology 67(10) 4399-4406.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. Journal of Applied Microbiology 123(3) 570-581.

Konopka, A., 2009. What is microbial community ecology? The ISME Journal 3(11) 1223-1230.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.L., 2014. Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. The ISME Journal 8(8) 1715-1726.

Leinster, T., Cobbold, C.A., 2012. Measuring diversity: the importance of species similarity. Ecology 93(3) 477-489.

Lidicker, W.Z., 1979. A clarification of interactions in ecological systems. BioScience 29(8) 475-477.

Ludwig, J.A., Quartet, L., Reynolds, J.F., Reynolds, J., 1988. Statistical ecology: a primer in methods and computing. John Wiley & Sons.

Lv, F., Xue, S., Wang, G., Zhang, C., 2017. Nitrogen addition shifts the microbial community in the rhizosphere of Pinus tabuliformis in Northwestern China. PloS one 12(2).

Lymer, D., Logue, J.B., Brussaard, C.P., BAUDOUX, A.C., Vrede, K., LINDSTRÖM, E.S., 2008. Temporal variation in freshwater viral and bacterial community composition. Freshwater Biology 53(6) 1163-1175.

McKenna Jr, J.E., 2019. The Laurentian Great Lakes: A case study in ecological disturbance and climate change. Fisheries Management and Ecology 26(6) 486-499.

Muturi, E.J., Donthu, R.K., Fields, C.J., Moise, I.K., Kim, C.-H., 2017. Effect of pesticides on microbial communities in container aquatic habitats. Scientific Reports 7.

Nogales, B., Lanfranconi, M.P., Piña-Villalonga, J.M., Bosch, R., 2011. Anthropogenic perturbations in marine microbial communities. FEMS Microbiology Reviews 35(2) 275-298.

O'Donnell, A.G., Young, I.M., Rushton, S.P., Shirley, M.D., Crawford, J.W., 2007. Visualization, modelling and prediction in soil microbiology. Nature Reviews Microbiology 5(9) 689-699.

Odonkor, S.T., Ampofo, J.K., 2013. Escherichia coli as an indicator of bacteriological quality of water: an overview. Microbiology Research 4(1) e2-e2.

Reutter, J.M., 2019. Lake Erie: Past, Present, and Future. Encyclopedia of Water: Science, Technology, and Society 1-15.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., 2010. Viral and microbial community dynamics in four aquatic environments. The ISME Journal 4(6) 739-751.

Saffo, M.B., 1992. Coming to terms with a field: words and concepts in symbiosis. Symbiosis 14(1-3) 17-31.

Singh, B.K., Quince, C., Macdonald, C.A., Khachane, A., Thomas, N., Al-Soud, W.A., Sørensen, S.J., He, Z., White, D., Sinclair, A., 2014. Loss of microbial diversity in soils is coincident with reductions in some specialized functions. Environmental Microbiology 16(8) 2408-2420.

Villaescusa, J.A., Jørgensen, S.E., Rochera, C., Velázquez, D., Quesada, A., Camacho, A., 2016. Carbon dynamics modelization and biological community sensitivity to temperature in an oligotrophic freshwater Antarctic lake. Ecological Modelling 319 21-30.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., 2010. Global threats to human water security and river biodiversity. Nature 467(7315) 555-561.

Watve, M.G., Gangal, R.M., 1996. Problems in measuring bacterial diversity and a possible solution. Applied and Environmental Microbiology 62(11) 4299-4301.

Wojciechowski, J., Heino, J., Bini, L.M., Padial, A.A., 2017. Temporal variation in phytoplankton beta diversity patterns and metacommunity structures across subtropical reservoirs. Freshwater Biology 62(4) 751-766.

DIEL DYNAMICS OF FRESHWATER BACTERIAL COMMUNITIES AT BEACHES IN LAKE ERIE AND LAKE ST. CLAIR, WINDSOR, ONTARIO [1]

2.1 Introduction

Freshwater ecosystems provide critical services, including drinking water, nutrient recycling, sport and commercial fisheries and recreation; however, they face many threats such as pesticide and fertilizer pollution, climate change, water extraction, habitat destruction, etc. (Dodds et al., 2013). The aquatic microbial community has fundamental direct and indirect roles in freshwater ecosystems, and they have served as biomarkers to quantify human-introduced stress on aquatic ecosystems for decades (Glasl et al., 2017). One of the major goals of microbial ecology is to improve our basic knowledge of bacterial community composition (BCC) and community dynamics; however, we have only a superficial knowledge of the factors that contribute to variation in BCC, both spatially and temporally – yet such information would allow microbial ecologists to better predict BCC dynamics, particularly in freshwater ecosystems.

A growing body of studies suggests that BCCs exhibit a wide range of temporal and spatial variation (Berry et al., 2017; Butler et al., 2019; Huntscha et al., 2018); however, the BCC of complex ecosystems such as freshwater are often assumed to be relatively stable over a very short periods (minutes to hours), but this assumption is seldom tested. Abiotic factors are known or suspected to affect BCCs, (e.g., light (Hölker et al., 2015), temperature (Villaescusa et al., 2016) and nutrients (Horton et al., 2019)), as well as biotic factors (e.g., abundances of grazers (Grubisic et al., 2017) and viral lysis (Lymer et al., 2008)) change on diel to seasonal cycles. How the BCCs are shaped by these episodic biotic and abiotic changes at the diel level is not known, yet they may ultimately influence the productivity and ecological services of freshwater BCCs.

The Laurentian Great Lakes (LGLs) are a critical source of drinking water and recreational activity (Dodds et al., 2013), hence monitoring water quality and safety is

---

mandatory and is currently conducted by enumeration of *Escherichia coli* as a fecal indicator bacteria (FIB) (Pachepsky and Shelton, 2011). Although daily variation in *E. coli* levels at freshwater public beaches (McPhedran et al., 2013) and cultivable indicator bacteria over 12-hour tidal cycle in an estuarine habitat (lotic ecosystems) (Mill et al., 2006) have been reported before, the potential for significant variation in *E. coli* abundance across fine temporal scales (e.g., diel cycle) is not well characterized. Such fine scale temporal variation is a major concern for agencies charged with monitoring of beach safety and human health risk.

As little is known about the freshwater BCC variation over fine temporal scales (diel cycle), the overall aim of this study was to measure freshwater BCC changes and *E. coli* level over hourly and day/night time scales (diurnal). Characterizing temporal variation of freshwater BCC at fine temporal scales would be critical to give a whole picture regarding community shift and consequently functional changes over a 24-hour period, while changes in *E. coli* levels over diel cycle are relevant for assessing human health risks. Here, we analyzed BCC diurnal dynamics at four public beaches on two lakes (Lake Erie and Lake St. Clair) during the summer (June, July and August) by metabarcoding and next-generation sequencing (NGS) of 16S rRNA gene. We also measured *E. coli* levels over diel cycles using qPCR.

2.2   Material and methods

2.2.1   Study sites and sample collection

Freshwater samples were collected every 2-hours over one diel cycle (24 h) from four public beaches located at Windsor-Essex County (Windsor, Ontario, Canada); Colchester Harbour (CH) and Holiday Conservation Beach (HB) in Lake Erie, and Lakeview Park (LP) and Sand Point (SP) beaches in Lake St. Clair in the summer of 2016 on three dates (June 10th, July 10th and August 10th). As LP and SP are located in the urban area, they have more visitors over summer than CH and HB which are located in the agriculture area. LP is proximal to an adjacent urban tributary (the Belle River connects with Lake St. Clair at LP beach), while SP and HB are near the inlet and outlet of the Detroit River respectively. CH, HB and SP beaches represented high-energy locations, while LP beach was influenced by restricted water flow due to

adjacent artificial piers and represented low-energy sites with low wave energy. (VanMensel et al., 2019) At each beach, we collected a total of 6 samples from two different spots; 3 samples from knee-deep (shoreline) and 3 samples from waist-deep (separated by approximately ~ 3-4 m for CH, HB and SP and ~ 30 m for LP) at 2-hour intervals over 24 h (8:00 AM, 10:00 AM, 12:00 PM, 14:00 PM, 16:00 PM, 18:00 PM, 20:00 PM, 22:00 PM, 24:00 PM, 2:00 AM, 4:00 AM & 6:00 AM). Thus, the total number of samples for each beach was 216 (12 bi-hourly samples x 2 sites x - 3 biological replicates x 3 months = 216 samples). Each water sample was 250 mL, collected at 0.5 m depth. Samples were transported on ice (4 ºC) to the laboratory within eight hours after collection. Upon arrival at the lab, water samples were immediately filtered using 0.2 µm polycarbonate membranes (Millipore, USA), and the filters stored at -20 ºC until DNA extractions were performed.

Environmental variables: As a proxy of environmental variables with known effect on BCC (Jørgensen et al., 1998; Shao et al., 2013), we collected water temperature every 2-hours after sample collection in the field. Additionally, bi-hourly data for precipitation, solar radiation and wind speed were collected from Environment Canada for each diel cycle (http://climate.weather.gc.ca/historical_data/search_historic_data_e.html). We controlled the collected data from Environmental Canada with the data received from local weather stations (Windsor Bell River, Windsor South and Windsor Riverside) to ensure the relevancy of data to local beach conditions.

2.2.2 DNA extraction, PCR, library preparation and sequencing

DNA was extracted from the filters by adding 400 mL of sucrose lysis buffer (400 mM NaCl, 750 mM sucrose, 20 mM ethylenediaminetetraacetic acid (EDTA), 50 mM Tris-HCl pH 9.0); a lysis buffer which we optimized before for environmental DNA extraction (Shahraki et al., 2019a). The samples were homogenized using Mini-beadbeater-16 (Lab Services BV, Nederland) for 1 min (three times) with an intensity of 3,450 oscillations/min-1 to break down microbial cell structure. Subsequently, magnetic bead DNA purification using an automated platform (Tecan Freedom Evo150 Liquid Handling Platform, Perkin Elmer, USA) was performed (Shahraki et al.,

2019a). Genomic DNA was suspended in 50 µL TE buffer and stored at -80 °C until use.

We amplified the V5-V6 region of the 16S rRNA gene (~350 bp) using V5F (acctgcctgccg-ATTAGATACCCNGGTAG) and V6R (acgccaccgagc-CGACAGCCATGCANCACCT) primers (He et al., 2017) and then sample barcode and adaptor sequences were added to each sample's PCR product by a second, ligation, PCR. Second-round PCR products were pooled and purified using the QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). The concentration of purified PCR product mix (library) was measured using an Agilent 2100 Bioanalyzer with a High Sensitive DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library was then diluted to 60 pmol/L and sequenced on an Ion PGM™ System (Thermo Fisher Scientific, Burlington, ON, Canada) (He et al., 2017).

2.2.3 qPCR and measuring *E. coli* level

We quantified *E. coli* level as a bioindicator of fecal contamination (Pachepsky and Shelton, 2011) using qPCR for all four beaches bi-hourly in diel cycles (only knee-deep samples) by targeting the *uidA* gene (Chern et al., 2011). SYBR green qPCR reactions were carried out triplicates for each sample, extracted DNA from blank filters and negative controls during DNA extraction in 20 µL including 10 µL SYBR green master mix (Applied Biosystems™, USA), 1 µL primers (combined forward and reverse primers, final concentration 10 pmol for each), 1 µL DNA and 8 µL ddH$_2$O. The PCR thermal cycle protocol was: 95 °C for 1 min followed by 40 cycles 95 °C for 10 s and 59 °C for 60 s.

We used *uidA* purified PCR product to prepare known concentrations (2, 2 x 10^1, 2 x 10^2, 2 x 10^3, 2 x 10^4, 2 x 10^5 and 2 x 10^6 copies/reaction) of the target gene to generate a standard curve for SYBR green qPCR for *uidA* as we described before (Shahraki et al., 2019b). We estimated the *uidA* gene copy number in our water samples based on the standard curve, assuming that the cell recovery and DNA extraction efficiency were 100% in all samples and that only one gene copy was present per cell. The *E. coli* concentration based on *uidA* gene was calculated for a 100 mL water sample. No-template control (NTC) reactions including SYBR green master

mix, forward and reverse primers of each marker and ddH$_2$O but no DNA template for assay.

2.2.4   Bioinformatics and statistical analyses

Sequence handling: The raw NGS data were de-multiplexed, quality filtered and trimmed of the adaptor, barcodes and primer sequences using the Quantitative Insights into Microbial Ecology (QIIME V. 1.9.1) bioinformatics pipeline (Caporaso et al., 2010). A minimum quality score of Q=20 and a molecular size cut-off of 200 bp was selected for quality assurance. Chimeras were removed using ChimeraSlayer in QIIME. Sequences were clustered into operational taxonomic units (OTUs) based on sequence similarity (97%) and then taxonomically assigned using BLAST against Greengenes 16S rRNA database version 13_8 as a reference data file (Edgar, 2010). Singleton and doubleton OTUs (across all samples) were removed from the OTU table and after testing for the effect of replicate and sampling site for each beach using alpha and beta diversity (see below) we combined replicates to simplify the analysis. OTU abundance is expressed as percent (%) relative abundance. We defined an OTU as "abundant" when it had a relative abundance above 1.00% of the community, "moderate" when the relative abundance was >0.10% and <1.00% and "rare" when the abundance was below 0.1% (Logares et al., 2014). The OTU table was rarefied to 4100 reads using QIIME for further analyses. Original fastq files with metadata are deposited in NCBI Sequence Read Archive (ID PRJNA595580).

Temporal and spatial effects: We used a generalized linear model (GLM) to measure whether or not replicates and sampling sites (knee-deep and waist-deep) had significant effects on BCC variation. Using GLM, replicates and sampling points for each sampling hour were considered as fixed factors. We also used a generalized linear mixed model (GLMM) to test for spatial effect (sampling beach) and temporal effects (month, day/night and sampling hour) on BCC variation. In the GLMM, sampling site at each beach, beach, month, day/night and sampling hours were included as fixed factors while the replicates for each sampling hour were considered as random factors. In both models, alpha diversity indices (Chao1 and Shannon) and the first (PCo1) and second (PCo2) coordinates of the principal coordinate analysis (PCoA) were used separately as dependent variable. SPSS version 19 (SPSS Inc, Chicago, Illinois) was

used to run both GLMM and GLM. We chose Chao1, Shannon, PCo1 and PCo2 individually as a simple measure of BCC variation. Chao1 and Shannon were calculated using the rarified OTU table. Bray–Curtis dissimilarity matrix of each diel cycle was calculated and PCoA was performed using the program Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). We selected PCo1 and PCo2 of the PCoA from each diel cycle with eigenvalues >1.0 for analysis. As our principal goal was to assess fine-scale temporal effects, we focused only on the effects of day/night and sampling hour as well as the interaction effect of sampling hour with other fixed factors on the BCC of diel cycles.

Environmental effects: Mean water temperature and wind speed during daylight hours (8:00 AM, 10:00 AM, 12:00 PM, 14:00 PM, 16:00 PM and 18:00 PM) were compared against night hours (20:00 PM, 22:00 PM, 24:00 PM, 2:00 AM, 4:00 AM and 6:00 AM) for each diel cycle using the Student t test. We applied RELATE analysis to correlate Bray- Curtis dissimilarity of relative abundance data of each diel cycle to the Euclidean distance of normalized environmental variables of each diel cycle as an environmental matrix by calculating Spearman's p correlation coefficient. A distance-based linear model (distLM) was used for analyzing and modeling the relationship between the BCC of each diel cycle and the environmental variables using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). *E. coli* level (qPCR data) for diel cycles were correlated to environmental variables using SPSS version 19 (SPSS Inc, Chicago, Illinois).

Hourly variation of the BCC: First, we applied permutational multivariate analysis of variance (PERMANOVA) analyses with Bonferroni correction (number of permutations; 9999) to compare the Bray-Curtis dissimilarity of BCC of replicates and BCC of two sampling sites at each beach (knee- and waist-deep) using PAST v3.12 software (Hammer et al., 2001). Then we used PERMANOVA with Bonferroni correction to compare the Bray-Curtis dissimilarity of BCCs across all sample hours in each diel cycle. We plotted the Bray-Curtis dissimilarity value of different hours to visualize the variations of the BCC of each hour in all diel cycles. Chao 1 was compared by one-way ANOVA using SPSS version 19 (SPSS Inc, Chicago, Illinois) between the sampling hours. The mean of the diversity indices was plotted against

sampling hour in each diel cycle. To characterize the diel variation at the OTU level, we plotted hourly OTU relative abundance (only the OTUs with >1% abundance in any sampling hour) over each diel cycle. The abundance of taxa at the phyla/class level for each diel cycle was also plotted by hour in diel cycle.

Day/night variation of the BCC: During a diel cycle, we considered the BCCs of sampling hours 8:00 AM, 10:00 AM, 12:00 PM, 14:00 PM, 16:00 PM & 18:00 PM as the day community and the BCCs of sampling hours 20:00 PM, 22:00 PM, 24:00 AM, 2:00 AM, 4:00 AM & 6:00 AM as the night community. Then we applied PERMANOVA analyses to compare the Bray-Curtis dissimilarity matrix of day BCC versus night BCC. PCoAs were used to characterize variation in BCC between day and night hours using PAST v3.12 software (Hammer et al., 2001). Alpha diversity indices of day hours were compared against night hours by one-way ANOVA. We applied linear discriminant analysis (LDA) (http://huttenhower.sph.harvard.edu/galaxy/LEfSe, with default settings; Kruskal-Wallis test) to test for significant fluctuation of OTUs (top 500 abundant OTUs) and taxa between day and night within each diel cycle. The abundance of taxa at the phyla/class level for each diel cycle was also plotted by hour in the diel cycle. All plots and figures were generated using OriginPro 2018 (OriginLab Corporation, Northampton, MA, USA) or R software version 3.6.1 using ggplot2 package (Wickham, 2011).

2.3   Results and discussion

Sequence handling: Before combing the replicates, the samples had a mean sequence read number of 3,389 (range: 2,102-8,509) after quality control. In total, 29,700 OTUs were identified across all samples after removing singleton and doubleton OTUs. We further trimmed the data set to exclude OTUs with fewer than 20 total reads across full diel cycles, which resulted in 4734 OTUs for analysis in all four locations over the three sampled diel cycles.

Temporal and spatial effects: Replicates and sample collection site (knee- and waist-deep) at each beach had no significant effect ($p<0.05$) on alpha or beta diversity measures, we, therefore, combined the data from the replicates for each site at each beach to simplify the statistical analysis. Using GLMM, we found that beach and month had strong and significant effects on alpha diversity indices, but no significant

effect on either PCo1 or PCo2 (Table 2.1). As our goal was specifically to characterize the hourly variation of BCCs over a diel cycle, we focused more on the day/night and sampling hour effects as well as the interaction effects of hourly sampling with beach, month and day/night on alpha diversity indices and PCo1/PCo2.

Our analyses showed that both day/night and sampling hour had significant effects on alpha diversity indices, PCo1 and PCo2 (Table 2.1). The day/night effect size was larger than the sampling hour effect. Day/night had a larger effect on Chao1 ($F=13.05$, $p<0.0001$) than the Shannon index ($F=6.25$, $p<0.0001$). Also, day/night had greater effect on PCo1 ($F=5.06$, $p=0.02$) than PCo2 ($F=3.89$, $p=0.04$). Sampling hour exhibited a similar pattern; which had larger effect size on Chao 1 ($F=4.93$, $p=0.004$) than Shannon ($F=4.01$, $p=0.001$), and greater effect size on PCo1 ($F=4.23$, $p=0.001$) than PCo2 ($F=3.11$, $p=0.003$). The interactions of beach*hours, month*hours, day/night*hours, all had significant effects for Chao1 and Shannon, with a larger effect size for Chao1 (Table 2.1). All interaction terms also had significant effects on PCo1, but only beach*hours showed a significant effect on PCo2 (Table 2.1). Overall, our GLMM showed high levels of diel variation in freshwater BCC; this variation was observed at the hourly level, as well as day versus night. This pattern of variation was consistent across four beaches from two lakes and in three summer months.

Table 2.1. GLMM analysis of the effects of the sampled beach and various temporal factors (month, day/night and hourly sampling) on freshwater BCC alpha diversity and Bray–Curtis dissimilarity principle coordinate axes (PCo1 & PCo2).

	Main fixed factors				Interactions			
	Beach	**Month**	**Day/Night**	**Sampling hours**	**Beach *Sampling hours**	**Month *Sampling hours**	**Day/night *Sampling hours**	**Beach*Month *Day/Night *Sampling hours**
Chao 1	F=18.05, **p<0.0001**	F=34.34, **p<0.0001**	F=13.05, **p<0.0001**	F=4.93, **p=0.004**	F=5.11, **p=0.01**	F=8.2, **p=0.006**	F=6.67, **p=0.001**	F=1.23, p=0.09
Shannon	F=16.5, **p<0.0001**	F=53.27, **p<0.0001**	F=6.25, **p<0.0001**	F=4.01, **p=0.001**	F=4.01, **p=0.01**	F=6.88, **p=0.004**	F=4.07, **p=0.02**	F=0.49, p=0.35
PCo1	F=0.24, p=0.76	F=0.13, p=0.85	F=5.06, **p=0.02**	F=4.23, **p=0.001**	F=3.46, **p=0.009**	F=1.75, **p=0.04**	F=1.8, **p=0.04**	F=0.8, p=0.49
PCo2	F=0.04, p=0.86	F=0.18, p=0.78	F=3.89, **p=0.04**	F=3.11, **p=0.003**	F=1.75, **p=0.03**	F=0.84, p=0.29	F=0.9, p=0.4	F=0.92, p=0.62

Environmental effects: No precipitation was recorded by Environment of Canada (http://climate.weather.gc.ca) and local weather stations not only during any of our diel sampling but also 24 h before sampling. Mean (day versus night hours) water temperatures and wind speed for each diel cycle showed substantial variation (Appendix A; Table S2.1) for each diel cycle. Solar radiation was close to zero for our night period (20:00 PM to 6:00 AM) but it ranged between 80-760 (W/m^2) for the day period (8:00 AM to 18:00 PM). The Euclidean distance of environmental variables (water temperature, solar radiation and wind speed) of each diel cycle was significantly correlated (r =0.14-0.3, p < 0.05) to the biological data (Bray–Curtis similarity matrix) across diel cycles. Also, distLM analysis shown that environmental variables had significant effect (p < 0.05) on the BCC of CH (June: r^2=0.12, July: r^2=0.13, August: r^2=0.11), HB (June: r^2=0.17, July: r^2=0.11, August: r^2=0.15), LP (June: r^2=0.11, July: r^2=0.21, August: r^2=0.2) and SP (June: r^2=0.12, July: r^2=0.12, August: r^2=0.14) (Appendix A; Table S2.2).

Freshwater BCC has been shown to be affected by various environmental factors such as sediment resuspension, salinity, water temperature, wind speed, precipitation and sunlight (Bryant et al., 2016; Shao et al., 2013). In our study, environmental variables (water temperature, solar radiation and wind speed) only explained 11-21% of the observed diel variation in BCC (mean across all diel cycles=19%). We did not include measures of nutrients in our bi-hourly study, and to the best of our knowledge, there is no study regarding the effect nutrients (such as phosphate and nitrogen) on fine-scale (24 hour period) temporal variation of freshwater BCC.

2.3.1 Bi-hourly variation of the BCCs within diel cycles

Analysis of the Bray-Curtis dissimilarity matrix showed that there was no significant variation ($p < 0.05$) between the BCC of the replicates at each sampling hours or between the BCC of the knee- and waist-deep sample sites across all diel cycles. PERMANOVA analysis shown significant variation (df= 11) among the BCC of the different sampling hours across all diel cycles as follows; CH (June: F=1.2, p=0.001, July: F=2.3, p=0.001 and August: F=2, p=0.001), HB (June: F=3.8, p=0.0001, July: F=2.3, p=0.001 and August: F=2.8, p=0.001), LP (June: F=3.15, p=0.0001, July: F=5.2, p=0.0001 and August: F=4.2, p=0.0001) and SP (June: F=2.5, p=0.001 July: F=1.8, p=0.001 and August: F=4.4, p=0.0001). Pairwise comparison of the BCC of the sampling hours showed that the BCCs of some sampling hours were significantly different from others.

Generally, we did not identify common patterns of hourly variation between all diel cycles which could be related to different environmental conditions in each beach zone (Cloutier et al., 2015). However, hourly variation within the diel cycles was noticeable (Figure 2.1). For example, the BCC of hours 16:00 PM, 18:00 PM, 20:00 PM and 24:00 PM in the BCC of HB (July) and the BCC of hours 8:00 AM, 14:00 PM, 16:00 PM and 18:00 PM in the BCC of SP (June) were significantly different from the BCC of other sampling hours in each of those diel cycles. Few studies have reported fine-scale (diel-level) temporal variation such as that reported here. Fine-scale temporal variation of the coastal marine microbial community (6-hour intervals) (Gilbert et al., 2010) and bacterial biomass and bacterial production of the BCC of Mediterranean Sea (sporadic temporal variation across 6-hours intervals) (Ghiglione et al., 2007; Mével et

al., 2008) have been reported before. Our study provided substantial fine-scale temporal BCC variation (2-hours intervals) over the 24-sampling period at all four locations and across all three sampling months in freshwater lakes.
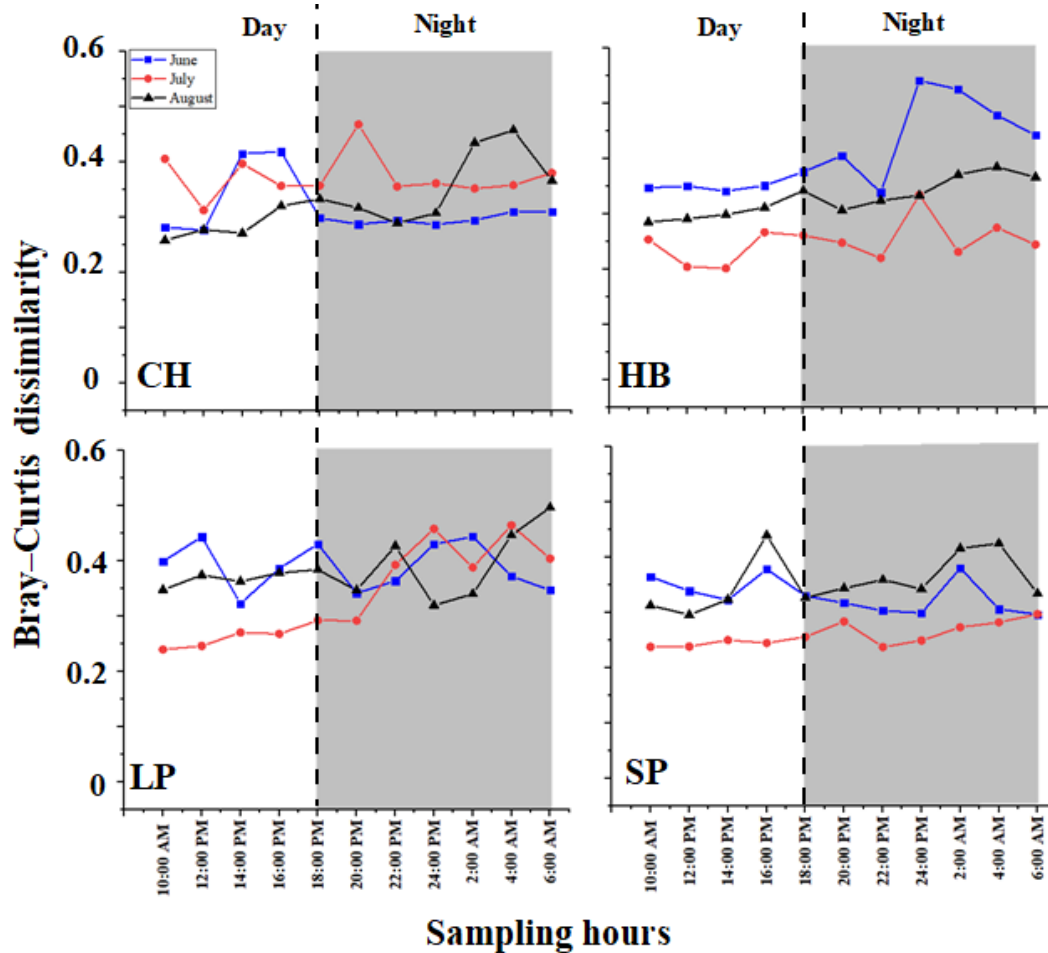
Figure 2.1. Variation in mean pair-wise Bray–Curtis dissimilarity of the BCC over different hours in all locations (CH, HB, LP and SP) over June, July and August. Gray area in each plot represents the night hours. As Bray–Curtis dissimilarity value was zero for 8:00 AM, we removed it from all graphs.

Bi-hourly variation of Chao1 index: No significant variation was detected among replicate or site samples (3 replicates x 2 sites =6 samples) using Student t test for alpha diversity indices within each diel cycle and we, therefore, combined the replicate data. We compared Chao1 index variation across sample hours within each diel cycle

using one-way ANOVA. We found that the mean of the Chao1 diversity index at 4:00 and 6:00 AM were significantly higher (p<0.05) than the mean for the other sampling hours in all diel cycles except HB July diel cycle (where the mean Chao1 index was significantly lower than other sampling hours).

Generally, the mean of Chao1 index increased from 8:00 AM at the beginning of the sampling to 6:00 AM of the next day in all four beaches over three sampling dates, except HB (July diel cycle) which showed the opposite trend (Appendix A; Figure S2.1). We noticed that the mean of Chao1 index remained constant (CH; June and HB; July), increased (CH; July and August, HB; June and August and LP; June and July) or declined (LP; August and SP; June, July and August) over day hours (Appendix A; Figure S2.1), but mostly increased over night hours. Shannon results plotted and presented in supplementary results (Appendix A; Figure S2.2).

Bi-hourly variation of OTUs and taxa: Within each diel cycle, 7-29 and 47-169 OTUs were identified as highly or moderately abundant, respectively. We only focused on the OTUs which were highly abundant in all or some of the sampling hours within each diel cycle. Only 2 OTUs belong to phylum *Actinobacteria* and family ACK-M1 (OTU1 and OTU2) and one OTU belongs to phylum *Proteobacteria* and family *Comamonadaceae* (OTU3) were abundant in all bi-hourly sampling across all diel cycles (Figure 2.2). There was also an obvious variation in the relative abundance of highly abundant OTUs across different hours of each diel cycle (Figure 2.2). For example, OTU1 showed fluctuations across the sampling hours of the diel cycle at CH (June: 6.2-7.4%, July: 11.6-17.5% and August: 6.7-9.7%), HB (June: 4-9.2%, July: 10.9-18.9% and August: 8.6-13%), LP (June: 4.6-11.4%, July: 9.4-16.8% and August: 4-15%) and SP (June: 5.5-15.9% , July: 16.3-21.9% and August: 6-20.5%). At the OTU level, our data showed that bi-hourly variation of freshwater BCC is mostly related to high abundant OTUs such as those belonging to family ACK-M1 of *Actinobacteria* (OTUs 1 and 2) and family *Comamonadaceae* of *Proteobacteria* (OTU3).
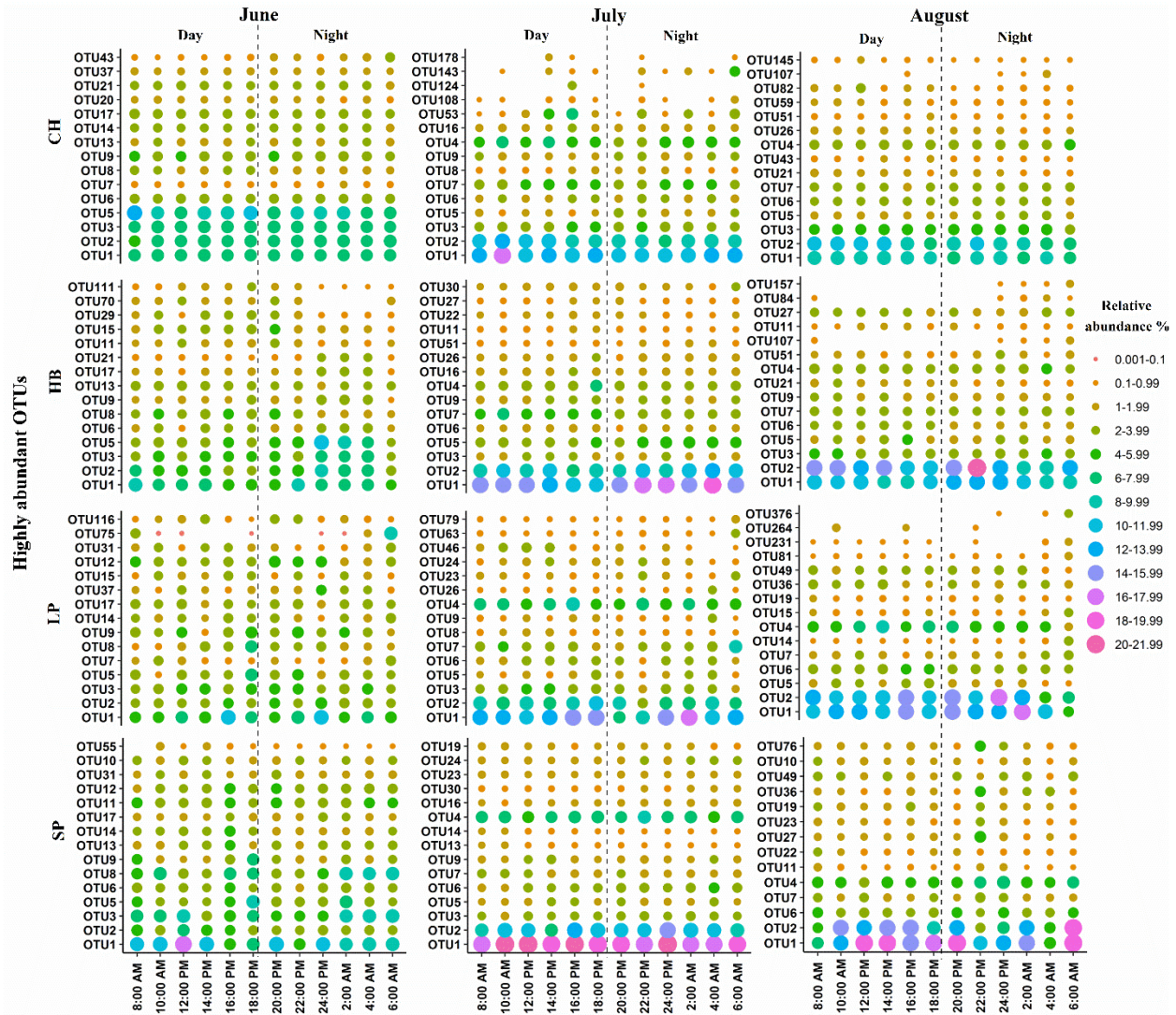
Figure 2.2. The relative abundance of the "abundant" OTUs across each diel cycle at the CH, HB, SP, and LP beaches over three sampling months (June, July, and August). OTUs which showed a relative abundance of > 1% even in one sampling hour were selected, and out of those OTUs, the top 15 OTUs with the highest standard deviation across the sampling hours were plotted. OTUs are listed on the Y-axis and the X-axis shows the sampling hour (started at 8:00 AM and continued to the next day at 6:00 AM) with a 2-h interval.

In each diel cycle, some OTUs were consistently highly abundant (with fluctuation in their relative abundance across different hours) overall sampling hours, however, in top of those OTUs, the relative abundance of some other OTUs belong to different taxa (family) exceeded 1% and became highly abundant in particular sampling hours. For

example, in the BCC of CH (August), families C111 (OTUs 7 and 27) and ACK-M1 (OTUs 1, 2, 4 and 10) belong to *Actinobacteria,* family *Comamonadaceae* of *Proteobacteria* (OTUs 3 and 6) and families *Cyclobacteriaceae* (OTU5) and *Sphingobacteriaceae* (OTU9) of *Bacteroidetes* were consistently highly abundant across all hours of the diel cycle. However, the BCC of hour 12:00 PM had 9 additional highly abundant OTUs; 6 of them belong to different families of *Proteobacteria* including *Comamonadaceae* (OTU20), *Methylophilaceae* (OTU21), *Oxalobacteraceae* (OTU145), *Pelagibacteraceae* (OTUs 52 and 59) and *Rhodobacteraceae* (OTU12), 2 OTUs belong to family *Chitinophagaceae* of *Bacteroidetes* (OTUs 26 and 82) and 1 OTU belong to family *Synechococcus* of *Cyanobacteria* (OTU42).

Actinobacteria (34.8±8.8%), *Betaproteobacteria* (18.5±5%) and *Bacteroidetes* (15.6±5.2%) were dominant phyla in the BCC of all diel cycles (Appendix A; Figure S2.3). Similar to our finding, *Actinobacteria* lineages has reported as a dominant phylum in the BCC of Lake Erie (Paver et al., 2020; VanMensel et al., 2019) and Lake St. Clair (VanMensel et al., 2019). *Actinobacteria* and *Betaproteobacteria* exhibited considerable variation in their relative abundance by sampling hours within each diel cycle in our study. For example, *Actinobacteria* changed in abundance dramatically at all sampled beaches: CH (June: 27.7-31.5%, July: 31.8-51.8% and August: 30.3-40.4%), HB (June: 17.3-30.8%, July: 33.5-48.4% and August: 33.5-53.5%), LP (June: 18.3-29.2%, July: 27-41.7% and August: 23.6-53.6%), SP (June: 19.6-36.6%, July: 39.7-51.5% and August: 21.5-55%) (Appendix A; Figure S2.3). The bi-hourly variation of *Betaproteobacteria* was noticeable (Appendix A; Figure S2.3). Fine temporal-scale variation in abundance of *Proteobacteria* in the wastewater microbiome (4-hour intervals) (Guo et al., 2019) has been reported before, however, our report provided more bi-hourly variation of most dominant phyla in a freshwater ecosystem.

2.3.2 Day/night variation of the BCCs

PERMANOVA analysis (df=1) showed the BCCs of day hours were significantly different from the BCCs of night hours for all diel cycles (Table 2.2). Multiple dimensional scaling by PCoA revealed distinct clustering of the day versus night BCC across all four beaches for all three sampling periods (Figure 2.3). In contrast, diel-

level (day/night) stability (no variation) of a maritime Antarctic lake (Pearce and Butler, 2002) and the community of a coastal microbial mat (sampled at 4-hours intervals during a 24-hours) (Cardoso et al., 2017) has been reported. It is surprising that day-night changes in the BCC have not been more widely reported, given the critical changes in abiotic and biotic factors from daylight to dark conditions (e.g. light energy, temperature, predators, etc.). While limited published evidence for day-night variation in BCC exists, day/night variation in the meta-transcriptome of microbial communities has been reported (Campbell and Kirchman, 2013; Kim et al., 2015; Poretsky et al., 2009). Our work showing significant and consistent day/night variation in the freshwater BCC corresponds to expected changes in functional BCC activity. Such functional shifts would fundamentally change the ecological processes that drive the BCC; however, more study is needed to determine the functional role of even greater fine-scale temporal variation in BCC in the provision of ecological services and the health risks of freshwater ecosystems.

Table 2.2. Day/night variation of the BCCs according to PERMANOVA analysis for all diel cycles.

		June	July	August
CH	F value	3.5	5.5	4
	p value	**0.0008**	**0.0001**	**0.0001**
HB	F value	7.7	6.1	4.4
	p value	**0.0001**	**0.0001**	**0.0001**
LP	F value	3.1	9.3	6.5
	p value	**0.0009**	**0.0001**	**0.0006**
SP	F value	2.1	3.8	3.8
	p value	**0.005**	**0.0002**	**0.0006**
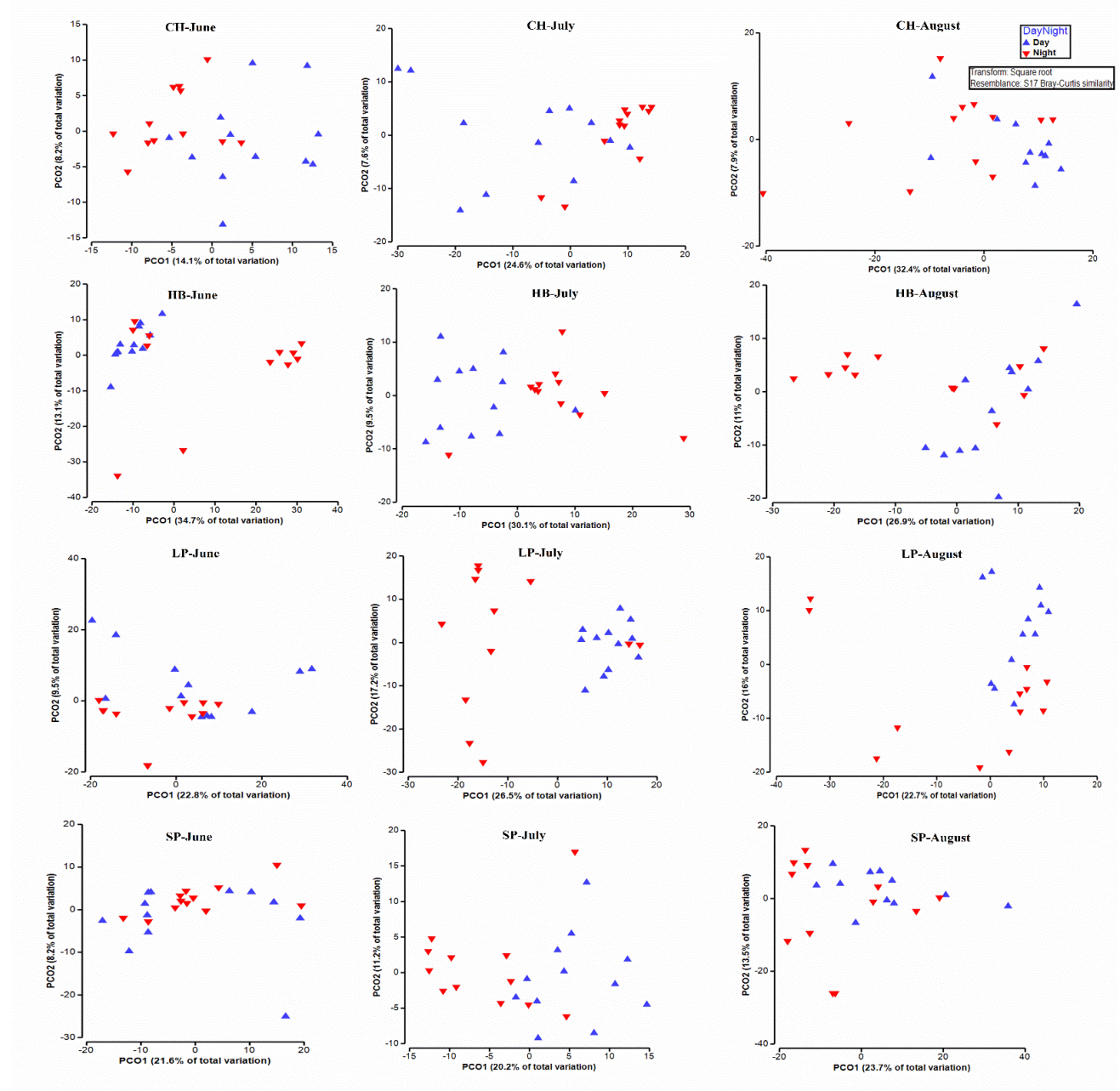
Figure 2.3. Scatterplots of PCo1 versus PCo2 from PCoA of the BCCs for day versus night hours for CH, HB, LP & SP beaches in three different sampling times (June, July & August) showing high-resolution small scale (day/night) temporal variation of the BCC in a diel cycle.

Day/night variation of Chao1 index: The Chao1 index was generally significantly elevated at night over most of the diel cycles. However, HB (July) showed the opposite effect, where the mean Chao1 index was significantly higher (df=1, F=14.34, p=0.001)

during daylight hours. There was also one diel sampling that showed no significant day-night variation (p>0.05) in Chao1 (SP (July); F=2.1, p=0.12; Appendix A; Table S2.3).

Overall, we detected significant day/night variation in both alpha and beta diversity in the freshwater bacterial community across all sampled diel cycles, with a general elevation of diversity in the night community. Similar to our finding, significant elevation of diversity at night hours (Gilbert et al., 2010) and high production and growth rate of freshwater BCC (Filippini et al., 2008) has been reported previously in aquatic ecosystems. Those differences were identified as potentially related to the absence of UV stress at nigh hours and its negative impact on rare taxa (Whitman et al., 2004).

Day/night variation of OTUs and taxa: We applied LDA to the top 500 abundant OTUs in the diel cycles to compare the BCCs of the day versus night community to identify the OTUs which show significant changes in their relative abundance across the three sampled beaches over the three sampled summer months. Generally, more OTUs showed significant increases in their relative abundance at night than those showing significant increases in their relative abundance during daylight hours. For example, at the CH beach, the relative abundances of 33 (June), 53 (July) and 44 (August) unique OTUs were significantly higher in the night community. This is in contrast to unique OTUs 13 (June), 19 (July) and 27 (August) which were more abundant during daylight hours. Generally, none of the 500 abundant OTUs showed consistent significant variation in their relative abundance in either the day or night community across all four beaches and all three sampling months potentially due to short temporal variation of abiotic variables (Bryant et al., 2016) (e.g. hydrology and the water chemistry, sunlight, etc.) and biotic variables (e.g. grazers, viruses, etc.) (Grubisic et al., 2017; Lymer et al., 2008).

Indeed, some OTUs showed opposite patterns in their relative day versus night abundances. However, we found little consistency in the day/night patterns of some OTUs across beaches and sampling months. For example, the relative abundances of 10 OTUs of phylum *Actinobacteria* belonging to families ACK-M1 (OTUs 2, 13, 18, 25, 58, and 102) and C111 (OTUs 7, 14, 62 and 68) were significantly higher in the

day community, while the relative abundances of 7 OTUs of *Actinobacteria* belonging to families ACK-M1 (OTUs 1 and 110), C111 (OTUs 134 and 135) and *Microbacteriaceae* (OTUs 41, 49, and 132) were significantly higher in the night communities (Appendix A; Table S2.4). High abundance of *Actinobacteria* particularly families ACK-M1 and C111 in the day community could be due to the presence of rhodopsins (actinorhodopsins) within these taxa, a potential mechanism for supplemental energy generation through light-harvesting, coupled with their UV stress resistance and potentially better contribution in biogeochemical cycling over day hours (Newton et al., 2011).

The relative abundance of OTUs belonging to phylum *Cyanobacteria* and family *Stramenopiles* (OTUs 42, 43, 89, 103 and 161) and phylum *Betaproteobacteria* and order *Flavobacteriales* (OTUs 33, 77, 80, 83, 99, 116 and 153) showed a significant increase only in the night community (Appendix A; Table S2.4). The apparent co-occurrence of *Cyanobacteria* and *Flavobacteriales* in our study especially at night may be due to the fact that members of *Flavobacteriales* playing a role in degrading cyanobacterial toxins or other problematic organic compounds, and thus enhancing proliferation of *Cyanobacteria* (Berg et al., 2009). The relative abundance of OTUs belonging to *Alphaproteobacteria* and order *Rhizobiales* (OTUs 22, 53, 54, 72 and 117) were significantly elevated in the night community, but only in some diel cycles such as CH and HB (August), LP (July) and SP (June). Order *Rhizobiales* (belong to phylum *Proteobacteria*) contain dominant nitrogen fixing taxa (Newton et al., 2011). We also found that OTUs assigned to family *Verrucomicrobiaceae* (OTUs 51, 87 and 114) were significantly high only in the BCC of day CH (August) and HB (June and July) (Appendix A; Table S2.4). Taxa belong to *Verrucomicrobiae* are not well studied in the BCC of freshwater lakes potentially due to rare occurrences in many aquatic ecosystems (Newton et al., 2011) however, this study provided fine temporal variation (day/night) of this taxon as well.

We applied LDA to compare the relative abundance of the 61 detected family-level taxonomically assigned bacteria between day and night assemblage across each diel cycle (Appendix A; Figure S2.4). Some diel cycles had limited taxonomic representation (4 families) but still showed a significant divergence between the BCC

of day and night (i.e., LP (July and August) and SP (June)). Other diel cycles had a large taxa diversity at the family level and also exhibited significant variation in relative abundance between day and night (i.e., CH (August; 19 families) and HB (June; 20 families). Overall, of the 61 identified bacterial families, 12 showed significant increases in their relative abundance only in the day (e.g. *Gemmatimonadaceae* (CH; August and HB; June and August)).

Moreover, the relative abundance of 21 bacterial families were significantly higher in the day communities for some diel cycles but were significantly higher in the night communities of other diel cycles. For example, the relative abundance of *Cytophagaceae* was significantly higher in the day for CH in June, but higher at night in August. Out of the 61 identified bacterial families; 17 showed significantly elevated abundance only at night (e.g., *Acidaminobacteraceae* at CH (August) and HB (July and August)). More interestingly, our metabarcoding analyses showed that the relative abundance of *Enterobacteriaceae* (*Gammaproteobacteria*) was significantly elevated in the night BCCs at CH (July and August), HB (June and July), LP (June and August) and SP (June). Some members of *Enterobacteriaceae* (enteric organisms) are waterborne pathogens (Bridge et al., 2010) and *E. coli* is one well-known member of this family as it is currently used as fecal indicator bacteria in water health monitoring (Edberg et al., 2000).

Our results showed strong day/night variation in the composition of freshwater BCC with the dominance of UV resistant and light-harvesting taxa such as families ACK-M1 and C111 (*Actinobacteria*) in the BCC of day versus enrichment of the BCC of night community with heterotrophic taxa mostly involves in biogeochemical cyclings such as *Stramenopiles* (*Cyanobacteria*), *Flavobacteriales* (*Betaproteobacteria*) and *Rhizobiales* (*Alphaproteobacteria*) (Berg et al., 2009; Newton et al., 2011) and taxa associated mostly with health risk such as *Enterobacteriaceae* (*Alphaproteobacteria*) (Bridge et al., 2010).

2.3.3 *E. coli* dynamic over diel cycle

The coefficient of variance between replicates (R^2) and qPCR efficiency was 0.99 and 103% for *uidA* gene. *E. coli* was not detected directly in our metabarcoding data as NGS usually does not allow for this taxonomic resolution, yet the diel variation in

Enterobacteriaceae made it possible that *E. coli* levels were also varying a fine temporal scale in these freshwater ecosystems. We used qPCR to monitor *E. coli* abundance in our samples and found *E. coli* consistently decreased from 8:00 AM through the day, and then increased over night for all three sampled months at CH, LP and SP beaches (Figure 2.4). In general, reductions in *E. coli* levels were most prominent at 10:00 AM, 12:00 PM, 14:00 PM and 16:00 PM over all diel cycles. In the majority of sampled diel cycles, *E. coli* levels were significantly higher in the night communities in comparison to the day communities, except CH (August) and LP (July and August), where there was no significant variation ($p>0.05$) in *E. coli* level between day and night communities (Figure 2.4). We found a high level of *E. coli* during the day in comparison to night (not significantly higher; $p>0.05$) only at LP beach (July) which could be related to pollution events (river discharge) or resuspending of sand by beachgoers or waves (Ge et al., 2012). Similarly, previous studies of river watershed and creeks ecosystems have also shown higher levels of *E. coli* in the morning relative to the afternoon (Stocker et al., 2016; Traister and Anisfeld, 2006). Moreover, we observed a consistent diel oscillation of *E. coli* abundance, with a steady decrease throughout the daylight hours and rising *E. coli* levels at night. Solar radiation was the only environmental variable that had a significant ($p=0.0001$) negative correlation (Pearson coloration=-0.65) with *E. coli* level in our diel cycles. This may be related to the germicidal action of solar radiation in shallow aquatic habitats (Maraccini et al., 2016; Nelson et al., 2018). We did not measure *E. coli* level by the culture-based method to compare with qPCR data, as culture-based *E. coli* data may have shown a stronger UV response than qPCR-based *E. coli* data. Due to considerable changes in *E. coli* level (bi-hourly and day/night variation) and extreme variability of *Enterococci* as FIB in coastal waters over 24 h (Boehm, 2007), it seems that water quality monitoring based on single grab sample should be replaced by multiple grab sample over different hours to gain high-resolution picture of water quality and human health risk.
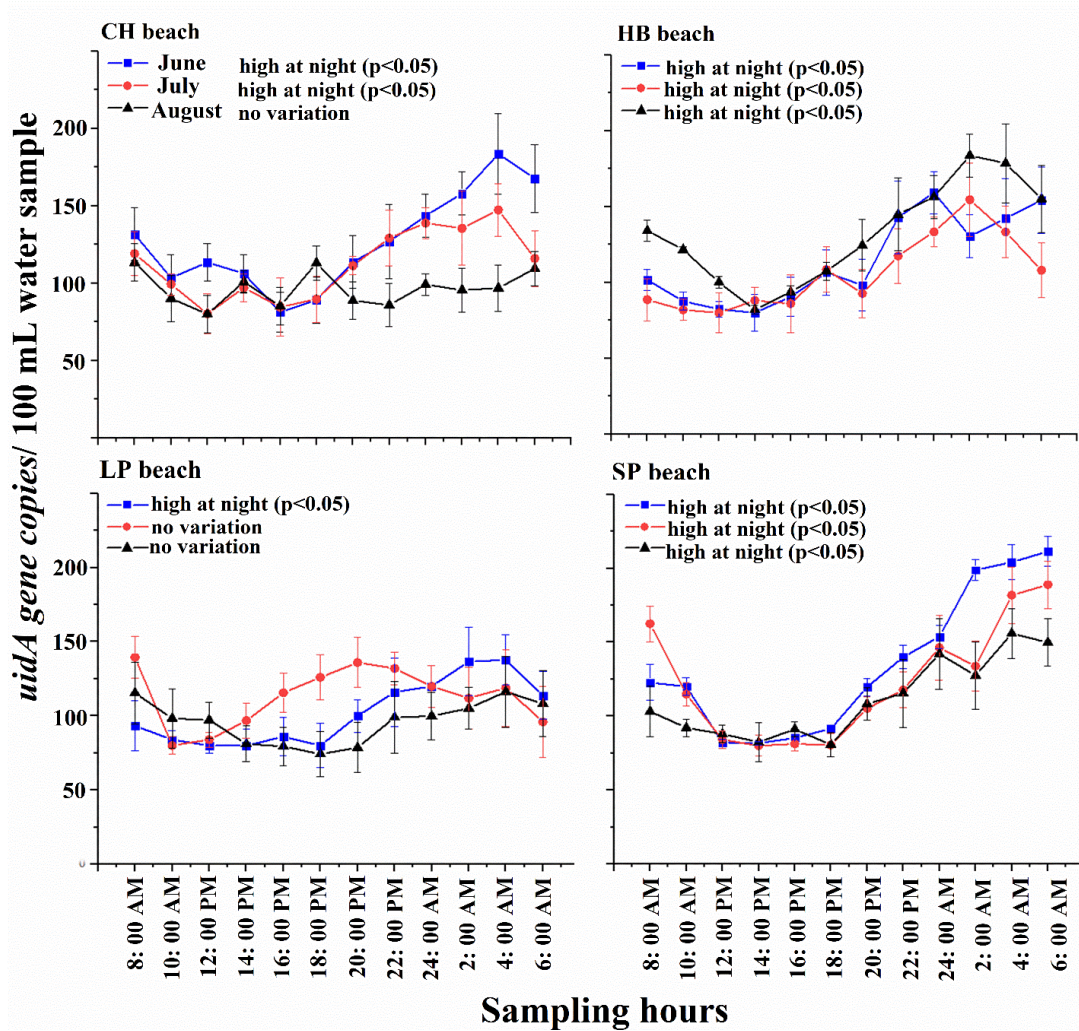
Figure 2.4. Line plots of hourly variation in E. coli mean abundance (error bars represented the standard deviation of *E. coli* level in each sampling hours) based on *uidA* gene qPCR assays. The panels show data for the four sampled beaches (CH, HB, LP & SP) over three sampling dates (June, July and August). *uidA* gene counts was estimated based on the standard curve and goodness of fitness in 100 mL water samples.

2.4   Conclusion

It is well known that the bacterial community provide ecological services such as biogeochemical processes in all aquatic ecosystems. Our study provides novel and exciting evidence for fine scale temporal (bi-hourly and day/night) variation of freshwater BCC in large lake ecosystems which could be mostly attributed to the

variation of dominant phyla such *Actinobacteria* and *Proteobacteria.* Such high levels of fluctuation of the BCC suggests fine-scale temporal variation in the biogeochemical cycles in large lakes may also be occurring. While the environmental variables we included in our analyses (water temperature, solar radiation and wind speed) only explained 19% of the observed diel variation in BCC, other key biotic and abiotic variables could be playing important roles as well. For example, we did not characterize viral and protist community variation as another potential driver of the BCC in this study; however, those communities have been reported as important factors (Grubisic et al., 2017; Lymer et al., 2008), and may have contributed to the BCC temporal variation observed in this study. As we detected low levels of *E. coli* during day hours with a significant negative correlation with solar radiation, combined with low levels of diversity in the day community versus the night community, it is likely that UV stress is a major environmental factor contributing to the fine-scale variation. We detected family-level variation of *Enterobacteriaceae* using metabarcoding data, indicative of possible important human-health related diel bacterial dynamics. Monitoring of *E. coli* using more sensitive qPCR assays confirmed the family-level variation patterns and substantial variation in *E. coli* abundance (day/night) over the diel cycle.

2.5   References

Berg, K.A., Lyra, C., Sivonen, K., Paulin, L., Suomalainen, S., Tuomi, P., Rapala, J., 2009. High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. The ISME journal 3(3) 314.

Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J., 2017. Cyanobacterial harmful algal blooms are a biological disturbance to western Lake Erie bacterial communities. Environmental Microbiology 19(3) 1149-1162.

Boehm, A.B., 2007. Enterococci Concentrations in Diverse Coastal Environments Exhibit Extreme Variability. Environmental Science & Technology 41(12) 8227-8232.

Bridge, J.W., Oliver, D.M., Chadwick, D., Godfray, H.C.J., Heathwaite, A.L., Kay, D., Maheswaran, R., McGonigle, D.F., Nichols, G., Pickup, R., 2010. Engaging with the water sector for public health benefits: waterborne pathogens and diseases in developed countries. Bulletin of the world Health Organization 88 873-875.

Bryant, J.A., Aylward, F.O., Eppley, J.M., Karl, D.M., Church, M.J., DeLong, E.F., 2016. Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. The ISME Journal 10(6) 1308.

Butler, T.M., Wilhelm, A.-C., Dwyer, A.C., Webb, P.N., Baldwin, A.L., Techtmann, S.M., 2019. Microbial community dynamics during lake ice freezing. Scientific Reports 9(1) 1-11.

Campbell, B.J., Kirchman, D.L., 2013. Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. The ISME Journal 7(1) 210.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5) 335.

Cardoso, D.C., Sandionigi, A., Cretoiu, M.S., Casiraghi, M., Stal, L., Bolhuis, H., 2017. Comparison of the active and resident community of a coastal microbial mat. Scientific Reports 7(1) 1-10.

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

Cloutier, D.D., Alm, E.W., McLellan, S.L., 2015. Influence of land use, nutrients, and geography on microbial communities and fecal indicator abundance at Lake Michigan beaches. Applied and Environmental Microbiology 81(15) 4904-4913.

Dodds, W.K., Perkin, J.S., Gerken, J.E., 2013. Human impact on freshwater ecosystem services: a global perspective. Environmental Science & Technology 47(16) 9061-9068.

Edberg, S., Rice, E., Karlin, R., Allen, M., 2000. Escherichia coli: the best biological drinking water indicator for public health protection. Journal of Applied Microbiology 88(S1) 106S-116S.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19) 2460-2461.

Filippini, M., Buesing, N., Gessner, M.O., 2008. Temporal dynamics of freshwater bacterio-and virioplankton along a littoral–pelagic gradient. Freshwater Biology 53(6) 1114-1125.

Ge, Z., Whitman, R.L., Nevers, M.B., Phanikumar, M.S., 2012. Wave-induced mass transport affects daily Escherichia coli fluctuations in nearshore water. Environmental Science & Technology 46(4) 2204-2211.

Ghiglione, J., Mevel, G., Pujo-Pay, M., Mousseau, L., Lebaron, P., Goutx, M., 2007. Diel and seasonal variations in abundance, activity, and community structure of particle-attached and free-living bacteria in NW Mediterranean Sea. Microbial Ecology 54(2) 217-231.

Gilbert, J.A., Field, D., Swift, P., Thomas, S., Cummings, D., Temperton, B., Weynberg, K., Huse, S., Hughes, M., Joint, I., 2010. The taxonomic and functional diversity of microbes at a temperate coastal site: a multi-omic study of seasonal and diel temporal variation. PloS One 5(11) e15545.

Glasl, B., Webster, N.S., Bourne, D.G., 2017. Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. Marine Biology 164(4) 91.

Grubisic, L.M., Bertilsson, S., Eiler, A., Heinrich, F., Brutemark, A., Alonso-Sáez, L., Andersson, A.F., Gantner, S., Riemann, L., Beier, S., 2017. Lake bacterioplankton dynamics over diurnal timescales. Freshwater Biology 62(1) 191-204.

Guo, B., Liu, C., Gibson, C., Frigon, D., 2019. Wastewater microbial community structure and functional traits change over short timescales. Science of The Total Environment 662 779-785.

Hammer, Ř., Harper, D., Ryan, P., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4 4-9.

He, X., Chaganti, S.R., Heath, D.D., 2017. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microbial Ecology 1-12.

Hölker, F., Wurzbacher, C., Weißenborn, C., Monaghan, M.T., Holzhauer, S.I., Premke, K., 2015. Microbial diversity and community respiration in freshwater sediments influenced by artificial light at night. Philosophical Transactions of The Royal Society 370(1667) 1-10.

Horton, D.J., Theis, K.R., Uzarski, D.G., Learman, D.R., 2019. Microbial community structure and microbial networks correspond to nutrient gradients within coastal wetlands of the Laurentian Great Lakes. FEMS Microbiology Ecology 95(4) fiz033.

Huntscha, S., Stravs, M.A., Bühlmann, A., Ahrens, C.H., Frey, J.r.E., Pomati, F., Hollender, J., Buerge, I.J., Balmer, M.E., Poiger, T., 2018. Seasonal Dynamics of Glyphosate and AMPA in Lake Greifensee: Rapid Microbial Degradation in the Epilimnion During Summer. Environmental Science & Technology 52(8) 4641-4649.

Jørgensen, N.O., Tranvik, L., Edling, H., Granéli, W., Lindell, M., 1998. Effects of sunlight on occurrence and bacterial turnover of specific carbon and nitrogen compounds in lake water. FEMS Microbiology Ecology 25(3) 217-227.

Kim, Y.-M., Nowack, S., Olsen, M.T., Becraft, E.D., Wood, J.M., Thiel, V., Klapper, I., Kühl, M., Fredrickson, J.K., Bryant, D.A., 2015. Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. Frontiers in Microbiology 6 209.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24(8) 813-821.

Lymer, D., Logue, J.B., Brussaard, C.P., BAUDOUX, A.C., Vrede, K., LINDSTRÖM, E.S., 2008. Temporal variation in freshwater viral and bacterial community composition. Freshwater Biology 53(6) 1163-1175.

Maraccini, P.A., Mattioli, M.C.M., Sassoubre, L.M., Cao, Y., Griffith, J.F., Ervin, J.S., Van De Werfhorst, L.C., Boehm, A.B., 2016. Solar inactivation of enterococci and Escherichia coli in natural waters: Effects of water absorbance and depth. Environmental Science & Technology 50(10) 5068-5076.

McPhedran, K., Seth, R., Bejankiwar, R., 2013. Occurrence and predictive correlations of Escherichia coli and Enterococci at Sandpoint beach (Lake St Clair), Windsor, Ontario and Holiday beach (Lake Erie), Amherstburg, Ontario. Water Quality Research Journal 48(1) 99-110.

Mével, G., Vernet, M., Goutx, M., Ghiglione, J.F., 2008. Seasonal to hour variation scales in abundance and production of total and particle-attached bacteria in the open NW Mediterranean Sea (0–1000 m). Biogeosciences 5(6) 1573-1586.

Mill, A., Schlacher, T., Katouli, M., 2006. Tidal and longitudinal variation of faecal indicator bacteria in an estuarine creek in south-east Queensland, Australia. Marine Pollution Bulletin 52(8) 881-891.

Nelson, K.L., Boehm, A.B., Davies-Colley, R.J., Dodd, M.C., Kohn, T., Linden, K.G., Liu, Y., Maraccini, P.A., McNeill, K., Mitch, W.A., 2018. Sunlight-mediated inactivation of health-relevant microorganisms in water: a review of mechanisms and modeling approaches. Environmental Science: Processes & Impacts 20(8) 1089-1122.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75(1) 14-49.

Pachepsky, Y., Shelton, D., 2011. Escherichia coli and fecal coliforms in freshwater and estuarine sediments. Critical Reviews in Environmental Science and Technology 41(12) 1067-1110.

Paver, S.F., Newton, R.J., Coleman, M.L., 2020. Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. Environmental Microbiology 22(1) 433-446.

Pearce, D., Butler, H., 2002. Short-term stability of the microbial community structure in a maritime Antarctic lake. Polar Biology 25(7) 479-487.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., Moran, M.A., 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environmental Microbiology 11(6) 1358-1375.

Shahraki, A.H., Chaganti, S.R., Heath, D., 2019a. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. Journal of Water and Health(17) 37-49.

Shahraki, A.H., Heath, D., Chaganti, S.R., 2019b. Recreational water monitoring: Nanofluidic qRT-PCR chip for assessing beach water safety. Environmental DNA 1(4) 305-315.

Shao, K., Gao, G., Tang, X., Wang, Y., Zhang, L., Qin, B., 2013. Low resilience of the particle-attached bacterial community in response to frequent wind-wave disturbance in freshwater mesocosms. Microbes and Environments 28 450-456.

Stocker, M., Rodriguez-Valentin, J., Pachepsky, Y., Shelton, D., 2016. Spatial and temporal variation of fecal indicator organisms in two creeks in Beltsville, Maryland. Water Quality Research Journal of Canada 51(2) 167-179.

Traister, E., Anisfeld, S.C., 2006. Variability of indicator bacteria at different time scales in the upper Hoosic River watershed. Environmental Science & Technology 40(16) 4990-4995.

VanMensel, D., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2019. Exploring bacterial pathogen community dynamics in freshwater beach sediments: A tale of two lakes. Environmental Microbiology 22 568-583.

Villaescusa, J.A., Jørgensen, S.E., Rochera, C., Velázquez, D., Quesada, A., Camacho, A., 2016. Carbon dynamics modelization and biological community sensitivity to temperature in an oligotrophic freshwater Antarctic lake. Ecological Modelling 319 21-30.

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

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3(2) 180-185.

TEMPORAL AND SPATIAL VARIATION IN MICROBIAL COMMUNITY DYNAMICS IN LARGE FRESHWATER LAKE ECOSYSTEMS: LAURENTIAN GREAT LAKES ERIE AND ST. CLAIR

3.1 Introduction

Freshwater ecosystems are unique in terms of the complexity of the ecosystem process and interactions across trophic levels. In addition, freshwater ecosystems regulate climate, support nutrient cycling, transport water and materials, maintain water quality and complex natural communities (Castello and Macedo, 2016). Freshwater bacterial communities have fundamental roles in nutrient cycling (Fisher et al., 2015), pollutant degradation (Singh and Walker, 2006), among other processes and thus have a critical influence on freshwater river and lake function. Despite these critical roles, the temporal and spatial dynamics of freshwater bacterial communities are not well characterized.

Microbes play fundamental roles in transforming organic carbon and reintroducing it into the food web, thus characterizing temporal and spatial changes in bacterial community composition (BCC) can provide deeper insight into the processes and mechanisms operating in lake ecosystems and ultimately improve our basic knowledge and ability to predict BCC dynamics and function. A growing body of literature suggests that BCC exhibit a remarkable range of temporal and spatial variation. BCC temporal variation can occur on a weekly (Berry et al., 2017), seasonal (Bush et al., 2017), interannual (Siles et al., 2017), and possibly decadal time scales (Fuhrman et al., 2015). Cyclic abiotic factors such as light (Hölker et al., 2015), temperature (Villaescusa et al., 2016) and nutrients characteristics (Lv et al., 2017), as well as biotic factors such as bacteriophages (Yoshida et al., 2018) may contribute to daily, weekly and seasonal cycles, but temporal variation goes beyond such straightforward cyclic relationships. BCC spatial variation also has reported within pounds (fine-scale) (Lear et al., 2014) within a lake (large scale) (Bouzat et al., 2013) and between the large water bodies (between the lakes) (Small et al., 2016), however, some studies also showed minor spatial variation particularly in great lakes (Rozmarynowycz et al., 2019).

The Laurentian Great Lakes (LGLs) in North America differ markedly in their hydraulic residence time, annual lake surface temperatures, ice cover and extent, and primary production levels (Sterner et al., 2017). The LGLs are warming rapidly, and thus are highly susceptible and responsive to any added anthropogenic induced stressors (Adrian et al., 2009). The LGLs serve as a powerful model to study the response of large complex ecosystems to climate change, for example, the timing and duration of ice cover, changes in the timing of the annual spring algal bloom, changes in energy and water fluxes, the expanse and influence of hypoxic zones and the response of biotic communities from microbes to fishes. Lake Erie, the smallest and shallowest of the LGLs, has undergone dramatic swings in water quality over the past century due to nutrient loading (primarily phosphates) from agricultural and urban sources (Davis, 1964). Phosphate removal programs ultimately resulted in significant improvement in the trophic state of Lake Erie (Schindler et al., 2016). However, key ecosystem services such as drinking water (for ~11 million people), important aquatic species habitat, water for the industrial sector and tourism/recreational activities (boating, shipping and fisheries; >$50 billion annually) are currently threatened by frequent cyanobacterial harmful algal blooms (cHABs) and hypoxia (Bullerjahn et al., 2016; Watson et al., 2016). Lake St. Clair, while technically not one of the five LGLs, is also heavily impacted by densely populated urban areas, and because of its location upstream, is also a potential source of nutrient load for Lake Erie (Scavia et al., 2016). Lake St. Clair is very shallow and highly affected by recurrent eutrophication symptoms (Casey, 1998). This study focusses on lakes Erie and St. Clair as two of the arguably most at-risk lakes within the LGL system. We predicted that the BCC in these two highly susceptible freshwater ecosystems would exhibit similar, but an unpredictable temporal and spatial variation.

Despite numerous studies addressing the biogeographical distribution of microbial species, microbial ecologists lack a basic understanding of the characteristic scales of temporal and spatial variation in aquatic BCC (Lindström and Langenheder, 2012); as BCC form the cornerstone of whole freshwater ecosystems. Arguably, this is a key gap in our basic understanding of aquatic bacterial diversity that hinders our ability to develop theories about how microbial mediated function and the stability of those

functions are maintained across space and time. Considering previous studies (Berry et al., 2017; Bush et al., 2017; Siles et al., 2017)'(Rozmarynowycz et al., 2019; Small et al., 2016), we must integrate long-term temporal sampling with large scale spatial sampling to allow not only assessment of change over time and space, but also the potential for the interaction between location and date of sampling. To address this knowledge gap, we sampled bi-weekly six recreational beaches locations in our two study lakes (Lake Erie and St. Clair) from June 2016 to August 2017 to characterize the BCC. We used 16S rRNA metabarcoding via next-generation sequencing (NGS) to insure accurate and complete BCC characterization. We hypothesized to detect significant temporal (bi-weekly, monthly and seasonal) and spatial (sampling location and lake) variations in the freshwater BCC. We predicted to observe strong temporal than spatial variation in the freshwater BCC, more specifically, we expected to observe more diversity and similarity in the BCC of two summers than other seasons. We expected to observe more diversity of the BCC in Lake Erie than St. Clair as Lake Erie is fed by Lake St. Clair. We also expected that environmental parameter such as water temperature has more influence on the BCC than other variables. The outcome of this study will increase our basic understanding of how freshwater BCC changes in different scales of time and space which is critical for monitoring the ecological service of BCC and aquatic ecosystem health.

3.2 Materials and methods

3.2.1 Study sites and sample collection

Freshwater samples were collected bi-weekly over 15 months from June 2016 to August 2017 from shorelines at six locations, including four locations from Lake Erie (Cedar Beach (CB), Colchester Harbour (CH) Beach, Holiday Beach (HB), and Point Pelee (PP) Beach) and 2 locations from Lake St. Clair (Lakeview Park (LP) Beach and Sand Point (SP) Beach) at Windsor-Essex County (Windsor, Ontario, Canada) (Figure 3.1). LP and SP are in urban areas, while CB, CH, HB and PP are located in agricultural areas. LP is near an urban tributary (the Belle River joins Lake St. Clair at LP beach), while SP and HB are near the inlet and outlet of the Detroit River respectively. CH, HB and SP are high-energy water movement locations, while LP has

restricted water flow due to adjacent artificial piers and thus is a low-energy site with low wave energy (VanMensel et al., 2020). Two water samples (each 250 mL) were collected at 0.5 m depth at each location (near-shore). In total, we collected 60 samples (2 replicates * 2 weeks/month * 15 months= 60 samples) at each location. Water samples were transported to the laboratory on ice and were filtered using 0.2 μm polycarbonate membranes (Millipore, USA), and the filter immediately stored at -20 °C until DNA extractions were performed. Water temperature was measured at the time of sample collection at each location. Other environmental variables, such as precipitation and daylight hours, were collected from Environment Canada (http://climate.weather.gc.ca/historical_data/search_historic_data_e.html) according to the sampling date. We also collected environmental data from local weather stations (Windsor Bell River, Windsor South and Windsor Riverside) to cross-check the relevancy of environmental data to local beach conditions.
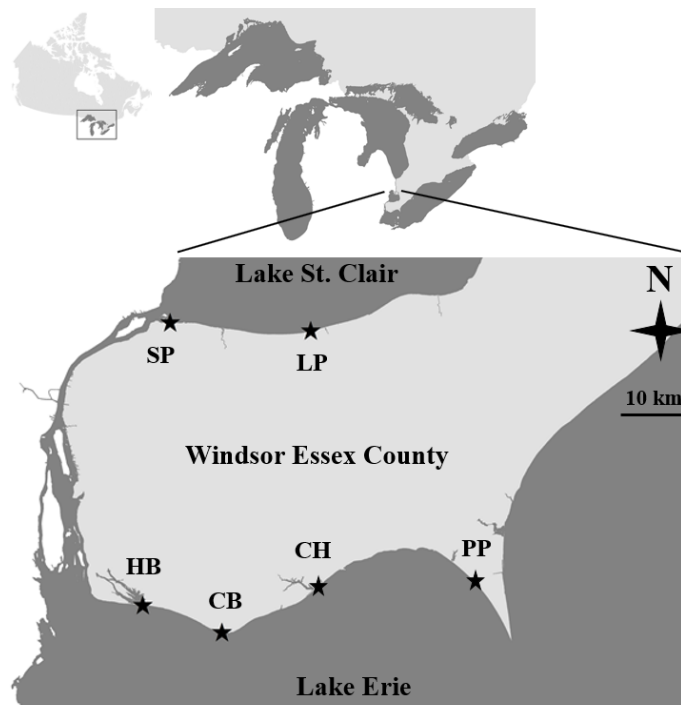
Figure 3.1. The sampling sites used for bacterial community composition in Lake Erie (Cedar Beach; CB, Colchester Harbour Beach; CH; Holiday Beach; HB and Point Pelee Beach; PP), and Lake St. Clair (Lakeview Park Beach; LP and Sand Point Beach; SP) in Windsor-Essex County.

3.2.2   DNA extraction, PCR, library preparation and parallel sequencing

DNA was extracted following Shahraki et al (Shahraki et al., 2019) by using sucrose lysis buffer (400 mM NaCl, 750 mM sucrose, 20 mM ethylenediaminetetraacetic acid (EDTA), 50 mM Tris-HCl pH 9.0) in combination with bead-beating (0.1 mm beads) and magnetic bead DNA purification. Purified genomic DNA was suspended in 50 µL TE buffer and stored at -80 °C. The extracted DNA was used as a template to amplify the V5-V6 region (~350 bp) of the 16S rRNA gene using V5F (acctgcctgccg-ATTAGATACCCNGGTAG) and V6R (acgccaccgagc-CGACAGCCATGCANCACCT) primers (He et al., 2017). Then sample barcode and adaptor sequences were ligated to each PCR product by a second, ligation, PCR(He et al., 2017). Second-round PCR products were pooled and purified using the QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). The concentration of purified PCR product mixture (library) was measured using an Agilent 2100 Bioanalyzer with a High Sensitive DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library was then diluted to 60 pmol/L and sequenced on an Ion PGM™ System (Thermo Fisher Scientific, Burlington, ON, Canada).

3.2.3   Bioinformatics and statistical analyses

Sequence handling: The raw sequence data was de-multiplexed, quality filtered and trimmed of the adaptor, barcode and primer sequences using the Quantitative Insights into Microbial Ecology (QIIME V. 1.9.1) bioinformatics pipeline (Caporaso et al., 2010). A minimum quality score of Q=20 and base-pair (bp) length cut-off of 200 bp was selected for quality assurance. Chimeras were removed using ChimeraSlayer in QIIME. The operational taxonomic units (OTUs) were assembled based on sequence similarity (97%) among the sequence reads and then taxonomically assigned using The Basic Local Alignment Search Tool (BLAST) against Greengenes 16S rRNA database version 13_8 as a reference data file (Edgar, 2010). The representative sequence for each OTU was selected using the most abundant method for assigning taxonomy in the Ribosomal Database Project (RDP) Classifier program with a minimum 80% confidence level (Wang et al., 2007). After removing the single and double read OTUs, the OTU table was rarefied to 2000 quality passed sequences for each sample to

calculate alpha diversity. Original OTU table (non-rarefied) was used to calculate relative abundance. We defined an OTU as "abundant" when it had a relative abundance above 1% of the community, "moderate" when the relative abundance was between 0.1-0.99% and "rare" when the abundance was below 0.1% (Logares et al., 2014).

Global spatial and temporal effects: We used nested ANOVA in R environment (version 3.1.1) (Team, 2013) to determine whether or not field replicates at each sampling event were significantly different in their BCC. Alpha diversity indexes (Chao1 and Shannon) and the first (PCo1) and second (PCo2) principal coordinates from the principal coordinates analysis (PCoA) across all samples (including replicates) were used as a dependent variable in the nested ANOVA (replicated were nested to week; week 1 and week 2 in each month) to test specifically for a replicate effect.

We used a generalized linear mixed model (GLMM) with Maximum Likelihood (ML) method implemented in the R package lme4 (Bates et al., 2012) to test for sampling location, lake (Erie and St Clair), week and month (and their interactions) effects on Chao1, Shannon, PCo1 and PCo2 as dependent variables (note, we found no significant effect of replicate, hence we dropped replicate from subsequent analyses). We chose the Chao1 and Shannon indexes of alpha diversity, and PCo1 and PCo2 as simple measures of BCC variation for our preliminary analysis. Chao1 and Shannon were calculated using the rarified OTU table. The Bray–Curtis dissimilarity matrix of each sample site over 15 months was calculated using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK) and after PCoA, we selected PCo1 (16.1%) and PCo2 (8.1%), which represented the most variation in the BCC PCoA, but did not incorporate PCo3 (6.5%), PCo4 (3.5%), PCo5 (3.4), PCo6 (2.2%) and PCo7 (1.5%) in the GLMM model as they explained a low level of variance. For each dependent variable, we run a global GLMM model by nesting sampling locations (6 locations) within lake (2 lakes) and sampling weeks (2 weeks/month) within sampling month (15 months) to determine the effect size of month and lake on the BCCs. We used replicates as a random factor and sampling month and lake as fixed factors. Model assumptions were assessed by examining the distribution of residuals and plotting

fitted values against residuals. Then we used ANOVA implemented in the car package (Fox et al., 2012) to evaluate the significance of fixed effects in the models using R (version 3.1.1) (Team, 2013). To evaluate the effect of the sampling location by month interaction, we ran GLMM models on data from each lake separately. In these models, we considered month (weeks nested within 15 sampling months) and sampling locations as a fixed factor and replicates as a random factor. We used diversity indexes as simple indicators of spatial and temporal variation of the BCCs as input for GLMM model, but we also measured the impact of spatial (sampling locations) and temporal (sampling months) variation on the relative abundance of each OTU (n=2100) using the Kruskal–Wallis one-way analysis of variance using MATLAB software.

Spatial variation: To characterize variation in BCC between the two sampled lakes as well as variation among the 6 sampling locations, we used four different approaches. They were: i) permutational multivariate analysis of variance (PERMANOVA) with 9999 permutations using the Vegan package (Dixon, 2003) in R environment (version 3.1.1) (Team, 2013) to compare the Bray–Curtis dissimilarity of the BCCs, ii) PCoA of the Bray–Curtis similarity matrix to visualize the pattern of the BCC variation using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK), iii) one-way ANOVA to compare the mean of diversity indexes and PCo1 and PCo2 of the BCCs using SPSS version 19 (SPSS Inc, Chicago, Illinois), and iv) linear discriminant analysis (LDA) using the LEfSe method (http://huttenhower.sph.harvard.edu/lefse/) (Segata et al., 2011) to compare the relative abundance of taxa (class level) in the bacterial communities.

Temporal variation: As we sampled two summers (2016 and 2017) but did not replicate sample any other seasons, we were not able to explicitly test for season effects. However, we applied hierarchical agglomerative clustering on the Bray–Curtis similarity matrix of the BCCs (top 300 highly abundant OTUs) using the group average method in PAST (Hammer et al., 2001) to explore the possibility of seasonal clustering in terms of BCC. Once we identified clear clusters, we tested for differences in BCC among the clusters and the 15 months of sampling following the same approach as we used for the spatial variation (above). Specifically, we used: i) PERMANOVA with 9999 permutations in the Vegan package (Dixon, 2003) in R environment (version

3.1.1) (Team, 2013) then compare the Bray–Curtis dissimilarity of the BCCs in the identified clusters and the 15 months of sampling, ii) SIMPER analysis on the Bray–Curtis dissimilarity matrix to compare the overall dissimilarity among clusters and the 15 months of sampling in the Vegan package (Dixon, 2003), iii) PCoA of the Bray–Curtis similarity matrix to visualize the pattern of BCCs among the clusters and the 15 months of sampling using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK), iv) one-way ANOVA to compare the mean of diversity indexes and PCo1 and PCo2 of the BCCs from the clusters using SPSS version 19 (SPSS Inc, Chicago, Illinois) and v) LDA (http://huttenhower.sph.harvard.edu/lefse/) (Segata et al., 2011) on the relative abundance of taxa (class level). Plots and graphs were generated using ggplot2 package (Wickham, 2011) in R (version 3.1.1) (Team, 2013) and Origin Pro 2019.

Environmental effects: To evaluate the relationship between the BCC at the various beaches and dates and environmental factors, we applied a RELATE analysis (Spearman's p correlation coefficient) on the Bray Curtis similarity matrix calculated from whole data sets collected bi-weekly from 6 different locations (beaches) located in two lakes over 15 months and the matrix of Euclidean distances calculated from normalized environmental data (daylight hours, precipitation and water temperature) as the environmental matrix. A distance-based linear model (distLM) was used for analyzing the relationship between the Bray-Curtis similarity matrix of the BCC of 15 sampling months and the environmental variables using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK).

3.3 Results

Sequence library: After quality control, 5.1 million Ion Torrent sequence reads remained across all 6 sample sites and 15 months. Each sample (replicate) had between 2,102-8,509 reads, with an average of 4,789 reads. In total, 27,643 OTUs were detected. After removing singleton and doubleton sequence reads, as well as OTUs with ≤ 20 reads from the data set, 2,100 OTUs were included in this study. The OTU table was rarefied to 2000 reads/sample.

Global spatial and temporal effect: Replicate had no significant effect on the variables (Chao1; df=1, F=0.01, p=0.75, Shannon; df=1, F=0.092, p=0.72, PCo1; df=1,

F=0.001, p=0.95 and PCo2; df=1, F=0.92, p=0.35), thus we combined sequence read data of the two replicates for each week at each location to increase the read depth for all further statistical analyses. Based on our global GLMM, we found that lake (as a broad spatial factor) had significant effects (p<0.05) only on alpha diversity indexes, however, month (as a broad temporal factor) had significant effects (p<0.05) on alpha diversity indexes, PCo1 and PCo2 (Table 2.1). In the lake-specific models (two models); sampling location had a significant effect (p<0.05) only on alpha diversity indexes but, month had a significant impact on alpha diversity indexes, PCo1 and PCo2 (Table 2.1). The interactions of sampling location with month also had significant (p<0.05) effects on the variables in two lake-specific models (Table 3.1).

Table 3.1. Results of GLMM analysis of bacterial community variation temporally and spatially. Dependent variables included alpha diversity indexes and Bray–Curtis dissimilarity principal coordinate analysis axes (PCo1 and PCo2). Degrees of freedom, F value and p values are shown (significant p values are highlighted).

Factors	df	Chao1		Shannon		PCo1		PCo2	
		F value	p value	F value	p value	F value	P value	F value	p value
Full model									
Lake (sampling location)	5	4.8	**0.000**	2.7	**0.02**	0.002	1.000	0.000	1.000
Month (week)	29	7.11	**0.000**	6.9	**0.000**	7.4	**0.000**	4.8	**0.000**
R^2		0.4		0.35		0.37		0.27	
Lake Erie									
Sampling location	3	6.5	**0.000**	5.8	**0.001**	0.006	0.99	0.000	1.000
Month	14	21.9	**0.000**	24.6	**0.000**	21.8	**0.000**	61.21	**0.000**
Month (week)	15	4.4	**0.000**	5.7	**0.000**	4.8	**0.000**	3.5	**0.000**
Sampling location x Month (week)	42	3.6	**0.002**	3.3	**0.003**	2.1	**0.005**	2.19	**0.001**
R^2		0.83		0.79		0.21		0.18	
Lake St. Clair									
Sampling location	1	10.3	**0.000**	9.3	**0.000**	0.003	0.95	0.000	1.000
Month	14	24.2	**0.000**	14.8	**0.000**	48.8	**0.000**	25.2	**0.000**
Month (week)	14	2.5	**0.001**	2.4	**0.003**	12.4	**0.000**	9.1	**0.000**
Sampling location x Month (week)	14	2.1	**0.01**	7.1	**0.000**	6.1	**0.000**	5.9	**0.000**
R^2		0.29		0.26		0.2		0.12	
Sampling location; 6 public beaches, Month; 15 sampling months, Week; bi-weekly sampling/month, Lake; two lakes. Parenthesis indicates nesting the variables and "x" indicates interaction effects.									

As our lake specific models showed that location (6 sampling locations) and month (15 sampling months) had significant effects on diversity indexes (Table 3.1), thus we also tested for the effects of those two main factors on the relative abundance of OTUs across all sites combined. Our Kruskal–Wallis analyses showed that out of 2,100 OTUs, the relative abundance of 336 (16%) of the OTUs were significantly affected by location (6 sampling locations), while the relative abundance of 1453 (69%) of the

OTUs were significantly affected by month (15 sampling months) (Figure 3.2). The interaction of location and month had a significant effect on the relative abundance of 311 (14%) OTUs (Figure 3.2). It is important to note that significance was not corrected for multiple simultaneous comparisons; however, the goal of this analysis was to show the pattern of effects on OTU relative abundance, highlighting the dominance of temporal effects relative to spatial and interaction effects (Figure 3.2).
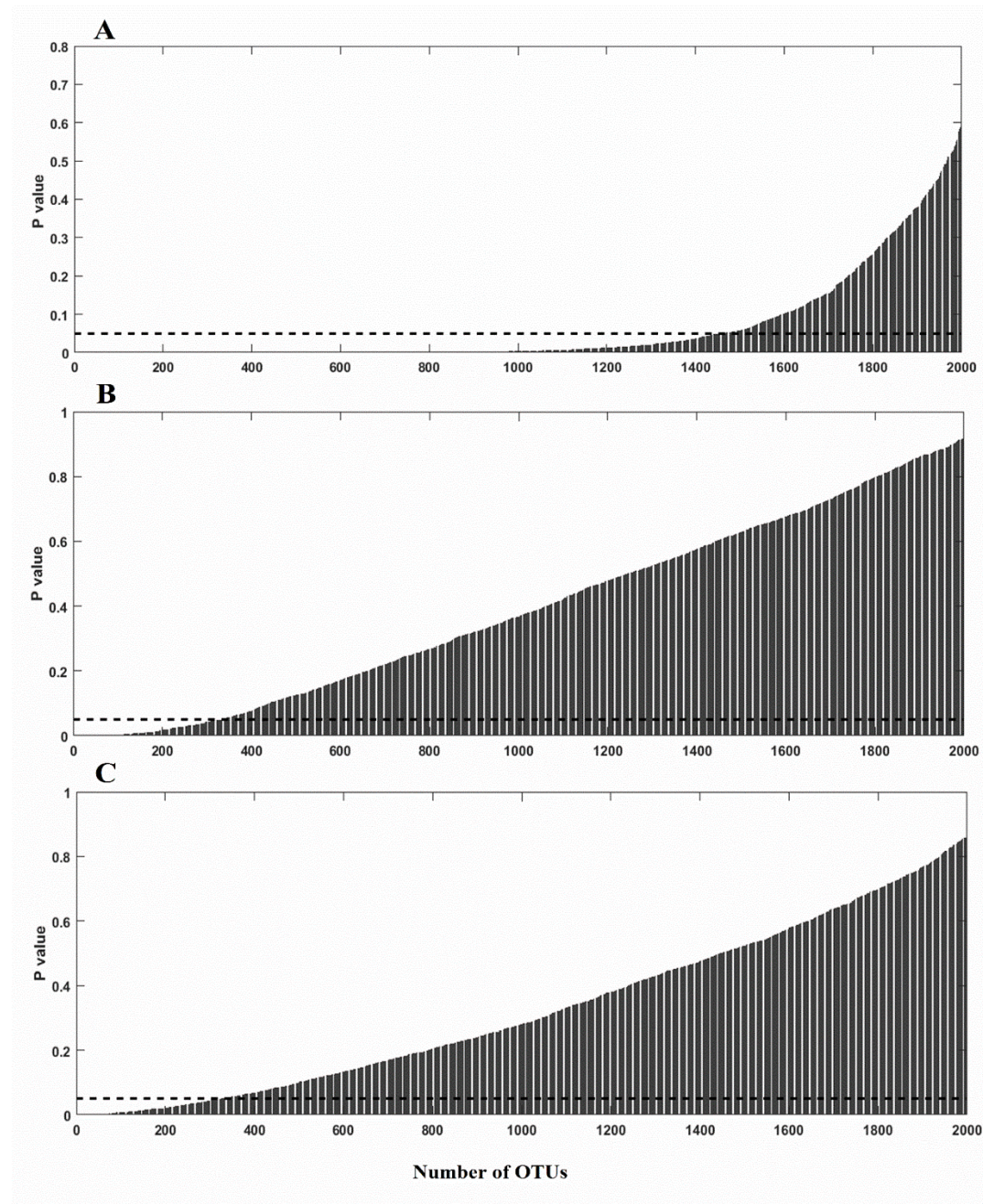
Figure 3.2. Histograms showing the effect (Kruskal–Wallis P-value) of location (6 sampling locations) and month (15 sampling months) and their interactions on the relative abundance of 2,100 bacterial OTUs sampled at 6 sites over 15 months. panel A; temporal effect (month), panel B; spatial effect (location) and panel C; the interaction of spatial and temporal effects (Location x Month). Uncorrected p values are shown on the in Y-axis and p=0.05 was used as cut-off of the significant effect (dashed line in each plot).

3.3.1 Bi-weekly diversity variation of the BCCs

We chose alpha diversity indexes and Bray–Curtis dissimilarity PCo1 and PCo2 to evaluate the spatial and temporal variation of the BCC across 6 different locations over 15 months of bi-weekly sampling. PCo1 (16.1%) and PCo2 (8.1%), which represented the most variation in the BCC PCoA, varied substantially over the 15 month sampling period (Figure 3.3), while PCo3 (6.5%), PCo4 (3.5%), PCo5 (3.4), PCo6 (2.2%) and PCo7 (1.5%) represented only minor levels of variation in the bi-weekly BCCs and are not shown. Inspection of the bi-weekly plots of alpha diversity (Chao1; Figure 3.3) and Shannon (Appendix B; Figure S3.1) also shows substantial temporal variation over the course of the 15-month sampling period. Shannon results were provided in the supplementary data. Although there is some bi-weekly variation in all three parameters (Figure 3.3 and Appendix B; Figure S3.1), the majority of variation is seen at longer temporal scales, consistent with our statistical analyses that showed, overall, sampling week (within the month) had no significant effect (p>0.05) on the BCCs (Table 3.1).
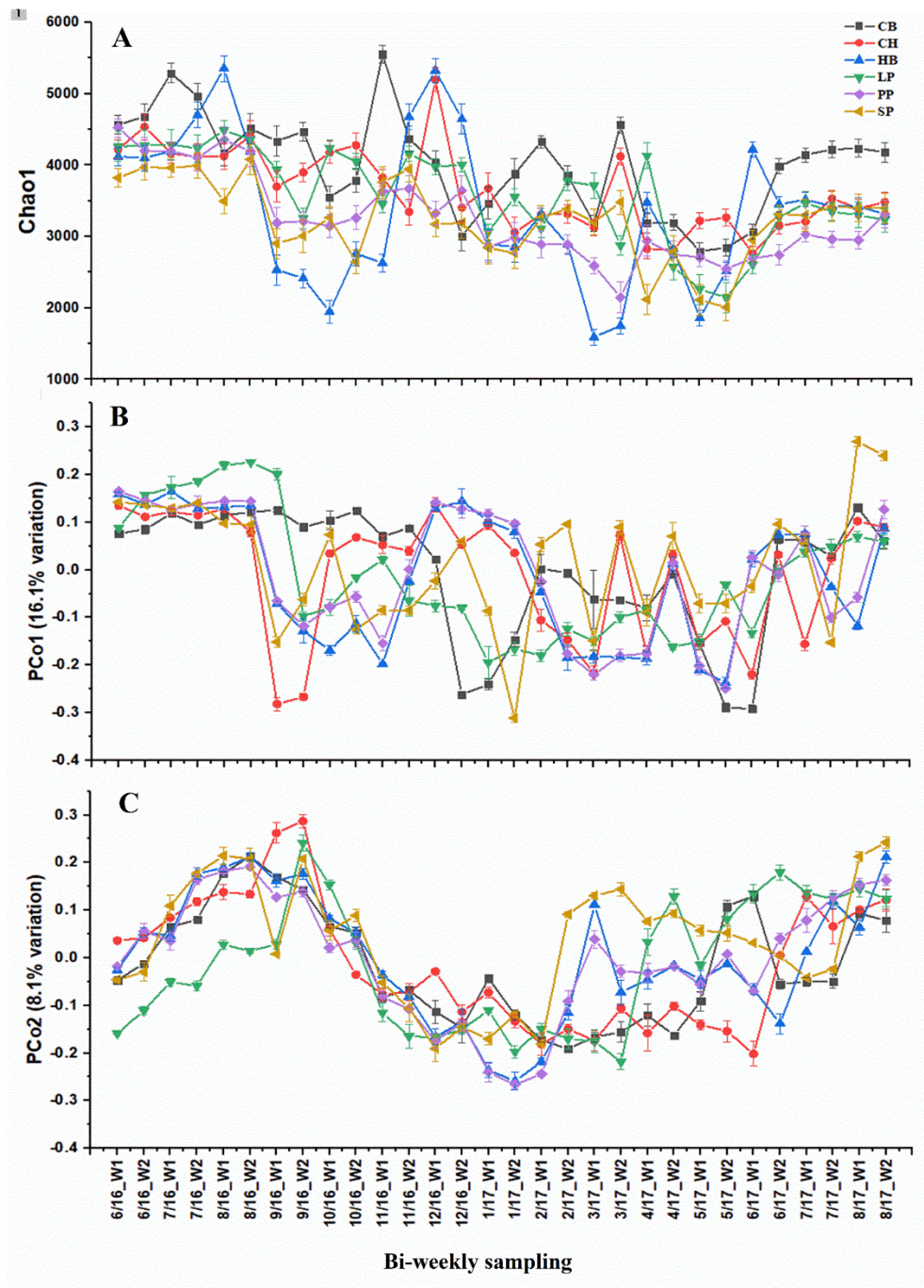
Figure 3.3. Line plots showing mean bi-weekly variation in Chao1 index (panel A) and Bray–Curtis dissimilarity PCo1 (panel B) and PCo2 (panel C) for the six different sampling locations (CB, CH, HB, LP, PP and SP) over 15 sampling months (June 2016 - August 2017). Error bars show standard deviation. In X-axis; numbers indicate the month in years, W shows weeks 1 and 2, 16 and 17 also indicate 2016 and 2017 years respectively.

3.3.2 Spatial variation

Broad spatial variation: There was no significant difference in the Bray–Curtis dissimilarity matrix for the BCCs of Lake Erie and Lake St. Clair across the full sampling period (PERMANOVA; df=1, F=2.064, p=0.07) (Appendix B; Figure S3.2, panel A). Similarly, we found no significant variation in the mean of the Chao1 (F=1.2), PCo1 (F=0.42) and PCo2 (F=0.92) values between two lakes by one-way ANOVA (df=1, p>0.05). Out of the 30 detected classes of bacteria across all samples, the relative abundance of only 5 of the classes including classes of *Acidobacteria*, *Acidimicrobiia* and *Thermoleophilia* (belong to *Actinobacteria*), *Chloroplast* (belong to *Cyanobacteria*) and *Saprospirae* (belong to *Bacteroidetes*).were significantly higher (LDA; p<0.05) in the BCCs of Lake Erie in compare to Lake St. Clair (Appendix B; Figure S3.3).

Spatial variation among different locations: There was no significant difference in the Bray–Curtis dissimilarity matrix for the BCCs of the six sampling locations (PERMANOVA; df=5, F=1.06, p=0.34) (Appendix B; Figure S3.2, panel B). One-way ANOVA showed a significant effect of sampling locations on the Chao1 index (F=5.32, p<0.001). Tukey post-hoc test revealed that only the mean of Chao1 index in CB was significantly higher (p<0.05) than HB, PP and SP but not from CH and LP. Using one-way ANOVA, there was no significant effect of sampling locations on PCo1 (df=5, F=0.644, p=0.66) and PCo2 (df=5, F=0.38, p=0.85). Out of the 2,100 OTUs, 336 OTUs (4 highly abundant, 21 moderately abundant and 311 rare OTUs) showed significant variation among the six sample locations (Figure 3.2). We only identified 1-5 classes of bacteria with significant divergence for some of the pairwise comparisons with no significant variation (p>0.05) at the class level between the BCC of HB and PP (Appendix B; Table S3.1). We identified a maximum 5 classes of bacteria with significant variation in their relative abundance between CB and CH, HB, LP, and SP; between CH and SP, between HB and SP and between LP and SP (Appendix B; Table S3.1). We only found a common pattern in the relative abundance of *Bacilli* across all sampling locations which was significantly (p<0.05) lower at CB (19%) compared to the other sampling locations (CH; 25.5%, HB; 29.23%, LP; 28.17%, PP; 28.39% and SP;24%).

3.3.3   Temporal variation

2.2.4.1 Broad temporal variation

BCCs variation: As week had no significant effect on the BCC (Table 3.1), we combined bi-weekly data from each location within each month. We thus had a total of 90 samples (15 months x 6 locations) for our cluster analysis. UPGMA clustering showed five major clusters diverging at between 50-60% similarity based on the Bray–Curtis similarity index (Figure 3.4). The BCCs of summer 2016 (June, July and August) were grouped in cluster 1. The BCCs of December (2016) and January (2017) were grouped as cluster 2. The BCCs of five months including February, March, April, May and June 2017 clustered together as cluster 3. The BCCs of July and August (summer 2017) were grouped as cluster 4 and the BCCs of fall 2016 (September, October and November) grouped into cluster 5 (Figure 3.4).
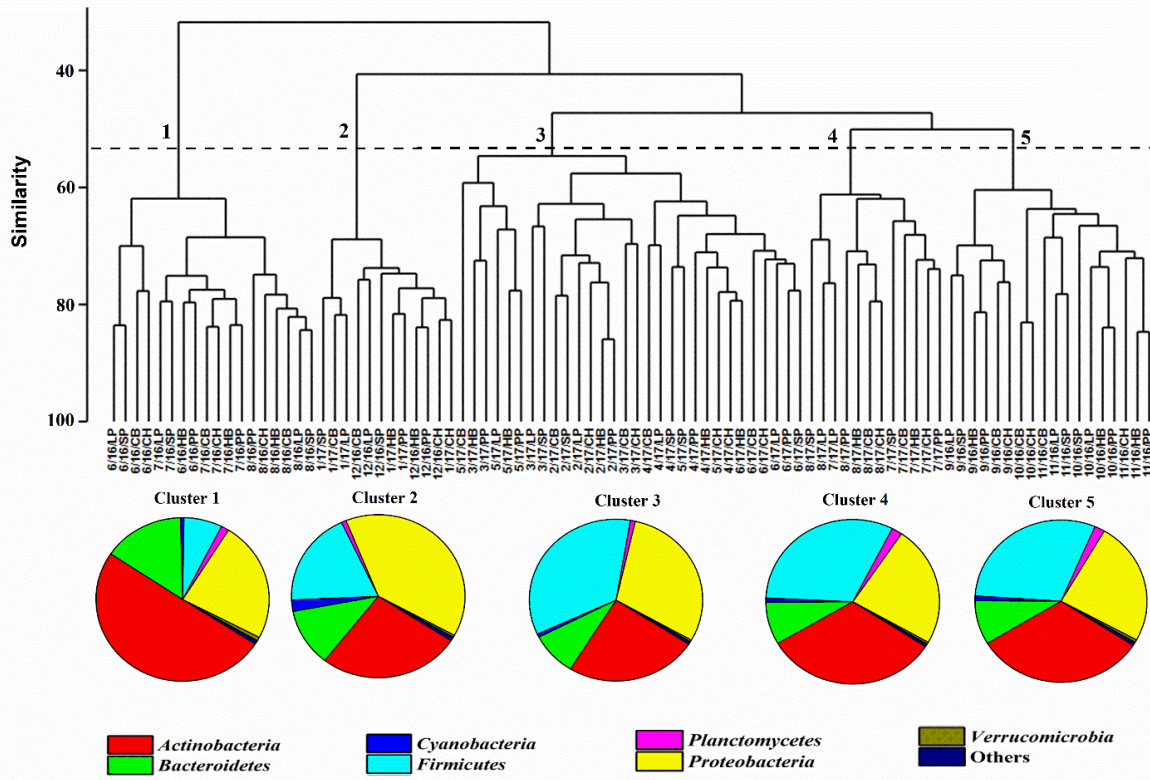
Figure 3.4. UPGMA tree showing cluster results based on Bray-Curtis similarity of the 15 monthly BCCs sampling across six sample sites (X-axis: 15 sampling months collected from 6 different location; CB, CH, HB, LP, PP and SP over 2016 and 2017. 2016 and 2017 showed as 16 and 17 respectively in the figure). The BCCs grouped into five broad temporal community clusters, all clusters were above 50- 60% Bray–Curtis similarity (Y-axis) and show considerable taxonomic divergence (pie charts showing phyla level composition). For this analysis, the Bray–Curtis similarity matrix was generated using the 300 most abundant OTUs.

There was significant variation among the BCCs of 5 clusters (PERMANOVA test; df=4, F=9.57, p=0.0001). Pairwise comparison of the BCCs of clusters 1-5 also showed significant variation (p<0.05) between the BCCs of all pairwise comparisons. The overall average dissimilarity of the BCCs of 5 broad clusters was 51% using SIMPER analysis (Appendix B; Table S3.2).

Diversity variation among clusters: The clusters differed significantly in Chao1 (df=4, F=27.42, p<0.0001), PCo1 (df=4, F=28.2, p<0.0001) and PCo2 (df=4, F=27.9, p<0.0001) using one-way ANOVA. Tukey post-hoc analysis showing that Chao1 and

PCo1 of cluster 1 (June, July and August; 2016) was significantly ($p<0.05$) higher than all 4 other clusters. PCo2 of cluster 2 (December and January) was significantly ($p<0.05$) lower than all 4 other clusters. More interestingly, the mean of Chao1 for cluster 3 (February, March, April, May and June; 2017) was significantly lower than all other clusters and the mean of PCo1 also was the lowest among all clusters but significantly lower than cluster 1 and 4 (Appendix B; Table S3.3). Variation in Shannon and PCo3-5 data are presented in supplementary data (Appendix B; Figure S3.4).

OTUs and taxonomic variation: At the OTU level, clusters 1, 2, 3, 4 and 5 had 7, 12, 11, 9 and 11 highly abundant (relative abundance > 1%) OTUs, respectively. Only four OTUs (2, 3, 4 and 6) were common among all 5 clusters. Only four of those OTUs; OTUs 2 and 3 (phylum *Actinobacteria* and family ACK-M1), OTU4 (phylum *Firmicutes* and family *Exiguobacteraceae*) and OTU6 (phylum *Proteobacteria* and family *Comamonadaceae*) were abundant across all clusters. The relative abundance of many OTUs significantly varied among clusters, as expected given that the clusters were defined based on variation in BCC. For example, the relative abundance of OTU2 (family ACK-M1) was significantly higher in the BCC of cluster 1 (9.5%) than all other clusters (2.5-3.8%). Out of 2100 OTUs; 75, 90, 79, 61 and 79 moderately abundant OTUs were detected in the BCCs of clusters 1, 2, 3, 4 and 5 respectively. Among the moderately abundant OTUs, only 11 OTUs (18, 20, 22, 27, 28, 37, 38, 43, 44, 52 and 63) were common among all 5 clusters.

While the BCCs of the clusters (1-5) were dominated by four phyla, the relative proportions of these phyla varied substantially (Figure 3.4). *Actinobacteria* was the most common (~50%) versus *Firmicutes* (7.16%) as the less common taxon in the BCC of cluster 1. In the BCC of clusters 4 and 5, *Actinobacteria* (~32% in both clusters 4 and 5) was dominant but *Firmicutes* (~32% in cluster 4 and ~30% in cluster 5) became the second most common taxon. In the BCC of cluster 2, *Proteobacteria* (~39%) was the most common taxon and *Firmicutes* (~19%) became the third most common phylum after *Actinobacteria* (~26%). More interestingly, *Cyanobacteria* levels were elevated in the BCC of cluster 2 (2.25%) compared to the other clusters (0.75±0.35%). The BCC of cluster 3 was enriched for *Firmicutes* (~35%) as the

dominant phylum and by *Proteobacteria* (~30%) as the second most common phyla. We observed substantial variation in the relative abundance of 30 classes of bacteria. For example, the relative abundance of *Bacilli* was significantly lower in the BCC of cluster 1 compared to all other clusters, while some classes such as *Actinobacteria* and *Thermoleophilia* (belonging to *Actinobacteria*) and *Cytophagia* and *Saprospirae* (belonging to *Bacteroidetes*) conversely had significantly higher relative abundance in the BCC of cluster 1 relative to all the other clusters (Appendix B; Table S3.4).

2.2.4.2 Monthly temporal variation

BCCs variation: As lake and sampling location had effects only on alpha diversity indexes of the BCC and sampling week had no significant effects (Table 3.1), we combined the data from the six sampling locations and the two within-month sampling dates for our temporal analysis. Thus, our temporal analyses focus on variation among the 15 months of sampling. There was a statistically significant effect of sampling month (df= 14, F=6.1 and p=0.0001) on the Bray–Curtis dissimilarity matrix of the BCCs based on PERMANOVA. The pairwise comparison of the BCCs of the months showed significant differences for most comparisons (Table 3.2).

Table 3.2. Pairwise dissimilarity (%; SIMPER) (above the diagonal) and PERMANOVA significance probabilities (below the diagonal) for the BCCs across the 15 months of sampling (numbers indicate months, 16 and 17 show 2016 and 2017 respectively. P values were adjusted using a Bonferroni correction for multiple comparisons.

Month /years	6/16	7/16	8/16	9/16	10/16	11/16	12/16	1/17	2/17	3/17	4/17	5/17	6/17	7/17	8/17
6/16		33.35	42.8	53.8	46.51	47.36	47.62	56.12	53.97	57.93	43.06	55	51.37	50.71	49.21
7/16	0.002		32.35	47.46	43.43	47.56	50.27	57.46	57.43	60.96	46.88	58.09	54.8	50.47	47.06
8/16	0.0022	0.0023		41.22	43.75	51.61	53.02	60.2	61.33	64.04	51.43	60.99	58.79	52.96	47.21
9/16	0.0026	0.0028	0.0013		41.33	47.22	52.45	55.33	55.36	55.3	50.17	54.85	53.32	49.02	48.02
10/16	0.002	0.0031	0.0013	0.0048		39.73	45.28	51.53	49.36	50.95	43.38	51.3	48.38	48.47	47.01
11/16	0.0028	0.0022	0.0025	0.0023	0.011		30.32	46.8	41.21	47.73	41.8	50.38	49.83	52.32	51.31
12/16	0.0024	0.002	0.0018	0.002	0.0018	0.0884		45.47	41.07	48.24	45.95	54.77	52.82	55.09	54.68
1/17	0.0021	0.0019	0.0031	0.0025	0.0024	0.0024	0.032		33.37	50.52	50.5	57.37	57.44	60.72	60.49
2/17	0.0018	0.0021	0.0021	0.0026	0.0028	0.002	0.011	0.087		34.33	47.31	54.69	53.79	58.25	59.35
3/17	0.0023	0.0017	0.0021	0.0028	0.0017	0.004	0.02	0.04	0.04		29.85	32.31	50.41	53.69	57.7
4/17	0.0025	0.0035	0.002	0.0023	0.0024	0.0027	0.0027	0.0024	0.0021	0.092		30.82	26.3	50.22	50.41
5/17	0.002	0.0015	0.0016	0.0024	0.0022	0.0029	0.0025	0.0027	0.0021	0.075	0.063		26.24	48.86	52.69
6/17	0.0024	0.0021	0.003	0.0026	0.0019	0.002	0.003	0.0028	0.0024	0.0399	0.062	0.058		39.16	46.87
7/17	0.0027	0.0024	0.0029	0.0018	0.0024	0.0027	0.002	0.0022	0.0024	0.0019	0.0014	0.0017	0.036		28.2
8/17	0.0024	0.0019	0.0029	0.0022	0.027	0.0017	0.0023	0.0027	0.0026	0.0035	0.0034	0.0103	0.039	0.262	

Diversity variation: We observed significant variation in Chao1 (df=14, F=8.97, p<0.0001), PCo1 (df=14, F=5.7, p<0.0001) and PCo2 (df=14, F=3.5, p=0.001) indexes among the 15 months of sampling using one-way ANOVA. Pairwise post-hoc tests showed that 30% of the comparisons among sampling months were statistically significant (p<0.05) for Chao1 among 15 months of sampling, while 10% of the pairwise post-hoc test comparisons of PCo1 and only 3% of the post-hoc test comparisons of PCo2 showed significant divergence (p<0.05) (Figure 3.5, Appendix B; Table S3.5).
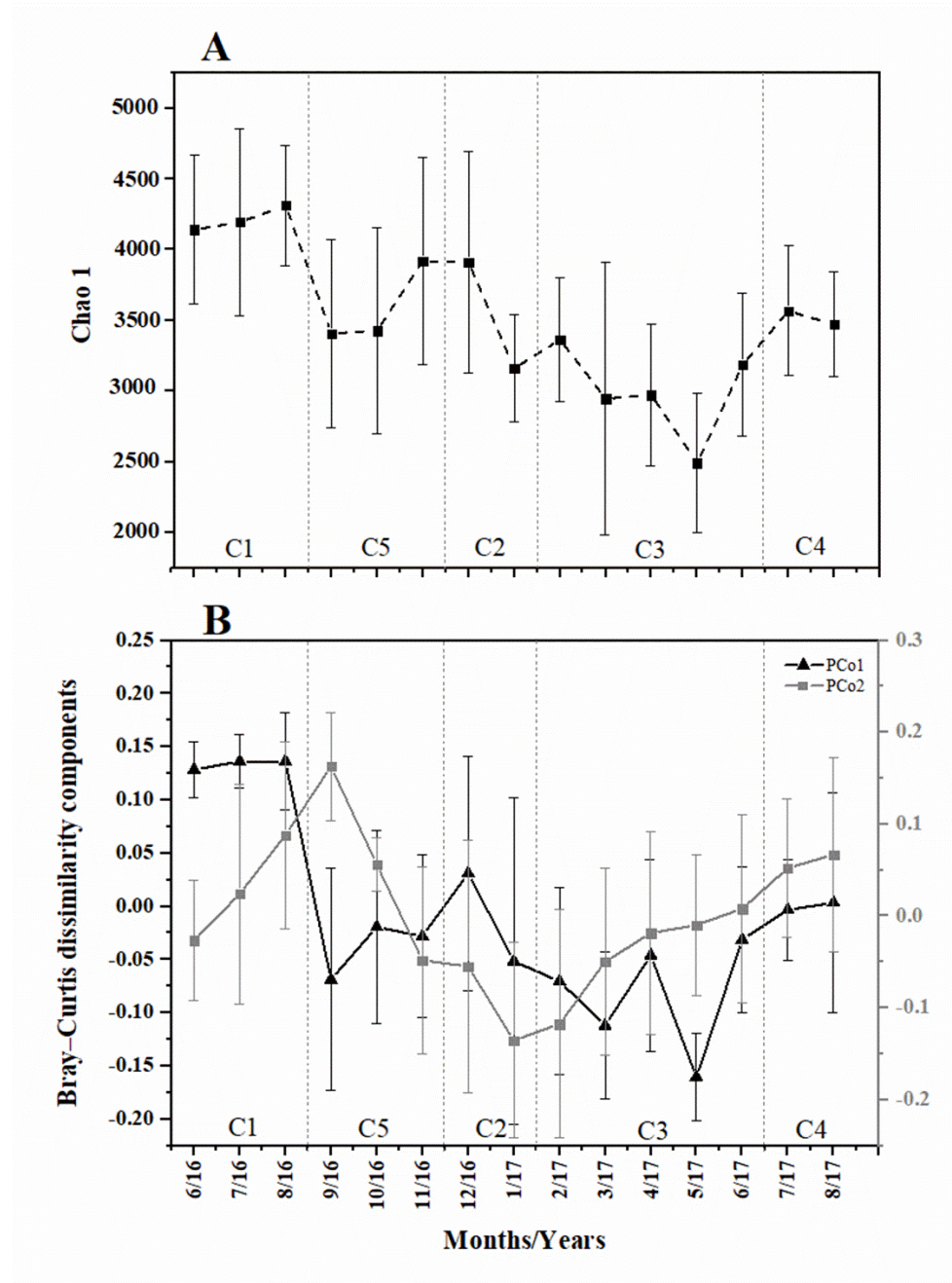
Figure 3.5. Line plots of monthly changes of Chao1 (top panel) and Bray–Curtis dissimilarity components (bottom panel) of 6 different locations over 15 months of sampling. C1-5: Cluster 1-5 is based on figure 4. PCo1 and 2 explained 22.9% and 14% variances of the BCC variation over 15 months respectively. X-axis: numbers indicate sampling months, 16 and 17 show sampling years 2016 and 2017 respectively.

OTUs and taxa variation: Out of 2100 OTUs, the relative abundance of 1453 (69%) OTUs changed significantly (Figure 3.2) across the months. The BCCs of July (2016) had the lowest number (6 OTUs) of highly abundant OTUs while the BCCs of February (2017) and June (2017) had the highest number of highly abundant OTUs (12 OTUs) among the 15 months of sampling. OTUs 2, 3 (family ACK-M1), 4 (family *Exiguobacteraceae*) and 6 (family *Comamonadaceae*) were highly abundant across all 15 months of sampling, but the relative abundance of others with high relative abundance dropped to ≤1% in some months (Figure 3.6). Some highly abundant OTUs showed highly variable patterns over 15 months of sampling. For example, the relative abundance of OTU2 (family ACK-M1) was 8.2%, 10.8% and 9.5% in June, July and August (2016) respectively but dropped to between 2.0% and 5.7% across all other months. In contrast, the relative abundance of OTU13 (phylum *Bacteroidetes* and family *Cyclobacteriaceae*) was 0.02-4.4% in the BCCs of June 2016 – June 2017 but increased up to 10.7% and 7.2% in the BCCs of July and August (2017) respectively (Figure 3.6, panel A). Many of even the highly abundant taxa exhibited unpredictable variation across the study period, highlighting the chaotic nature of BCCs.

Across the 15 months of sampling, 60-90 OTUs had moderate relative abundance (0.999-0.01%), but only 2 of those OTUs (OTU37; phylum *Actinobacteria*, family ACK-M1 and OTU42; phylum *Bacteroidetes,* family *Chitinophagaceae*) were consistently in the moderate relative abundance range, as the abundance of most of the OTUs with moderate abundance dropped to <0.01% (rare abundance) while a few increased to above 1% (abundant OTUs). Interestingly, the relative abundance of family *Flavobacteriaceae* belonging to *Bacteroidetes* (OTU50) and *Oscillatoriaceae* belonging to *Cyanobacteria* (OTUs 65 and 90) were elevated in the January BCCs relative to all other months (Figure 3.6, panel B).
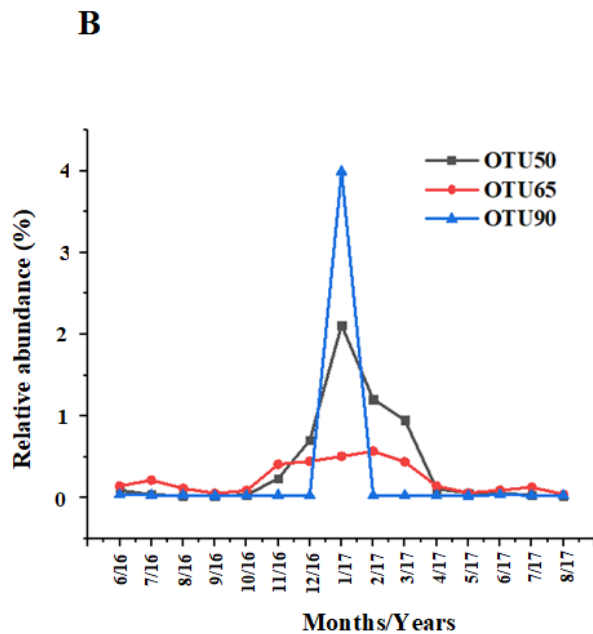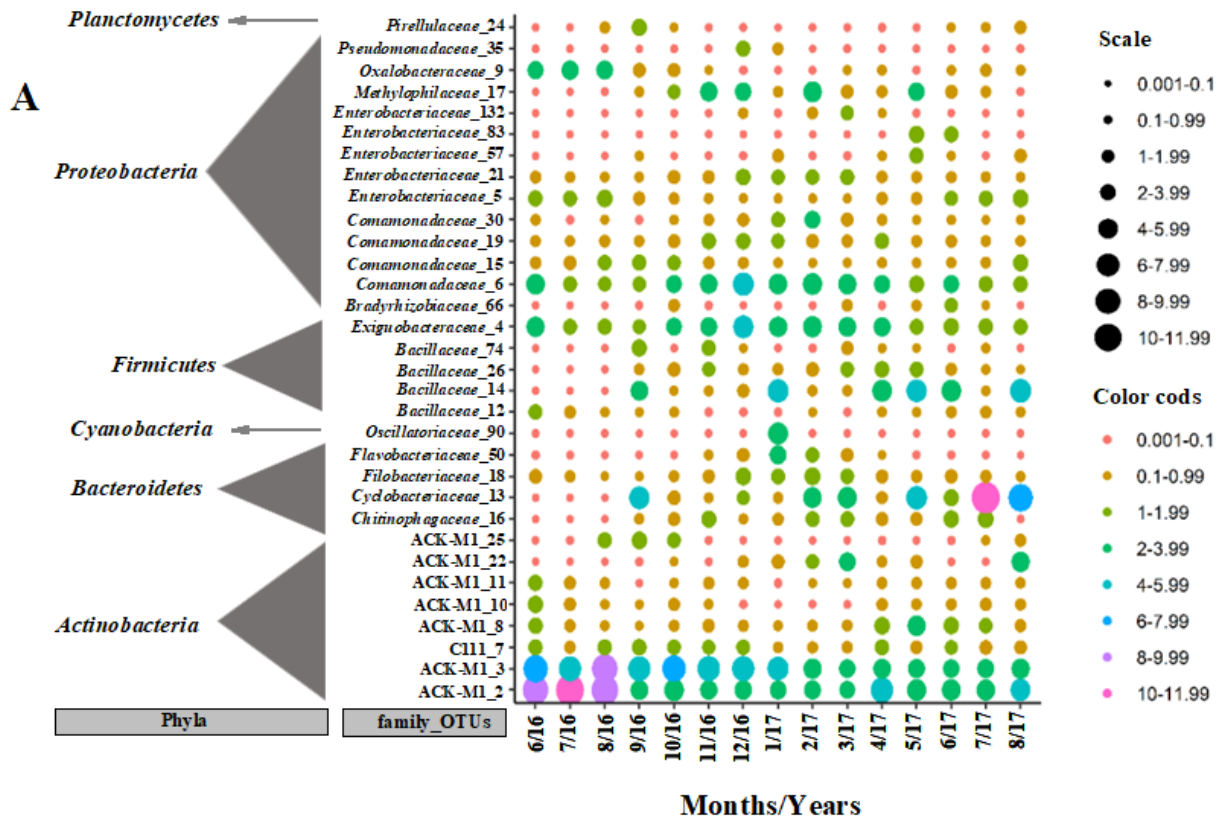
Figure 3.6. The relative abundance of highly abundant OTUs are (panel A) and *Flavobacteriaceae* (OTU50) and *Oscillatoriaceae* (OTUs 65 and 90) (panel B) over 15 months of sampling.

At the phyla level, the BCCs of June (47.56%), July (49.83%) and August (53.37%) in the summer of 2016, were enriched by *Actinobacteria* compared to June (27.47%), July (30%) and August (34.78%) in summer 2017 and all other months. In contrast, taxa belong to *Firmicutes* was not common in the BCCs of June (6.93%), July (7.24%) and August (7.31%) in summer 2016 while they were consistently common in the composition of the BCCs in other months (30.43 ± 6.88%). The relative abundance of taxa belonging to class *Proteobacteria* increased in January (34.72%) and February (43.34%) compared to other months (26.21 ± 3.84%).

At the class level, significant variation (p<0.05) was detected among months by pairwise LDA among the 30 identified taxonomic classes. Only 2 bacterial classes showed significant differences between the BCCs of June and July (2016), meanwhile, 24 bacterial classes showed significant variation between the BCCs of July 2016 and June 2017. Classes *Actinobacteria*, *Acidimicrobiia*, *Thermoleophilia* (phylum *Actinobacteria*) and *Saprospirae* and *Cytophagia* (phylum *Bacteroidetes*) showed high relative abundance in the June and July and August (2016) BCCs compared to other months (Figure 3.7). Conversely, class *Bacilli* had low relative abundance in the June and July and August (2016) BCCs compared to all other sampling months (Figure 3.7). Class *Chloroplast* became a highly abundant taxa in the BCCs of cold months (November and December; 2016 and January; 2017). We also observed a noticeable increase in the relative abundance of classes *Beta*, *Delta* and *Gammaproteobacteria* in the BCC of January (2017) and a shift in the relative abundance of class *Planctomycetia* from highly to moderately abundant in the cold months (November-May) (Figure 3.7).
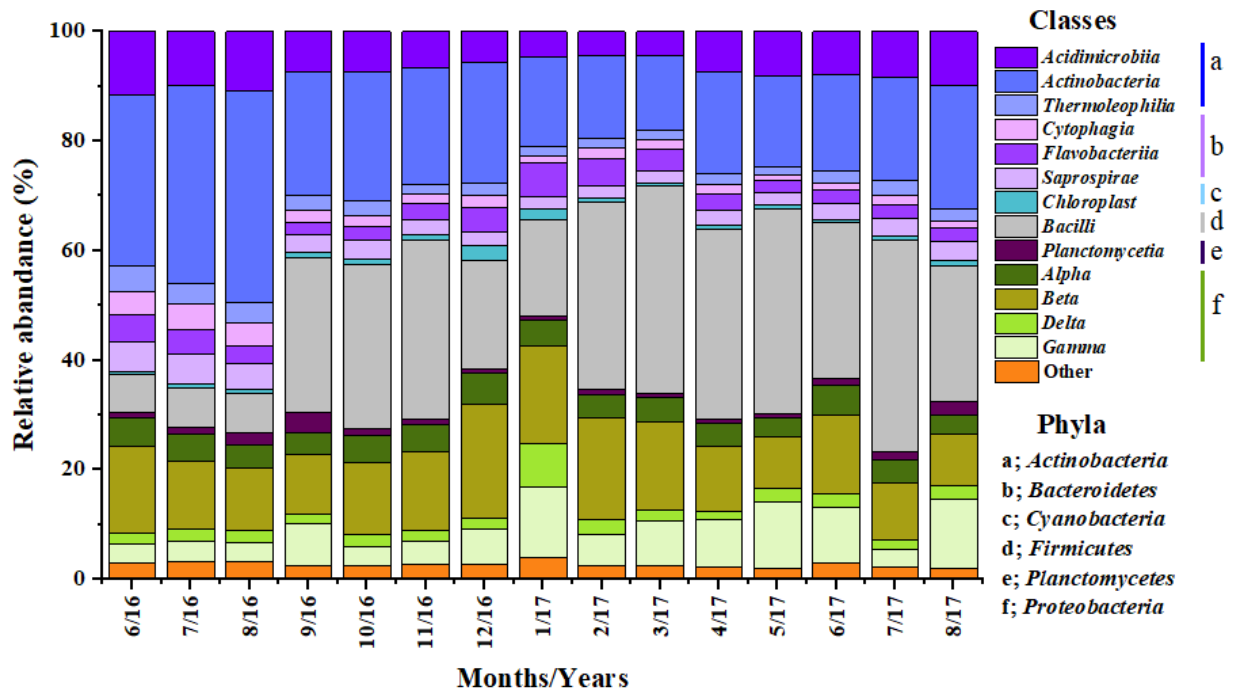
Figure 3.7. Bar chart showing the relative abundance of the BCCs at the class-level for combined sampling locations and bi-weekly sampling over the 15 months of sampling.

We also separated OTUs belonging to putative heterotrophic (2100 OTUs) from phototrophic OTUs (57 OTUs) over 15 months of sampling (Appendix B; Figure S3.5) and combined the relative abundance of OTUs belonging to same taxa at the family level and focused on the dominant OTUs. Surveys of heterotrophic bacteria revealed significant changes in the composition of heterotrophic bacteria between the two summers (2016 and 2017). The change was characterized by significant reductions in the relative abundance of families such as C111 and ACK-M1 (*Actinobacteria*) in summer 2017 relative to summer 2016, coupled with increased abundance of *Cyclobacteriaceae* (*Bacteroidetes*), *Enterobacteriaceae* (*Gammaproteobacteria*) and *Bacillaceae* (*Bacilli*) in summer 2017 (Appendix B; Figure S3.5). More interestingly, *Oscillatoriaceae* (*Cyanobacteria*), a phototrophic family, exhibited noticeably higher abundance in colder, lower sunlight months (Appendix B; Figure S3.5, panel B) which coincided with an increase in the abundance of *Flavobacteriaceae* family (a

heterotrophic bacteria) in the colder months; November - March (Appendix B; Figure S3.5).

Environmental variables effect: The mean water temperature varied considerably, with summer highs of 24.3 ± 0.4 °C and 22 ± 0.2 °C for summer 2016 and 2017 respectively, with the highest and lowest water temperatures in August 2016 (28 ± 0.72 °C) and January 2017 (0.05 ± 0.070 °C) (Appendix B; Figure S3.6). Environmental parameters (daylight, precipitation and water temperature) correlated with the BCCs of 6 sampling locations over 15 sampling months (Spearman Rho=0.32, p < 0.05). Daylight (4%), precipitation (4%) and water temperature (13%) together explained 21% of the total variation of the biological data (Appendix B; Figure S3.7).

3.4 Discussion

Our results showed minor spatial variation (between lakes and among sampling locations). It has been reported that variation of environmental parameters such as salinity, redox conditions and dissolved organic matters (DOM), etc. (Beck et al., 2017), as well as habitat variation (Lear et al., 2014), among sampling locations, are strong drivers of the spatial variation of the microbial communities. We did not measure abiotic parameters such as nutrient levels at our sampling locations; however, due to connectivity of the two lakes by the Detroit River (Burniston et al., 2018), the short distances among the sampling locations and the eutrophication of Lake Erie (Watson et al., 2016) and St. Clair (Bocaniov et al., 2019), our sampling locations might have similar habitat features which consequently resulted in the little spatial variation of the BCCs in our study. For example, even comparisons of upper Great Lakes (lakes Superior and Huron) BCC data with Lake Erie returned only a few OTUs (383 out of ~13000 OTUs) with significant differences in their abundance (Rozmarynowycz et al., 2019), consistent with our observation of only relatively minor spatial variation effects on BCC sampled at the same time. A report of only minor spatial variation in the metabolic profiles in 8 carbon substrates out of 31 in the sediment microbial communities of Lake St. Clair (Oest et al., 2018) highlighted the lack of substantial spatial variation among sites even at the functional level. On the other hand, a study of BCC variation across long distances (~3000 km) showed two separate communities; upper versus lower Missouri River (Henson et al., 2018).

Additionally, three distinctly different bacterial assemblages were reported in the upper, middle and lower Yenisei River (1800 km) (Kolmakova et al., 2014) – that study also reported matching nutrient spatial variation. The reported presence of distinct BCCs across large spatial scales versus our weak spatial effects across short distances may reflect nutrients gradients that may occur across long distance and drives microbial habitat variation.

We observed strong temporal variation over our sampling effort that included more than a year (15 months), which captured vast temporal (seasonal) variation. Clustering of the BCCs for the 15 sampling months resulted in five highly divergent BCC clades, which closely matched to seasonal patterns. Many studies have reported high prokaryotic microbial diversity in summer relative to winter (Hao et al., 2017), which we also observed. More interestingly, we found significant variation between the BCC of summer 2016 and 2017, with a significant decrease in the diversity indexes of the BCC of summer 2017. As the water temperature of two summers was not significantly different, the variation of the two summer BCCs is likely related to other abiotic (such nutrients bioavailability) and biotic factors that differed between the two summers (Bižić-Ionescu et al., 2014; Newton et al., 2011). Perhaps not surprisingly, we found significant correlations between selected environmental factors (daylight hours, precipitation and water temperature) and the BCC; however, all three environmental factors only explained 21% of the total variation in BCC across the 15 months. This study was not designed to test for annual effects; however, significant differences in the BCC in the two sampled summers (2016 and 2017) highlights the potential for unpredicted annual temporal variation along with a seasonal and monthly temporal variation of freshwater BCCs. Few studies have characterized monthly temporal variation in freshwater BCC. However, one study reported monthly monitoring of BCC of Lake Taihu over 3 years (2009–2011) at four different sites, and showed significant monthly (and consequently seasonal) variation of diversity indexes of the BCC (Peng et al., 2018). In line with our observation of strong seasonal variation in BCC, drastic seasonal transitions of microbial abundance and diversity have been reported in lakes (Bižić-Ionescu et al., 2014; Butler et al., 2019). Reported high levels of variability in BCC of bog lakes over 5 years (with unique communities in each year of sampling)

(Linz et al., 2017) was also consistent with our limited results regarding annual (summer) BCC variation in large freshwater lake ecosystems.

Few microbial taxa showed higher abundance in Lake Erie relative to Lake St. Clair, despite Lake Erie is a substantially larger and deeper lake. The few that were more abundant in Lake Erie mostly belong to *Actinobacteria*. *Actinobacteria* are often the numerically dominant phylum in lakes (Newton et al., 2011), but their abundance decreases with oxygen limitation (Taipale et al., 2009) and overloading of the nutrients (Haukka et al., 2006). Low abundance of this phylum in Lake St. Clair is likely due to the low level of oxygen or higher loading of nutrients in the smaller, shallower Lake St. Clair. We found little evidence for taxonomic variation (at the class level) among sampling locations in Lake Erie perhaps reflecting relatively uniform microbial habitat characteristics among the three sampling locations (all public beaches). Interestingly, the abundance of *Bacteroidetes and Verrucomicrobia (two taxa* associated with high-nutrient environments) (Newton et al., 2011; Wagner and Horn, 2006) and *Planctomycetes* (a key taxon in anaerobic ammonium oxidation) (Wagner and Horn, 2006) were significantly different between the two sampling locations in Lake St. Clair potentially due to overloading of nutrients from an adjacent urban tributary near one of the sampling sites (LP).

Previous microbial community work in Lake Erie (Mou et al., 2013; Sharma et al., 2009; Wilhelm et al., 2014) suggested that *Actinobacteria* are important components of the microbial community, particularly in the summer, which matched to our finding for 2016 but not in 2017. The summer 2017 BCC showed a significant reduction of *Actinobacteria* compared to summer 2016, potentially due to changes in abiotic variables such as reduced oxygen levels (Taipale et al., 2009) or nutrient overloading (Haukka et al., 2006) in summer 2017. OTUs belonging to *Proteobacteria* were in the top three most abundant phyla in the BCCs across different months and seasons. For example, *Proteobacteria* was the most abundant phylum in January and June (2017); the second most common in summer 2016 (June, July and August) and third most common in July (2017). *Proteobacteria* are reported as very abundant in many different freshwater lake habitats, but their relative abundance varies among lakes, within lakes and over time (Newton et al., 2011). Curiously, one ubiquitous group of

metabolically versatile bacterial was observed at high abundance in the coldest months of our study; order *Pseudomonadales* (families *Moraxellaceae* and *Pseudomonadaceae*) in January and February and *Enterobacteriaceae* (*Gammaproteobacteria*) in January, March, May, June and July (2017). Freshwater lake *Bacteroidetes* are often found in high abundance during periods following cyanobacterial blooms (Newton et al., 2011). It has been reported that *Flavobacterium* spp. (belonging to *Bacteroidetes*) is the dominant taxa of the winter community in Lake Erie (Wilhelm et al., 2014), and while we also found a significant elevation of *Flavobacteriaceae* over the cold months, we identified *Proteobacteria* as the dominant phylum in the winter (cluster 2). In our data set, the relative abundance of *Cyanobacteria* (family *Oscillatoriaceae*) and *Bacteroidetes* (family *Flavobacteriaceae*) exhibited correlated abundance in the BCC of November to March, likely reflecting the dependency of *Bacteroidetes* on the organic matter loading by Cyanobacteria (Eiler and Bertilsson, 2007). Previous studies have also noted high levels of *Cyanobacteria* during winter months in freshwater reservoirs (Valério et al., 2008). It has been suggested that high concentrations of overwintering vegetative *Cyanobacteria* cells may provide a large inoculum for blooms during warmer seasons (Ma et al., 2016), but the impact of family *Oscillatoriaceae* on algal bloom dynamics in Lake Erie and St. Clair is not well known.

Cytophaga is well known to be proficient in degrading biopolymers such as cellulose and chitin, part of the high molecular mass fraction of DOM (Kirchman, 2002; Newton et al., 2011). In our study, the relative abundance of *Cytophaga* (phylum *Bacteroidetes*) changed from being a rare component of the community in June, July and August (summer 2016) to high abundance in May, July and August (2017), indicating potentially elevated availability of DOM in the summer of 2017. Although *Firmicutes* is generally a minor freshwater lake community taxon (Newton et al., 2011), in our study the relative abundance of OTUs belonging to *Firmicutes* increased over time. Indeed, this phylum became one of the most dominant phyla across all sampling points after summer 2016. Similar to our finding, a high abundance of *Firmicutes* (23%) was reported from water samples collected from freshwater public beaches (Ohio, Madison lake) (Lee et al., 2016). In that study, *Exiguobacterium* and

Paenisporosarcina were the most dominant *Firmicutes* genera (Lee et al., 2016), while in our study *Bacillaceae* (September 2016 - August 2017) and *Exiguobacteraceae* (December, February and March 2016-17) were the most abundant genera in Lake Erie and St. Clair.

We noted a composition shift of the freshwater BCC from a community enriched by *Actinobacteria* (sensitive to nutrient overloading and low oxygen level) to one enriched by *Proteobacteria* (adapted to nutrient overloading) (Newton et al., 2011), *Bacteroidetes* (proficient in the degradation of complex biopolymers and DOM) (Newton et al., 2011) *and Firmicutes* (diverse metabolic capabilities and resistant to oxygen limitation) (Martiny et al., 2006) over time. Although we did not measure nutrients in our study, the observed pattern of BCC change indicated likely increases in the loading of nutrients into both lakes from fall 2016 onwards. However, the mechanism(s) responsible for the observed BCC shift requires further investigation. Furthermore, we observed temporal variation in *Enterobacteriaceae* abundance; a family that includes many waterborne pathogens and fecal indicator bacteria (FIBs) (Ramírez-Castillo et al., 2015), *Pseudomonadales*; a taxon which may act as an opportunistic pathogen in fish (Su et al., 2020) and humans (Malhotra et al., 2019) and *Oscillatoriaceae (Cyanobacteria),* all of which reflect variation in potential pathogens and health risk, particularly over the summer.

3.5   Conclusion

Our results showed that although freshwater BCC may have a cyclic seasonal or annual variation component, the details of the composition of the community can change unpredictably over the temporal and spatial scales included in our study. The observed BCC variation could be linked to the functional activity of the community, making additional studies necessary to characterize the consequence of this variation on the ecological services of BCC in a large freshwater ecosystem. Our results also showed that long term monitoring of the bacterial community could serve as a sensitive proxy of freshwater ecosystems health and perhaps even function.

3.6 References

Adrian, R., O'Reilly, C.M., Zagarese, H., Baines, S.B., Hessen, D.O., Keller, W., Livingstone, D.M., Sommaruga, R., Straile, D., Van Donk, E., 2009. Lakes as sentinels of climate change. Limnology and Oceanography 54(6part2) 2283-2297.

Bates, D., Maechler, M., Bolker, B., 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Beck, M., Reckhardt, A., Amelsberg, J., Bartholomä, A., Brumsack, H.-J., Cypionka, H., Dittmar, T., Engelen, B., Greskowiak, J., Hillebrand, H., 2017. The drivers of biogeochemistry in beach ecosystems: a cross-shore transect from the dunes to the low-water line. Marine Chemistry 190 35-50.

Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J., 2017. Cyanobacterial harmful algal blooms are a biological disturbance to western Lake Erie bacterial communities. Environmental Microbiology 19(3) 1149-1162.

Bižić-Ionescu, M., Amann, R., Grossart, H.-P., 2014. Massive regime shifts and high activity of heterotrophic bacteria in an ice-covered lake. PloS One 9(11).

Bocaniov, S.A., Van Cappellen, P., Scavia, D., 2019. On the role of a large shallow lake (Lake St. Clair, USA-Canada) in modulating phosphorus loads to Lake Erie. Water Resources Research.

Bouzat, J.L., Hoostal, M.J., Looft, T., 2013. Spatial patterns of bacterial community composition within Lake Erie sediments. Journal of Great Lakes Research 39(2) 344-351.

Bullerjahn, G.S., McKay, R.M., Davis, T.W., Baker, D.B., Boyer, G.L., D'Anglada, L.V., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae 54 223-238.

Burniston, D., Dove, A., Backus, S., Thompson, A., 2018. Nutrient concentrations and loadings in the St. Clair River–Detroit River Great Lakes Interconnecting Channel. Journal of Great Lakes Research 44(3) 398-411.

Bush, T., Diao, M., Allen, R.J., Sinnige, R., Muyzer, G., Huisman, J., 2017. Oxic-anoxic regime shifts mediated by feedbacks between biogeochemical processes and microbial community dynamics. Nature Communications 8(1) 789.

Butler, T.M., Wilhelm, A.-C., Dwyer, A.C., Webb, P.N., Baldwin, A.L., Techtmann, S.M., 2019. Microbial community dynamics during lake ice freezing. Scientific Reports 9(1) 1-11.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5) 335.

Casey, G.D., 1998. National water-quality assessment of the Lake Erie-Lake St. Clair Basin, Michigan, Indiana, Ohio, Pennsylvania, and New York: Environmental and hydrologic setting. US Department of the Interior, US Geological Survey.

Castello, L., Macedo, M.N., 2016. Large-scale degradation of Amazonian freshwater ecosystems. Global Change Biology 22(3) 990-1007.

Davis, C.C., 1964. Evidence for the eutrophication of Lake Erie from phytoplankton records. Limnology and Oceanography 9(3) 275-283.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of Vegetation Science 14(6) 927-930.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19) 2460-2461.

Eiler, A., Bertilsson, S., 2007. Flavobacteria blooms in four eutrophic lakes: linking population dynamics of freshwater bacterioplankton to resource availability. Applied and Environmental Microbiology 73(11) 3511-3518.

Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa (Washington, DC) 3.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., 2012. Package 'car'. Vienna: R Foundation for Statistical Computing.

Fuhrman, J.A., Cram, J.A., Needham, D.M., 2015. Marine microbial community dynamics and their ecological interpretation. Nature Reviews Microbiology 13(3) 133.

Hammer, Ř., Harper, D., Ryan, P., 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4 4-9.

Hao, C., Wei, P., Pei, L., Du, Z., Zhang, Y., Lu, Y., Dong, H., 2017. Significant seasonal variations of microbial community in an acid mine drainage lake in Anhui Province, China. Environmental Pollution 223 507-516.

Haukka, K., Kolmonen, E., Hyder, R., Hietala, J., Vakkilainen, K., Kairesalo, T., Haario, H., Sivonen, K., 2006. Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial Ecology 51(2) 137-146.

He, X., Chaganti, S.R., Heath, D.D., 2017. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microbial Ecology 1-12.

Henson, M.W., Hanssen, J., Spooner, G., Fleming, P., Pukonen, M., Stahr, F., Thrash, J.C., 2018. Nutrient dynamics and stream order influence microbial community patterns along a 2914 kilometer transect of the Mississippi River. Limnology and Oceanography 63(5) 1837-1855.

Hölker, F., Wurzbacher, C., Weißenborn, C., Monaghan, M.T., Holzhauer, S.I., Premke, K., 2015. Microbial diversity and community respiration in freshwater sediments influenced by artificial light at night. Philosophical Transactions of the Royal Society B 370(1667) 20140130.

Kirchman, D.L., 2002. The ecology of Cytophaga–Flavobacteria in aquatic environments. FEMS Microbiology Ecology 39(2) 91-100.

Kolmakova, O.V., Gladyshev, M.I., Rozanov, A.S., Peltek, S.E., Trusova, M.Y., 2014. Spatial biodiversity of bacteria along the largest Arctic river determined by next-generation sequencing. FEMS Microbiology Ecology 89(2) 442-450.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.L., 2014. Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. The ISME Journal 8(8) 1715-1726.

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

Lindström, E.S., Langenheder, S., 2012. Local and regional factors influencing bacterial community assembly. Environmental Microbiology Reports 4(1) 1-9.

Linz, A.M., Crary, B.C., Shade, A., Owens, S., Gilbert, J.A., Knight, R., McMahon, K.D., 2017. Bacterial community composition and dynamics spanning five years in freshwater bog lakes. MSphere 2(3) e00169-00117.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24(8) 813-821.

Lv, F., Xue, S., Wang, G., Zhang, C., 2017. Nitrogen addition shifts the microbial community in the rhizosphere of Pinus tabuliformis in Northwestern China. PloS one 12(2) e0172382.

Ma, J., Qin, B., Paerl, H.W., Brookes, J.D., Hall, N.S., Shi, K., Zhou, Y., Guo, J., Li, Z., Xu, H., 2016. The persistence of cyanobacterial (M icrocystis spp.) blooms throughout winter in L ake T aihu, C hina. Limnology and Oceanography 61(2) 711-722.

Malhotra, S., Hayes, D., Wozniak, D.J., 2019. Cystic fibrosis and Pseudomonas aeruginosa: the host-microbe interface. Clinical Microbiology Reviews 32(3) e00138-00118.

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4(2) 102-112.

Mou, X., Jacob, J., Lu, X., Robbins, S., Sun, S., Ortiz, J.D., 2013. Diversity and distribution of free-living and particle-associated bacterioplankton in Sandusky Bay and adjacent waters of Lake Erie Western Basin. Journal of Great Lakes Research 39(2) 352-357.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75(1) 14-49.

Oest, A., Alsaffar, A., Fenner, M., Azzopardi, D., Tiquia-Arashiro, S.M., 2018. Patterns of change in metabolic capabilities of sediment microbial communities in river and lake ecosystems. International Journal of Microbiology 2018.

Peng, Y., Yue, D., Xiao, L., Qian, X., 2018. Temporal variation and co-occurrence patterns of bacterial communities in eutrophic Lake Taihu, China. Geomicrobiology Journal 35(3) 186-197.

Ramírez-Castillo, F., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F., Harel, J., Guerrero-Barrera, A., 2015. Waterborne pathogens: detection methods and challenges. Pathogens 4(2) 307-334.

Rozmarynowycz, M.J., Beall, B.F., Bullerjahn, G.S., Small, G.E., Sterner, R.W., Brovold, S.S., D'souza, N.A., Watson, S.B., McKay, R.M.L., 2019. Transitions in microbial communities along a 1600 km freshwater trophic gradient. Journal of Great Lakes Research 45(2) 263-276.

Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. Journal of Great Lakes Research 42(6) 1139-1150.

Schindler, D.W., Carpenter, S.R., Chapra, S.C., Hecky, R.E., Orihel, D.M., 2016. Reducing phosphorus to curb lake eutrophication is a success. ACS Publications.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic biomarker discovery and explanation. Genome biology 12(6) R60.

Shahraki, A.H., Chaganti, S.R., Heath, D., 2019. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. Journal of Water and Health 17(1) 37-49.

Sharma, A.K., Sommerfeld, K., Bullerjahn, G.S., Matteson, A.R., Wilhelm, S.W., Jezbera, J., Brandt, U., Doolittle, W.F., Hahn, M.W., 2009. Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater Actinobacteria. The ISME Journal 3(6) 726-737.

Siles, J.A., Cajthaml, T., Filipova, A., Minerbi, S., Margesin, R., 2017. Altitudinal, seasonal and interannual shifts in microbial communities and chemical composition of soil organic matter in Alpine forest soils. Soil Biology and Biochemistry 112 1-13.

Singh, B.K., Walker, A., 2006. Microbial degradation of organophosphorus compounds. FEMS Microbiology Reviews 30(3) 428-471.

Small, G.E., Finlay, J.C., McKay, R., Rozmarynowycz, M.J., Brovold, S., Bullerjahn, G.S., Spokas, K., Sterner, R.W., 2016. Large differences in potential denitrification and sediment microbial communities across the Laurentian great lakes. Biogeochemistry 128(3) 353-368.

Sterner, R.W., Ostrom, P., Ostrom, N.E., Klump, J.V., Steinman, A.D., Dreelin, E.A., Vander Zanden, M.J., Fisk, A.T., 2017. Grand challenges for research in the Laurentian Great Lakes. Limnology and Oceanography 62(6) 2510-2523.

Su, L., Xu, C., Cai, L., Qiu, N., Hou, M., Wang, J., 2020. Susceptibility and immune responses after challenge with Flavobacterium columnare and Pseudomonas fluorescens in conventional and specific pathogen-free rare minnow (Gobiocypris rarus). Fish & Shellfish Immunology 98 875-886.

Taipale, S., Jones, R.I., Tiirola, M., 2009. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquatic Microbial Ecology 55(1) 1-16.

Team, R.C., 2013. R: A language and environment for statistical computing.

Valério, E., Faria, N., Paulino, S., Pereira, P., 2008. Seasonal variation of phytoplankton and cyanobacteria composition and associated microcystins in six Portuguese freshwater reservoirs, International Journal of Limnology. EDP Sciences, pp. 189-196.

VanMensel, D., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2020. Exploring bacterial pathogen community dynamics in freshwater beach sediments: A tale of two lakes. Environmental Microbiology 22(2) 568-583.

Villaescusa, J.A., Jørgensen, S.E., Rochera, C., Velázquez, D., Quesada, A., Camacho, A., 2016. Carbon dynamics modelization and biological community sensitivity to temperature in an oligotrophic freshwater Antarctic lake. Ecological Modelling 319 21-30.

Wagner, M., Horn, M., 2006. The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Current Opinion in Biotechnology 17(3) 241-249.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73(16) 5261-5267.

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56 44-66.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3(2) 180-185.

Wilhelm, S.W., LeCleir, G.R., Bullerjahn, G.S., McKay, R.M., Saxton, M.A., Twiss, M.R., Bourbonniere, R.A., 2014. Seasonal changes in microbial community structure and activity imply winter production is linked to summer hypoxia in a large lake. FEMS Microbiology Ecology 87(2) 475-485.

Yoshida, T., Nishimura, Y., Watai, H., Haruki, N., Morimoto, D., Kaneko, H., Honda, T., Yamamoto, K., Hingamp, P., Sako, Y., 2018. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. The ISME Journal 12 1287–1295.

FUNCTIONAL AND VIRULENCE GENE TRANSCRIPTION VARIATION IN BACTERIAL METATRANSCRIPTOMES IN LARGE FRESHWATER LAKE ECOSYSTEMS

4.1 Introduction

Freshwater ecosystems are remarkable in terms of the complexity of their ecosystem processes and the interactions across trophic levels (Geist, 2011). While freshwater ecosystems provide critical services, including drinking water, nutrient recycling, sport and commercial fisheries and recreation, they face many threats such as pesticide and fertilizer pollution, climate change, water extraction, and habitat destruction, among others (Castello and Macedo, 2016; Dodds et al., 2013). The Laurentian Great Lakes (LGLs) collectively are the largest freshwater ecosystems in the world, and arguably represent the single most valuable natural resource on the North American continent (McKenna Jr, 2019). Roughly 14% of the total US population and 60% of the total Canadian population (~40 million Canadians and Americans) live within the LGL watershed (MacKenzie, 1997). The well-documented degradation of the LGL's water resources dates back more than 100 years (Beeton, 1965), but recent studies report ongoing degradation of the LGLs ecosystem due to intensive fishing, loss of habitat, introduction of exotic species, nutrient enrichment, and chemicals (Jenny et al., 2020; Wattigney et al., 2017).

Bacteria communities (BCs) are an important component of the microbial community (Berdjeb et al., 2011) and have fundamental direct and indirect roles in all ecosystems, but are particularly important in providing freshwater ecosystem services (Blaser et al., 2016). For example, *Cyanobacteria* are photoautotrophic, with some diazotrophic properties, and are one of the main players in inorganic carbon (C) and nitrogen (N) fixation (Vahtera et al., 2010). Heterotrophic bacteria also have a critical role in regenerating and mobilizing nutrients in freshwater food webs, and it has become clear that aquatic heterotrophic bacteria drive transformations and the cycling of most biologically active elements in the aquatic ecosystems (Newton et al., 2011). A variety of abiotic and biotic factors are known, or suspected, to affect BC composition and function. Abiotic factors such as light levels (day versus night) (Poretsky et al.,

2009), water temperature (Lei and Lai, 2019) and nutrient loading (Ren et al., 2019) can have significant temporal effects on BC composition (Shahraki et al., 2020) and hence the community activity and function. On the other hand, the adaptation of BC to local conditions, habitats (Lear et al., 2014) and spatial variation in nutrient gradients (Davis et al., 2015) may also result in species sorting, which could lead to spatial variation in the function of the BC in aquatic ecosystems. High throughput sequencing of the metatranscriptome has provided an unprecedented opportunity to quantify gene activity in natural microbial communities (Moran et al., 2013) and hence roughly characterize community function. The study of functional variation of the BC in relation to abiotic and biotic parameters provides insights into how these parameters are interlinked with BC dynamics and ecological activity (Singh et al., 2009). We have only a superficial knowledge of the factors that contribute to variation in BC function, both spatially and temporally, yet such information would allow us to better predict BC function, particularly in freshwater ecosystems.

Lake Erie, the smallest and shallowest of the LGLs, has been impacted by poor water quality as far back as the 1960s (Davis, 1964). Phosphate removal programs ultimately resulted in significant improvement in the trophic state of Lake Erie (Schindler et al., 2016). However, with changes in the ecology, climate, and the now, nonpoint P sources, the extent and duration of Lake Erie's cyanobacteria harmful algae blooms (cHABs) and hypoxia has increased dramatically since the mid-1990s (Scavia et al., 2014). Increased the hypoxic areas (up to a maximum daily extent of 11,600 km^2 (Karatayev et al., 2018)), and toxic *Microcystis* blooms led to a "do not drink" advisory for 500,000 people living in the Toledo, Ohio, area (Ho and Michalak, 2015). Lake St. Clair (technically not one of the five LGLs) also is very shallow and highly affected by recurrent eutrophication (Healy et al., 2008), and is a potential source of nutrient loading for Lake Erie (Scavia et al., 2016). Lake Erie and St. Clair ecosystems are currently threatened by frequent water degradation, such as cHABs (Bullerjahn et al., 2016; Watson et al., 2016). More studies are needed to address the temporal and spatial variation of bacterial function associated with variation in water quality, as well as the basic ecology, of these two connected water bodies.

The BC composition (Glasl et al., 2017) is used as a biomarker to indirectly quantify human-induced stress on aquatic ecosystems, and thus it is also used to monitor water quality and safety. However, BC composition as a health biomarker has some limitations, as sometimes altered BC (not resilient) might result in process rates similar to the original community if the members of the community are functionally redundant (Allison and Martiny, 2008). Thus, an altered BC may not accurately reflect community function and may result in a false positive for the functional breakdown. Since RNA is more transient in nature, and hence more responsive to temporal change, monitoring the BC metatranscriptome allows the identification of changes in the transcriptional activity of the whole BC as well as selected biomarkers (Yosef and Regev, 2011).

Indicator organisms such as *Escherichia coli* (Soller et al., 2016) and toxin-producing cyanobacteria (Bartram and Chorus, 1999) serve as proxies for human health risk in aquatic ecosystems; however, these indicators have limitations. *E. coli* monitoring is time-consuming and shows poor correlation with the occurrence of actual human pathogens (Ishii and Sadowsky, 2008; Jang et al., 2017; Pattis et al., 2017). On the other hand, the World Health Organization (WHO) developed a three-level guideline approach (measuring cyanobacteria abundance, chlorophyll-a, and microcystins) (Bartram and Chorus, 1999) to evaluate the level of health risk causing by cHABs in freshwater ecosystems. However, quantifying approximately 90 known variants of microcystin and the related toxins is difficult and time-consuming. Moreover, Loftin et al, (Loftin et al., 2016) argued that the presence of chlorophyll-a does not always equate to cHABs, since chlorophyll is common amongst all phytoplankton. Clearly more studies are needed to develop microbial biomarkers (at the community or species levels) as reliable and practical indicators of water quality and human risk. To this, we need standardized reference data to provide baseline metrics that are indicative of healthy microbial community variation and function.

A recent study of a freshwater lake during ice-free and ice-covered periods (November – April), showed high nitrification activity of freshwater BC under ice-covered conditions (Butler et al., 2019). As cHABs are the main ecosystem health challenge of the lower LGLs in recent years, many studies have focused on the activity

of the cyanobacteria community (Steffen et al., 2015) and the interaction of cyanobacteria and heterotrophic bacteria during bloom events (Briland et al., 2020). However, little is known about the temporal and spatial variation of the freshwater BC ecological activity in non-bloom conditions, and how the abundance of harmful microbes might vary naturally. To address these knowledge gaps, we sampled three recreational beach locations in lakes Erie (two beaches) and St. Clair (one beach), in the winter (January), summer (July) and fall (October) in 2019. The first aim of the study was to generally characterize the temporal (winter, summer and fall) and spatial (three sampling locations; public beaches) variation across the whole transcriptome of the freshwater BC under non-bloom conditions. The second aim was to examine temporal and spatial variation of physiologically and ecologically (i.e. ecosystem function) relevant genes. The third aim was to explore the "natural" temporal and spatial variation in the transcript abundance of genes associated with human health risk. 1) We hypothesized that the freshwater BC exhibits high temporal (seasonal) variation but limited spatial variation (due to conservation of function). 2) We also hypothesized high temporal (seasonal) variation in the transcription profiles of the physiologically and ecologically relevant genes, coupled with limited spatial variation in the freshwater BC. 3) We also hypothesized that the community of harmful bacterial and the transcripts of their related virulence genes "naturally" have temporal and spatial variation. This study provides a temporal and spatial snapshot of the physiological and ecological activity the freshwater BC in non-bloom conditions and the genes associated with health risk which could increase our understanding of microbial ecology in an ecosystem affected by human anthropogenic activity.

4.2 Material and Methods

4.2.1 Study sites and sample collection

Freshwater samples were collected from three public beaches; Colchester Harbour (CH) and Seacliff Beach (SL) in Lake Erie, and Sand Point (SP) in Lake St. Clair on January 10th (winter), July 10th (summer) and October 10th (fall) in 2019. SP is located in an urban area and is close to the outlet of Lake St. Clair into the Detroit River, while CH and SL are located in agriculture areas (Figure 4.1). At each beach, we collected 2

replicate water samples offshore at ~0.5 m depth, for a total of 18 samples (2 replicates x 3 beaches x 3 seasons; winter, summer and fall). For each replicate, approximately 10 liters of freshwater were collected, and 6 liters were filtered immediately in the field using 6 separate filters (polycarbonate membranes, 0.2 µm pore size; Millipore, USA). All six filters/sample were cut into small pieces using sterile scissors and then transferred into a 15 mL Falcon tube (RNase/ DNase free). The Falcon tubes were stored immediately on dry ice in the field, then transferred to the lab and stored at -80 °C until RNA extractions were performed.
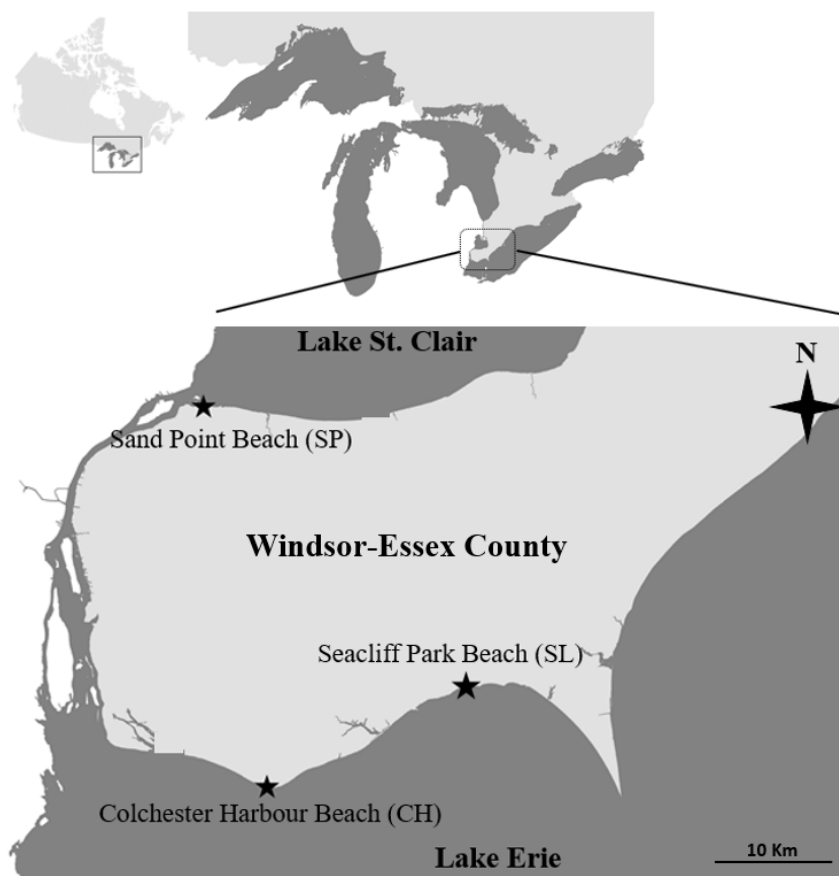
Figure 4.1. Map of Windsor-Essex County and three beaches in Lake Erie (2) and Lake St. Clair (1) sampled for this research.

Environmental variables: Water temperature was measured in the field immediately after sample collection. Precipitation and daylight hours data were from Environment Canada (http://climate.weather.gc.ca/historical_data/search_historic_data_e.html). We also

measured total nitrogen (TN = NO$_2$+NO$_3$) and total phosphate (TP) levels from the samples according to the Environmental Protection Agency (EPA) methods (method 365.1 and EPA 353.2) using a SmartChem 170 discrete analyzer (Westco Scientific Instruments, Canada).

4.2.2 RNA Extraction, quality control and sequencing

RNA was extracted from the filters using PowerSoil Total RNA Isolation kits (MoBio), following the manufacturer's protocol with slight modifications. DNase/RNase-free reagents, tubes and pipet tips were kept chilled on ice when practical; exceptions include reagents that required room temperature (to avoid precipitation) and sample transfers. RNA precipitation was extended to >12 hours at −20°C to increase yield, and the final pellet was re-suspended in 80 μL RNase-free water. Aliquots of extracted RNA were kept at −80°C until further analyses. The quality and quantity of extracted RNA samples were assessed in-house using an Agilent 2100 Bioanalyzer (Agilent Technologies, Mississauga, ON, Canada) before RNA sequencing. All samples (with their replicates) with RNA integrity numbers (RIN) over 7.0 and concentrations >100 ng μL^{-1} were sent to the Genome Quebec Innovation Center at McGill University in Quebec, Canada for RNA sequencing for metatranscriptomic analysis. All 18 collected samples (2 replicates x 3 beaches x 3 seasons) passed QC and were sequenced by Illumina HiSeq 4000.

4.2.3 Bioinformatic analysis

From the Illumina platform, we obtained paired-end reads in fastq format (Phred +33), separated into individual files for each single-end read library. The quality of each sequence file was evaluated by FastQC (Wingett and Andrews, 2018). Paired-end reads were merged using SortMeRNA version 2.0 (Kopylova et al., 2012) with default parameters. Ribosomal RNA sequences (5S, 5.8S, 16S, 18S, 23S, and 28S) were identified and excluded from the merged reads but used for taxonomic identification (next section). Low quality reads (length > 30 nt) and adaptors were also removed from the non-rRNA data for each sample using Trimmomatic version 0.39 (Bolger et al., 2014). The non-rRNA reads were *de novo* assembled by Trinity version 2.0.6 after data cleaning (Grabherr et al., 2011). The functional annotation of the *de novo* assembled

transcripts was completed using the KO [Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology] database assigned by TransDecoder and Trinotate implemented in Trinity version 2.0.6 (Grabherr et al., 2011) and BLASTX against SwissProt database (uniprot_sprot.trinotate_v2.0). To obtain gene-level transcript read counts, we mapped the non-rRNA read file of each sample to the *de novo* assembled transcriptome using Bowtie version 2.2.4 (Langmead et al., 2009). We used the default setting in Bowtie version 2.2.4 for the allowed number of mismatches in a seed alignment during alignment (default: 0). The gene transcript read counts were calculated by Coreset version 1.06 (Davidson and Oshlack, 2014) for each replicate. The reads belonging to archaea, eukaryotes and viruses were removed from the data to focus on bacterial transcriptional activity. Moreover, the data of different isoforms of the same gene was combined before differential expression analysis.

4.2.4 Transcriptome analysis

Taxonomic assignment: 16S rRNA encoding sequences that were filtered from non-rRNA data by SortMeRNA version 2.0 (Kopylova et al., 2012) were identified to taxon by performing the Basic Local Alignment Search Tool (BLAST) against the small subunit ribosomal ribonucleic acid (SSU rRNA) SILVA non-redundant reference database with a stringent e-value cut off (e-40) (Pruesse et al., 2007). The best matches were accepted for taxonomic assignment. Replicate SSU rRNA data were combined and normalized to the total number of SSU rRNA matches for each sample. We defined taxa as "abundant", "moderate" and "rare" when the relative abundance (read count) of that taxon was $\geq$ 1.00%; 1.00% < and > 0.1%; $\leq$ 0.1% respectively (Logares et al., 2014). Here we present only the composition data for the taxa associated with water quality degradation (i.e., cyanobacteria and waterborne pathogens) assigned at the species level.

Transcriptome description: The Bioconductor R package, edgeR, was used for transcriptome analysis (Robinson et al., 2010). The genes that had <40 reads overall 18 samples were included in differential transcription analysis. Then, gene cluster transcription was normalized using the TMM (Trimmed Mean of M-values) method in which the library size (total read count; RC) was corrected by the estimated normalization factor and scaled to per million reads (Robinson and Oshlack, 2010).

Heat maps were generated using the normalized reads of the metatranscriptome sequence reads for all samples and their biological replicates (18 samples in total) to visualize the pattern of gene transcription among replicates and different samples over winter, summer and fall in the three sampling locations (CH, SL and SP beaches). Differences in transcript proportions among the samples were explored using multi-dimensional scaling (MDS) plots based on the average squared distance to the mean of the transcripts between each pair of samples.

4.2.5 Statistical analyses

Whole transcriptome data: We generated a Bray–Curtis dissimilarity matrix from the sequence read count of transcriptomic data for each sample using Primer-e software version 7.0 (Primer-E Ltd., Plymouth, UK). The first (PCo1), second (PCo2) and third (PCo3) coordinates of the principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity matrices were used separately as dependent variables in a nested one-way ANOVA (in R software version 3.6.1) to test the effect of sampling location (replicates nested within locations), season and the interaction of sampling location and season on the metatranscriptome. We selected the first three PCos as they explained, combined, 83% of the total variance and all had eigenvalues greater than five. As replicate was found to be not significant, then we used a one-way ANOVA in R software version 3.6.1 to compare PCo1, PCo2 and PCo3 among sampling seasons, sampling locations (beaches) and the interaction of season and sampling location. We used a Kruskal-Wallis test in R software version 3.6.1 to test for significant temporal (season) and spatial (different locations; CH, SL and SP) variation in the KEGG pathways.

Temporal and spatial differential transcription analyses: The change in the differential transcription patterns over the three seasons was visualized with fitted model Mean-Difference (MD) plots. To control for the effects of multiple simultaneous tests, we corrected the false discovery rate (FDR) using the Benjamini-Hochberg method (Chen et al., 2016). Gene transcripts with $\log_2$-fold-changes greater than 1.5 fold, present in at least two different species in the BC, with FDR ≤ 0.001 and P-values ($p < 0.05$) were considered as significantly differentially expressed among different seasons. The magnitude of the differential transcription over three sampling locations was visualized with a fitted model MD plots. We also performed spatial differential

transcription analysis using the metatranscriptomic data of the three sampling locations (CH, SL and SP) by determining transcripts with greater than 1.5 log$_2$-fold-change that were present in at least two different species of BC with FDR $\leq$ 0.001 and p < 0.05 as evidence for significant variation among sampling locations. The transcripts of the genes which showed significant variation in the differential expression analysis and were detected at least in two different species of bacteria were normalized to an internal control housekeeping gene (DNA-directed RNA polymerase beta' subunit; *rpoB*) (Nieto et al., 2009) in each sample. The transcript levels were further compared using a Kruskal-Wallis test (R software version 3.6.1).

Virulence gene transcripts and taxa associated with water degradation: As we detected a high number of virulence gene transcripts (cyanotoxins and bacteria virulence genes), as well as transcripts for taxa associated with water degradation (cyanobacteria and waterborne pathogens), we chose to not perform individual gene transcript tests for temporal and spatial variation. Instead, the abundance of cyanotoxin genes (normalized based on *rpoB* gene sequence read number) was combined for each season and location and we used the one-way ANOVA test in R software version 3.6.1 to determine significant temporal (seasonal) and spatial (different beaches). We also combined the abundance of virulence genes (normalized based on *rpoB* gene sequence read number) for each season and location to determine significant temporal (seasonal) and spatial (different beaches) variation using the one-way ANOVA. Taxonomic data of the *Cyanobacteria* community were also combined by season and location. Taxonomic data of waterborne pathogens (no *Cyanobacteria*) were also combined for each season and location. We determined significant temporal (seasonal) and spatial (different beaches) variation of the *Cyanobacteria* community (combined data) and waterborne pathogens (no *Cyanobacteria*) using one-way ANOVA. We report the number of virulence gene transcripts/sample for winter, summer and fall over three sampling locations (beaches). All plots and figures were generated using OriginPro 2018 (OriginLab Corporation, Northampton, MA, USA) or R software version 3.6.1 using ggplot2 package (Wickham, 2011).

Environmental data: Paired t-tests were used to compare environmental variable values (water temperature, daylight, TN and TP) among sampling locations (CH, SL,

SP) and seasons (winter, summer and fall) using IBM® SPSS statistics 19.0 (Somer, NY, USA). Metatranscriptomic data of the replicates were combined for each sample to test the effect of the environmental variables on transcript abundance. To test for the relationship between the environmental variables and variation in the metatranscriptome, we used a RELATE analysis. For this analysis, the Bray-Curtis dissimilarity of the metatranscriptomic data of the samples was correlated to the Euclidean distance of normalized environmental variables of the samples (combined variables as an environmental matrix) to calculate Spearman's p correlation coefficient. A distance-based linear model (distLM) was used to analyze and model the relationship between the metatranscriptomic data of the samples and the environmental variables using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK).

4.3 Results

4.3.1 Metatranscriptome sequencing and *de novo* assembly

The mean sequence length of the pair-end transcript reads was 119 ± 26 bp. Sequencing statistics of the metatranscriptomic profiles obtained from the Illumina HiSeq 4000 run for all samples are provided in Table S4.1 (Appendix C). Altogether, the metatranscriptomic run resulted in 29–35 million reads for each of the 18 samples. After removing low-quality transcripts and rRNA sequences, $268,704 \pm 82,600$ transcript reads/replicate were processed for functional annotation and assigned through the KEGG database (Appendix C; Table S4.1). After removing the transcripts belonging to archaea, eukaryotes and viruses, and the transcripts with unknown function, 4,574 genes (combined all isoforms from a unique gene) clusters (~2.9 million combined) were identified in 2,142 different bacterial species and analyzed in detail.

4.3.2 Season and location effects

The nested one-way ANOVA showed that season had a strong significant effect on the PCo1, PCo2 and PCo3 of the Bray-Curtis dissimilarity matrix form the BC metatranscriptome (Table 4.1). Sampling location also had a significant effect but compared to season the effect size was small (Table 4.1). The interaction effect of season and sampling location was significant, but also explained a small proportion of

the variance in the metatranscriptome (Table 4.1), indicating that season had similar effects across the three locations.

Table 4.1. Effect of season, sampling location and their interaction on first three principle coordinates of the Bray-Curtis dissimilarity matrices for the bacterial meta-transcriptome. PCo1, PCo2 and PCo3 explained a combined 83% of the total variance.

Variables		Season	Sampling location (replicate nested)	Sampling location* season
	df	2	2	4
PCo1	F value	9456	89.9	8
	P value	**1.12e-15**	**1.13e-06**	**0.009**
	R^2 value	0.77	0.09	0.11
PCo2	F value	221	25	11
	P value	**2.22e-08**	**0.0001**	**0.001**
	R^2 value	0.41	0.1	0.08
PCo3	F value	356	85	49
	P value	**1.57e-06**	**2.68e-09**	**4.05e-06**
	R^2 value	0.32	0.14	0.09

4.3.3   Bacterial community composition

Proteobacteria were significantly more common in the metatranscriptome (based on 16S rRNA sequence data) in winter (60%) compared to summer (46%, p=0.007) and fall (38%, p=0.001), and it was higher, but not significantly different (p=0.08) in the summer compared to fall (Appendix C; Figure 4.2A). *Cyanobacteria* was significantly more abundant in summer (13%, p=0.001) and fall (14%, p=0.001) relative to winter (4%) (Appendix C; Figure 4.2A). The abundance of *Actinobacteria* was significantly lower in winter (11%, p=0.02) and summer (12%, p=0.02) relative to fall (25%), while the relative abundance of *Bacteroidetes* was higher but not significantly so in winter (18%, p=0.07) and summer (17%, p=0.07) compared to fall (11%). *Proteobacteria* and *Cyanobacteria* constituted 48-49% and 9.8-11% of the BC community based on 16SrRNA in the metatranscriptome across all three locations (CH, SL and SP) with no significant variation (p>0.05) among the sampling locations. The abundance of *Actinobacteria* was marginally higher, but not significantly different, in SL (18%) relative to CH (16%, p=0.87) and SP (14%, p=0.8) while the abundance of

Bacteroidetes was lower but also not significantly different, in SL (12%) relative to CH (16%, p=0.8) and SP (17%, p=0.75) (Appendix C; Figure 4.2A).

We identified 22 different species of *Cyanobacteria* (Figure 4.2B); however, we combined the relative abundance of the different species for each location and season to evaluate their temporal and spatial variation on a broad scale (see above). Our analyses of the *Cyanobacteria* community revealed no significant (p>0.05) spatial variation; however, temporal (seasonal) variation was high. There was no significant difference in the relative abundance of *Cyanobacteria* species between summer and fall (p=0.55); however, the relative abundance of *Cyanobacteria* species was significantly higher in summer (p=0.003) and fall (p=0.004) relative to winter (Figure 4.2B).

We identified 22 (potential) pathogens in the SSU rRNA reads; however, 7 of them (*Acinetobacter johnsonii, A. lwoffii, Campylobacter coli, C. fetus, Francisella novicida, Shigella dysenteriae* and *S. sonnei*) had relative abundances of less than <0.1 in the collected samples. We combined the relative abundance of the different species for each sample to test their temporal and spatial variation on a broad scale. There was no spatial variation (p>0.05) in relative abundance of detected pathogens (combined data) among the three different sampling locations (CH, SL and SP beaches). The relative abundance of detected pathogens (combined data) was significantly higher in summer (p=0.001) and fall (p=0.003) relative to winter, while there was no significant differences between the abundance of combined pathogens between summer and fall (p=0.43) (Figure 4.2C).
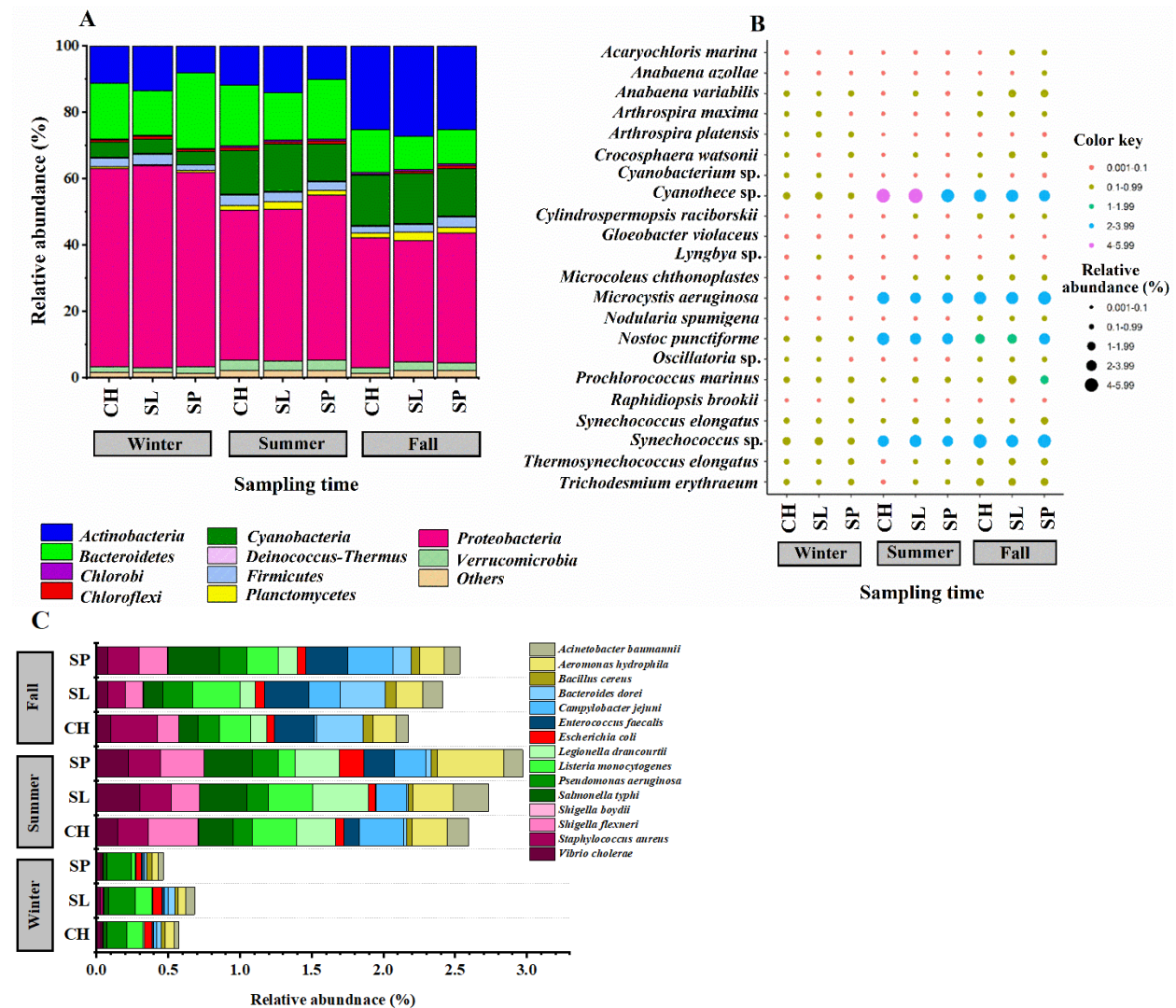
Figure 4.2. Distribution and composition of active bacteria based on 16S rRNA sequence reads in the metatranscriptome of the collected freshwater samples. Panel A: Relative abundance across all locations and seasons sampled (phyla with <1% relative abundance are combined and displayed as "others"). Panel B: Relative abundance of different species of genus *Cyanobacteria* across the samples from three different locations (CH, SL and SP) over winter, summer and fall 2019. Panel C: Relative abundance of 15 pathogens detected in CH, SL and SP over winter, summer and fall. Average values of the relative abundance of replicates are shown (C).

4.3.4   Metatranscriptome description

Our primary analysis based on differential transcription analyses revealed strong seasonal clustering of the metatranscriptomic data of the collected samples from

different locations with no obvious variation between biological replicates of each sample (Figure 4.3A and 4.3B). Differential transcription analyses of summer, relative to winter showed 2,362 upregulated and 1,292 downregulated genes (Figure 4.3C), consistent with greater activity in the summer season. Differential transcription analyses of summer, relative to fall revealed 676 upregulated and 315 downregulated genes by comparing the metatranscriptome data of these two seasons, again consistent with seasonal-related differences in BC activity. Most of the transcripts (3,097 out of 4,574) did not show significant variation between summer and fall (Figure 4.3C). Differential transcription analyses of fall, relative to winter showed 1,738 upregulated and 1,270 downregulated genes (Figure 4.3C). Pairwise differential expression analysis of the metatranscriptomic data of CH versus SP showed significant upregulation ($p<$ 0.05, FDR <0.001 and 1.5-fold change) of 31 and 68 unique transcripts in the transcriptomes of CH and SP respectively. Pairwise differential expression analysis of the metatranscriptomic data of SL versus SP showed significant upregulation of 72 and 115 unique transcripts in SL and SP respectively (Figure 4.3D). There was no significant variation in the treansciption level of different genes between SL and CH (Figure 4.3D).
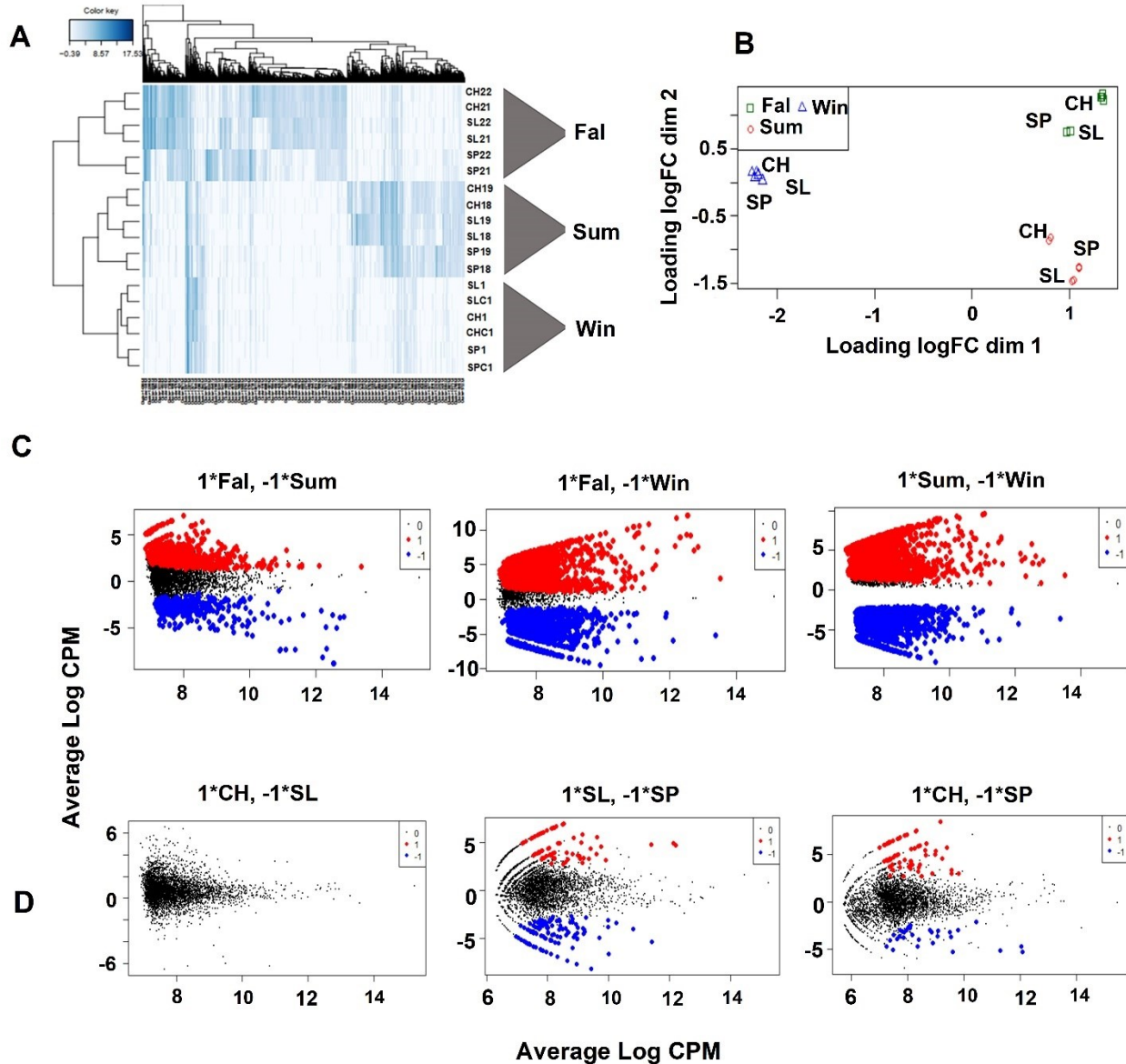
Figure 4.3. Comparison of the transcription levels across the metatranscriptome of seasonally (winter, summer and fall) sampled bacterial communities across three sampling locations (CH, SL and SP). Panel A: Heat map of the normalized transcripts reads of metatranscriptomic datasets for three different sampling locations over winter, summer and fall. Panel B: The MDS plot of the data sets showing a seasonal variation of the collected samples transcriptome. Panel C: Pairwise MD plot showing the temporal (seasonal) log-fold change and average abundance of each transcript (log CPM) of fall versus summer, fall versus winter and summer versus winter. Panel D:

Pairwise MD plot showing the spatial log-fold change and average abundance of each gene (log CPM) of CH versus SL, SL versus SP and CH versus SP. Significantly upregulated (red; 1) and downregulated (blue; -1) DE genes among pairwise contrasts along with the genes with no significant variation (black, 0) are highlighted (C). Abbreviations; Fal; Fall, Sum; Summer and Win; Winter.

4.3.5 Temporal variation

Temporal variation of the BC transcriptome: Principal coordinates analysis (PCoA) based on the Bray–Curtis dissimilarity matrix generated using the metatranscriptomic data (transcript counts) showed three distinct clusters matching the three sampling seasons (Appendix C; Figure S4.1A). PCo1, PCo2 and PCo3 explained 61.1%, 15.4% and 6.4% respectively, of the total variation (83%). There were significant seasonal effects for PCo1 (df=2, F=575, p<0.00001) and PCo2 (df=2, F=30, p=0.0001) but not for PCo3 (df=2, F=1.5, p=0.56) (Appendix C; Figure S4.1B).

We identified 107 KEGG pathways in the metatranscriptome; however, we focused on the top 50 KEGG pathways. Notably, photosynthesis, ribosome activity and oxidative phosphorylation (the pathways with the top three most abundant transcripts) were transcribed significantly higher in both summer and fall in comparison with winter (Figure 4.4). In total, 17 pathways were significantly upregulated in summer and fall compared to winter (Figure 4.4) – most of these were involved in metabolism, cell cycle, biosynthesis, photosynthesis, oxidative phosphorylation and transport. Only two pathways were significantly upregulated in winter relative to summer and fall (bacterial secretion system and systemic lupus erythematosus) and only two pathways were significantly upregulated in fall (i.e. glycosyl phosphatidy linositol biosynthesis and focal adhesion) relative to winter and summer (Figure 4.4).
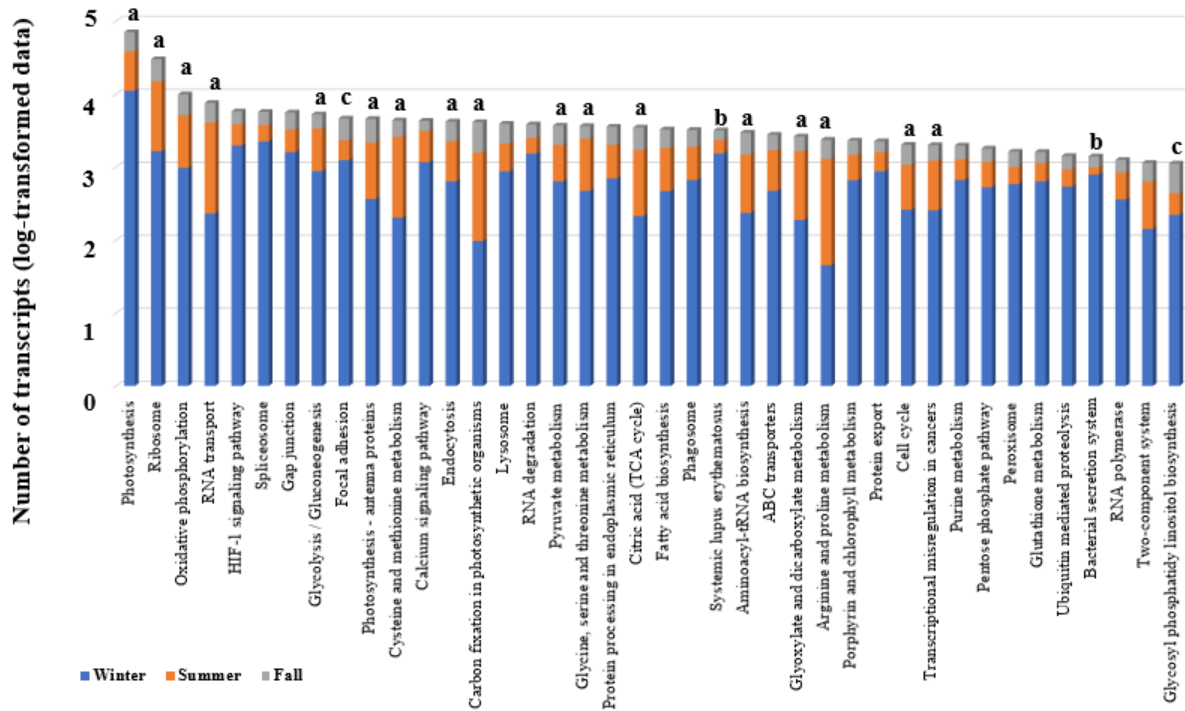
Figure 4.4. Out of top 50 abundant KEGG pathways, the 40 most abundant KEGG pathways in the metatranscriptomic data of winter (blue), summer (green) and fall (red) are displayed. The pathways marked with the letters were over-expressed (p<0.05) in the; summer and fall versus winter (a), winter versus fall and summer (b) and fall versus winter and summer (c).

Temporal variation in physiological and ecological activity-related gene transcription: In total 4,574 genes were detected in the metatranscriptome data. We selected 52 genes for transcription analyses that fulfilled the criteria which we defined before. Transcripts for the selected genes showed significant temporal variation in their transcription (Table 4.2). Out of the 52 selected genes, the transcription level of 23 (44%) genes were higher in summer versus fall and winter (Table 4.2), and the transcription of 19 (37%) genes was higher in both summer and fall versus winter. More interestingly, out of the 52 selected genes; the transcription of only 3 genes showed significant upregulation in the fall relative to summer (as well as being upregulated in summer relative to winter). Those transcripts were identified as nitrogen pathway genes, specifically, *narB* (assimilatory nitrate reductase)*, nirA* (nitrite reductase) and *nifH* (nitrogenase) (Table 4.2). Differential expression analyses of the

52 selected genes also revealed that the summer BC was significantly over-transcribing the genes involved in the metabolism of amino acids such as arginine, glutamate, histidine and homocysteine compared to winter and fall (Table 4.2). Also, the freshwater BC significantly over-expressed the transcripts of genes involved in the biosynthesis of arginine, cysteine, methionine, proline and serine in summer and fall relative to winter. More interestingly the BC metatranscriptome of summer exhibited significantly elevated levels of the transcripts of the *lpdA* and *sdhA* genes (citric acid cycle), *gpmA*, *gap* and *moxF* (glycolysis), *gltA2* and *gcl* (glyoxylate cycle) and *fdsD*, *moxF*, *metF* and *argD* (methylotrophy) relative to winter and fall (Table 4.2). Differential expression analysis also showed that the transcripts of *psaA* (photosystem I) and *psbA*, *psbA1*, *psbA2*, *psbD1*, *psb27* and *psbO* (photosystem II) genes were significantly elevated in summer and fall, compared to winter. Different species of *Cyanobacteria* such as *Anabaena* sp., *Gloeobacter* sp., *Microcystis* sp., *Prochlorococcus* sp., *Synechococcus* sp., *Synechocystis* sp. and *Thermosynechococcus* sp. were identified as the primary sources of photosynthesis. Out of 52 transcripts, we only identified overexpression of the transcripts of the *exeD* (Type II secretion system protein D), *pufA* (Light-harvesting protein) and *cspA*, *cspC* and *cspG* (cold shock-like proteins) genes in winter, relative to summer and fall (Table 4.2). Out of 52 transcripts, 31 (60%) genes which were part of physiological pathways (i.e. amino acid metabolism, Calvin cycle, glycolysis, methylotrophy, phototrophy, polyamine degradation, transport and translation) and nutrient cycling (nitrogen, phosphate assimilation) were detected, at low levels, in the winter metatranscriptomic data; however, transcripts of the genes involved in citric acid cycle and glyoxylate cycle were not detected in the winter at all (Table 4.2).

Table 4.2. Temporal variation in transcript abundance in the bacterial meta-transcriptome for biological/ecological relevant genes across three sampling seasons. We only present genes with known function that were part of well-characterized physiological and ecological pathways and were expressed in at least two different species of bacteria.

Pathways	Function	Gene(s)	Winter	Summer	Fall
Amino acid metabolism	Glutamate synthase	*gltB*	-	+ [s]	+
	L-histidine	*hisG*	-	+ [s]	+
	Cysteine synthase	*cysK*	+	+ [sf]	+ [sf]
	L-homocysteine	*ahcY*	-	+ [s]	+
	L-arginine	*argG*	+	+ [sf]	+ [sf]
	S-adenosyl methionine synthase	*metK*	+	+ [sf]	+ [sf]
	Proline-tRNA ligase	*proS*	+	+ [sf]	+ [sf]
	Serine-glyoxylate aminotransferase	*argD*	+	+ [sf]	+ [sf]
	Iron-sulfur cluster biosynthesis	*iscS*	+	+ [s]	+
Bacterial secretion system	Type II secretion system protein D	*exeD*	+ [w]	+	+
Calvin cycle	Ribulose bisphosphate carboxylase	*cbbL*	+	+ [sf]	+ [sf]
Citric acid cycle	Dihydrolipoyl dehydrogenase	*lpdA*	-	+ [s]	+
	Succinate dehydrogenase	*sdhA*	-	+ [s]	+
	Succinate dehydrogenase iron-sulfur subunit	*sdhB*	-	+ [sf]	+ [sf]
	Succinyl-CoA ligase	*sucD*	-	+ [sf]	+ [sf]
Heat shock	Cold shock-like protein	*cspA, cspC, cspG*	+ [w]	+	+
Glycolysis	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	*gpmA*	-	+ [s]	+
	Glyceraldehyde-3-phosphate dehydrogenase	*gap*	-	+ [s]	+
	Methanol dehydrogenase	*moxF*	+	+ [s]	+
	Phosphoglycerate kinase	*pgk*	-	+ [sf]	+ [sf]
	Pyruvate dehydrogenase	*pdhA*	+	+ [sf]	+ [sf]

	Pyruvate kinase	*pyk*	+	+^{sf}	+^{sf}
Glyoxylate cycle	Citrate synthase 1	*gltA2*	-	+^{s}	+
	Glyoxylate carboligase	*gcl*	-	+^{s}	+
	Malate synthase	*mls*	-	+^{sf}	+^{sf}
	Serine glyoxylate aminotransferase	*sgaA*	-	+^{sf}	+^{sf}
Methylotrophy	Formate dehydrogenase	*fdsD*	+	+^{s}	+
	Methanol dehydrogenase	*moxF*	+	+^{s}	+
	Methylene tetrahydrofolate reductase	*metF*	+	+^{s}	+
	Serine-glyoxylate aminotransferase	*argD*	+	+^{s}	+
Nitrogen	Assimilatory nitrate reductase	*narB*	+	+^{s}	+^{ff}
	Nitrite reductase	*nirA*	+	+^{s}	+^{ff}
	Nitrogenase (N fixation)	*nifH*	+	+^{s}	+^{ff}
Phosphate assimilation	8-phosphate phosphatase KdsC	*kdsC*	+	+^{s}	+
	Alkaline phosphatase synthesis	*phoP*	+	+^{s}	+
Phototrophy	Light-harvesting protein	*pufA*	+^{w}	+	+
	Photosystem I	*psaA*	+	+^{sf}	+^{sf}
	Photosystem II	*psbA, psbA1, psbA2, psbD1, psb27, psbO*	+	+^{sf}	+^{sf}
Polyamine degradation	Deoxyhypusine synthase	*dys2*	+	+^{s}	+
	Spermidine/putrescine transport system permease	*potC*	+	+^{s}	+
Serine cycle	Serine hydroxymethyltransferase	*glyA*	+	+^{sf}	+^{sf}
Transport	Vitamin B12 transporter BtuB	*btuB*	+	+^{sf}	+^{sf}
	Biopolymer transport protein ExbB	*exbB*	+	+^{sf}	+^{sf}
	Sulfate transport system permease protein CysT	*cysT*	+	+^{sf}	+^{sf}
Translation	Translation initiation factor IF-1	*infA1*	+	+^{s}	+

'+' indicates occurrence in winter, summer and fall. Letters indicate significantly higher transcript frequency based on differential expression analysis. sf; high expression in summer and fall versus winter, w; high expression in winter versus summer and fall, s; high expression in summer versus winter and fall, ff; high expression in fall versus summer and winter.

Temporal dynamic of the transcripts associated with water degradation: The transcript of three cyanotoxin genes belong to two different *Cyanobacteria* species, specifically, *M. aeruginosa* (*mcyA* gene) and *Synechococcus* sp. (*mcyD* and *mcyE*) were detected in the metatranscriptomic data of three seasons. There was a significant temporal variation in the level of cyanotoxin transcripts (df=2, F=203, p<0.00001). Post-hoc test revealed that the level of cyanotoxin transcripts (combined data for each season) was significantly higher in fall relative to summer (p=0.001) and winter (p<0.00001) but there was no significant (p=0.16) variation in the level of cyanotoxin transcripts between in summer relative to winter (Figure 4.5A).
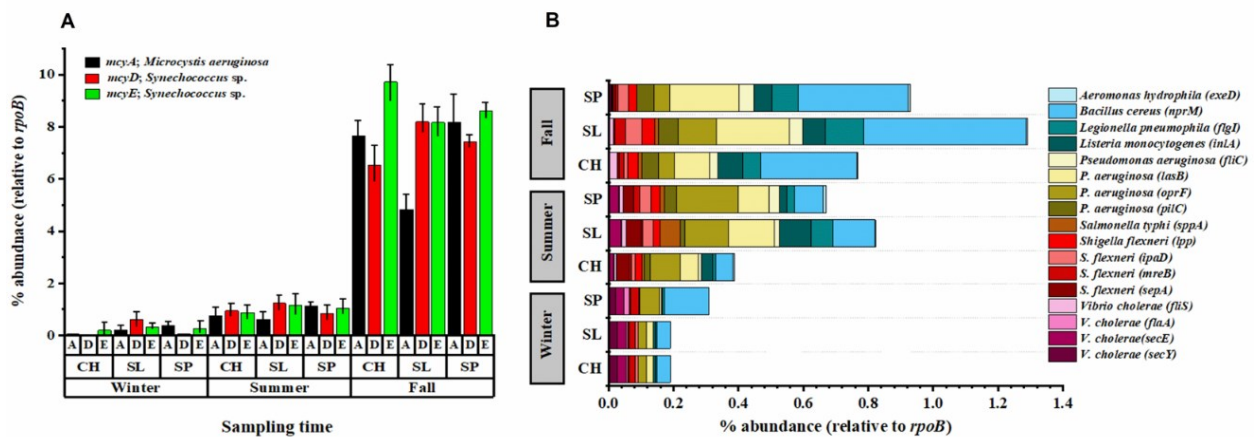
Figure 4.5. The abundance of transcripts (relative to *rpoB*) of three different cyanotoxins (*mcyA*, *mcyB* and *mcyC*) detected in the transcriptome data of CH, SL and SP beaches over winter, summer and fall are displayed in panel A. Error bars in panel A show the standard deviation of the transcript abundance between the replicates. The abundance of transcripts (relative to *rpoB*) of the genes associated with virulence belong to waterborne bacterial pathogens which were detected in the metatranscriptomic datasets of three locations (CH, SL and SP) over winter, summer and fall are presented in panel B. The average relative abundance of the transcripts read from the triplicates are shown.

In the metatranscriptomic data, 17 different unique transcripts assigned to virulence genes belonging to 8 different pathogens were detected. Out of the 17 unique transcripts, five (*exeD* (*Aeromonas hydrophila*), *fliC* and *pilC* (*P. aeruginosa*), *spaA* (*Salmonella typhi*) and *fliS* (*Vibrio cholerae*)) genes were not detected in winter at any location. However, the transcripts of the 17 different virulence genes were detected in

both summer and fall (Figure 4.5B). We did not test temporal variation for transcripts of the 17 different virulence genes individually (due to statically power limitations); however, when we combined the transcript data of all virulence genes detected in three sampling locations within each season, we found significant variation in the abundance of virulence genes transcripts (df=2, F=10.3, p<0.0001). Post-hoc test revealed that abundance was significantly higher in summer relative to winter (p<0.0001) but not to fall (p=0.4). Additionally, the transcript abundance (combined data) was also significantly higher in fall relative to winter (p=0.008) (Figure 4.5B).

4.3.6 Spatial variation

Spatial variation of the BC transcriptome: Principal coordinates analysis (PCoA) based on the Bray–Curtis dissimilarity matrix generated using the metatranscriptomic data (transcript sequence read abundance) showed no distinct clusters for sampling locations (Appendix C; Figure S4.1A). There were no significant location effects (p>0.05) for PCo1 (df=2, F=0.25, p=0.56), PCo2 (df=2, F=0.1, p=0.87) and PCo3 (df=1, F=0.08, p=0.92) (Appendix C; Figure S4.1B). Pairwise comparisons of the 50 top KEGG pathways between sampling locations (CH, SL and CH) showed significantly location effects on transcription (p<0.05) in only 8 of the top 50 KEGG pathways (Appendix C; Figure S4.2). Indeed, pairwise comparison of the metatranscriptomic data of SL and CH showed no significant differences among the top 50 KEGG pathways. The photosynthesis pathway (top ranked KEGG pathway) was over-transcribed (p<0.05) in SP versus CH and SL. Moreover, aminoacyl-tRNA biosynthesis, fatty acid biosynthesis, glycolysis (pyruvate metabolism), porphyrin metabolism and purine metabolism pathways were significantly over-transcribed in SP relative to CH and SL (Appendix C; Figure S4.2). There were only two KEGG pathways (i.e. ribosome and two-component system) which showed a significant elevation in the transcript read number in CH versus SP (Appendix C; Figure S4.2). Overall, very few of the pairwise spatial comparisons resulted in significant location effects on the transcription of the top 50 KEGG pathways – indicating little functional variation among the three sample site BCs.

Spatial transcriptional variation at genes data associated with physiological and ecological activity: We identified 17 transcripts assigned to genes within the

selected 52 ecologically/physiologically relevant genes with significant spatial variation. None of those transcripts showed significant differences between the metatranscriptomes of SL and CH (Appendix C; Table S4.2). In total, the transcripts of 14 genes involved in physiological pathways in the BC metatranscriptomes were significantly overexpressed in SP relative to CH and SL. None of the transcripts of the genes which were identified as contributing to nutrient cycling (nitrogen and phosphate) showed significant spatial variation. The transcripts of only 3 genes (*rpoB*, *phoP* and *rstA*) were overexpressed in CH in comparison to SP (Appendix C; Table S4.2).

Spatial dynamic of the transcripts associated with water degradation: There was no significant spatial variation in the transcript abundance of the combined cyanotoxin genes (*mcyA, mcyD* and *mcyE*) among the three different sampling locations in winter (df=2, F=0.07, p=0.65), summer (df=2, F=3.1, p=0.07) and fall (df=2, F= 3.6, p=0.05). There was also no significant variation in the transcript abundance of the virulence genes among locations in winter (df=2, F=0.04, p=0.96), summer (df=2, F=0.3, p=0.65) and fall (df=2, F=0.05, p=0.98) when the transcripts data of the virulence genes were combined in each season (Figure 4.5).

Environmental Effects: Our analysis showed that season had a strong effect on the freshwater BC metatranscriptome. We collected environmental variable information to test the effect of these variables on the BC metatranscriptome to perhaps explain some of the temporal variation observed in this study. No precipitation was recorded by Environment of Canada (http://climate.weather.gc.ca) either during any of the sampling days, or in the 24 hours prior to sampling. Mean (±SD) water temperatures at the time of sampling were 1.33±0.15 °C, 23±0.5 °C and 12.5±0.5 °C for winter, summer and fall respectively (Appendix C; Table S4.3). The mean water temperature was significantly higher in summer relative to fall (p=0.02) and winter (p=0.0001). Also, the mean water temperature was significantly higher in fall (p=0.0009) relative to winter. The level of total nitrogen (TN) was significantly higher in summer (mean=0.9±0.08; p= 0.02) and fall (mean=1.9±0.5; p= 0.04), relative to the winter (mean=0.2±0.02), while TN level was not significantly different between summer and fall (p=0.06). The level of total phosphorus (TP) was significantly higher in fall

(mean=0.1±0.03; p= 0.04) and summer (mean=0.09±0.4; p=0.03) relative to winter (mean=0.03±0.005), while TP between fall and summer was not significantly different (p=0.05). The Euclidean distance of environmental variables (water temperature, daylight, TN and TP) was significantly correlated (r=0.47, p=0.01) with the Bray-Curtis dissimilarity matrix of metatranscriptomic data. Also, our distLM analysis showed that environmental variables had a significant effect (r^2=0.82, p < 0.05) on the Bray-Curtis dissimilarity matrix of metatranscriptome. Among the environmental variables, TP explained only 3% of total variance explained by environmental variables (82%) and had no significant effect (p=0.22) on the metatranscriptome data. On the other hand, water temperature (49%), daylight (15%) and TN (15%) all had significant effects (p< 0.05) on the metatranscriptome (Appendix C; Table S4.3, Figure S4.2). There was no significant variation (p > 0.05) of water temperature, daylight, TN and TP among three sampling locations (CH, SL and SP) within any season.

4.4 Discussion

In this study, we characterized the temporal and spatial variation of the freshwater BC metatranscriptome with a particular focus on gene transcripts associated with ecosystem and human health in recreational fresh water. We detected strong temporal variation associated with seasonal changes (winter, summer and fall), limited spatial variation (sampling locations) and minor interaction effects of season and sampling location on the metatranscriptome data. This pattern may be attributed to the expected high seasonal variation associated with environmental variables critical to BC composition and function. Spatial variation, on the other hand, may require a greater distribution of sampling sites to have a large effect on the critical environmental variables, and hence on the BC metatranscriptome. This study provides unique information regarding the combination of spatial and temporal effects on the freshwater BC which is critical for ecological and human health reasons (Antwis et al., 2017). In our study, environmental variables (water temperature, daylight, TN and TP) explained 82% of the observed variation in the metatranscriptomic data. The strong "season" effect detected in this study is likely driven by environmental variables such as water temperature, day/night cycle length and nitrogen levels. In line with our study, others also suggested that temporal variation of environmental variables such as light (day

versus night) (Poretsky et al., 2009), water temperature (Lei and Lai, 2019) and nutrients (Ren et al., 2019) have a significant effect on the BC metatranscriptome and function.

A key goal of this study was to characterize the temporal variation of the freshwater BC metatranscriptome over winter (no ice-covered condition), summer and fall under non-cHAB conditions. Few studies have addressed metatranscriptomic temporal variation of the freshwater BC at such a large temporal scale. For example, Huntscha et al, reported temporal variation of the BC metatranscriptome (Lake Greifensee) from March to August with a high level of activity (defined as upregulated transcription) of the BC in warm versus cold months, particularly high degradation and consumption of glyphosate as an alternative phosphorus source in warm temperature (Huntscha et al., 2018). This is consistent with our observed high levels of transcription at genes encoding physiological and ecological function in the summer and fall, relative to the colder winter sampling. The 52 physiologically and ecologically relevant genes were selected as they were either part of key physiological pathways or involved in the basic ecological services/functions of the freshwater BC. Transcription levels of nitrogen cycle genes in our study were high in the warmer seasons (summer and fall) versus cold season (winter), again consistent with published reports from Lake Chaohu (China) (Fan et al., 2019) and a constructed wetland (Columbia, MO) (Sims et al., 2012). In our study, season also had a large influence on the transcription profiles of the ecologically relevant genes. For example, we found that the transcription level of nitrogen cycle genes was significantly higher in fall compared to summer.

Out of the 52 ecologically relevant genes, only 31 (60%) were detected in the winter metatranscriptome, highlighting the low ecological activity at very cold temperatures and low light levels. Butler et al, (2019) reported high nitrification activity of freshwater BCs under ice-covered conditions (Butler et al., 2019); however, we only detected the transcripts of different genes involved in nitrogen and phosphate cycles in winter (although we had no ice coverage) indicating low activity relative to other seasons. This may be due to the role played by ice covering - indeed ice duration drives the accumulation of nitrate, and potentially other nutrients such as phosphate, in north temperate lakes (Powers et al., 2017). Taken together, our data show that lack of ice

cover and environmental conditions such as low temperature, low light and limited inputs of nutrients are the main drivers of low activity of freshwater BC in winter. On the other hand, low transcription of the biologically related genes in winter might be due to a transition of many taxa of the freshwater BC into a dormant, viable but non-culturable state (VBNC) (McDougald et al., 1998). Transcription of the genes encoding cold-shock proteins (*cspA, cspC, cspG*) were significantly elevated in our winter metatranscriptome, which could explain how the freshwater BC shifts their transcriptomic profile toward genes which may protect freshwater BC at cold temperatures (Chen et al., 2012; De Maayer et al., 2014). Transcription of the genes associated with arginine and proline biosynthesis were significantly downregulated in winter, perhaps due to the reported lower proline/arginine content in proteins during the winter, making the downregulation a possibly adaptive response to low-temperature conditions in psychrophilic taxa (Ayala-del-Río et al., 2010; Chen et al., 2012).

We found *Cyanobacteria* as one of the most important taxa in our sampled BCs, with roles both in ecological and physiological community activities (e.g., photosynthesis, nitrogen cycle, etc.). In fact, the majority of photosynthesis genes (*psaA, psbA, psbA1, psbA2, psbD1, psb27, psbO*) were expressed by the *Cyanobacteria* community. Interestingly, many nitrogen cycle genes (*narB, nirA* and *nifH*) were also predominantly expressed by *Cyanobacteria*, highlighting the importance of *Cyanobacteria* activity in the ecology and function of freshwater ecosystems (Scott and Marcarelli, 2012), at least under non-cHAB conditions.

Our study supports limited spatial variation in aquatic metatranscriptomes, at least at the scale of our three sampling locations (CH, SL and SP beaches). Consistent with our observations, gene transcription analysis of sediment BCs collected from Lake Erie and Lake St. Clair showed only minor spatial variation of BC activity between two lakes (VanMensel et al., 2020). This limited spatial variation in both water column and sediment BCs may be attributed to the connection between Lake Erie and St. Clair through the Detroit River; however, that connection should not have much effect on sediment BCs. This connection may have, however, resulted in a very similar water conditions due to mixing and hence with similar environmental conditions (i.e., no significant spatial variation of water temperature, daylight and nutrient levels) which

could lead to the low spatial variation in BC function as measured by gene transcription. Composition analysis of freshwater BC (bi-weekly sampling for 15 months) of four public beaches located in Lake Erie and other two public beaches located in Lake St. Clair (Chapter 2 data, unpublished) also showed limited spatial variation. This pattern may be due to proximity of our sampling locations, and generally similar environmental conditions (water temperature, nutrient levels, etc.). Nevertheless, we still were able to capture some spatial variation in the gene transcription profiles. For example, the upregulation of the genes involved in aminoacyl-tRNA biosynthesis, fatty acid biosynthesis, glycolysis, phototrophy, porphyrin metabolism and purine metabolism at the SP site (urban impacted area) compared to the CH and SL sites (agriculture impacted area) could be due to higher loading of nutrients to SP from the urbanized area as reported previously for Lake Michigan (Fisher et al., 2015).

It was not our intention to study the BC composition with metabarcoding in this study; however, we did use 16S rRNA transcripts to assess the taxonomic profile of the BCs. Our previous studies using metabarcoding and 16S rRNA NGS did not provide high-resolution taxonomic information (at the species level) to capture pathogen community composition in freshwater samples (Shahraki et al., 2020). However, by analyzing the transcript sequences from the 16S rRNA gene in the metatranscriptome data, we were able to identify cyanotoxin producing *Cyanobacteria* species as well as 22 different human pathogens and gene transcripts for cyanotoxin genes and 17 human virulence genes. Both taxa and gene transcripts associated with water degradation and human health risk showed substantial temporal (seasonal) variation. Upregulation of microcystin genes transcript in the fall (August to October) versus summer (June and July) during a cHBA bloom has been reported before (Tang et al., 2018); however, our results also showed elevated transcription of cyanotoxin genes in the fall compared to summer under non-cHBA bloom conditions. Out of the 17 different virulence-associated genes detected in our metatranscriptomic data, 5 of them did not occur in the winter, highlighting the negative impact of low water temperature on the transcription of virulence genes, as previously reported (Feng et al., 2016). Studies have shown that beach sediment could be a reservoir for a wide range of pathogens (Mohiuddin et al.,

2017) and may explain the consistent presence of some of the pathogens reported here. For example, *Salmonella* sp. virulence genes (*pipB2* and *sspH2*) and *Bordetella pertussis* (*hlyB* and *cyaB*) have been reported in the metatranscriptome of the sediment at public beaches in Lake Erie (Kingsville beach) and Lake St. Clair (Bell River beach) (VanMensel et al., 2020). However, to the best of our knowledge, this study is the first to report seasonal variation in the transcription of bacterial virulence genes associated with water quality degradation in a natural system. Some of the pathogens detected in our study, such as *Salmonella* and *Shigella*, are not commonly believed to grow in freshwater environments, instead they are thought to be typically introduced into aquatic ecosystems by surface and subsurface runoff, wastewater and agricultural discharge, or avian/animal excrement (Bibi et al., 2015; Field and Samadpour, 2007). The presence of *B. dorei* (a human-specific microbial source tracking marker) (Haugland et al., 2010) in our metatranscriptomes also suggests potential pollution of freshwater samples with sewage. However, most of the other pathogens detected in this study (such as *A. hydrophila, P. aeruginosa, Listeria monocytogenes* and *V. cholera*) are known to survive both in sewage and in aquatic ecosystems (Bridle, 2014). Our data also suggest that toxin-producing *Cyanobacteria* and pathogenic organisms were present and active in the water column of freshwater samples collected from three public beaches at the time of sampling, particularly in summer and fall, but they can survive and exhibit transcriptional activity even at very cold temperatures.

4.5   Conclusion

Our study provides insight into temporal (three seasons; winter, summer and fall 2019) and spatial (three sampling locations) variation in the freshwater BC metatranscription profiles. We also showed that seasonal variation in environmental variables such as water temperature (the main driver of the transcriptomes variation) and nutrients level can contribute to the regulation of gene transcription of the freshwater BC, and can have a strong impact on the temporal variation of the BC metatranscriptome in freshwater ecosystems. Our transcriptomic analysis revealed that ecologically relevant genes and the genes associated with water quality degradation are very sensitive to seasonal changes, but less so to spatial effects, at least at our spatial scales (maximum distance; ~50 km between SP and SL). This study is a pioneering

study focusing on temporal and spatial variation in the function of critical physiological and ecological genes in freshwater BC, which will serve to direct future research into BC function in relatively unstressed ecosystems. Our data provide more information regarding the fact that gene transcription profiles are useful to not only evaluate the ecological activity in the BC, but also can capture the underlying pathogenic potential of aquatic ecosystems when considering human health risks at public beaches for recreation water use.

4.6   References

Allison, S.D., Martiny, J.B., 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105(Supplement 1) 11512-11519.

Antwis, R.E., Griffiths, S.M., Harrison, X.A., Aranega-Bou, P., Arce, A., Bettridge, A.S., Brailsford, F.L., de Menezes, A., Devaynes, A., Forbes, K.M., 2017. Fifty important research questions in microbial ecology. FEMS Microbiology Ecology 93(5) fix044.

Ayala-del-Río, H.L., Chain, P.S., Grzymski, J.J., Ponder, M.A., Ivanova, N., Bergholz, P.W., Di Bartolo, G., Hauser, L., Land, M., Bakermans, C., 2010. The genome sequence of Psychrobacter arcticus 273-4, a psychroactive Siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. Applied and Environmental Microbiology 76(7) 2304-2312.

Bartram, J., Chorus, I., 1999. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. CRC Press.

Beeton, A.M., 1965. Eutrophication of the St. Lawrence Great Lakes. Limnology and Oceanography 10(2) 240-254.

Berdjeb, L., Ghiglione, J.F., Domaizon, I., Jacquet, S., 2011. A 2-year assessment of the main environmental factors driving the free-living bacterial community structure in Lake Bourget (France). Microbial Ecology 61(4) 941-954.

Bibi, F., Qaisrani, S., Ahmad, A., Akhtar, M., Khan, B., Ali, Z., 2015. Occurrence of Salmonella in freshwater fishes: A review. Journal of Animal and Plant Sciences 25(3) 303-310.

Blaser, M.J., Cardon, Z.G., Cho, M.K., Dangl, J.L., Donohue, T.J., Green, J.L., Knight, R., Maxon, M.E., Northen, T.R., Pollard, K.S., 2016. Toward a predictive understanding of Earth's microbiomes to address 21st century challenges. mBio, pp. e00714-00716.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15) 2114-2120.

Bridle, H., 2014. Overview of waterborne pathogens, Waterborne Pathogens. Elsevier, pp. 9-40.

Briland, R.D., Stone, J.P., Manubolu, M., Lee, J., Ludsin, S.A., 2020. Cyanobacterial blooms modify food web structure and interactions in western Lake Erie. Harmful Algae 92 101586.

Bullerjahn, G.S., McKay, R.M., Davis, T.W., Baker, D.B., Boyer, G.L., D'Anglada, L.V., Doucette, G.J., Ho, J.C., Irwin, E.G., Kling, C.L., 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. Harmful Algae 54 223-238.

Butler, T.M., Wilhelm, A.-C., Dwyer, A.C., Webb, P.N., Baldwin, A.L., Techtmann, S.M., 2019. Microbial community dynamics during lake ice freezing. Scientific Reports 9(1) 1-11.

Castello, L., Macedo, M.N., 2016. Large-scale degradation of Amazonian freshwater ecosystems. Global Change Biology 22(3) 990-1007.

Chen, Y., Lun, A.T., Smyth, G.K., 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research 5.

Chen, Z., Yu, H., Li, L., Hu, S., Dong, X., 2012. The genome and transcriptome of a newly described psychrophilic archaeon, M ethanolobus psychrophilus R 15, reveal its cold adaptive characteristics. Environmental Microbiology Reports 4(6) 633-641.

Davidson, N.M., Oshlack, A., 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. Genome Biology 15(7) 1-14.

Davis, C.C., 1964. Evidence for the eutrophication of Lake Erie from phytoplankton records. Limnology and Oceanography 9(3) 275-283.

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during Planktothrix blooms in Sandusky Bay, Lake Erie. Environmental Science & Technology 49(12) 7197-7207.

De Maayer, P., Anderson, D., Cary, C., Cowan, D.A., 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO Reports 15(5) 508-517.

Dodds, W.K., Perkin, J.S., Gerken, J.E., 2013. Human impact on freshwater ecosystem services: a global perspective. Environmental Science & Technology 47(16) 9061-9068.

Fan, Y.-Y., Li, B.-B., Yang, Z.-C., Cheng, Y.-Y., Liu, D.-F., Yu, H.-Q., 2019. Mediation of functional gene and bacterial community profiles in the sediments of eutrophic Chaohu Lake by total nitrogen and season. Environmental Pollution 250 233-240.

Feng, B., Guo, Z., Zhang, W., Pan, Y., Zhao, Y., 2016. Metabolome response to temperature-induced virulence gene expression in two genotypes of pathogenic Vibrio parahaemolyticus. BMC Microbiology 16(1) 1-10.

Field, K.G., Samadpour, M., 2007. Fecal source tracking, the indicator paradigm, and managing water quality. Water Research 41(16) 3517-3538.

Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa (Washington, DC) 3.

Geist, J., 2011. Integrative freshwater ecology and biodiversity conservation. Ecological Indicators 11(6) 1507-1516.

Glasl, B., Webster, N.S., Bourne, D.G., 2017. Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. Marine Biology 164(4) 91.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology 29(7) 644.

Haugland, R.A., Varma, M., Sivaganesan, M., Kelty, C., Peed, L., Shanks, O.C., 2010. Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. Systematic and Applied Microbiology 33(6) 348-357.

Healy, D., Chambers, D., Rachol, C., Jodoin, R., 2008. Water quality of the St. Clair River, Lake St. Clair, and their US tributaries, 1946− 2005: US Geological Survey Scientific Investigations Report 2007− 5172. Reston, VA.

Ho, J.C., Michalak, A.M., 2015. Challenges in tracking harmful algal blooms: A synthesis of evidence from Lake Erie. Journal of Great Lakes Research 41(2) 317-325.

Huntscha, S., Stravs, M.A., Bühlmann, A., Ahrens, C.H., Frey, J.r.E., Pomati, F., Hollender, J., Buerge, I.J., Balmer, M.E., Poiger, T., 2018. Seasonal Dynamics of Glyphosate and AMPA in Lake Greifensee: Rapid Microbial Degradation in the Epilimnion During Summer. Environmental Science & Technology 52(8) 4641-4649.

Ishii, S., Sadowsky, M.J., 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes and Environments 23(2) 101-108.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. Journal of Applied Microbiology 123(3) 570-581.

Jenny, J.-P., Anneville, O., Arnaud, F., Baulaz, Y., Bouffard, D., Domaizon, I., Bocaniov, S.A., Chèvre, N., Dittrich, M., Dorioz, J.-M., 2020. Scientists' Warning to Humanity: Rapid degradation of the world's large lakes. Journal of Great Lakes Research 46 686-702.

Karatayev, A.Y., Burlakova, L.E., Mehler, K., Bocaniov, S.A., Collingsworth, P.D., Warren, G., Kraus, R.T., Hinchey, E.K., 2018. Biomonitoring using invasive species in a large lake: Dreissena distribution maps hypoxic zones. Journal of Great Lakes Research 44(4) 639-649.

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28(24) 3211-3217.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10(3) R25.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.L., 2014. Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. The ISME Journal 8(8) 1715-1726.

Lei, K.-H., Lai, H.-T., 2019. Effects of sunlight, microbial activity, and temperature on the declines of antibiotic lincomycin in freshwater and saline aquaculture pond waters and sediments. Environmental Science and Pollution Research 26(33) 33988-33994.

Loftin, K.A., Graham, J.L., Hilborn, E.D., Lehmann, S.C., Meyer, M.T., Dietze, J.E., Griffith, C.B., 2016. Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. Harmful Algae 56 77-90.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24(8) 813-821.

MacKenzie, S.H., 1997. Toward integrated resource management: lessons about the ecosystem approach from the Laurentian Great Lakes. Environmental Management 21(2) 173-183.

McDougald, D., Rice, S.A., Weichart, D., Kjelleberg, S., 1998. Nonculturability: adaptation or debilitation? FEMS Microbiology Ecology 25(1) 1-9.

McKenna Jr, J.E., 2019. The Laurentian Great Lakes: A case study in ecological disturbance and climate change. Fisheries Management and Ecology 26(6) 486-499.

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

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.-K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A., 2013. Sizing up metatranscriptomics. The ISME Journal 7(2) 237-243.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75(1) 14-49.

Nieto, P.A., Covarrubias, P.C., Jedlicki, E., Holmes, D.S., Quatrini, R., 2009. Selection and evaluation of reference genes for improved interrogation of microbial transcriptomes: case study with the extremophile Acidithiobacillus ferrooxidans. BMC molecular biology 10(1) 1-11.

Pattis, I., Moriarty, E., Billington, C., Gilpin, B., Hodson, R., Ward, N., 2017. Concentrations of Campylobacter spp., Escherichia coli, Enterococci, and Yersinia spp. in the Feces of Farmed Red Deer in New Zealand. Journal of environmental quality 46(4) 819-827.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., Moran, M.A., 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environmental Microbiology 11(6) 1358-1375.

Powers, S.M., Labou, S.G., Baulch, H.M., Hunt, R.J., Lottig, N.R., Hampton, S.E., Stanley, E.H., 2017. Ice duration drives winter nitrate accumulation in north temperate lakes. Limnology and Oceanography Letters 2(5) 177-186.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35(21) 7188-7196.

Ren, Z., Qu, X., Zhang, M., Yu, Y., Peng, W., 2019. Distinct bacterial communities in wet and dry seasons during a seasonal water level fluctuation in the largest freshwater lake (Poyang Lake) in China. Frontiers in Microbiology 10 1167.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1) 139-140.

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11(3) 1-9.

Scavia, D., Allan, J.D., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S.B., Briland, R.D., Daloğlu, I., DePinto, J.V., 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central basin hypoxia. Journal of Great Lakes Research 40(2) 226-246.

Scavia, D., DePinto, J.V., Bertani, I., 2016. A multi-model approach to evaluating target phosphorus loads for Lake Erie. Journal of Great Lakes Research 42(6) 1139-1150.

Schindler, D.W., Carpenter, S.R., Chapra, S.C., Hecky, R.E., Orihel, D.M., 2016. Reducing phosphorus to curb lake eutrophication is a success. ACS Publications.

Scott, J.T., Marcarelli, A.M., 2012. Cyanobacteria in freshwater benthic environments, Ecology of Cyanobacteria II. Springer, pp. 271-289.

Shahraki, A.H., Chaganti, S.R., Heath, D.D., 2020. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microbial Ecology 1-13.

Sims, A., Gajaraj, S., Hu, Z., 2012. Seasonal population changes of ammonia-oxidizing organisms and their relationship to water quality in a constructed wetland. Ecological Engineering 40 100-107.

Singh, B.K., Dawson, L.A., Macdonald, C.A., Buckland, S.M., 2009. Impact of biotic and abiotic interaction on soil microbial communities and functions: A field study. Applied Soil Ecology 41(3) 239-248.

Soller, J.A., Eftim, S., Wade, T.J., Ichida, A.M., Clancy, J.L., Johnson, T.B., Schwab, K., Ramirez-Toro, G., Nappier, S., Ravenscroft, J.E., 2016. Use of quantitative microbial risk assessment

to improve interpretation of a recreational water epidemiological study. Microbial Risk Analysis 1 2-11.

Steffen, M.M., Belisle, B.S., Watson, S.B., Boyer, G.L., Bourbonniere, R.A., Wilhelm, S.W., 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. Applied and Environmental Microbiology 81(9) 3268-3276.

Tang, X., Krausfeldt, L.E., Shao, K., LeCleir, G.R., Stough, J.M., Gao, G., Boyer, G.L., Zhang, Y., Paerl, H.W., Qin, B., 2018. Seasonal gene expression and the ecophysiological implications of toxic Microcystis aeruginosa blooms in Lake Taihu. Environmental Science & Technology 52(19) 11049-11059.

Vahtera, E., Autio, R., Kaartokallio, H., Laamanen, M., 2010. Phosphate addition to phosphorus-deficient Baltic Sea plankton communities benefits nitrogen-fixing Cyanobacteria. Aquatic Microbial Ecology 60(1) 43-57.

VanMensel, D., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2020. Exploring bacterial pathogen community dynamics in freshwater beach sediments: A tale of two lakes. Environmental Microbiology 22(2) 568-583.

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56 44-66.

Wattigney, W.A., Li, Z., Ragin-Wilson, A., 2017. The biomonitoring of Great lakes populations program. Journal of Environmental Health 79(8) 42.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3(2) 180-185.

Wingett, S.W., Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research 7.

Yosef, N., Regev, A., 2011. Impulse control: temporal dynamics in gene transcription. Cell 144(6) 886-896.

NUTRIENT STRESS DRIVES ADAPTIVE CHANGES IN FRESHWATER BACTERIA COMMUNITIES: RESPONSE TO SECONDARY STRESS MEDIATED BY HORIZONTAL GENE FLOW WITH REDUCED COMMUNITY CHANGE

5.1   Introduction

A stressor (perturbation or disturbance) in the environment can change the bacterial community (BC) composition drastically either permanently or temporarily (Shade et al., 2012). However, the form and magnitude of the BC response at the community and individual taxon level might differ depending on the stressor. Following a change in the environment, ecosystem function might be disrupted either because the species composition changes, or because component species fail to adapt to the new environment (Lawrence et al., 2012). Specifically, the microbial composition might change to reflect 1) resistance (no changes in composition and function), 2) resilience (short-term changes in composition but quickly recover to its initial composition), 3) functional redundancy (permanent changes in the composition but still providing the same function) or 4) alterations in both composition and function (Allison and Martiny, 2008). Irrespective of the type of response to the stressor, the stress process might prepare the community to be more resilient to further, and possibly more severe, stressors: this is the basis of adaptive bacterial community change.

Eutrophication of freshwater lakes such as the Laurentian Great Lakes (LGLs) ecosystems due to anthropogenic activity is one of the most challenging of the global environmental problems (Bhagowati and Ahamad, 2019). Nutrient enrichment has had many undesirable effects on LGL ecosystems, including changes in microbial species composition, microbial community shifts to bloom-forming algal species, reductions in species diversity, water degradation, oxygen depletion, etc. (Smith and Schindler, 2009). Many studies report nutrient enrichment can cause shifts in the freshwater ecosystem structure (Haukka et al., 2006) and function, which persists over extended periods of time (regime shifts) and may be abrupt (Hecky et al., 2010) and sometimes catastrophic (Scheffer et al., 2001). Environmental field surveys identified significant shifts in the BC composition in response to a range of environmental changes such as methane enrichment (Crevecoeur et al., 2015), solar UVR (Sarmento et al., 2015),

urbanization (Li et al., 2020) among others. Those environmental stressors may drive changes in the nature of the BC species interactions, biodiversity and the stability of the microbial ecosystems. Freshwater monitoring has also shown that nutrient enrichment can change the composition of the *cyanobacteria* community from *Microcystis* in the low-phosphate regions to *Anabaena* and *Planktothrix* in the high-phosphate regions of Lake Erie (Harke et al., 2016). An experimental study showed that adding nutrients (Nitrogen; N and Phosphate; P) to freshwater (collected from Sandusky Bay; Lake Erie) can cause toxic cyanobacterial blooms dominated by *Planktothrix* (Davis et al., 2015). Additionally, nutrient overloading has been reported to cause negative interactions among species of the BC (made up of 8 different species of bacteria) which led to less diverse communities (Ratzke et al., 2020). However, the nature of potentially negative impacts of different concentrations of nutrients on natural freshwater BC composition is not clear.

The response of microbes to disturbance is mediated by various strategies at the individual (cellular) level (such as plasticity and adaptive gene expression), the population level (such as adaptive composition changes, modified meta-gene expression and changes in dispersal patterns), and the community level (such as diversity, compositional and functional changes) (Shade et al., 2012). All of these changes are facilitated by the fast population growth rates and generally high mutation rates of microorganisms (Allison and Martiny, 2008). Additionally, horizontal gene transfer (HGT) through mechanisms such as conjugation, transformation, transduction (Johnsen and Kroer, 2007) and mutation (Swings et al., 2017) is generally believed to be important components of bacterial adaptation to environmental stress. Mobile genetic elements such as plasmids (Tschäpe, 1994), conjugative transposons (Wozniak and Waldor, 2010) and integrons (Gillings, 2014), among others, are known to play important roles in HGT in the majority of aquatic ecosystems. Selection pressures imposed by anthropogenic pollutants such as chemicals, antibiotics and heavy metals result in rapid BC evolution and the spread of class 1 integrons among taxa in both clinical and environmental settings (Gillings et al., 2015). In fact, HGT can serve as a proxy metric for human impacts on natural environments. On the other hand, gene transfer agents (GTAs), virus-like particles, are now maintained in the genomes of

some bacteria (well characterized in *alphaproteobacteria*) and archaea (Lang et al., 2017). GTA-mediated gene transfer frequencies have been documented to be a thousand to a hundred million times higher than estimates of HGT in the oceans, with as high as 47% of the culturable natural microbial community confirmed as gene recipients (McDaniel et al., 2010). However, the impact of nutrient loading on HGT rate in freshwater BCs is not well known.

As freshwater BCs have fundamental roles in biogeochemical cycling, organic matter decomposition, food web function, and others, they influence the aquatic ecosystem function at a fundamental level (Cotner and Biddanda, 2002). Characterizing the effects of nutrient loading on BC composition and their potential adaptive responses can provide a deeper insight into the processes and mechanisms operating in lake ecosystems. This would improve our basic knowledge regarding the flexibility and tolerance of BCs against anthropogenic stressors. Although some studies have addressed the negative impact of nutrient overloading on freshwater ecosystems, they mainly focused on *Cyanobacteria* communities and *Cyanobacteria* harmful algal blooms (cHABs) (Gobler et al., 2016; Steffen et al., 2017). However, little is known about possible adaptive responses of the entire freshwater BC to stressors, and how this adaptation can help the adapted BC to deal with the highest levels of stress encountered in severely impacted ecosystems. In this study, we used nutrient loading as a stressor in two experiments; the phase 1 experiments (adaptation phase) consisted of exposure of freshwater BCs to low and high levels of nutrients, followed by the phase 2 experiments (challenge phase) where the potentially adapted communities resulting from phase 1 exposure were challenged by exposure to very high levels of nutrients to assess BC response. This study aimed to study the adaptive response of freshwater BC to the different level of nutrients (low and high) as stressors by testing whether BCs exposed to different levels of stress (low and high nutrient loads) have different responses to the challenge (very high level of nutrient load). We hypothesized that: 1) BCs exposed to the high level of nutrient loading, relative to the low levels, will exhibit a reduced response to the very high challenge level of nutrient loading in the second phase experiments (where the reduced response would be indicated by reduced loss of diversity and reduces BC community composition change), 2) changes in biodiversity

and community composition resulting from the phase 1 experimental exposure to high and low nutrient loading would be irreversible, indicative of genetic shifts in the community and 3) the nutrient stressors would increase the levels of HGT marker expression in the BCs. This study thus provides a unique approach to developing predictive applications of adaptive response theory in natural freshwater bacterial communities.

5.2   Materials and methods

5.2.1   Study design, sample collection and treatments

This study had two experimental phases (adaptation and challenge phases). In the adaptation phase experiments (phase 1); we used a low nutrient dose (LND) and a high nutrient dose (HND) separately to treat natural BCs collected from two locations (one in Lake Erie and another one in Lake St. Clair, Ontario, Canada). The LND was 5X above the normal level of the measured nutrients (ambient) at the sampling sites, while the HND was almost 100X above the normal level of the measured nutrients at the sampling sites (Table 1). Phosphate (P) and nitrogen (N) levels in the HND were equivalent to the nutrient level reported for agriculturally (but not greenhouse) impacted water in Windsor Essex-County (Maguire et al., 2018). In the response challenge phase experiments (phase 2), we challenged the adapted BCs (post phase 1) with the highest level of nutrients; the challenge nutrient dose (CND) was designed to quantify the response of the adapted BCs. To make the CND, we used the highest level of nutrients recorded from multiple sources to the environment in Windsor Essex-County and are known to impact the freshwater ecosystem. We used P (2500X ambient) and N (820X ambient) concentrations for the CND, these levels are reported from Lake Erie tributaries influenced by greenhouses and agriculturally impacted areas in Windsor Essex-County (Maguire et al., 2018), combined with NH_4Cl (460X ambient), a typical concentration found in Windsor municipal wastewater treatment effluent (1999) (Table 5.1).

Freshwater samples were collected from from two sites; Lakeview Park (LP) beach in Lake St. Clair, which is proximal to an urban tributary (the Belle River connects with Lake St. Clair at LP beach; 42°17'51.8"N 82°42'42.6"W) and Seaclift beach (SL;

42°01'45.9"N 82°36'21.2"W) in Lake Erie, which is located in the greenhouse impacted area of Lake Erie. Sampling was performed on July 10th, 2017. From each site, 50 L water was collected from 0.5 m depth, and delivered to the laboratory on ice. The water samples were divided into two 25 L sub-samples, and one (25 L sub-sample from each location) was filtered using 0.2 μm polycarbonate membrane (Millipore, USA) to remove bacteria, and the filtered water was immediately stored at -20 ° C to use in the phase 2 experiment. We used the second sub-sample (unfiltered water sample; 25 L) from each location for the experiments in phase 1. All experiments were performed with 1.5 L of lake water in clear glass 2.0 L bottles, in triplicate. We used orthophosphate (KH_2PO_4; referred to as P from here on) and various forms of nitrate (a combination of $Ca(NO_3)_2$ plus KNO_3; ratio 1:1, referred to as N from here on) and ammonium (NH_4Cl) to make treatment doses. P (total) and total N (total) levels were measured at the two sample sites (LP and SL) on the day of sampling to define the ambient nutrient levels.

Table 5.1. Ambient nutrient levels at the sampling sites at SL (Lake Erie) and LP (Lake St. Clair), with the doses used for the low (LND) and high (HND) phase 1 exposures plus the challenge (CND) nutrient concentrations. Abbreviation: SD; standard deviation, LND; low nutrient dose, HND; high nutrient dose, CND; challenge nutrient dose, ND; not determined.

Nutrients	LP	SL	Mean (SD), mg/L	Adaptive phase (nutrients level; mg/L)		Challenge phase (nutrients level; mg/L)
				LND	HND	CND
P (total)	0.003	0.001	0.002±0.001	0.01	0.187	5.17 [b]
N (total)	0.09	0.13	0.11±0.05	0.58	11.57	89.97 [b]
NH₄Cl	ND	ND	0.06 [a]	0.3	6	27.7 [c]

[a] We did not measure NH_4Cl levels in the lake samples, we used NH_4Cl level measured by Windsor Essex-County (https://www.ontario.ca/data/drinking-water-surveillance-program) in drinking water (2012). [b] Levels measured in the streams in a greenhouse influenced area of Windsor Essex-County (Maguire et al., 2018). [c] Levels measured in Windsor municipal wastewater treatment effluent; (Canadian Environmental Protection Act (1999).

Phase 1 experiments (adaptation phase): This phase consisted of nine 1.5 L glass bottles filled with fresh water collected from each location (Total N=18). The freshwater BCs in the bottles were exposed to; 1) low nutrients dose (LND community; N = 3), high nutrients dose (HND community; N = 3), no addition of nutrient - control condition (P1-Co; N = 3), replicated for the two sample sites (Figure 5.1). Phase 1 experiments were run in a growth chamber programmed for 22 $^{\circ}$C with 15:9 hours light: dark to match average ambient conditions at the two sample sites. From each glass bottle, we collected a 150 mL sample and filtered it using 0.2 µm polycarbonate membrane (Millipore, USA) after 1, 2, 3 and 4 weeks of incubation, and the filters were stored at -20 $^{\circ}$C. We also filtered the initial water sample for each location before the experiment. At the end of the phase 1 experiment (at week 4), we transferred 50 mL of each incubated sample to new glass bottles filled with 1.45 L of the filtered lake water (stored at -20 $^{\circ}$C).

Phase 2 experiments (challenge phase): The experimental microcosms created at the end of the phase 1 experiments (see above) consisted of two types of BCs: 1) low nutrient adapted (LND), and 2) high nutrient adapted (HND). The phase 2 experiments were designed to test the "adapted" BCs for the altered response to additional, extreme nutrient stress, as well as relaxed selection (i.e., back to ambient conditions). Thus the LND BCs were transferred into two sets of the microcosms (glass bottles); the first set (three replicates) were treated with very high levels of nutrients (Table 5.1) as a challenge treatment (challenged community; C-LND), while the second set (three replicates) were transferred to the filtered and frozen lake water from the initial sampling, with no nutrient enhancement (relaxed community; R-LND). This was to study the BC dynamics after removing the stressor and thus constitute the relaxation of stress treatment (Figure 5.1). We used the same experimental design for the HND BCs from the phase 1 experiment to create triplicates of the nutrient challenge (challenged community; C-HND) and stress-relaxed (relaxed community; R-HND). We also transferred 50 mL of the control treatment microcosms from the phase 1 experiments to new microcosms (in triplicate) and filled to 1.5 L with the filtered sample site water with no nutrient treatment as the control in the phase 2 experiment (control community; P2-Co). For the phase 2 experiment challenge, we used a very high nutrient level dose,

described in Table 5.1. Sub-samples (150 mL) from each microcosm were collected from at the end of weeks 1, 2, 3 and 4 and filtered using 0.2 μm polycarbonate membrane (Millipore, USA) and the filters were stored at -20 ℃.
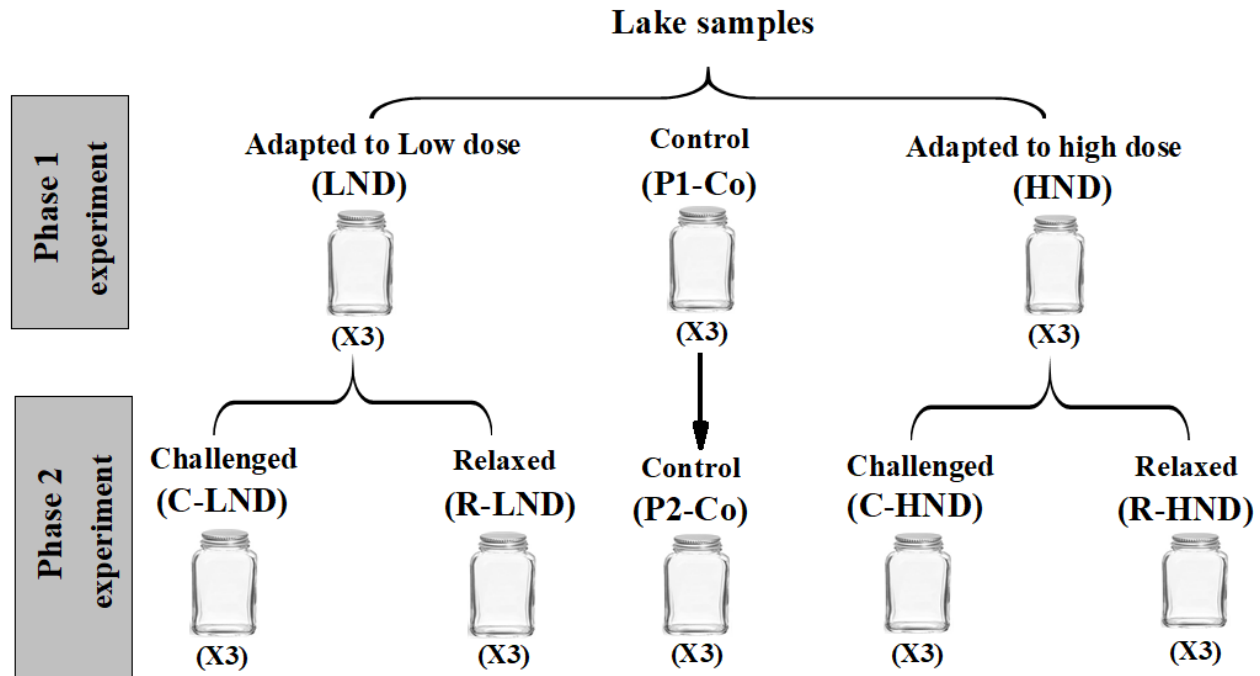
Figure 5.1. Schematic diagram of the designed experiment for the current study. Phase 1 experiment was designed to adapt the freshwater BCs to low (LND) and high (HND) doses of nutrients and phase 2 was designed to test the impact of very high nutrients dose (CND) on the adapted BCs.

5.2.2 DNA extraction, PCR, library preparation and sequencing

DNA was extracted from the filters using the sucrose lysis buffer method according to Shahraki et al, (Shahraki et al., 2018). The Genomic DNA was suspended in 50 μL TE buffer and kept at -80 °C until use. The extracted DNA samples (two out of three replicates; one replicate not extracted as backup) for each week over the two experimental phases were amplified for the V5-V6 region of the 16S rRNA gene using V5F (acctgcctgccg-ATTAGATACCCNGGTAG) and V6R (acgccaccgagc-CGACAGCCATGCANCACCT) primers (He et al., 2017) for metabarcoding. In total we PCR-amplified 124 BCs; 64 samples for each sample site (2 replicates x 4 weeks x 8 treatment communities; Figure 5.1). PCR products were cleaned using Sera-Mag

Magnetic Beads (GE, Healthcare Life Science, UK) and then used as templates for a second short-cycle PCR to ligate adaptor for next-generation sequencing (NGS) and barcode (for sample identification) sequences to create a library for NGS according to a previously described protocol (He et al., 2017). The products from the second, ligation, PCR for all samples were pooled and purified using QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). The concentration of purified PCR product mix (library) was measured on an Agilent 2100 Bioanalyzer with a High Sensitive DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library was then diluted to 60 pmol/L and sequenced on an Ion PGM™ System using the Ion PGM™ Sequencing 400 Kit and an Ion 318™ Chip (Thermo Fisher Scientific, Burlington, ON, Canada).

5.2.3 Transcription of horizontal gene transfer markers

Integrons are mobile genetic elements (and a key component of HGT) which have three core features including an integron-integrase gene (*int1*), a recombination site (*attI*) and a promoter (Gillings et al., 2015). We measured transcription of the *int1* gene using a 148 bp region of the gene (int1-Forward: 5'-GGCTTCGTGATGCCTGCTT-3' and int1-Reverse 5'-CATTCCTGGCCGTGGTTCT-3') (Di Cesare et al., 2016b) for qPCR. We also measured transcription of the GTA large subunit terminase gene using a 168 bp of the gene (GTAtermi-Forward: 5'-TGGTCTGGCCSAACGGSGC-3', GTAtermi-Reverse:5'-CCCAGCCGCAGCGCRAAYT-3') (McDaniel et al., 2010) for qPCR. Ten-fold serial dilutions were made from the original PCR product (5 dilutions) for each marker and used for qPCR assays to generate standard curves for each marker. The standard curve was created by linear regression of the threshold cycle (C_T) value obtained by qPCR versus known concentration to allow the calculation of the efficiency of the two assays (Bustin et al., 2009). As our original water samples collected from SL and LP were negative for *Vibrio cholerae ctxA* gene and the *Escherichia coli* O26 *manC* gene (confirmed by qPCR, details not shown), we added known concentrations of the amplified product of these two markers (2000 copies of each gene/reaction) to our extracted DNA samples as internal controls to evaluate the presence of inhibitors in the qPCR assays. As our goal was to determine the level of GTA terminase and *int1* transcription as a proxy for HGT activity, we compared the C_T values among the different treatment microcosms at week 4 in the phase 2 experiments;

this gave us an overview of the HGT response to challenge. SYBR green qPCRs were carried out on the DNA extracted from all replicate and treatment microcosms from the LP and SL samples (Shahraki et al., 2019). All qPCR reactions (with no template controls; NTC) were run in triplicate using MicroAmp Fast 96-well reaction plates and on a QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems). The mean C_T values of the different treatments were compareed using one-way ANOVA by using SPSS version 19 (SPSS Inc, Chicago, Illinois).

5.2.4 Sequence handling

The raw sequence data from the ION Torrent platform were de-multiplexed, quality filtered and trimmed of the adaptor, barcode and primer sequences using the Quantitative Insights into Microbial Ecology (QIIME V. 1.9.1) bioinformatics pipeline (Caporaso et al., 2010). A minimum quality score of Q=20 and base-pair cut-off of 200 bp were selected for quality assurance. Chimeras were identified and removed using ChimeraSlayer in QIIME. Operational taxonomic units (OTUs) were assembled based on sequence similarity (97%) among the sequence reads and then taxonomically assigned using BLAST (Edgar, 2010). The representative sequence for each OTU was selected using the most abundant method for assigning taxonomy in the RDP Classifier program with a minimum 80% confidence level (Edgar, 2010). After removing the singleton and doubleton OTUs, we defined an OTU as "abundant" when it had a relative abundance ≥1% of the community, "moderate" when the relative abundance was between 0.1-0.99% and "rare" when the abundance was < 0.1% (Logares et al., 2014). The OTU table was rarefied to 2000 quality-passed sequences for each sample to calculate alpha diversity values.

5.2.5 Statistical analyses

Microcosm replicates and sampling site effects: We used a generalized linear mixed model (GLMM) in SPSS version 19 (SPSS Inc, Chicago, Illinois) to determine whether the microcosm replicates and sampling sites (LP and SL) had significant effects on the BC variation in phase 1 and 2 separately. Alpha diversity indexes (Chao1 and Shannon) and the first (PCo1), second (PCo2) and third (PCo3) coordinates of the principal coordinate analysis (PCoA) of the Bray-Curtis dissimilarity matrices were

used separately as dependent variables for each model (phase 1 or phase 2), with replicate, sampling site and treatment (replicates nested within treatment) as fixed factors and week as a random factor in the GLMM. Bray–Curtis dissimilarity matrices based on the BCs were calculated and PCoA was performed using the program Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). We selected PCo1, PCo2, and PCo3 of the PCoA, as they had eigenvalues >5.0, for the analysis. To test for significant differences in the BC of two sampling locations (LP and SL), we compared the Bray–Curtis dissimilarity matrix of the BC of the original lake samples collected from two locations (LP and SL), and the BCs of LP and SL across all weekly samples in phase 1 and phase 2 experiments separately (4 weeks for each phase) using permutational multivariate analysis of variance (PERMANOVA) with 9999 permutations and Bonferroni correction in Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). To run PERMANOVA for each phase, replicates were nested within treatment, and week was considered as a random effect.

Adaptive (phase 1) experiment: We used a principle response curve (PRC) analysis (Van den Brink and Braak, 1999) with a Monte Carlo permutation test (with 9999 permutations) to evaluate overall changes in the community over time (4 weeks) for each sampling location individually (LP and SP) and the pooled data of the two locations (LP+SL) in the phase 1 experiment using Canoco 5.0 (Biometrics). We then focused on the BCs of week 4 in the adaptive phase 1 experiment. To test for adaptive changes (phase 1 experiment; week 4), we compared the BCs of P1-Co, LND, and HND using PERMANOVA. We compared Chao1 and PCo1, PCo2, and PCo3 of the week 4 phase 1 experiment microcosm BCs using one-way ANOVA in SPSS version 19 (SPSS Inc, Chicago, Illinois). A heat map was generated by agglomerative hierarchical clustering (unweighted pair group method with arithmetic mean; UPGMA) method for high abundant OTUs detected in the BCs week 4 using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). We also compared the relative abundance of the "abundant" (> 1%) OTUs of the BCs (week 4) using Linear discriminant analysis (LDA) as an effect size (LEfSe) method (http://huttenhower.sph.harvard.edu/lefse/) (Segata et al., 2011). Plots and graphs were generated using ggplot2 package (Wickham, 2011) in R and Origin Pro 2019.

Challenge response (phase 2) experiment: The overall response of the adapted BCs to challenge were evaluated by principal response curve (PRC) analysis (Van den Brink and Braak, 1999) over phase 2 (4 weeks) for each sampling location individually (LP and SP) and the pooled data of two locations (LP+SL). Then we focused more on the BCs of week 4 to test for the overall response of the adapted BCs to the relaxed and extreme nutrient challenge treatments (phase 2; week 4). We compared the BCs of the different challenge treatments (P2-Co, C-LND, R-LND, C-HND and R-HND) over 4 weeks by using PERMANOVA. We also compared the Chao1 and PCo1, PCo2, PCo3 of the week 4 phase 2 experiment microcosm BCs using one-way ANOVAs. A heat map was generated by UPGMA method for high abundant OTUs detected in the BCs week 4 using Primer-e software version 7.0.13 (Primer-E Ltd., Plymouth, UK). We also compared the relative abundance of abundant OTUs of the BCs (week 4) using LDA (Segata et al., 2011).

We also performed a network analysis of the BCs of phase 2 microcosms (week 4) to identify the keystone taxa and determine the pattern of correlations among the OTUs in each community. For the network analysis, we computed positive and negative interaction effects between all pairs of OTUs (>0.05% relative abundance) using SparCC, a python module for computing correlation coefficients (Friedman and Alm, 2012). Only the significant ($P<0.05$) and strongly positive and negative correlations ($|r| >0.7$) were included in creating edges in the networks. We identified those taxa that play central roles in community interactions, commonly referred as keystone species (Banerjee et al., 2018), as taxa whose nodes had a high degree of closeness (average distance of a node to any other nod; >70), and centrality (identify important nodes in a network; >0.36) and low betweenness centrality (the number of shortest paths between any two nodes in the graph passing through a node; <0.07) (Berry and Widder, 2014). These metrics illustrate both the number of connections and how important those connections are to the overall network (Banerjee et al., 2018; Berry and Widder, 2014). We used Cytoscape version 3.7.2 with the group attributes layout (Kohl et al., 2011) to visualize the OTU network OTUs. Plots and graphs were generated using ggplot2 package (Wickham, 2011) in R and Origin Pro 2019.

5.3 Results

Sequence handling: After quality control, the remaining Ion Torrent sequence reads for all samples were close to two million (1,916,287 reads). We removed the unassigned taxa sequences from the data set, as well as all OTUs with fewer than three reads across all samples. This resulted in individual replicate samples having 2142-8309 reads. When the OTU table was rarefied to 2000 reads/sample, a total of 4567 OTUs were retained. We mostly focused on the BCs of week 4 of phase 1 and 2 experiments in this study. In the adaptation (phase 1) experiment; we identified 787, 682 and 572 OTUs in the BCs of P1-Co, LND and HND respectively. In the challenge response (phase 2) experiment, we identified 620, 502, 670, 534 and 560 OTUs in the BCs of P2-Co, C-LND, R-LND, C-HND and R-HND respectively.

Replicate and sample site (Lake Erie and Lake St. Clair) effects: The GLMM model showed that the microcosm replicates had no significant effect on Chao1 (df=1, F=0.52, p=0.21), Shannon (df=1, F=0.42, p=0.25), PCo1 (df=1, F=0.12, p=0.69), PCo2 (df=1, F=0.01, p=0.92) and PCo3 (df=1, F=0.002, p=0.99). Sampling site also had no significant effect on Chao1 (df=1, F=0.82, p=0.14), Shannon (df=1, F=0.77, p=0.18), PCo1 (df=1, F=0.1, p=0.72), PCo2 (df=1, F=0.016, p=0.92) and PCo3 (df=1, F=0.008, p=0.98).

The original BCs of the two sampling sites (SL and LP) showed no significant difference in overall composition based on the Bray-Curtis dissimilarity matrix comparisons (PERMANOVA; df=1, F=0.25, p=0.92). There was also no significant difference between the BCs of the LP and SL across the phase 1 experiments (df=1, F=1.7, p=0.11) or the phase 2 experiments (df=1, F=2.03, p=0.07) by PERMANOVA. The PCoA analysis of the Barry-Curtis dissimilarity matrix of the BCs of LP and SL over two phases of the experiment (4 weeks) did not show a clear separation of the communities of two locations (Appendix D; Figures S5.1). Based on these results, we combined replicates and sampling sites for further analyses, except where noted.

5.3.1 Adaptation of the BC to the low and high level of nutrients

Adaptation at the community level: Principal response curve (PRC) analyses of the community of each location (LP and SL) separately showed that nutrient as stressor

had a significant effect on the BC of LP (F=2.6, p=0.009) and SL (F=2.2, p=0.01). As there were no significant differences between the BCs among two locations (LP and SL) across all treatments, we combined the data of two locations for each treatment; however, we show PRC results for sample sites separately and combine to illustrate the similarity of the two sample site responses. PRC analysis of combined data (LP+SL) also showed that nutrient as stressor had a significant effect (p<0.05) on the BC (F=2.3, p=0.009) (Figure 5.2). Monte Carlo permutation test indicated that only the effect size of first PRC was significant which explained 61%, 53% and 57% all variance for LP, SL and combined (LP+SL) data respectively. Second PRC only explained 11.9%, 9.8% and 10.3% of the variance for LP, SL and combined (LP+SL) data respectively and the effect size of treatment was not significant (p>0.05) (Figure 5.2).
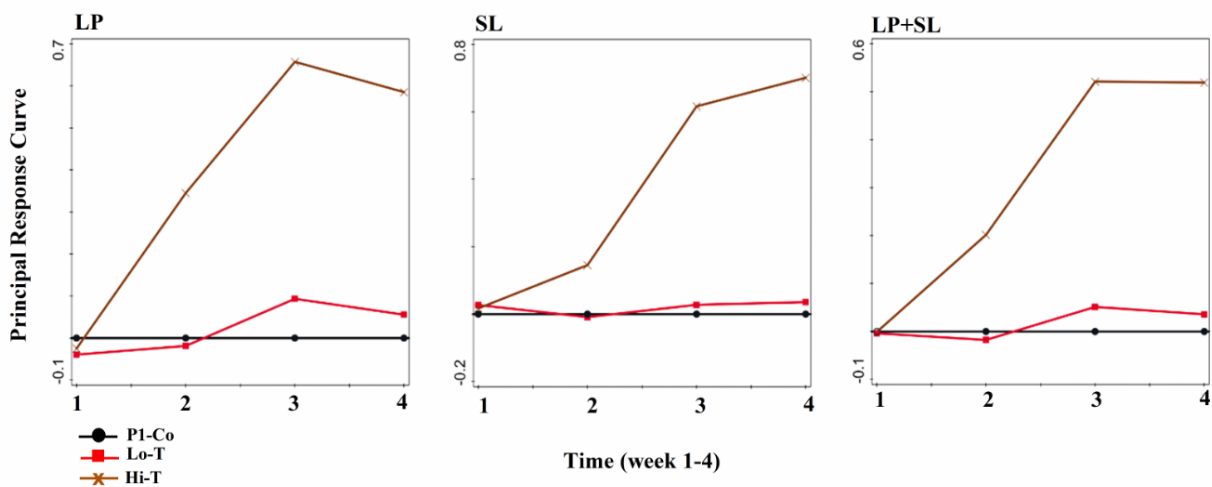
Figure 5.2. Principal response curves (PRC) resulting from the analysis of the BC response to low and high levels of nutrients in the adaptation (phase 1) experiments. The curves show the effects of nutrient additions as a stressor, relative to the control treatment (P1-Co), for the BC adapted to LND and HND of nutrients. The Y-axis indicates the difference from the control and X-axis indicates the time of sampling.

To test for the effects of the adaptive (phase 1) experiment, we compared the BC of the LND, HND and the control (P1-Co) for week 4 using PERMANOVA. There was a significant difference among the three BCs at the end of the phase 1 experiment (df=2,

F= 7.7, p=0.001). Pairwise comparisons of BCs for the two treatments and the control using PERMANOVA showed that the BC of HND was significantly different from the BC of both LND (F= 3.1, p=0.04) and P1-Co (F=3.7, p=0.02), while there was no significant difference between the BC of LND and P1-Co (F=1.5, p=0.08).

Adaptation at OTU and taxa levels: As the sampling site had no significant effect on diversity indexes or principle coordinates axes (PCoA1-PCoA3; see above), we combined the data for the two sampling sites for the analyses of Chao 1, PCo1, PCo2 and PCo3. PCo1, PCo2 and PCo3 explained 40, 22 and 7% of total variation (cumulative variation=69%) for the BCs of week 4. There were no significant differences between P1-Co and LND for Chao1, PCo1 and PCo2 (p > 0.05), while Chao1, PCo1 and PCo2 showed significantly (p < 0.05) lower values for HND compared to the other two microcosms using one-way ANOVA. The mean of PCo3 was significantly lower for HND in comparison to P1-Co but was not significantly different from LND (Appendix D; Figure S5.2).

In total, 13 OTUs were highly abundant (≥ 1%) in the BC of P1-Co which were assigned to different phyla. Out of those 13, only 3 (23%; OTUs 9, 27 and 129) belong to *Proteobacteria* (Figure 5.3). In the BC of LND, 15 OTUs were abundant of which 5 (34%; OTUs 23, 28, 30, 92, 129 and 145) belong to *Proteobacteria*. Nine out of 15 highly abundant OTUs of LND were also abundant in the BC of P1-Co. In the BC of HND, 20 OTUs were abundant of which 12 (60%; OTUs 5, 7, 19, 30, 32, 55, 56, 66, 85, 105, 121 and 452) belong to *Proteobacteria*. None of the abundant OTUs were common between P1-Co and HND, while OTU30 (phylum *Proteobacteria*) was the only highly abundant OTU shared between the LND and HND BCs. Taxonomic affiliations of highly abundant OTUs assigned to *Proteobacteria* were different in the HND from the *Proteobacteria* detected in the BC of P1-Co and LND (Figure 5.3).
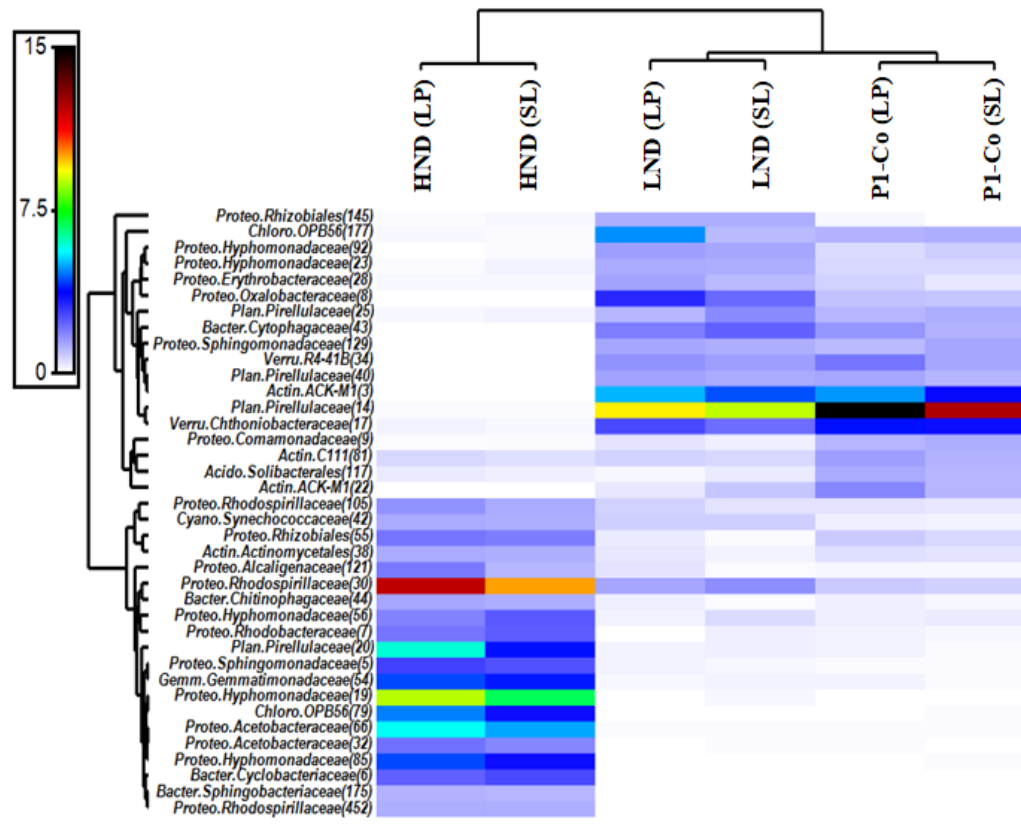
Figure 5.3. Hierarchical clustering heatmap of the abundant OTUs (OTU numbers are given in parenthesis) with their taxonomic information (phyla and family). Vertical clustering was based on the similarity in the abundance of OTUs. Horizontal clustering was based on the similarly in the relative abundance of OTUs for each sampling location and treatment. The color code indicates the relative abundance of each OTU. Although there were no significant differences between the community of two sampling locations at the whole community level – minor differences exist and that is why both locations are presented in the heat map. Taxonomic abbreviations: Acido; *Acidiobacteria*, Actin; *Actinobacteria*, Bacter; *Bacteroidetes*, Chloro; *Chloroflexi*, Cyano; *Cyanobacteria*, Gemm; *Gemmatimonadetes*, Plan; *Planctomycetes*, Proteo; *Proteobacteria* and Verru; *Verrucomicrobia*.

LDA analysis of the highly abundant OTUs for the week 4 community in the adaptive (phase 1) experiment revealed substantial, and significant ($p < 0.05$), variation among the BCs of P1-Co and LND (8 OTUs), P1-Co and HND (17 OTUs) and LND and HND (16 OTUs). More specifically, the relative abundances of OTUs 5, 8, 23 and

92 (all assigned to *Proteobacteria*) were significantly higher (p < 0.05) in the BC of LND treatment compared to P1-Co (Appendix D; Figure S5.3). Additionally, the relative abundances of OTUs 5, 56, 65, 85 and 452 (all assigned to *Proteobacteria*) also were significantly higher in the BC of HND treatment compared to P1-Co. More interestingly, the relative abundance of the OTUs assigned to *Actinobacteria* were significantly higher in the BC of P1-Co compared to that of the LND treatment (OTUs 22 and 81) and the HND treatment (OTUs 3, 22 and 81). Moreover, the relative abundance of the OTUs assigned to *Planctomycetes* were significantly higher in the BC of P1-Co compared to the BC of the LND (OTU 14) and HND (OTUs 14 and 40) treatments (Appendix D; Figure S5.3). The relative abundance of *Cyanobacteria* (OTU42; family *Synechococcaceae*, genus *Synechococcus*) was elevated relative to the BC of P1-Co (0.18%) 0.6% and 1.2% in the BC of LND and HND respectively (Figure 5.3).

5.3.2 Response of the adapted BCs to the challenge

Response at the community level: The adapted BCs from the phase 1 experiments (LND and HND) were exposed to very high levels of nutrient stress (Table 5.1) to study how previous exposure of low and high nutrient levels (phase 1 experiments) could modify the response of the BC to very high level of nutrients in the challenge (phase 2) experiments. Additionally, we cultured the adapted communities; LND and HND, under ambient conditions at the time of sampling the two sites (control) using preserved water from the initial sampling events. We predicted that the BCs exposed to high levels of nutrient stress in the adaptive experimental conditions would show reduced community change (diversity and composition) than those exposed to low levels of nutrient stress. PRC analysis of the BC for each location (LP and SL) separately showed that the challenge with very high levels of nutrient stress had a significant effect on the BCs of LP (F=3.5, p=0.002) and SL (F=3.1, p=0.003) (Figure 5.4). As there were no significant differences between the BC of two locations (LP and SL) in response to the challenge in the phase 2 experiments (df=1, F=1.08, p=0.1), we combined the data from the two locations for each treatment. PRC analysis of the combined data of week 4 of the phase 2 experiments (LP+SL) also showed that the challenge had a significant effect on the BCs (F=3.2, p=0.003) (Figure 6.4). Monte

Carlo permutation test indicated that only the effect size of the first PRC was significant which explained 52%, 48% and 44% all variance for LP, SL and polled (LP+SL) data. Second PRC only explained 10.1%, 8.2% and 8.9% of the variance for LP, SL and polled (LP+SL) data and the effect size of treatment was not significant (p>0.05).
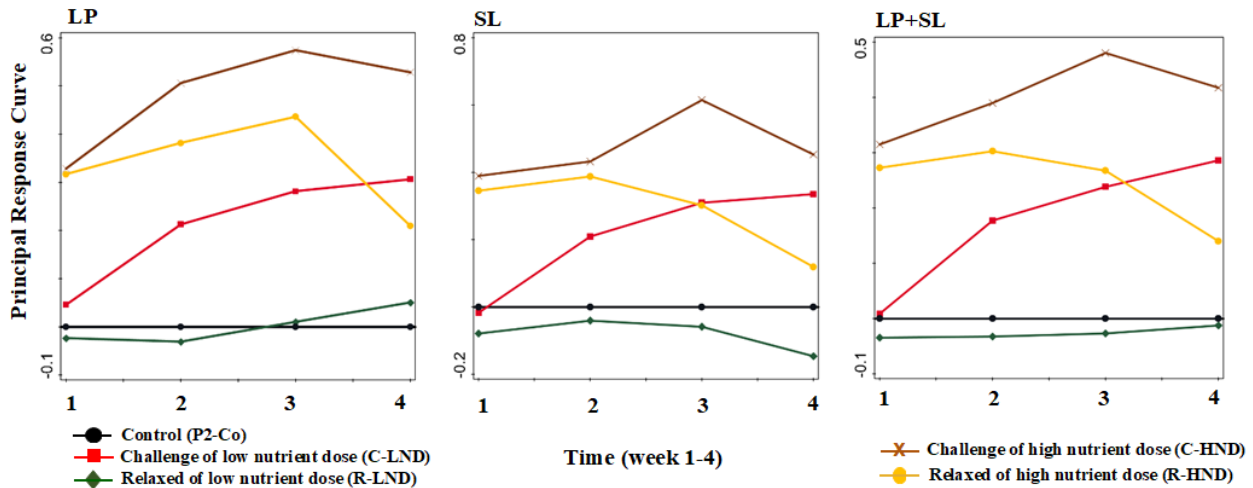
Figure 5.4. Principal response curves (PRC) resulting from the adapted BCs response to the very high levels of nutrients in the challenge phase, indicating the effects of nutrient additions as a stressor on the adapted BCs. The challenge of LND (C-LND) and HND (C-HND) along with their relaxed communities (R-LND and R-HND) were studied. The Y-axis indicates the difference from the control and X-axis indicates the time of sampling.

We also found significant differences between the challenged (phase 2) BCs (C-LND and C-HNT), the control BC (P2-Co) and the relaxed BCs (R-LND and R-HND) of week 4 by PERMANOVA (df=5, F=7.7; p=0.001). Pairwise PERMANOVA analysis of the communities showed that there was a significant difference among all treatments (Table 5.2).

Table 5.2. PERMANOVA analysis (pairwise) of the BC of different communities of week 4 (phase 2). Bonferroni corrected p values are given in above the diagonal and F values are provided below the diagonal.

	P2-Co	C-LND	R-LND	C-HND	R-HND
P2-Co		**0.002**	**0.032**	**0.006**	**0.004**
C-LND	3.8		**0.001**	**0.006**	**0.002**
R-LND	1.8	2.9		**0.01**	**0.008**
C-HND	3.2	2.3	3.3		**0.006**
R-HND	2.5	2.7	2.5	2.9	

Response at the taxonomic and OTU level: Due to no significant effect (p>0.05) of sampling site on diversity indices and the PCoAs, we combined their sequence data, re-calculate the values of Chao1, PCo1, PCo2 and PCo3 and compared the mean of the indexes among treatments using one-way ANOVA (treatment was the only effect). PCo1, PCo2 and PCo3 explained 33, 19.5 and 9% (cumulative variation=61.5%) of the total variation within the Bray-Curtis dissimilarity matrix for the BCs of week 4 in the phase 2 experiment The combined data of week 4 showed significant differences in the mean of Chao1 (df=4, F=161.2, p<0.00001), PCo1 (df=4, F=231, p<0.00001), PCo2 (df=4, F=132, p<0.00001) and PCo3 (df=4, F=136, p<0.00001) among P2-Co, C-LND, C-HND, R-LND and R-HND (Appendix E; Figure S6.4). The mean of Chao1, PCo1 and PCo3 were not significantly different (p>0.05) between the two challenged communities; however, the mean of Chao1, PCo1 and PCo3 were higher in the C-HND than the C-LND community. Meanwhile, the mean of PCo2 was significantly higher (p< 0.05) for C-HND compared to C-LND (Appendix D; Figure S5.4).

In week 4 of the challenge response (phase 2) experiment, 20 OTUs were highly abundant in the BCs of P2-Co and they belonged to diverse taxa (6 different phyla), but mostly *Proteobacteria* (7 OTUs; 35%), *Actinobacteria* (4 OTUs; 20%) and *Planctomycetes* (4 OTUs; 20%) (Figure 6.5). Out of those 20 highly abundant OTUs, only 4, 5, 4 and 3 OTUs were also abundant in the BCs of C-LND, R-LND, C-HND and R-HND respectively. In the BC of C-LND, we identified 19 highly abundant OTUs (assigned to 7 different phyla) which mostly belonged to *Proteobacteria* (9

OTUs; ~50%) and *Cyanobacteria* (3 OTUs; 16%). Out of those 19 abundant OTUs in the BC of C-LND; 3, 9 and 3 OTUs were also abundant in the BCs of R-LND, C-HND and R-HND respectively. The BC of C-HND was dominated by 16 abundant OTUs (assigned to 7 different phyla) of which *Proteobacteria* (6 OTUs, 38%) and *Bacteroides* (3 OTUs; 20%) constituted the highest number of OTUs. Out of those 16 abundant OTUs, only 3 were common and abundant with the BC of R-HND (Figure 5.5). The BC of R-LND had 11 abundant OTUs (assigned to 7 different phyla) which 3 and 5 of them also were abundant in the BCs C-LND and P2-Co respectively. The BC of R-HND had 12 abundant OTUs (assigned to 7 different phyla) which 3 of them were abundant in the BC of C-HND and another 3 OTUs were also abundant in the BC of P2-Co.
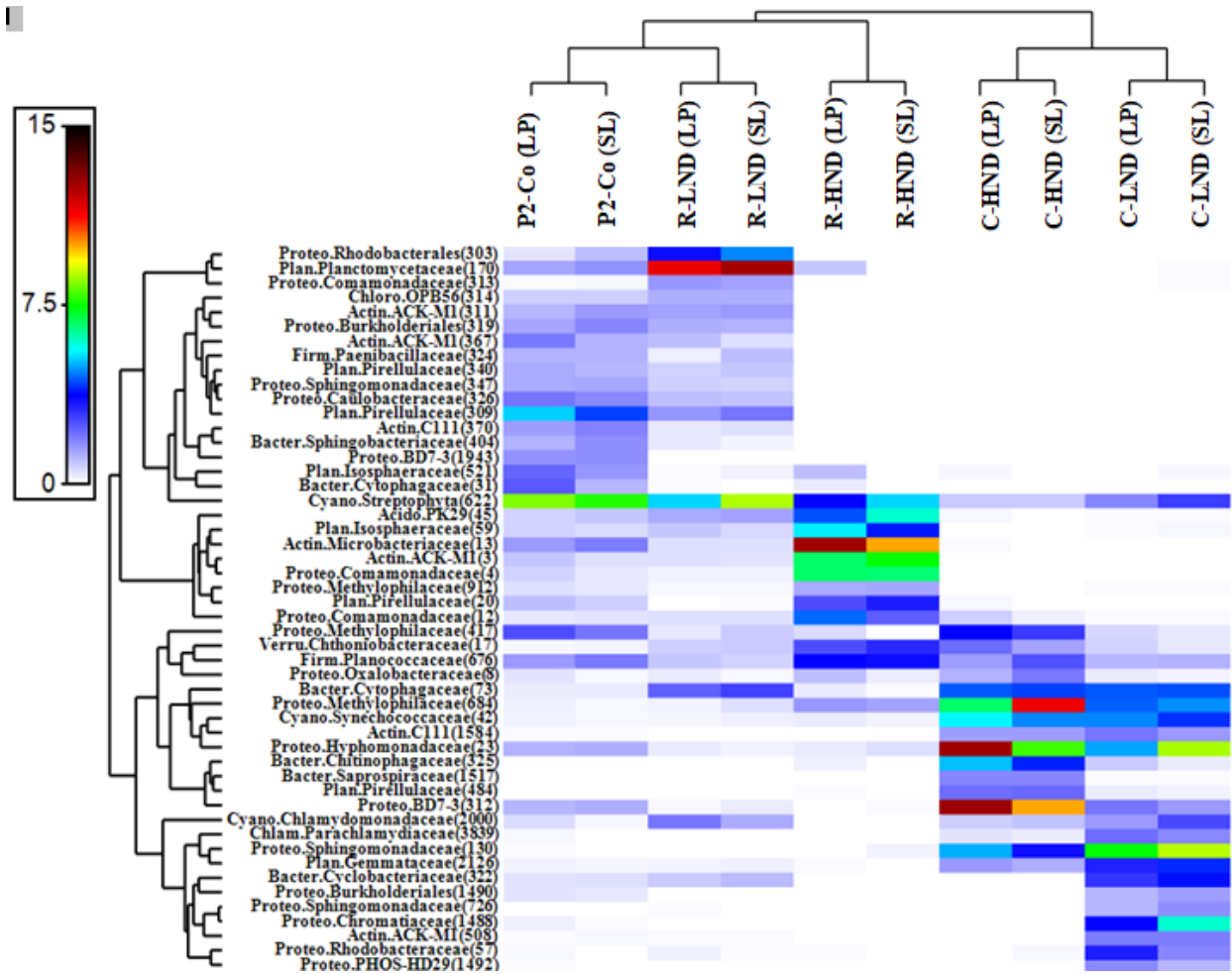
Figure 5.5. Hierarchical clustering heatmap of the abundant OTUs in the BC of control and treatments microcosms at week 4 of the challenge response (phase 2) experiment. Vertical clustering was based on the similarity in the abundance of OTUs. Horizontal clustering was based on the similarly in the relative abundance of OTUs for each sampling location and treatment. OTU numbers are given in parenthesis with their taxonomic assignment (phyla and family). Although there were no significant differences between the community of two sampling locations at the whole community level – minor differences exist and that is why both locations are presented in the heat map. The color code indicates the relative abundance of each OTU. The lines in the heatmap represent the relative abundance of each OTU across the BC of P2-Co, C-LND, C-HND, R-LND and R-HND of LP and SL. Taxa abbreviation: Acido; *Acidiobacteria*, Actin; *Actinobacteria*, Bacter; *Bacteroidetes*, Chlam; *Chlamydiae*, Chloro; *Chloroflexi*, Cyano; *Cyanobacteria*, Firm; *Firmicutes*, Plan; *Planctomycetes*, Proteo; *Proteobacteria* and Verru; *Verrucomicrobia*.

Exposure of the adapted BCs to very high levels of nutrients led to significant increases of the relative abundance of some taxa such as *Actinobacteria* (family C111; OTU1584) and *Proteobacteria* (families *Sphingomonadaceae*; OTU130 and *Hyphomonadaceae*; OTU23) in the challenge BCs of both C-LND and C-HND (Figure S5). The challenge of the BCs adapted to the low level of nutrients (LND), resulted in significant increases of the abundance of *Cyanobacteria* (OTUs 622 and 2000), a different family of *Planctomycetes* such as *Gemmataceae* (OTU2126), *Isosphaeraceae* (OTU59) and *Planctomycetaceae* (OTU 170) as well as specific families of *Proteobacteria* (such as *Burkholderiales* (OTU 1490), *Chromatiaceae* (OTU1488) and *Sphingomonadaceae* (OTUs 130 and 726) in the BC of C-LND, compared to the BC of C-HND. In contrast, the challenge of the BC adapted to the high levels of nutrients (HND) mostly resulted in a significant elevation in the abundance of unique families of *Proteobacteria* (such as *Oxalobacteraceae* (OTU 8) and *Methylophilaceae* (OTUs 417 and 684), *Firmicutes* (OTU676) and *Verrucomicrobia* (OTU17)) in the BC of C-HND compared to C-LND (Appendix D; Figure S5.5). We noticed that the relative abundance of *Proteobacteria* was still significantly higher in the BCs of C-LND (OTUs 23, 130, 726, 1488 and 1490) and C-HND (OTUs 23, 130, 312 and 417) compared to their corresponding relaxed communities (Appendix D; Figure S5.5). The relative abundance of only a few taxa of *Actinobacteria* and *Acidiobacteria* was significantly higher in the BC of the relaxed treatment (R-LND and R-HND) compared to their corresponding challenge microcosms (Appendix D; Figure S5.5). Taxonomic composition (at the phyla level) of the BCs of the various microcosm treatments in phase 1 (adaptation phase) and phase 2 (challenge phase) experiments are presented in Figure S5.6 (Appendix D).

5.3.3 Microbial network analysis and keystone taxa

The microbial networks were constructed using all OTUs with a relative abundance of greater than 0.05% across all treatments of the challenge phase week 4 (Figure 5.6, Appendix D; Table S5.1). Using thresholds for keystone taxa definition i.e., (taxa with degree > 70, closeness > 0.3, and betweenness < 0.07), we identified 50, 41, 34, 26 and 18 keystone taxa (OTUs) for the BCs of P2-Co, R-LND, C-HND, C-LND and R-HND at week 4 respectively. The keystone taxa (50 OTUs) in the BC of P2-Co belonged to

27 bacterial families of 7 different phyla which mostly classified as *Proteobacteria* (17 OTUs; 34%), *Planctomycetes* (10 OTUs; 20%), *Bacteroidetes* (9 OTUs; 18%). Keystone taxa belonged to *Actinobacteria, Actinobacteria, Chloroflexi* and *Firmicutes* also were identified (Appendix D; Table S5.2). Keystone taxa (26 OTUs) of C-LND were affiliated to 17 bacterial families of 4 phyla mostly including *Proteobacteria* (20 OTUs; 77%) and other phyla such as *Actinobacteria, Bacteroidetes* and *Cyanobacteria* (Appendix D; Table S5.2). Compare to C-LND, more keystone taxa (34 OTUs) were identified in the C-HND community which were affiliated to the 19 bacterial families of 6 phyla mostly including *Proteobacteria* (22 OTUs; 65%). Keystone taxa belonged to *Actinobacteria, Bacteroidetes, Cyanobacteria* and *Planctomycetes* also were identified in the BC of C-HND (Appendix D; Table S5.2).

Some keystone taxa of the challenge communities (C-LND and C-HND) assigned to the same bacterial families (i.e. C111, *Comamonadaceae, Cytophagaceae, Rhizobiaceae, Rhodobacteraceae, Synechococcaceae, Xanthomonadaceae*), but only one keystone taxa (OTU42; *Cyanobacteria*, family *Synechococcaceae*) was common between two challenges BCs. Some bacterial families with their keystone taxa such as *Saprospiraceae* (OTU1517, *Bacteroidetes*) and *Methylophilaceae* (OTUs 417, 684, 912; *Proteobacteria*) and BD7-3 (OTU312; *Proteobacteria*) were only detected in the C-HND community (Appendix D; Table S5.2). In contrast, some bacterial families of *Proteobacteria* with their keystone taxa such as *Bdellovibrionaceae* (OTUs 174 and 317), *Burkholderiales* (OTU1490) and *Erythrobacteraceae* (OTUs 425 and 449) were only identified in the C-LND community (Appendix D; Table S5.2). Both relaxed communities had keystone taxa belong to *Acidobacteria*, *Firmicutes Planctomycetes*, and *Proteobacteria*. However, keystone taxa classified in *Actinobacteria* (5 OTUs) and *Bacteroidetes* (11 OTUs) and *Verrucomicrobia* (1 OTU) were only detected in the R-LND community and keystone taxa belong to *Cyanobacteria* (1 OTU) and *Gemmatimonadetes* (2 OTUs) were only identified in the R-HND community.
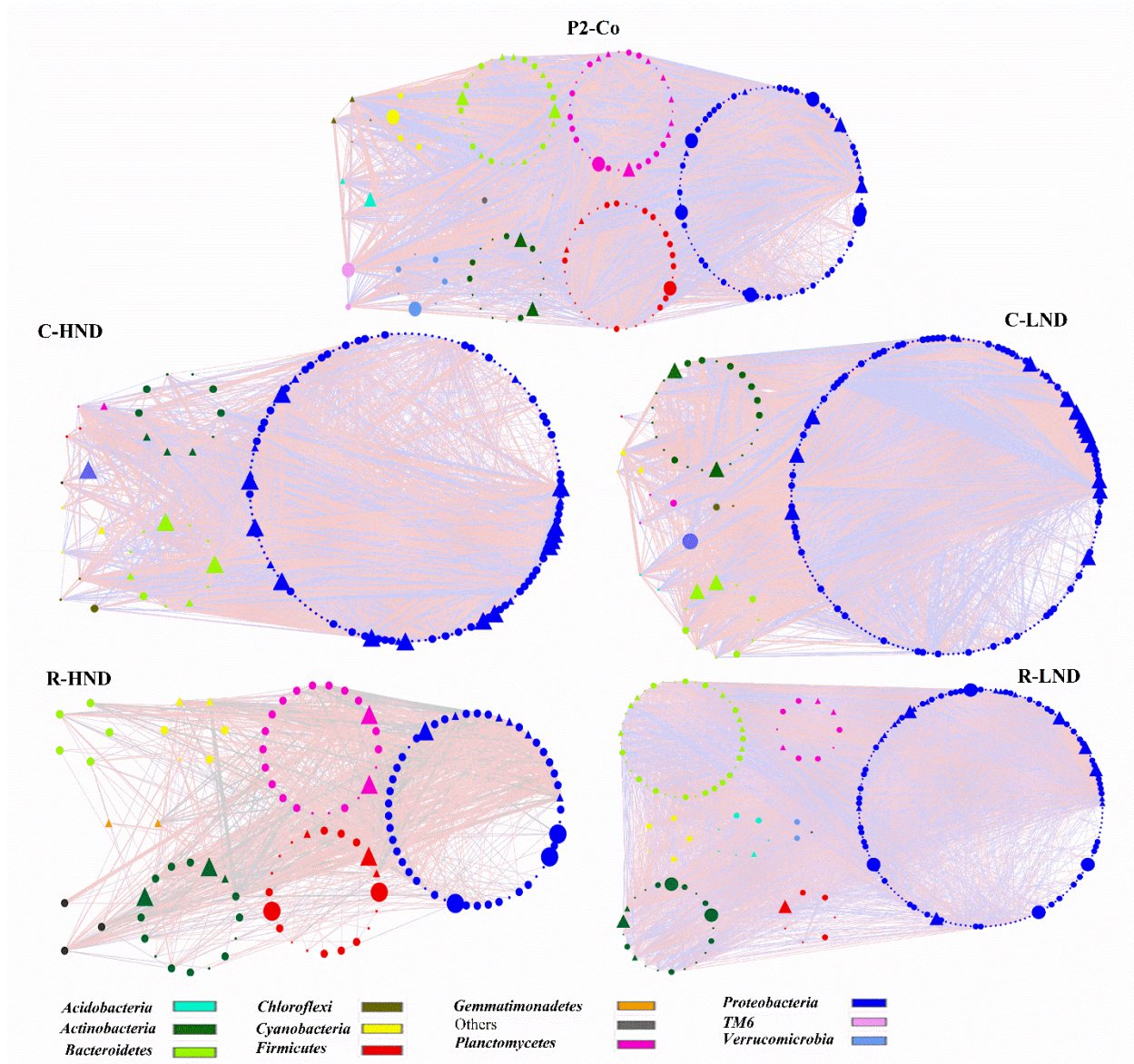
Figure 5.6. Network analysis diagrams of the BC compositions (phyla level) for the different treatment and control microcosms in the challenge experiment (challenged communities; C-LND and C-HND, relaxed communities; R-LND and R-HND and control; P2-Co) at week 4 (end of the experiment). Circles (nodes) are the OTUs belongs to each phylum while triangles (nodes) show keystone OTUs in each phylum for different treatments. Pink and gray lines (edges) represent positive and negative correlations respectively.

5.3.4 Horizontal gene transfer rate in the challenge phase

We measured gene transcription at two well-documented HGT markers (GTA terminase and *Int1*) at week 4 in the challenge (phase 2) experiments. The goal was to determine if the challenge triggered potentially adaptive elevated transcription levels of HGT-related genes. As there was no significant effect of sample site on gene transcription at either marker locus under any treatment in the response to challenge (phase 2) experiment (p=0.9), we combined the transcription data for the two sample sites within each treatment. Both makers showed similar patterns among the treatments in the experimental phase 2 (week 4). There was a significant difference in transcription of GTA terminase (df=4, F=123, p<0.00001) and *Int1* (df=4, F=69, p<0.00001) among treatments. The transcription level of GTA terminase and *Int1* markers in the BCs of C-LND and C-HND were significantly higher (p < 0.05) than their transcription level for BCs of P2-Co, R-LND and R-HND (Figure 5.7). The transcription of GTA terminase (C_T mean= 19.2±0.5) and *Int1* (C_T mean= 21.2±0.8) in C-HND were higher than the transcription of GTA terminase (C_T mean= 20.1±0.6) and *Int1* (C_T mean= 22.4±0.5) for the C-LND however, Post-hoc test did not show significant differences in the transcription level of both marks under the C-LND and C-HND treatments (Figure 5.7).
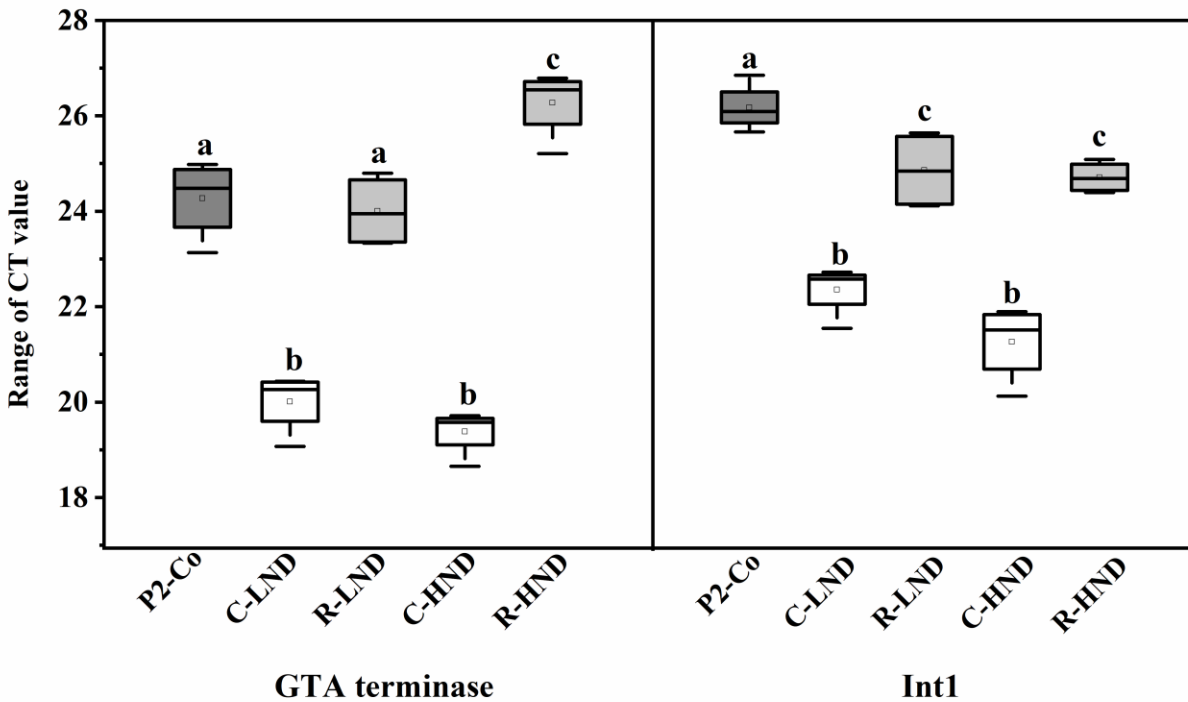
Figure 5.7. Box and whisker plot showing mean gene transcription (C_T values) for HGT markers (GTA terminase and *Int1*) in different BCs of phase 2; P2-Co, C-LND, R-LND, C-HND and R-HND (week 4). The thick bar is median, upper and lower quartiles represent 75% and 25% of the data respectively. Whiskers are used to indicate variability outside the upper and lower quartiles.

5.4   Discussion

Given that freshwater ecosystems are undergoing rapid and major environmental changes globally due to a variety of stressors, including eutrophication, obtaining a quantitative and process-level understanding of the mechanisms that affect BC response to such stressors is pivotal for predicting the responses of ecosystems to novel or changing selective forces. Although much research has been completed on the dose-response of nutrient loading on microbial community composition (Gobler et al., 2016; Ratzke et al., 2020; Steffen et al., 2017), less is known about the potential for microbial community "evolution" or adaptive change that could mitigate future changes - the implications of such responses are critical at the local, regional and global scales. Here, we test the effect of different levels of nutrients as anthropogenic stressors as selection

pressures on freshwater BCs of lakes Erie and St. Clair in a two-phase study (phase 1; adaptation experiment; and phase 2; response to challenge experiment). Specifically, we evaluated nutrient stress (low and high levels of nutrient stress) effects on the adaptive capacity of freshwater BCs and showed how this adaption might help mitigate the effects of extreme nutrient stress in freshwater BCs (as well as on particular taxa and horizontal gene transfer capacity).

In our adaptation phase experiments, we found the low level of nutrient stress (5 times above ambient at the time of sampling) only resulted in slight changes (no significant effects at the community level) in the BC, while high levels of stress, induced significant community shifts in the BC. Our results also showed that increasing the level of the stressor (nutrients from low to high level) resulted in an elevated abundance of OTUs belong to *Proteobacteria*; a fast-growing and nutrient loving phylum of freshwater BC (Newton et al., 2011). However the same stressor resulted in a reduced abundance of *Actinobacteria*; the numerically dominant phylum in freshwater lakes (Newton et al., 2011; Shahraki et al., 2020), which prefer oligotrophic ecosystems (Haukka et al., 2006) and are sensitive to eutrophication as a stressor. Similar to that reported before (Lee et al., 2017), we found that eutrophication not only resulted in the dominance of *Proteobacteria* in the BC, but also a significant loss of diversity. Thus, nutrient overloading could result in some taxa, such as *Proteobacteria*, to dramatically alter the chemical microenvironment, driving further negative interactions among bacterial species. This process would eventually exclude bacterial species from the community, ultimately resulting in a permanent loss of biodiversity (Ratzke et al., 2020), as we observed. Such a loss of diversity would be expected to alter BC function as well (Banerjee et al., 2018). More interestingly, our work also showed an increase in the relative abundance of *Cyanobacteria* in response to both low and high levels of nutrient stress. Previous work on Lake Erie has shown that elevated nutrient levels are associated with increases in the relative abundance of *Anabaena* and *Planktothrix* (close to the mouth of the Maumee River), *Microcystis* (the western basin of Lake Erie) (Harke et al., 2016) and *Planktothrix* (Sandusky Bay) (Davis et al., 2015) within the *Cyanobacteria* community. However, in our adaptation

118

(phase 1) experiments, nutrient overloading increased the level of *Synechococcus* (OTU42), another taxon capable of producing cHABs (Berry et al., 2015).

Although our phase 1 experiments produced interesting changes in the BC of our study sites, perhaps more interesting are the results of the challenge response (phase 2) experiments. For this study, we interpreted the reduced impact of the extreme challenge on key taxa abundance, which drives community function, as an adaptive response. Increases in the abundance of a single taxon may reflect elevated performance (and fitness) of that particular taxon in a stressful environment, however, increases in the abundance of different taxa (OTUs) belonged to higher taxa level such as phylum was interpreted as evidence of better adaptation at the community level in this study. The nutrient challenge experiments with extreme levels of nutrients led to significant shifts in the composition of the pre-adapted communities relative to the control microcosm BCs. Conversely, we found that under the nutrient stress relaxed conditions, the altered BCs from the phase 1 experiments did not return to the control community composition or diversity, indicating that re-culturing of the adapted communities under ambient conditions could not restore the altered BCs to their original state. We did not monitor the functionality of the BC in either the adaptive or challenge phases of our study; however, a significant shift in the BC composition, significant loss of diversity in the challenge phase and changes in keystone taxa profile make it likely that a significant shift in the BC function occurred as well - as reported in various natural ecosystems (Banerjee et al., 2018; Peter et al., 2011; Philippot et al., 2013; Schimel, 1995).

Microbes must acclimate to immediate stress by altering their allocation of resources from growth to survival pathways; however, extreme stress may force them into dormancy (Suzina et al., 2004) or kill them. Death and dormancy both remove the microbial function from the ecosystem; however, whereas dormant organisms can regain activity when conditions improve; dead ones do not (Schimel et al., 2007). Extensive microbial death due to stressors may result in substantial time lags in recovery (or complete failure to recover) and, ultimately, breakdown of ecosystem function (Clein and Schimel, 1994). In our study, we monitored BC recovery and the potential resumption of dormant bacterial activity by re-culturing the adapted BCs under ambient conditions. We found that the BCs that were pre-adapted to the high

levels of nutrients (HND) exhibited reduced recovery under the relaxed conditions, relative to the BCs exposed to the low level of nutrients stress (LND). This was reflected in reduced recovery of diversity (Chao1, PCo1 and PCo2) in the adapted BC to the high level of nutrients in comparison to the adapted BC to the low level of nutrients (Appendix D; Figure S5.4).

Significant increases in the abundance of some taxa belonging to *Proteobacteria* (i.e. *Burkholderiales*; OTU1490 and *Chromatiaceae;* OTU1488) in the challenge (phase 2) experiment using the BC adapted to the low level of nutrients (LND), coupled with the dominance of some taxa belonging to *Proteobacteria* (i.e. BD7-3; OTU312 and *Methylophilaceae*; OTU684) in BC adapted to the high level of nutrients (HND) demonstrated that different taxa within a single phylum (*Proteobacteria*) adapted differently and generated different responses to the extreme nutrient challenge in the phase 2 experiment. Such broad community responses are critical for BC function and highlight the complexity of BC. Some studies reported the correlation of stressors with some bacterial taxa due to better adaption of those taxa to stressors. For example, Liu et al, reported a positive significant correlation and better adaptation of *Firmicutes* and *Bacteroidetes* to total Hg content (Liu et al., 2018), however, our study for the first time demonstrate a substantial variation in the adaptive response of different bacterial taxa of phyla (i,e. *Proteobacteria*) of the "pre-adapted" communities against a very high level of the stressor. Our study shows that different taxa might exhibit different adaptive responses even if they are belonging to the same bacterial class or phylum.

While the ecological importance of the dramatically responding taxa described above is not well known, *Chromatiaceae* (phototrophic purple sulfur bacteria) have a high capacity to use ammonia and adapt to heterotrophic growth under nutrient-depleted (oligotrophic) conditions (Imhoff, 2014). Thus *Chromatiaceae* may be able to exploit even the extreme levels of nutrients we used as a challenge and achieve high-population densities (Guyoneaud et al., 1996). Thus, the exposure of a natural BC to low levels of nutrient stress may increase the adaptive response of *Chromatiaceae* against our challenge (extreme levels of nutrients). On the other hand, it has been suggested that methanol (as one metabolic node) could be a major carbon source

produced by members of methanotrophs such as *Methylophilaceae* for community cross-talking and to support satellite communities (Krause et al., 2017). Adaptation to the high level of nutrients in the phase 1 experiments, followed by the challenge with extreme levels of nutrients in the phase 2 experiment resulted in elevated and high abundance of *Methylophilaceae*. Pre-stressing the freshwater BCs in this study with the high levels of nutrients in phase 1 appears to allow methanotrophs to become an important component of the freshwater BC, likely due to methanotrophs' capacity for nitrogen fixation (Bowman, 2006). Such changes in abundance likely facilitated the challenged BCs to make connections between nitrogen fixation and the carbon cycle in our anthropogenically affected microcosm ecosystems.

As expected, the BC of the control microcosm in the challenge (phase 2) experiment had the highest number of keystone taxa with the highest diversity at the phyla level (7 different phyla). Our detailed analysis of the keystone taxa (identified by our network analyses) revealed that adaptation to the high level of nutrient (HND) versus the low level of nutrients (LND) allowed the BC to maintain a higher number of keystone taxa (34 OTUs affiliated to the 19 bacterial families of 6 phyla) with more diversity in the challenge phase. This indicates that the HND pre-adaptation treatment likely improved the functional response of the BC to the challenge. We identified diverse keystone taxa (26 OTUs identified to 17 bacterial families) belonging to 4 different phyla in the BCs adapted to low levels of nutrient (LND) after the challenge (phase 2) experiments. However, 77% (20 OTUs) of them were assigned to *Proteobacteria*, indicative of a poor adaptive response to the challenge as it resulted in the elimination of keystone taxa in other phyla (maladaptation). This finding revealed that higher levels of stress in the adaptation phase (nutrients in this study) provide better outcomes for the BC against very high levels of nutrient stress.

Various taxa belonging to *Proteobacteria* (77 and 65% keystone taxa in C-LND and C-HND respectively) were the main part of keystone taxa in adapted BCs after challenge with a very high level of nutrients. *Proteobacteria* are a fast-growing and nutrient loving part of freshwater BCs (Newton et al., 2011); however, this study for the first time provides more details regarding the wide range of adaptive response of diverse freshwater *Proteobacteria* to different levels of stress (low versus high nutrient

doses in this study). Perhaps more importantly, we showed that "pre-adaption" exposure helps particular taxa in this phylum to deal with very high levels of nutrient stress. Some keystone taxa *Planctomycetes* (nitrogen fixation and autotrophic metabolism) (Delmont et al., 2018) and *Verrucomicrobia* (diverse metabolic strategies and association with the high-nutrient environment) (Lindström et al., 2004)) in the BC of microcosms adapted to high levels of nutrient stress (HND) responded to the challenge treatment with better performance and adaptation under a very high level of stress (nutrient overloading in this study). This was likely due to their ability to exploit the large nutrient resources available in the challenge phase.

Interestingly, three OTUs affiliated with family C111 of *Actinobacteria* achieved keystone taxon status in the BC adapted to the high level of nutrient stress (HND), after the challenge phase, compared to only one family C111 OTU for the BC adapted to low levels of nutrients (LND). *Actinobacteria* is one of the main components of freshwaters BC (Newton et al., 2011) and is sensitive to nutrient overloading (Haukka et al., 2006). It is thus apparent that pre-treatment with high versus low level of nutrients resulted in an improvement in the role of family C111 taxa after exposure to high levels of nutrient stress in the challenge (phase 2) experiments. Generally, keystone taxa analysis revealed a noticeable reduction in the diversity of keystone taxa in the BCs adapted to the LND (4 phyla, mostly *Proteobacteria*; up to 77%) versus HND (6 phyla). As keystone taxa are, by definition, the drivers of microbiome structure and function (Banerjee et al., 2018), our analyses show that pre-treatment with high levels of the nutrient can facilitate the maintenance more diverse keystone taxa and thus likely better functionality relative to the LND pre-treatment, after exposure to very high levels of nutrient stress.

Our study showed a substantial BC composition shift in the microcosms after the adaptive (phase 1) and the challenge (phase 2) experiments, primarily in those under high-stress treatments. While we detected a few significant changes in the LND BC in phase 1 experiments, we do not know if the BC was responding in other ways. While our experiments were not designed to test for genetic change within individual taxa, our measurement of transcription of the HGT marker genes allowed us to explore responses involving complex community-level genetic interactions. Indeed, some

studies have suggested that HGT is one of the main mechanisms involved in the adaptive response in bacterial communities (Springael and Top, 2004), shaping microbial community structure in response to heavy metals as stressors (Hemme et al., 2016) and microbial community response to antibiotic and heavy metals (Di Cesare et al., 2016a). We observed higher transcription of two HGT markers in the BC adapted to the high level of nutrients relative to the BC adapted to the low level of nutrients when they challenged with a very high level of nutrients (Figure 5.7). This finding shows that pre-adaptation of the BC to the high level of stress relative to the low level of stressor would increase the HGT rate and potentially help the entire community to better survive against a very high level of the stressor by higher-level of genetic exchanges. Our results presented in this study provide explicit evidence that HGT plays a crucial role in the adaptive response of freshwater BC against stressors such as eutrophication.

5.5   Conclusion

Our two-phase adaptation-challenge response study showed that different levels of stress pre-treatment have very different effects on the potential for adaptive responses in bacterial communities. High levels of stress pre-treatment provide a better adaptive response to secondary extreme stress by allowing the retention of diversity in the community, especially diversity in the keystone taxa. The HND pre-treatment also resulted in a higher rate of HGT than in the microcosms exposed to the LND pre-treatment. Our results highlight the role of nutrient overloading as a stressor in driving a BC composition shift, diversity loss and potentially change the function of the BC by altering keystone taxa. On the other hand, different levels of nutrient stress can provide better adaptive responses (e.g., growth conditions) for taxa such as those in phylum *Proteobacteria*, while low levels of pre-treatment of nutrient stress exclude some critical members of the community (such as *Planctomycetes* and *Verrucomicrobia*) due to poor adaptive response. It is interesting to note that the elimination of stressors (relaxed nutrient stress) does not allow complete recovery of the BC to the original state under even ambient nutrient conditions. This study also revealed that BCs may employ HGT as one of the major adaptive mechanism for microbial populations against environmental stressors. This study provides a linkage between environmental

stressors with the BC composition and its response to the different levels of stressors which is critical in advancing our understanding of ecosystem ecology.

5.6 References

1999. Canadian Environmental Protection Act, Ammonia in the Aquatic Environment. Environmental Canada.

Allison, S.D., Martiny, J.B., 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 105(Supplement 1) 11512-11519.

Banerjee, S., Schlaeppi, K., van der Heijden, M.G., 2018. Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology 16(9) 567-576.

Berry, D., Widder, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology 5 219.

Berry, D.L., Goleski, J.A., Koch, F., Wall, C.C., Peterson, B.J., Anderson, O.R., Gobler, C.J., 2015. Shifts in cyanobacterial strain dominance during the onset of harmful algal blooms in Florida Bay, USA. Microbial Ecology 70(2) 361-371.

Bhagowati, B., Ahamad, K.U., 2019. A review on lake eutrophication dynamics and recent developments in lake modeling. Ecohydrology & Hydrobiology 19(1) 155-166.

Bowman, J., 2006. The methanotrophs—the families Methylococcaceae and Methylocystaceae. The prokaryotes 5 266-289.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 55(4) 611-622.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5) 335.

Clein, J.S., Schimel, J.P., 1994. Reduction in microbial activity in Birch litter due to drying and rewetting event. Soil Biology and Biochemistry 26(3) 403-406.

Cotner, J.B., Biddanda, B.A., 2002. Small players, large role: microbial influence on biogeochemical processes in pelagic aquatic ecosystems. Ecosystems 5(2) 105-121.

Crevecoeur, S., Vincent, W.F., Comte, J., Lovejoy, C., 2015. Bacterial community structure across environmental gradients in permafrost thaw ponds: methanotroph-rich ecosystems. Frontiers in Microbiology 6 192.

Davis, T.W., Bullerjahn, G.S., Tuttle, T., McKay, R.M., Watson, S.B., 2015. Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during Planktothrix blooms in Sandusky Bay, Lake Erie. Environmental Science & Technology 49(12) 7197-7207.

Delmont, T.O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S.T., Rappé, M.S., McLellan, S.L., Lücker, S., Eren, A.M., 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nature Microbiology 3(7) 804-813.

Di Cesare, A., Eckert, E., Corno, G., 2016a. Co-selection of antibiotic and heavy metal resistance in freshwater bacteria. Journal of Limnology 75(Suppl 2) 59-66.

Di Cesare, A., Eckert, E.M., D'Urso, S., Bertoni, R., Gillan, D.C., Wattiez, R., Corno, G., 2016b. Co-occurrence of integrase 1, antibiotic and heavy metal resistance genes in municipal wastewater treatment plants. Water research 94 208-214.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19) 2460-2461.

Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey data. PLoS Computational Biology 8(9).

Gillings, M.R., 2014. Integrons: past, present, and future. Microbiology and Molecular Biology Reviews 78(2) 257-277.

Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M., Zhu, Y.-G., 2015. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. The ISME Journal 9(6) 1269.

Gobler, C.J., Burkholder, J.M., Davis, T.W., Harke, M.J., Johengen, T., Stow, C.A., Van de Waal, D.B., 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae 54 87-97.

Guyoneaud, R., Matheron, R., Baulaigue, R., Podeur, K., Hirschler, A., Caumette, P., 1996. Anoxygenic phototrophic bacteria in eutrophic coastal lagoons of the French Mediterranean and Atlantic coasts (Prévost Lagoon, Arcachon Bay, Certes fishponds). Hydrobiologia 329(1-3) 33-43.

Harke, M.J., Davis, T.W., Watson, S.B., Gobler, C.J., 2016. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. Environmental Science & Technology 50(2) 604-615.

Haukka, K., Kolmonen, E., Hyder, R., Hietala, J., Vakkilainen, K., Kairesalo, T., Haario, H., Sivonen, K., 2006. Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial Ecology 51(2) 137-146.

He, X., Chaganti, S.R., Heath, D.D., 2017. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microbial Ecology 75 1-12.

Hecky, R., Mugidde, R., Ramlal, P., Talbot, M., Kling, G., 2010. Multiple stressors cause rapid ecosystem change in Lake Victoria. Freshwater Biology 55 19-42.

Hemme, C.L., Green, S.J., Rishishwar, L., Prakash, O., Pettenato, A., Chakraborty, R., Deutschbauer, A.M., Van Nostrand, J.D., Wu, L., He, Z., 2016. Lateral gene transfer in a heavy metal-contaminated-groundwater microbial community. MBio 7(2) e02234-02215.

Imhoff, J.F., 2014. The family Chromatiaceae. Springer.

Johnsen, A.R., Kroer, N., 2007. Effects of stress and other environmental factors on horizontal plasmid transfer assessed by direct quantification of discrete transfer events. FEMS Microbiology Ecology 59(3) 718-728.

Kohl, M., Wiese, S., Warscheid, B., 2011. Cytoscape: software for visualization and analysis of biological networks, Data Mining in Proteomics. Springer, pp. 291-303.

Krause, S.M., Johnson, T., Karunaratne, Y.S., Fu, Y., Beck, D.A., Chistoserdova, L., Lidstrom, M.E., 2017. Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions. Proceedings of the National Academy of Sciences 114(2) 358-363.

Lang, A.S., Westbye, A.B., Beatty, J.T., 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. Annual Review of Virology 4 87-104.

Lawrence, D., Fiegna, F., Behrends, V., Bundy, J.G., Phillimore, A.B., Bell, T., Barraclough, T.G., 2012. Species interactions alter evolutionary responses to a novel environment. PLoS Biology 10(5).

Lee, Z.M.-P., Poret-Peterson, A.T., Siefert, J.L., Kaul, D., Moustafa, A., Allen, A.E., Dupont, C.L., Eguiarte, L.E., Souza, V., Elser, J.J., 2017. Nutrient stoichiometry shapes microbial community structure in an evaporitic shallow pond. Frontiers in Microbiology 8 949.

Li, Y., Fan, L., Zhang, W., Zhu, X., Lei, M., Niu, L., 2020. How did the bacterial community respond to the level of urbanization along the Yangtze River? Environmental Science: Processes & Impacts.

Lindström, E.S., Vrede, K., Leskinen, E., 2004. Response of a member of the Verrucomicrobia, among the dominating bacteria in a hypolimnion, to increased phosphorus availability. Journal of Plankton Research 26(2) 241-246.

Liu, Y.-R., Delgado-Baquerizo, M., Bi, L., Zhu, J., He, J.-Z., 2018. Consistent responses of soil microbial taxonomic and functional attributes to mercury pollution across China. Microbiome 6(1) 1-12.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J.R., Dunthorn, M., 2014. Patterns of rare and abundant marine microbial eukaryotes. Current Biology 24(8) 813-821.

Maguire, T.J., Wellen, C., Stammler, K.L., Mundle, S.O., 2018. Increased nutrient concentrations in Lake Erie tributaries influenced by greenhouse agriculture. Science of the Total Environment 633 433-440.

McDaniel, L.D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K.B., Paul, J.H., 2010. High frequency of horizontal gene transfer in the oceans. Science 330(6000) 50-50.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75(1) 14-49.

Peter, H., Beier, S., Bertilsson, S., Lindström, E.S., Langenheder, S., Tranvik, L.J., 2011. Function-specific response to depletion of microbial diversity. The ISME Journal 5(2) 351-361.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., Maron, P.-A., 2013. Loss in microbial diversity affects nitrogen cycling in soil. The ISME Journal 7(8) 1609-1619.

Ratzke, C., Barrere, J., Gore, J., 2020. Strength of species interactions determines biodiversity and stability in microbial communities. Nature Ecology & Evolution 4(3) 376-383.

Sarmento, H., Casamayor, E.O., Auguet, J.-C., Vila-Costa, M., Felip, M., Camarero, L., Gasol, J.M., 2015. Microbial food web components, bulk metabolism, and single-cell physiology of piconeuston in surface microlayers of high-altitude lakes. Frontiers in Microbiology 6 361.

Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. Nature 413(6856) 591-596.

Schimel, J., 1995. Ecosystem consequences of microbial diversity and community structure, Arctic and alpine biodiversity: patterns, causes and ecosystem consequences. Springer, pp. 239-254.

Schimel, J., Balser, T.C., Wallenstein, M., 2007. Microbial stress-response physiology and its implications for ecosystem function. Ecology 88(6) 1386-1394.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., Huttenhower, C., 2011. Metagenomic biomarker discovery and explanation. Genome biology 12(6) R60.

Shade, A., Peter, H., Allison, S.D., Baho, D., Berga, M., Bürgmann, H., Huber, D.H., Langenheder, S., Lennon, J.T., Martiny, J.B., 2012. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology 3 417.

Shahraki, A.H., Chaganti, S.R., Heath, D., 2018. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. Journal of Water and Health 17(1) 37-49.

Shahraki, A.H., Chaganti, S.R., Heath, D.D., 2020. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microbial Ecology 1-13.

Shahraki, A.H., Heath, D., Chaganti, S.R., 2019. Recreational water monitoring: Nanofluidic qRT-PCR chip for assessing beach water safety. Environmental DNA 1 305-315.

Smith, V.H., Schindler, D.W., 2009. Eutrophication science: where do we go from here? Trends in Ecology & Evolution 24(4) 201-207.

Springael, D., Top, E.M., 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends in Microbiology 12(2) 53-58.

Steffen, M.M., Davis, T.W., McKay, R.M.L., Bullerjahn, G.S., Krausfeldt, L.E., Stough, J.M., Neitzey, M.L., Gilbert, N.E., Boyer, G.L., Johengen, T.H., 2017. Ecophysiological examination of the Lake Erie Microcystis bloom in 2014: linkages between biology and the water supply shutdown of Toledo, OH. Environmental Science & Technology 51(12) 6745-6755.

Suzina, N., Mulyukin, A., Kozlova, A., Shorokhova, A., Dmitriev, V., Barinova, E., Mokhova, O., Duda, V., 2004. Ultrastructure of resting cells of some non-spore-forming bacteria. Microbiology 73(4) 435-447.

Swings, T., Van den Bergh, B., Wuyts, S., Oeyen, E., Voordeckers, K., Verstrepen, K.J., Fauvart, M., Verstraeten, N., Michiels, J., 2017. Adaptive tuning of mutation rates allows fast response to lethal stress in Escherichia coli. Elife 6 e22939.

Tschäpe, H., 1994. The spread of plasmids as a function of bacterial adaptability. FEMS Microbiology Ecology 15(1-2) 23-31.

Van den Brink, P.J., Braak, C.J.T., 1999. Principal response curves: Analysis of time-dependent multivariate responses of biological community to stress. Environmental Toxicology and Chemistry: An International Journal 18(2) 138-148.

Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics 3(2) 180-185.

Wozniak, R.A., Waldor, M.K., 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nature Reviews Microbiology 8(8) 552.

RECREATIONAL WATER MONITORING: NANOFLUIDIC QRT-PCR CHIP FOR ASSESSING BEACH WATER SAFETY ONTARIO [2]

6.1   Introduction

Human fecal, and consequently pathogen, contamination is a serious environmental and public health problem that affects water bodies around the world. Fecal pollution has become more widespread due to population growth and climate change and this is correlated with increases in waterborne disease (Levy et al., 2016). Recreational water is the second most common source of waterborne disease outbreaks worldwide, after contaminated drinking water (McClung et al., 2017). Degradation of freshwater quality also significantly reduces human recreational opportunities and results in economic loss (e.g., decreased tourism, reduced fishing activity, etc.) (Vörösmarty et al., 2010). In the US, for example, epidemiologic data indicate the risk for developing acute gastrointestinal illness symptoms is as high as 15 cases per 1000 swimmers, with a total of 90 million recreational water use related illnesses nationwide, which translates into \$2.2-\$3.7 billion in losses annually (DeFlorio-Barker et al., 2018). Consequently, improving beach monitoring and regulation technologies, coupled with beach remediation efforts guided by appropriate source tracking methods, would improve public health, recreation opportunities and economic wellbeing.

Culturing and counting fecal indicator bacteria (FIBs), such as *Escherichia coli*, remains the most common approach for monitoring pathogenic water pollution (Soller et al., 2016). Although coliform culture has been the standard for water safety monitoring, it has several important limitations, including; 1) it is time-consuming, 2) some FIBs can survive and grow in aquatic habitats (Ishii and Sadowsky, 2008), 3) there is generally a poor correlation between FIBs and the occurrence of pathogens (Jang et al., 2017) and 4) there is a lack of coliform host specificity (Pattis et al., 2017). Bacterial culturing is the gold standard for pathogen detection and has served us well for more than 150 years, but it is not applicable for routine water monitoring programs due to cost and time of culture method. Only well-equipped laboratories with approved

---

biosafety clearance are able to culture pathogenic bacteria complex samples compared to new molecular methods (Jacovides et al., 2012). Improved molecular technologies will be able to evaluate water quality and safety rapidly, more precisely and less expensively, particularly when more than one pathogen is targeted (Wolk and Hayden, 2011).

Quantitative real-time PCR (qRT-PCR or qPCR) is an emerging tool for water pathogen monitoring and has been recommended by the US Environmental Protection Agency (EPA) for detection and quantification of *E. coli* as a FIB (Chern et al., 2011). qPCR methods have been used to detect individual waterborne pathogens such as *E. coli* O157 (Wu et al., 2015) and *Yersinia enterocolitica* (Cheyne et al., 2010).

Multiplex qPCR is widely applied to quantify groups of waterborne pathogens (Fan et al., 2008; Nhung et al., 2007). Some of those studies utilized microfluidic plates (Ishii et al., 2013; Ramalingam et al., 2010). However, most studies do not validate and assess probe sensitivity, nor do they generally include source tracking markers. The inclusion of Microbial Source Tracking (MST) to determine the specific fecal source (e.g. human, dog, cattle, wildlife, etc.) is critical as the source will influence human health risk (Harwood et al., 2014). However, the simultaneous detection and quantification of many water safety-related target loci is challenging due to the technical difficulties associated with multiplexed qPCR. High density, low volume qPCR (nanofluidic) platforms are becoming more readily available (Morrison et al., 2006) and they allow multiple target detection with very high sensitivity (Friedrich et al., 2016).

A robust, high-throughput quantitative tool that targets FIBs, MST and pathogen virulence markers is required for effective real-time monitoring of recreational water and subsequent risk management. Consequently, the objectives of this study were to 1) design and validate a rapid nanofluidic qRT-PCR (TaqMan® assays in the OpenArray® platform) assays to monitor FIBs, MST markers and waterborne pathogens simultaneously and 2) evaluate the application of the designed method for monitoring of FIBs, MST and waterborne pathogens in different environmental samples. We combine the OpenArray platform with robotic DNA extraction to dramatically improve our ability to detect, monitor and quantify key waterborne health risks in less than four

hours. We apply our assay to a variety of samples (sewage, fecal control samples, environmental water samples and beach sand samples) and show the wide utility of our approach for both regulatory and research applications.

6.2   Materials and methods

6.2.1   Target selection and designing of the primers

Fifteen key waterborne pathogenic species (primary causes of bacterial waterborne disease (Ramírez-Castillo et al., 2015)) two FIBs including *E. coli* (two independent genes; 23S rRNA and *uidA* gene) and *Enterococci* spp.*,* one general fecal pollution MST marker and six host-specific MST markers were selected for inclusion on our OpenArray® plate (Appendix E; Table S6.1). We used existing qPCR primers and probes for the detection and quantification of *E. coli* (targeting 23S rRNA) (Chern et al., 2011) as a FIB, along with new primers and probe for the *uidA* gene; another universal gene conserved among different strains of *E. coli*. To assess the occurrence of inhibition during qPCRs, we also used previously suggested marker (amplification of 70 bp fragment of NH8B_3641 gene encoding NosZ-like protein of *Pseudogulbenkiania* spp. NH8B) as an internal positive control (Ishii et al., 2013).

Single copy virulence genes, which are present in all pathogenic strains of each waterborne pathogen, were selected as targets for primers and probe design (Appendix E; Table S6.1). For MST of fecal pollution, 16S rRNA sequence from host-specific *Bacteroides* for Canada goose (*Branta canadensis*) (Vogt et al., 2018)*,* dog (*Canis lupus* familiaris) and pig (*Sus scrofa domesticus*) (Okabe et al., 2007) and 16S rRNA sequence from *Methanobrevibacter smithii* (human MST marker) (Ufnar et al., 2006), *Catellicoccus marimammalium* (seagull MST marker) (Lee et al., 2013) and C40 mitochondria gene (human MST marker) (Caldwell et al., 2011) were used to design specific primers for qPCR. A universal primer set targeting 16S rRNA gene of *Bacteroides* spp. was designed as a general MST marker. We used published sequence data of each target to design specific primers and probes targeting virulence genes using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). We used the following parameters of Primer3Plus software to design the primers; no

CG clamp, max poly-X (score: 3), max self complementarity (score: 3), max 3' self complementarity (score: 0). Other parameters kept as default.

6.2.2 Primer validation

The specificity of the designed primers were checked *in silico* using the Basic Local Alignment Search Tool (BLAST) tool (Altschul et al., 1997) against the GenBank database. We used primer specificity stringency of at least 2 nucleotide mismatches to unintended targets, including at least 2 nucleotide mismatches within the last 5 nucleotides at the 3' end of the primers as default parameters.

Further, each designed primer set was tested in the laboratory against the host-specific and pooled non-specific fecal matter DNA (Canada goose, dog, human, pig, seagull and sewage) and the DNA of target and non-target species of bacteria. For each marker (Appendix E; Table S6.1), PCR master mix was prepared as described previously and each target was amplified with 59 ˚C annealing temperature in 35 cycles (Shahraki et al., 2019). The amplicon for each marker was visualized on an agarose gel and the amplicon size was estimated along with 50 bp DNA Ladder (Invitrogen™, USA). To test the sensitivity, we used purified PCR product - PCR product of each primer was purified using QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instruction. The concentration of each PCR product was measured using Agilent 2100 Bioanalyzer with a High Sensitive DNA chip (Agilent Technologies, Mississauga, ON, Canada). The PCR product for each marker, with estimated copy numbers, were combined and then diluted (10-fold) to make seven known concentrations of all targets (2,000,000, 200,000, 20,000, 2,000, 200, 20, 2 and 1 copies of each target/reaction) for SYBR green qPCR assays to determine the sensitivity of the designed primers.

6.2.3 Sample preparation and DNA extraction

We collected a variety of environmental samples, including beach water (6 X 250 mL), beach pore water (6 X 250 mL) and sand (6 X 20 g) from six public beaches (Windsor-Essex County, Ontario, Canada). We also collected stream and pond water samples (3 X 250 mL, each) near the sampled beaches. We collected sewage samples (3 X 250 mL) and fecal samples (N = 5 from Canada goose, dog, pig and seagull as

well as two healthy human fecal samples) for use as control samples. The collected water and sewage samples were delivered to the laboratory on ice within four hours of collection. Water and sewage samples were filtered with 0.22-micron pore size, 47 mm diameter polycarbonate filters (IsoporeTM, Millipore, MA). After filtration, each filter was placed in a 2 mL sterile tube containing 0.5 g glass beads (0.1 mm diameter, Bio-Spec Products, Bartesville, US). For sand samples, 20 g of sand from each location was added to a tube containing 250 mL distilled sterile water and the mixture was vigorously shaken for 2 min. The supernatant was transferred into a sterile bottle and the sample was filtered as described above. For fecal samples, 2 g of each sample was added to a 2 mL sterile tube with 0.5 g glass beads.

For cell lysis, 400 µL of sucrose lysis buffer (400 mM NaCl, 750 mM, 20 mM ethylenediaminetetraacetic acid (EDTA), 50 mM Tris-HCl pH 9.0) was added to each tube. After bead beating (1 min, 3 times), each sample was treated with lysozyme (Sigma-Aldrich, USA); 10 mg/mL, 37 °C for 20 min and proteinase K (Thermo Scientific, USA); 0.2 mg/mL, 50 °C for 20 min. Proteinase K was deactivated in 95 °C for 5 min and then DNA was purified from the samples using the Solid Phase Reversible Immobilization (SPRI) paramagnetic bead-based method on an automated liquid handling platform (Tecan Freedom Evo150 Liquid Handling Platform, Perkin Elmer, USA) (Shahraki et al., 2019). The purity and concentration of the extracted DNA from samples were checked (ND-1000, Nanodrop, USA) (Shahraki et al., 2019).

6.2.4 SYBR green qPCR assay

We used SYBR green qPCR to test the efficiency of the primers before designing the OpenArray® plate. Furthermore, our validation of the SYBR green qPCR allows applications of our markers where the OpenArray® plate is not feasible due to lack of access to the OpenArray® equipment, or where single marker vs multiple markers are interested. We applied SYBR green qPCR to the control (fecal and sewage) and environmental samples and validated the assays using dilutions of positive controls. SYBR green qPCR reactions were carried out on environmental and fecal sample DNA and all known-concentration target template DNAs in 20 µL including 10 µL SYBR green master mix (Applied Biosystems™, USA), 1 µL primers (combined forward and reverse primers, final concentration 10 pmol for each), 1 µL/ reaction DNA and 8 µL

ddH$_2$O. We also used SYBR green assays of two pathogen virulence genes (*ctxA* and *manC*) and NH8B_3641 gene (internal positive control) to evaluate the presence of PCR inhibitors over qPCR assays. Extracted DNA of water samples were negative in SYBR green qPCR assays for *ctxA*, *manC* and NH8B_3641 genes. For this experiment, extracted DNA from each water sample was inoculated with known-concentration of *ctxA*, *manC* and NH8B_3641 genes (2,000,000, 200,000, 20,000, 2,000, 200, 20, 2 and 1 copies of each gene/µL) in fresh tubes and were subjected to SYBR green qPCR. No-template control (NTC) reaction including SYBR green master mix, forward and reverse primers of each marker and ddH$_2$O but not DNA template was run for each assay. Real time program consisted of 95 °C for 1 min followed by 40 cycles 95 °C for 10 s, 60 °C for 60 s. All reactions were duplicated in MicroAmp Fast 96-well reaction plates and were run on QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems, USA).

6.2.5   Nanofluid OpenArray®

The goal of this research was to develop a high-throughput panel of qPCR assays in 33-nL reactions to monitor recreational water safety. We chose the nanofluidic TaqMan® assays in the OpenArray® platform (gene expression platform, Thermo Fisher Scientific, USA), as it is capable of performing duplicate qPCR assays for 25 marker loci, including 3 FIB markers (two assays for *E. coli*; 23S rRNA and *uidA* gene, and one for *Enterococcus* spp.), 7 MST markers (one general MST marker, two human-specific and host-specific for Canada goose, dog, pig and seagull) and 15 waterborne pathogen virulence gene markers simultaneously. The OpenArray® plate was preloaded with primers and probes labeled with carboxyfluorescein (FAM) at the 5´-end and Black Hole Quencher® 1 (BHQ1) at the 3′-end. The efficiency of our TaqMan® assays in the OpenArray® platform was evaluated using DNA from our selected control (sewage and fecal) and environmental samples. The extracted DNA from each sample (2.5 µL) was mixed with 2.5 µL TaqMan® OpenArray® Real-Time PCR Master Mix (Applied Biosystems™) for loading on the OpenArray® instrument. Also, seven combined known concentrations (2,000,000, 200,000, 20,000, 2,000, 200, 20, 2 and 1 copies of each target/hole (OpenArray® well) were loaded on the OpenArray® plate. The mixture was loaded using the QuantStudio 12K Flex Accufill

System (Applied Biosystems™, USA) and the OpenArray® plate was run on the QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems, USA). Two no template controls (NTCs) were run along with fecal and environmental samples. After the initial denaturation at 95 °C for 3 min, 40 cycles of the following program were used for amplification; denaturation at 95 °C for 10 s, annealing/extension at 60 °C for 10 s.

6.2.6   Data analysis

The most common method of quantitation associated with qPCR assays is based on the measurement of the threshold cycle (C_T) and is the C_T (baseline threshold) method (Bustin et al., 2009). Relative threshold (Crt) is an alternative method that is more robust for analyzing data generated on the Applied Biosystems™ QuantStudio™ 12K Flex Real. As these two methodologies give comparable quantitation cycle (Cq) values and fold change (FC) values in large data sets (Applied Biosystems QuantStudio™ 12K Flex Real-Time PCR System, Application Note), we used C_T method to analyze our data. The assay was classified as a true positive when two replicates were positive. The standard curve for each marker was generated by linear regression analysis of the C_T value versus the amounts of the template DNA for SYBR green (log copies/μL) and TaqMan® assays (log copies/hole) qPCR assays. The efficiency of each assay was calculated using a standard curve with the equation: Eff = 10(−1/slope) − 1, where slope corresponds to the slope of the standard curve (Bustin et al., 2009). If all target sequences double during each replication cycle of qPCR, amplification efficiency considered as 100% (Bustin et al., 2009).

We estimated the template copy number of the target locus in environmental and fecal samples for each assay using the TaqMan® assays C_T value based on the standard curve. The lowest concentration of each target which generated a C_T value in all replicates was identified as the detection limit. We converted gene copies/250 mL water or gene copies/2 g of fecal matters, assuming that the cell recovery and DNA extraction efficiency were 100% in all samples and that only one gene copy was present per cell. Zero was given to non-detects assays for statistical analysis. Pearson's correlation was used to correlate the concentration of *Bacteroides* spp. (general MST marker), *E. coli* and host-specific markers (Canada goose and seagull) individually,

with detected waterborne pathogens (combined as a single variable). Also, we correlated combined host-specific MST markers concentration as a single variable with combined waterborne pathogens as another variable in all environmental samples. Then we excluded the environmental samples which were negative for waterborne pathogens from the data and the correlation was repeated between individual MST markers, *E. coli* and the combined host-specific marker with combined waterborne pathogens as a single variable. Heatmaps were generated in R version 3.5.3 with ggplot2 on $\log_{10}$ transformed gene copies measured by the OpenArray® plate (Warnes et al., 2009).

6.3 Results

DNA extraction: All the samples had more than 50 ng/µL DNA concentration with the optimum ration of 260/280 (~1.8) and 260/230 (~2.0-2.2) (Lucena-Aguilar et al., 2016) (Appendix E; Table S6.2).

6.3.1 Assay validation

All the designed amplicons had a length of 66-90 bp (Appendix E; Table S6.1) and melting temperature (Tm) of 59±2 °C. We used the Basic Local Alignment Search Tool (BLAST) tool to predict the specificity of the primers and possible non-specific amplification. All the designed probes and primers were specific for their intended targets with a few exceptions; the designed probe for *Enterococcus* spp. showed possible annealing with *Sphingobacterium psychroaquaticum,* however, due to the specificity of the forward and reverse primers for the *Enterococcus* genus, we used the designed probe in the OpenArray® plate. Also, the designed probe and primers for *Bacteroides* spp. as a general MST marker, showed possible annealing with *Prevotella* spp. As genus *Bacteroides* and *Prevotella* belongs to the same phyla (*Bacteroidetes*) and both have high sequence similarity of 16S rRNA and have fecal source (Bernhard and Field, 2000), the designed primer set was *Bacteroides-Prevotella* taxon-specific, however, in the manuscript we considered *Bacteroides* spp. (general MST marker). Also, as *E. coli* and *Shigella* spp. are sister species and consequently have high similarity in their genome sequence (Zuo et al., 2013) the designed probe and primers

for *uidA* gene of *E. coli* also showed possible annealing with *Shigella* spp. based on our "*in silico*" sequence-based analysis.

In PCR trials, the virulence gene and FIBs primers amplified the appropriate target (Appendix E; Table S6.1) in the target bacterial species and no false positives or false negatives were detected by conventional PCR (Appendix E; Figure S6.1-S6.2). The virulence genes and FIBs also were amplified in sewage samples. FIBs amplicons were amplified in different fecal matters (Appendix E; Figure S6.1-S6.2). Some virulence genes were amplified by their corresponding specific primers in pooled fecal samples of different animals and humans. For example, both *C. coli* and *C. jejuni* were amplified in pooled fecal samples of Canada goose and seagull (Appendix E; Figure S6.2). General MST marker (*Bacteroides* spp.) was detected in all fecal and sewage samples (Appendix E; Figure S6.2). No PCR amplification was detected for any of the host-specific MST markers (i.e., no amplification of non-target host fecal sample DNA), while all host-specific MST assays detected their specified host fecal matter DNA (Appendix E; Figure S6.1-S6.2). As *E. coli* (different strains) and *Shigella* spp. are closely related, the detection of *Shigella* spp. by *uidA* and 23S rRNA genes, PCRs were considered as true positive.

While the goal of this study was to develop a TaqMan® qPCR assays in the OpenArray® platform, we also tested SYBR green qPCR performance for applications where the OpenArray® plate is not feasible. The mean SYBR green qPCR C_T values across all 25 marker loci with known template DNA concentrations of 2,000,000, 200,000, 20,000, 2,000, 200 and 20 copies/reaction were 17.42 ± 0.36, 21.17 ± 0.45, 24.12 ± 0.56, 27.4 ± 0.50, 30.6 ± 0.59 and 33.8 ± 0.57 qPCR (Appendix C; Table S4.3), a pattern of increasing C_T consistent with the dilutions. Two copies/reaction were detected in 6 (24%) of the assays, but none of the assays generated a C_T value for one copy/reaction (Appendix E; Table S6.3). Mean amplification efficiency (across all 25 assays) was 93 – 115% for the SYBR green qPCR (Table 6.1). No C_T value was detected for NTC for each SYBR green assay. Also, SYBR green qPCR assays of three genes (*ctxA*, *manC* and NH8B_364) with know concentration of each gene in water samples shown a linear relationship between the inoculated gene copies of each marker and the those quantified by qPCR assays (Appendix E; Figure S6.4).

The mean C_T values across all assays for the known template DNA concentrations 2,000,000, 200,000, 20,000, 2,000, 200, 20 and 2 copies/hole (OpenArray® well) were 16.63 ± 0.5, 19.85 ± 0.64, 22.93 ± 0.66, 25.91 ± 0.69, 29.06 ± 0.5, 32.68 ± 0.77 and 35.62 ± 0.79 respectively for the TaqMan® assays (Appendix E; Table S6.3). All assays generated C_T values for 2 copies/hole; however, only 6 (24%) of the assays generated C_T values for one copy/hole (Appendix E; Table S6.3). Mean amplification efficiency (across all 25 assays) was 92 - 115% for TaqMan® assays (Table 6.1). No C_T value was detected for NTC for each TaqMan® assay.

Table 6.1. Performance of the SYBR green qPCR and TaqMan® assays in the OpenArray® platform

Species/Target	Genes	TaqMan® assays in the OpenArray® plate			SYBR green qPCR		
		Slope	r2	Efficiency (%)	Slope	r2	Efficiency (%)
Acinetobacter baumannii	*gltA*	-3.1687	0.98	106	-3.0267	0.99	114
Aeromonas hydrophila	*lip*	-3.3741	0.99	97	-3.1509	0.99	107
Bacterioides spp.	16S rRNA	-3.0304	0.98	113	-3.182	0.99	106
Bacterioides sp. clone CGOF52 (Goose marker)	16S rRNA	-3.1662	0.99	106	-3.1183	0.98	109
Bacteriodes A3 (Dog marker)	16S rRNA	-3.3384	0.99	99	-3.1981	0.99	105
Bacteriodes Cluster 1, PigA4 (Pig marker)	16S rRNA	-3.0957	0.99	110	-3.0081	0.98	115
Campylobacter coli	*glyA*	-3.2852	0.99	101	-3.0625	0.99	112
Campylobacter jejuni	*hipo*	-3.1499	0.99	107	-3.1499	0.99	107
Catellicoccus marimammalium (Seagull marker)	16S rRNA	-3.1158	0.99	109	-3.3625	0.99	98
Enterococcus spp.	23S rRNA	-3.5028	0.99	92	-3.3171	0.98	100
Escherichia coli	23S rRNA	-3.3234	0.99	99	-3.1791	0.99	106
Escherichia coli	*uidA*	-3.0039	0.99	115	-3.4497	0.99	94
Escherichia coil O157:H7	*manC*	-3.0807	0.99	111	-3.2806	0.99	101
Escherichia coli O111	*manC*	-3.2033	0.99	105	-3.359	0.99	98
Escherichia coli O26	*manC*	-3.1909	0.99	105	-2.9999	0.98	115
Human C40 mitochondria (Human marker)	MT-ND2	-3.2345	0.99	103	-3.0903	0.99	110
Klebsiella pneumoniae	*phoE*	-3.1186	0.99	109	-3.0449	0.99	113
Legionella pneumophila	*mipA*	-3.1292	0.99	108	-3.4601	0.99	94
Listeria monocytogenes	*Hly*	-3.3589	0.98	98	-3.4888	0.99	93
Methanobrevibacter smithii (Human marker)	*nifH*	-3.1292	0.99	108	-3.4705	0.99	94
Pseudomonas aeruginosa	*regA*	-3.1515	0.99	107	-3.0936	0.97	110

Salmonella typhimurium	*invA*	-3.2793	0.99	101	-3.3565	0.97	98
Shigella spp.	*ipaH*	-3.1852	0.99	106	-3.1458	0.99	107
Staphylococcus aureus	*gyrA*	-3.2261	0.99	104	-3.2081	0.98	104
Vibrio cholerae	*ctxA*	-3.1595	0.99	107	-3.3796	0.99	97

6.3.2 TaqMan® assays in the OpenArray® plate applications

Five different types of environmental samples (beach water, beach sand, beach pore water, stream water and pond water), sewage samples (known as contaminated environmental samples) and fecal samples from various sources (Canada goose, dog, human, pig and seagull) were used to evaluate the effectiveness of the designed OpenArray® plate. We successfully quantified FIBs and MST markers in all fecal samples and sewage samples (as known as contaminated environmental samples) using our newly developed TaqMan® assays. Virulence genes of waterborne pathogens were detected in sewage samples (Figure 6.1). FIBs and MST markers (except the pig marker), and the virulence genes of some waterborne pathogens, including *A. baumannii, A. hydrophila, P. aeruginosa, K. pneumoniae* and *E. coli* O26, were quantified in environmental samples (Figure 6.2).
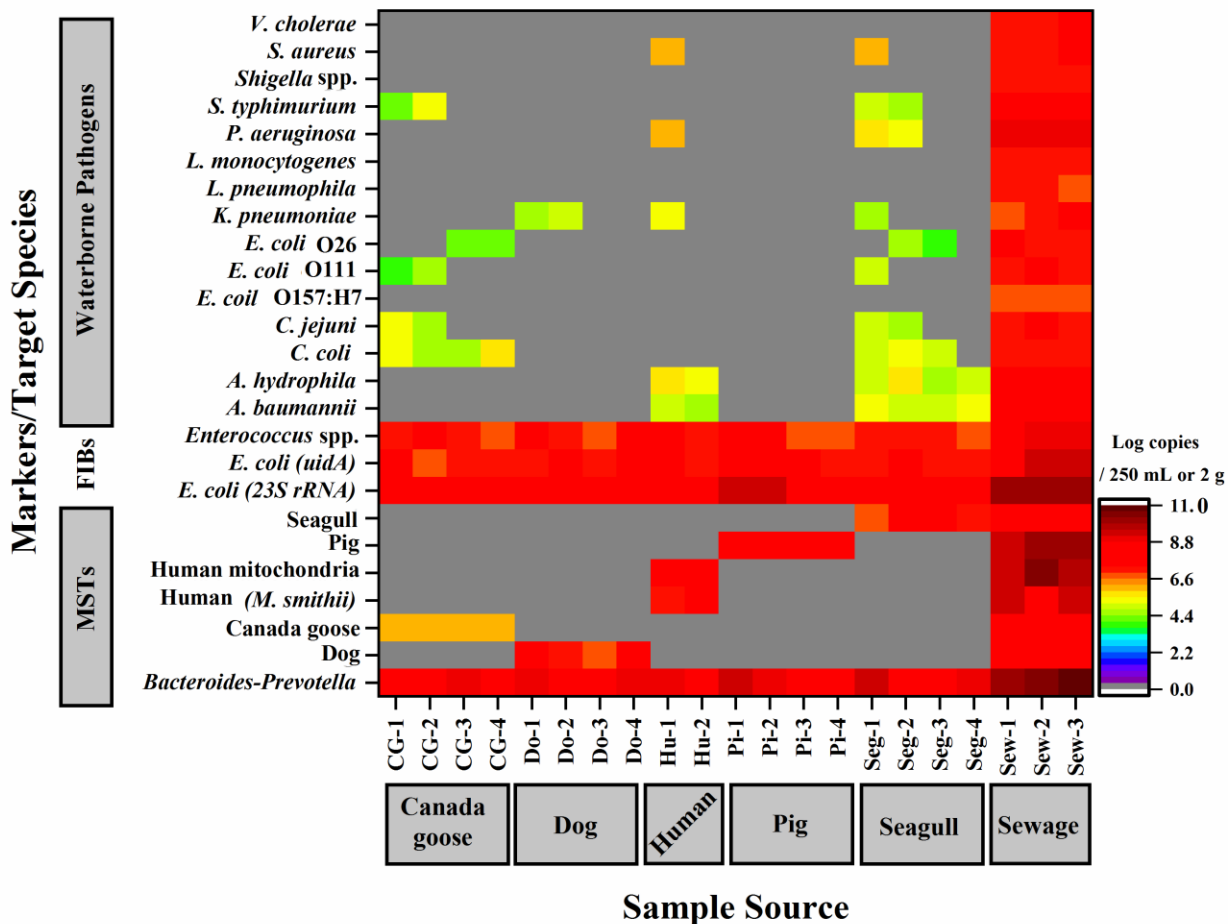
Figure 6.1. Heatmap of qPCR amplification (based on $\log_{10}(C_T)$) using TaqMan® assays in the OpenArray® platform to detect targets in fecal samples. Seven MST, 2 FIBs and 15 waterborne pathogen markers are presented on the Y-axis. Non-detection is shown by gray color in the heatmap.
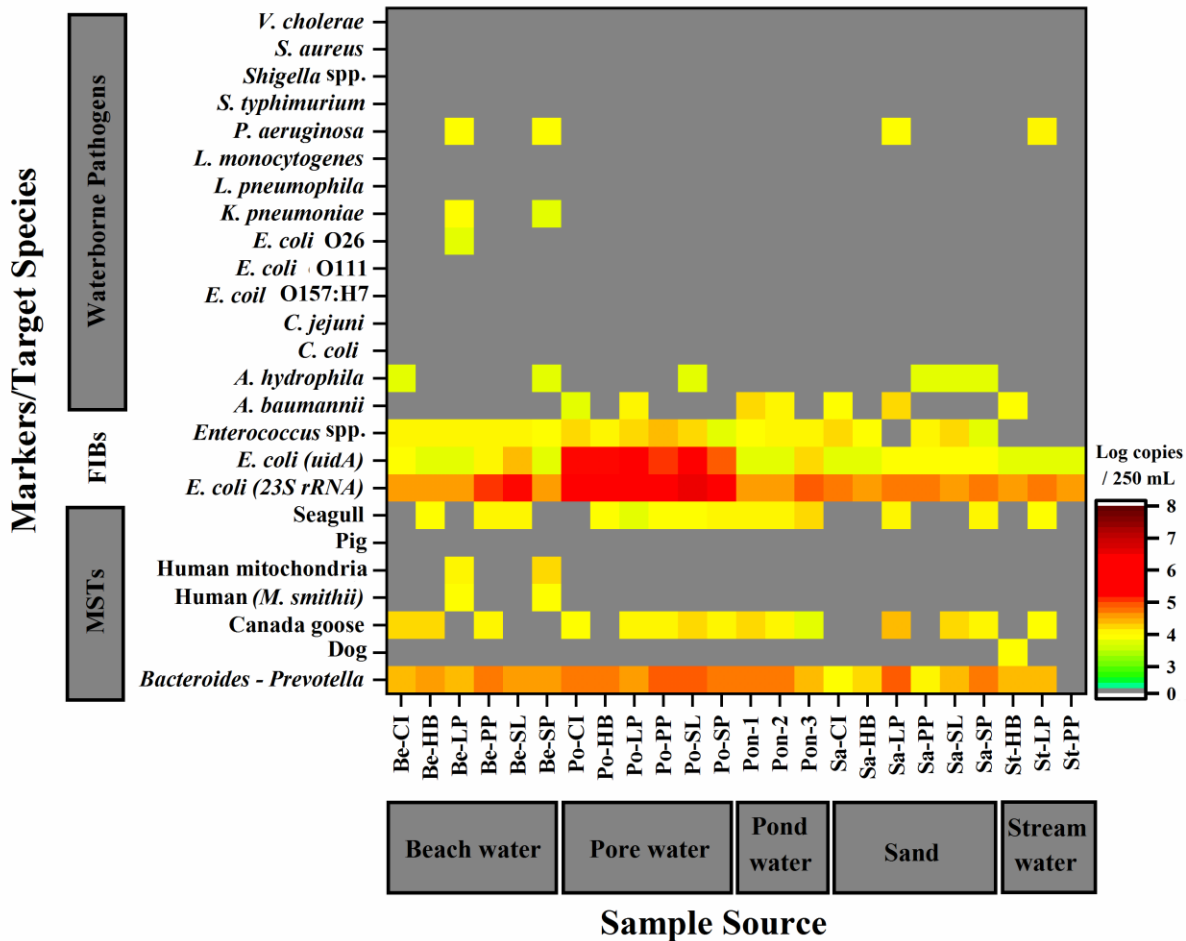
Figure 6.2. Heatmap of qPCR amplification (based on $\log_{10}(C_T)$) using TaqMan® assays to detect targets in the environmental samples. Seven MST, 2 FIBs and 15 waterborne pathogen markers are presented on the Y-axis. Environmental samples were collected from 6 public beaches PP=Point Pelee beach; SL=Seacliff beach; CI=Cedar Island beach; HB=Holiday beach; LP=Lakeview Park beach and SP=Sand Point beach. Non-detection is shown by gray color in the heatmap.

E. coli as a FIB was detected in all sewage samples, the appropriate fecal samples and in the environmental samples. The estimated concentration of *E. coli* was high in our study using the 23S rRNA marker compared to the single-locus *uidA* gene marker due to the presence of multiple rRNA loci in *E. coli* cells (Brosius et al., 1981) and potentially low taxonomic specificity for environmental samples (Figure 6.1-6.2) (Chern et al., 2011). We thus focus on *E. coli* levels based on the single-copy *uidA*

gene as all virulence genes targeted in this study are single-copy genes. However, the inclusion of the 23S rRNA marker allows comparisons to other studies that use that marker. Based on *uidA* qPCR data among 24 environmental samples, 9 (37.5%) had *E. coli* levels >4 log copies/250 mL; including 2 beach water samples and 1 pond water sample and 6 pore water samples. The remaining 15 (62.5%) environmental samples had *E. coli* levels <4 log copies/250 mL level. *Enterococcus* spp., the other tested FIB, was present in most of the tested samples except four environmental samples (one sand and three stream samples) – despite *E. coli* being detected in those samples (Figure 6.2). *E. coli* was not correlated with waterborne pathogens across all 24 environmental samples (R^2=0.005, p=0.87), nor was it correlated in the samples which were positive for waterborne pathogens (15 samples, R^2=0.07, p=0.33; Appendix D; Figure 6.3A).

All seven MST markers were detected in sewage samples and their appropriate host fecal samples (Figure 6.1). Out of 24 environmental samples, two beach water samples (8%) were positive for human MST markers (C40 mitochondria and *M. smithii*). Canada goose and seagull MST markers were detected individually in 15 (62.5%) and 14 (58%) of the environmental samples respectively. Canada goose and seagull MST markers amplified simultaneously in 12 (54%) of the environmental samples. The dog MST marker amplified in one environmental sample (stream), while the pig marker did not amplify in any of the environmental samples. The samples with *E. coli* levels >4 log *uidA* copies/250 mL (6 pore water samples, 2 beach water samples and 1 pond water sample), were positive for one of the bird species (Canada goose and seagull) MST markers. Out of 15 environmental samples with *E. coli* levels <4 log *uidA* gene copies/250 mL); 11 samples (73%) were positive for general and host-specific MST markers (human and birds) and 4 (27%) samples showed none of the specific MST marker signatures but were positive for general MST marker (*Bacteroides* spp.).

All 15 waterborne pathogens were detected in sewage samples (Figure 6.1). Interestingly, virulence genes of *C. coli*, *C. jejuni*, *E. coli* O26 and O111, *S. typhimurium* were detected in Canada goose and seagull fecal samples, while seagull samples additionally contained *A. baumannii*, *A. hydrophila*, *K. pneumoniae*, *P. aeruginosa,* and *S. aureus* (Figure 6.1). *A. baumannii*, *A. hydrophila*, *P. aeruginosa*, *K. pneumoniae* and *E. coli* O26 were detected in 29%, 25%, 16%, 8% and 4% among 24

environmental samples respectively. The *A. baumannii* virulence gene (*gltA*) was detected in 7 environmental samples (pond (2), pore (2), sand (2), stream (1); Figure 6.2). Also, five environmental samples (beach (1), pore (1), sand (3)) were positive for the *A. hydrophila* virulence gene (*lip*). *P. aeruginosa* virulence gene (*regA*) was positive in 4 environmental samples (beach (2), sand (1) and stream (1)). Two beach water samples were positive for *K. pneumoniae* virulence gene (*phoE*) and one of those samples was also positive for *E. coli* O26 virulence gene (*manC*). Virulence genes of *C. coli*, *C. jejuni*, *E. coil* O157:H7, *E. coli* O111, *L. pneumophila*, *L. monocytogenes*, *S. typhimurium*, *Shigella* spp., *S. aureus* and *V. cholerae* were not detected in any of the collected environmental samples (Figure 6.2).

Out of 24 tested environmental samples, 20 (83%) samples were positive for host-specific MST markers (either of birds, human or dog markers) of which 13 (63%) were positive for 1-3 virulence genes of waterborne pathogens. Out of 24 samples, 2 (8%) samples with pathogens were negative for host-specific MST markers but were positive for general MST marker (*Bacteroides* spp.). Out of 24 samples, 2 samples were negative for host-specific MST markers and pathogens but one of them was positive for general MST marker. Among all 24 environmental samples, there was no significant correlation between any MST markers with waterborne pathogens in this study (Canada goose marker: $R^2=0.02$, p=0.62; seagull marker: $R^2=0.008$, p=0.72; general MST marker: $R^2=0.027$, p=0.33). However, there was a better correlation ($R^2=0.159$, p=0.05) between combined host-specific MST markers (Canada goose, dog, human and seagull) with waterborne pathogens (Figure 6.3B).
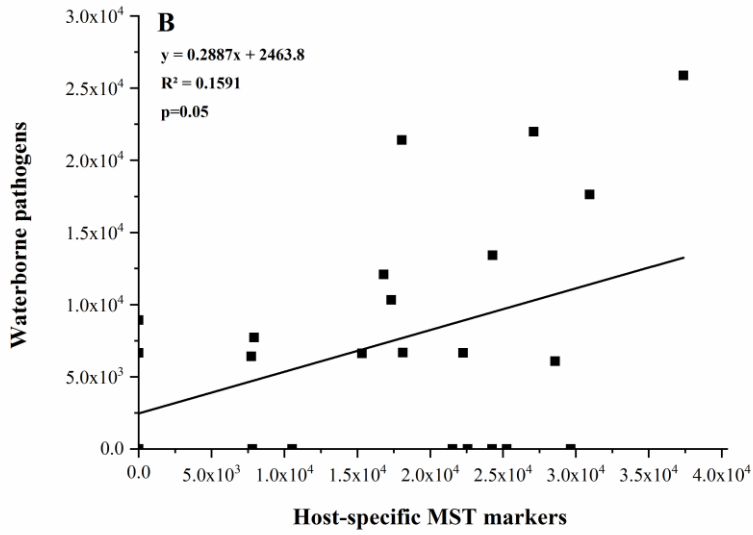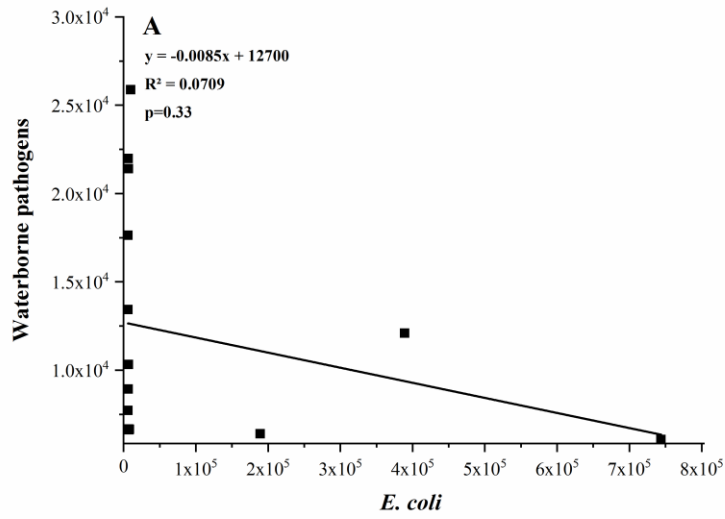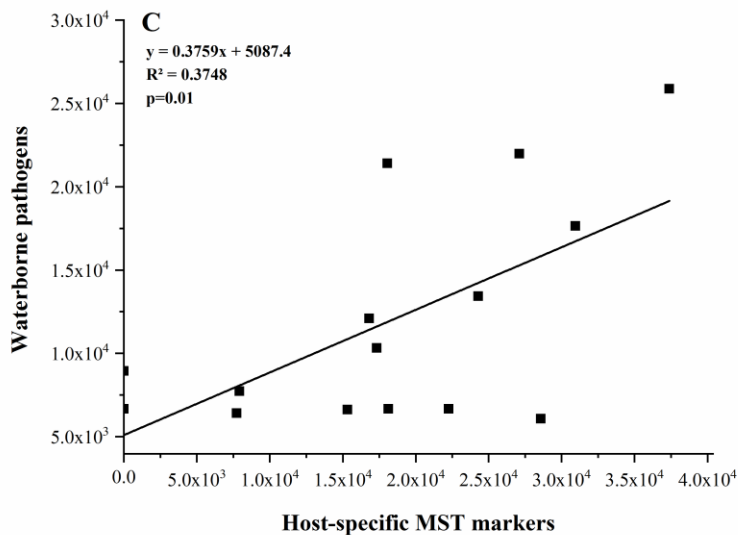
Figure 6.3. Correlations between combined gene concentrations (copies/250 mL) of host-specific MST markers (Canada goose, dog, human and seagull), *E. coli* and waterborne pathogens among the environmental samples measured by TaqMan® assays in the OpenArray® plate. A; correlation between waterborne pathogens (combined concentration of pathogens as a single variable) with *E. coli* (*uidA* gene) among 15 environmental samples positive for pathogens. B; the correlation between host-specific markers (combined concentration of markers as a single variable) with waterborne pathogens (combined concentration of detected pathogens as a single variable) among all 24 environmental samples. C; the correlation between host-specific markers (combined concentration of markers as a single variable) with waterborne pathogens (combined concentration of detected pathogens as a single variable) among 15 environmental samples positive for pathogens.

Out of the 15 environmental samples positive for waterborne pathogens, 3 (20%) samples were pore water with the highest *E. coli* level (>4 log *uidA* copies/250 mL) but 12 (80%) samples had *E. coli* levels <4 log *uidA* copies/250 mL. Out of 15 environmental samples which were positive for waterborne virulence genes, 13 (87%) samples were positive for host-specific MST markers (10 positives for the two bird markers, 2 for the human marker and 1 for the dog marker) however, 2 (sand) samples were negative for all host-specific MST markers but still were positive for the general MST marker (*Bacteroides* spp.). For the environmental samples which were positive

for pathogens (15 samples), there was no significant correlation between any MST markers individually with waterborne pathogens (Canada goose marker: $R^2=0.009$, p=0.56; seagull marker: $R^2=0.07$, p=0.38; general MST marker: $R^2=0.049$, p=0.47) but, there was a significant correlation ($R^2=0.37$, p=0.01) between the combined host-specific MST markers with waterborne pathogens (Figure 6.3C).

6.4   Discussion

In serially diluted positive controls, our TaqMan® assays in the OpenArray® platform were sensitive to the detection of 2 copies/hole (OpenArray® well) for all assays and to 1 copy/hole (OpenArray® well) for 24% of the assays. This compares well to SYBR green qPCR (positive in 24% of assays with 2 copies/reaction and negative in all 1 copies/reaction). However, due to the much larger input volume of DNA (~60 times) in SYBR green qPCR (1 µL) compared to TaqMan® assays (16.5 nL), SYBR green qPCRs would be more efficient for detection. This sensitivity challenge for the TaqMan® assays in the OpenArray® platform was due to the small volume of input DNA could be overcome by increasing water sample volume combined by the concentration of the assay DNA using DNA precipitation protocols or commercially available filters (up to 100X concentration using Microcon® Centrifugal Filters; Millipore, MA) or by specific target amplification (Ishii et al., 2013) before loading the DNA on the OpenArray® plate. Courdray-Meunier et al., found SYBR green qPCR to be more sensitive than TaqMan® assays in the OpenArray® platform for quantifying human pathogenic viruses (Coudray-Meunier et al., 2016). We found that 2 copies/hole was the robust detection limit across all 25 of our target loci using TaqMan® assay qPCR. Detection limits ranging from 2.0 to 2.0×10^6 copies/µL were previously reported for microfluidic qPCR assays designed for FIBs and waterborne pathogens (Ishii et al., 2013). High sensitivity of assays reported by Ishii et al (2013), maybe a result of the pre-amplification step of input temples used for microfluidic qPCR assays (Ishii et al., 2013).

Evaluation of the presence of PCR inhibitors in qPCR experiments is crucial (Schrader et al., 2012). High quality extracted DNA from the samples (Appendix E; Table S6.2) as well as the presence of a linear relationship between the inoculated know concentration of *ctxA*, *manC* and NH8B_3641 (internal control) to the DNA of

water samples in SYBR green qPCR assays (Appendix E; Figure S6.4) indirectly confirmed accurate quantification of genes in TaqMan assays.

Pore water samples had the highest *E. coli* levels (4.9-5.8 log copies/250 mL), showing that the pore water may act as a reservoir and nonpoint source of *E. coli* for surface waters (Vogel et al., 2016). Generally, in our environmental samples, *E. coli* was not consistently associated with waterborne pathogens both across all 24 environmental samples (R^2=0.005, p=0.87) as well as in the 15 waterborne pathogen positive samples (R^2=0.07, p=0.33, Figure 6.3A), potentially indicating that *E. coli* is a poor health risk indicator for recreational waters – a pattern previously reported (Jang et al., 2017; Odonkor and Ampofo, 2013). However, Thoe et al., reported that *E. coli* is a suitable FIB in their weekly study of 37 Hong Kong beaches over four years (Thoe et al., 2018). This highlights the fact that *E. coli* as a FIB should be evaluated in the context of local environmental conditions – our OpenArray® plate will facilitate the efficient local assessment of FIBs.

Out of 24 environmental samples, 9 (37.5%) samples (6 pore water, 2 beach water and 1 pond water sample) with *E. coli* levels >4 log *uidA* gene copies/250 mL, showed positive bird MST marker signatures. On the other hand, out of 15 (62.5%) environmental samples with *E. coli* levels <4 log *uidA* gene copies/250 mL, 11 samples (73%) were positive for host-specific MST markers (birds, human and dog markers). Presence of host-specific MST markers in 73% of the samples with *E. coli* level <4 log *uidA* gene copies/250 mL, along with the presence of general MST marker (*Bacteroides* spp.) in 93% those samples, indicating the likely presence of other sources of fecal pollution (Rogers et al., 2018). Similar to our findings, inconsistent relationships between FIBs and MST markers have been reported previously (Bradshaw et al., 2016).

The presence of pathogen virulence genes in sewage samples in our study and others (García-Aljaro et al., 2018), indicates that sewage is potentially an important source of waterborne pathogens. Detection of *C. coli*, *C. jejuni*, *E. coli* O26 and O111, *S. typhimurium* in Canada goose fecal samples and the presence of those pathogens plus *A. baumannii*, *A. hydrophila*, *K. pneumoniae*, *P. aeruginosa,* and *S. aureus* (Figure

6.2) in seagull fecal samples highlights the potential role of birds as vectors for pathogenic bacteria at beaches (Vogt et al., 2018).

The detection of our host-specific MST markers in 87% of the environmental samples which were positive for pathogen virulence genes shows that in our study host-specific MST markers are much better (R^2=0.159, p=0.05) than FIBs for determining recreational water safety, while also defining the source of the pollution. This correlation of the MST markers with the pathogen markers was significant for 15 environmental samples which were positive for waterborne pathogens (R^2=0.37, p=0.01, Figure 6.3C), showing that combination of host-specific MST markers (in our setting; Canada goose, dog, human and seagull) are potentially better predictors of water safety. Replacing MST with FIBs would be an excellent solution based on our data. However, as FIBs are in practice for a long time instead of eliminating FIBs completely, incorporating MSTs along with FIBs will help in making more accurate decisions. Our data also showed that combining the multiple host-specific MST markers would be the best tool for monitoring recreational water quality and potentially could avoid false positive or false negatives.

Prior to this study, several approaches have been proposed for monitoring recreational water safety. Li et al., developed a microarray chip to detect waterborne pathogens and MST (bovine, pig and human but not birds such as Canada goose and seagull markers) (Li et al., 2015); however, their microarray sensitivity and specificity is relatively low compared to our TaqMan® assays, as well as previously published qPCR assays (Ishii et al., 2013). Additionally, Ishii et al., and Ramalingam et al., developed TaqMan assays (microfluidic platform) that allowed the simultaneous detection of 14 and 4 waterborne pathogens respectively (Ishii et al., 2013; Ramalingam et al., 2010); however, neither incorporated MST markers. To the best of our knowledge, this is the first molecular genetic water safety tool that incorporates human and animal MST markers along with waterborne pathogens and FIBs in a high sensitivity parallel assay. While specificity and sensitivity are critical for recreational water safety monitoring, timelines are also important. We thus incorporated robotic DNA extraction (Shahraki et al., 2019) to minimize human error and assay run time. Using such a platform, we completed environmental sample preparation, DNA

extraction, TaqMan® assays (OpenArray® platform) analyses and data processing in less than four hours.

We used the TaqMan probe assays in our OpenArray® plate to achieve higher specificity than possible with SYBR green qPCR method. Because only sequence-specific amplification is measured in TaqMan assays, the detection of false positives is greatly reduced (Tajadini et al., 2014). Also, we ran duplicates for each sample and the results were highly consistent; however, false negatives are still possible due to the low concentration of pathogens expected in most environmental samples (Johnson et al., 2013) coupled with stochastic sampling effects (Peccoud and Jacob, 1996). We detected few positive hits for waterborne pathogens in our environmental samples (Figure 6.2) potentially due to the low concentration of the pathogens in our collected samples. For actual risk assessment, increasing the volume of water sampled would increase the likelihood of waterborne pathogen detection.

The running cost for TaqMan® assays in the OpenArray® platform (48 subarrays, 66 wells) for 25 markers (duplicate run in each subarray) for this study was \$11/sample and \$0.44/marker, including the reagent costs for the assays and the plate but excluding costs for labor and equipment, which is comparable to conventional qPCR (\$0.45/assay/sample) previously estimated (Ishii et al., 2014).

6.5   Conclusion

Through rapid and simultaneous quantification of waterborne pathogens, MST markers and FIBs, our OpenArray® plate will contribute to determining the concentration of the target pathogens with significant health risks. In properly equipped laboratories, this OpenArray® plate would be able to quantify FIBs, MSTs to define the source of fecal pollution and waterborne pathogens in less than four hours, which would allow close to real-time monitoring of water safety and hence facilitate effective risk management.

6.6   References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25(17) 3389-3402.

Bernhard, A.E., Field, K.G., 2000. A PCR assay to discriminate human and ruminant feces on the basis of host differences in Bacteroides-Prevotella genes encoding 16S rRNA. Applied and environmental microbiology 66(10) 4571-4574.

Bradshaw, J.K., Snyder, B.J., Oladeinde, A., Spidle, D., Berrang, M.E., Meinersmann, R.J., Oakley, B., Sidle, R.C., Sullivan, K., 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 101 498-509.

Brosius, J., Dull, T.J., Sleeter, D.D., Noller, H.F., 1981. Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli. Journal of molecular biology 148(2) 107-127.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical chemistry 55(4) 611-622.

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

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

Cheyne, B.M., Van Dyke, M.I., Anderson, W.B., Huck, P.M., 2010. The detection of Yersinia enterocolitica in surface water by quantitative PCR amplification of the ail and yadA genes. Journal of Water and Health 8(3) 487-499.

Coudray-Meunier, C., Fraisse, A., Martin-Latil, S., Delannoy, S., Fach, P., Perelle, S., 2016. A novel high-throughput method for molecular detection of human pathogenic viruses using a nanofluidic real-time PCR system. PloS one 11(1) e0147832.

DeFlorio-Barker, S., Wing, C., Jones, R.M., Dorevitch, S., 2018. Estimate of incidence and cost of recreational waterborne illness on United States surface waters. Environmental Health 17(1) 3.

Fan, H., Wu, Q., Kou, X., 2008. Co-detection of five species of water-borne bacteria by multiplex PCR. Life Science Journal 5(4) 47-54.

Friedrich, S.M., Zec, H.C., Wang, T.-H., 2016. Analysis of single nucleic acid molecules in micro- and nano-fluidics. Lab on a Chip 16(5) 790-811.

García-Aljaro, C., Blanch, A.R., Campos, C., Jofre, J., Lucena, F., 2018. Pathogens, faecal indicators and human-specific microbial source-tracking markers in sewage. Journal of applied microbiology 126 701-707.

Harwood, V.J., Staley, C., Badgley, B.D., Borges, K., 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) 1-40.

Ishii, S., Kitamura, G., Segawa, T., Kobayashi, A., Miura, T., Sano, D., Okabe, S., 2014. Microfluidic quantitative PCR for simultaneous quantification of multiple viruses in environmental water samples. Applied and environmental microbiology 80(24) 7505-7511.

Ishii, S., Sadowsky, M.J., 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes and Environments 23(2) 101-108.

Ishii, S., Segawa, T., Okabe, S., 2013. Simultaneous quantification of multiple food and waterborne pathogens by use of microfluidic quantitative PCR. Applied and environmental microbiology AEM. 00205-00213.

Jacovides, C.L., Kreft, R., Adeli, B., Hozack, B., Ehrlich, G.D., Parvizi, J., 2012. Successful identification of pathogens by polymerase chain reaction (PCR)-based electron spray ionization time-of-flight mass spectrometry (ESI-TOF-MS) in culture-negative periprosthetic joint infection. Journal of Bone and Joint Surgery 94(24) 2247-2254.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. Journal of Applied Microbiology 123(3) 570-581.

Johnson, G., Nolan, T., Bustin, S.A., 2013. Real-time quantitative PCR, pathogen detection and MIQE, PCR Detection of Microbial Pathogens. Springer, pp. 1-16.

Lee, C., Marion, J.W., Lee, J., 2013. Development and application of a quantitative PCR assay targeting Catellicoccus marimammalium for assessing gull-associated fecal contamination at Lake Erie beaches. Science of the Total Environment 454 1-8.

Levy, K., Woster, A.P., Goldstein, R.S., Carlton, E.J., 2016. Untangling the impacts of climate change on waterborne diseases: a systematic review of relationships between diarrheal diseases and temperature, rainfall, flooding, and drought. Environmental science & technology 50(10) 4905-4922.

Li, X., Harwood, V.J., Nayak, B., Staley, C., Sadowsky, M.J., Weidhaas, J., 2015. A novel microbial source tracking microarray for pathogen detection and fecal source identification in environmental systems. Environmental science & technology 49(12) 7319-7329.

Lucena-Aguilar, G., Sánchez-López, A.M., Barberán-Aceituno, C., Carrillo-Avila, J.A., López-Guerrero, J.A., Aguilar-Quesada, R., 2016. DNA source selection for downstream applications based on DNA quality indicators analysis. Biopreservation and biobanking 14(4) 264-270.

McClung, R.P., Roth, D.M., Vigar, M., Roberts, V.A., Kahler, A.M., Cooley, L.A., Hilborn, E.D., Wade, T.J., Fullerton, K.E., Yoder, J.S., 2017. Waterborne disease outbreaks associated with environmental and undetermined exposures to water—United States, 2013–2014. MMWR. Morbidity and mortality weekly report 66(44) 1222.

Morrison, T., Hurley, J., Garcia, J., Yoder, K., Katz, A., Roberts, D., Cho, J., Kanigan, T., Ilyin, S.E., Horowitz, D., 2006. Nanoliter high throughput quantitative PCR. Nucleic Acids Research 34(18) e123-e123.

Nhung, P.H., Ohkusu, K., Miyasaka, J., Sun, X.S., Ezaki, T., 2007. Rapid and specific identification of 5 human pathogenic Vibrio species by multiplex polymerase chain reaction targeted to dnaJ gene. Diagnostic microbiology and infectious disease 59(3) 271-275.

Odonkor, S.T., Ampofo, J.K., 2013. Escherichia coli as an indicator of bacteriological quality of water: an overview. Microbiology research 4(1) 2.

Okabe, S., Okayama, N., Savichtcheva, O., Ito, T., 2007. Quantification of host-specific Bacteroides–Prevotella 16S rRNA genetic markers for assessment of fecal pollution in freshwater. Applied Microbiology and Biotechnology 74(4) 890-901.

Pattis, I., Moriarty, E., Billington, C., Gilpin, B., Hodson, R., Ward, N., 2017. Concentrations of Campylobacter spp., Escherichia coli, Enterococci, and Yersinia spp. in the Feces of Farmed Red Deer in New Zealand. Journal of environmental quality 46(4) 819-827.

Peccoud, J., Jacob, C., 1996. Theoretical uncertainty of measurements using quantitative polymerase chain reaction. Biophysical journal 71(1) 101-108.

Ramalingam, N., Rui, Z., Liu, H.-B., Dai, C.-C., Kaushik, R., Ratnaharika, B., Gong, H.-Q., 2010. Real-time PCR-based microfluidic array chip for simultaneous detection of multiple waterborne pathogens. Sensors and Actuators B: Chemical 145(1) 543-552.

Ramírez-Castillo, F., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F., Harel, J., Guerrero-Barrera, A., 2015. Waterborne pathogens: detection methods and challenges. Pathogens 4(2) 307-334.

Rogers, S.W., Shaffer, C.E., Langen, T.A., Jahne, M., Welsh, R., 2018. Antibiotic-Resistant Genes and Pathogens Shed by Wild Deer Correlate with Land Application of Residuals. EcoHealth 15(2) 409-425.

Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors–occurrence, properties and removal. Journal of applied microbiology 113(5) 1014-1026.

Shahraki, A.H., Chaganti, S.R., Heath, D., 2019. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. Journal of Water and Health 17 37-49.

Soller, J.A., Eftim, S., Wade, T.J., Ichida, A.M., Clancy, J.L., Johnson, T.B., Schwab, K., Ramirez-Toro, G., Nappier, S., Ravenscroft, J.E., 2016. Use of quantitative microbial risk assessment to improve interpretation of a recreational water epidemiological study. Microbial Risk Analysis 1 2-11.

Tajadini, M., Panjehpour, M., Javanmard, S.H., 2014. Comparison of SYBR Green and TaqMan methods in quantitative real-time polymerase chain reaction analysis of four adenosine receptor subtypes. Advanced Biomedical Research 3.

Thoe, W., Lee, O.H., Leung, K., Lee, T., Ashbolt, N.J., Yang, R.R., Chui, S.H., 2018. Twenty five years of beach monitoring in Hong Kong: A re-examination of the beach water quality classification scheme from a comparative and global perspective. Marine pollution bulletin 131 793-803.

Ufnar, J., Wang, S.Y., Christiansen, J., Yampara-Iquise, H., Carson, C., Ellender, R.D., 2006. Detection of the nifH gene of Methanobrevibacter smithii: a potential tool to identify sewage pollution in recreational waters. Journal of applied microbiology 101(1) 44-52.

Vogel, L.J., O'Carroll, D.M., Edge, T.A., Robinson, C.E., 2016. Release of Escherichia coli from foreshore sand and pore water during intensified wave conditions at a recreational beach. Environmental science & technology 50(11) 5676-5684.

Vogt, N.A., Pearl, D.L., Taboada, E.N., Mutschall, S.K., Janecko, N., Reid-Smith, R., Bloomfield, B., Jardine, C.M., 2018. Epidemiology of Campylobacter, Salmonella and antimicrobial resistant Escherichia coli in free-living Canada geese (Branta canadensis) from three sources in southern Ontario. Zoonoses and public health 65 873-886.

Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., 2010. Global threats to human water security and river biodiversity. Nature 467(7315) 555-561.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., 2009. gplots: Various R programming tools for plotting data. R package version 2(4) 1.

Wolk, D.M., Hayden, R.T., 2011. Quantitative molecular methods, Molecular Microbiology. American Society of Microbiology, pp. 83-105.

Wu, S.-Y., Hulme, J., An, S.S.A., 2015. Recent trends in the detection of pathogenic Escherichia coli O157: H7. BioChip Journal 9(3) 173-181.

Zuo, G., Xu, Z., Hao, B., 2013. Shigella strains are not clones of Escherichia coli but sister species in the genus Escherichia. Genomics, proteomics & bioinformatics 11(1) 61-65.

GENERAL CONCLUSION

7.1 Summary

Climate change, changes in precipitation and run-off patterns, eutrophication, physical alteration of habitats, pollutants and invasive species are global stressors that are threatening the health and biodiversity of all ecosystems. Microbial communities constitute the majority of the Earth's biomass and catalyze many ecosystem processes and deliver ecosystem services in virtually all ecosystems. Microbial communities are thus essential for sustaining life. The Laurentian Great Lakes (LGLs) constitute an enormous freshwater ecosystem and as such regulate climate, support nutrient cycling, transport water and materials, maintain water quality, support fisheries and recreational water activities (Castello and Macedo, 2016). The LGL bacterial communities (BCs) have fundamental roles in all LGL ecosystem biological processes (Fisher et al., 2015; Singh and Walker, 2006) and consequently, have a critical influence on the LGL's function. The recycling of organic material through bacteria and microzooplankton to higher trophic levels, known as the "microbial loop", is an important process in aquatic ecosystems. Monitoring abundance, biomass and the functional activities (enzymatic hydrolysis, production and respiration) of aquatic prokaryotes could provide useful information on the impact of both natural and anthropogenic effects on the microbial loop in aquatic environments, and to assess the environmental status from a holistic point of view (Caruso et al., 2016).

Due to the interaction of biotic and abiotic variables with the BC, and complex interactions among the BC taxa, multi-disciplinary research approaches that combine ecological, evolutionary, microbiological and genetic research fields are necessary to better understand how multiple stressors are likely to impact aquatic ecosystems and to provide decision-making tools to tackle the consequences of stress effects. This Ph.D. thesis consists of five research projects (chapters) and those projects were designed to focus on various key topics in freshwater BC ecology and human health-related issues. Throughout my doctoral research, I incorporated multidisciplinary and statistical approaches that allowed me to address critical knowledge gaps in freshwater microbial ecology, including; 1) tempo-spatial variation of freshwater BC composition and gene

transcription profile at various different scales, 2) quantitative monitoring of waterborne pathogens and indicators using a new method, 3) assessing virulence and other human-health related genes in the BC metatranscriptome, and lastly 4) characterizing the adaptive response of "pre-exposed" freshwater BCs to a very high level of nutrient stress.

Chapter 2 was designed to test for fine-scale temporal variation in the freshwater BC composition. Significant bi-hourly and day/night variation in the composition of the freshwater BC was detected, with a significant elevation of bacterial diversity at night. We also noted fine-scale temporal variation of *Escherichia coli* (a fecal indicator bacteria, or FIB) with the lowest level between 10:00 – 16:00, and significant elevation at night versus day (Shahraki et al., 2020). Chapter 3 of this thesis focused on temporal (large-scale; monthly and seasonal) and spatial (six different sampling locations in Lake Erie and St. Clair) variation in BC composition. We identified substantial monthly and seasonal variation in the composition of the BCs; however, we only captured weak spatial variation. Interestingly, the BC from the samples taken in the two summers (2016 and 2017) not only were significantly different, but also the BC diversity in summer 2017 was significantly lower than in summer 2016. In chapter 4, strong temporal (seasonal) variation was detected in the transcription profile of physiologically and ecologically relevant genes of the freshwater BC; however, spatial variation was limited. We detected the highest transcriptional activity in summer versus the lowest level in winter. Chapter 5 was designed to explore the response of "pre-adapted" freshwater BCs (i.e., exposed to different levels of a nutrient stressor, specifically low and high doses) to very high levels of the nutrient stressor (the challenge). Pre-exposure of the BC to high doses of the stressor (nutrients) resulted in a more adaptive response of the BC to the challenge, characterized by reduced loss of taxonomic diversity and a higher rate of horizontal gene transfer. In chapter 6, a nanofluidic quantitative real-time PCR (qRT-PCR; TaqMan® OpenArray®) chip was designed and validated to monitor 2 FIBs, 7 Microbial Source Tracking (MST) markers and 15 bacterial waterborne pathogens simultaneously (Shahraki et al., 2019b) in recreational fresh water. The OpenArray chip was able to successfully detect the targeted genes in fecal and environmental samples with a high sensitivity (for all

markers; 2 copies of target/hole of OpenArray®) and specificity, all with a turn-around time of less than 4 hours.

Three research chapters (chapters 2, 3 and 4) of my thesis focused on temporal and spatial variation of the freshwater BC composition and gene transcription profile at different scales. Adding to the tempo-spatial studies for microbial communities across different ecosystems will provide essential information about the expected scales of variability, allowing for better biological and ecological interpretations of deviations from normal ranges of variability. This is particularly important in the current century where all ecosystems are heavily impacted by human-introduced stressors (Fierro et al., 2019; Halpern et al., 2008). To predict or manipulate microbially mediated processes, we also need to understand the temporal and spatial patterns of the BC and their diversity at multiple levels (Figure 7.1). Moreover, analysis of temporal and spatial variation at different scales (broad to fine scales) and different levels (individual taxa to the whole community) and the interaction effects of sampling time and location, have allowed us to detect patterns of microbial distribution and their associated ecological activities. Characterizing tempo-spatial changes in the composition and ecological activity of the freshwater BC can provide deeper insight into the processes and mechanisms operating in lake ecosystems (i.e., LGLs) and ultimately improve our basic knowledge and ability to predict BC dynamics and function.
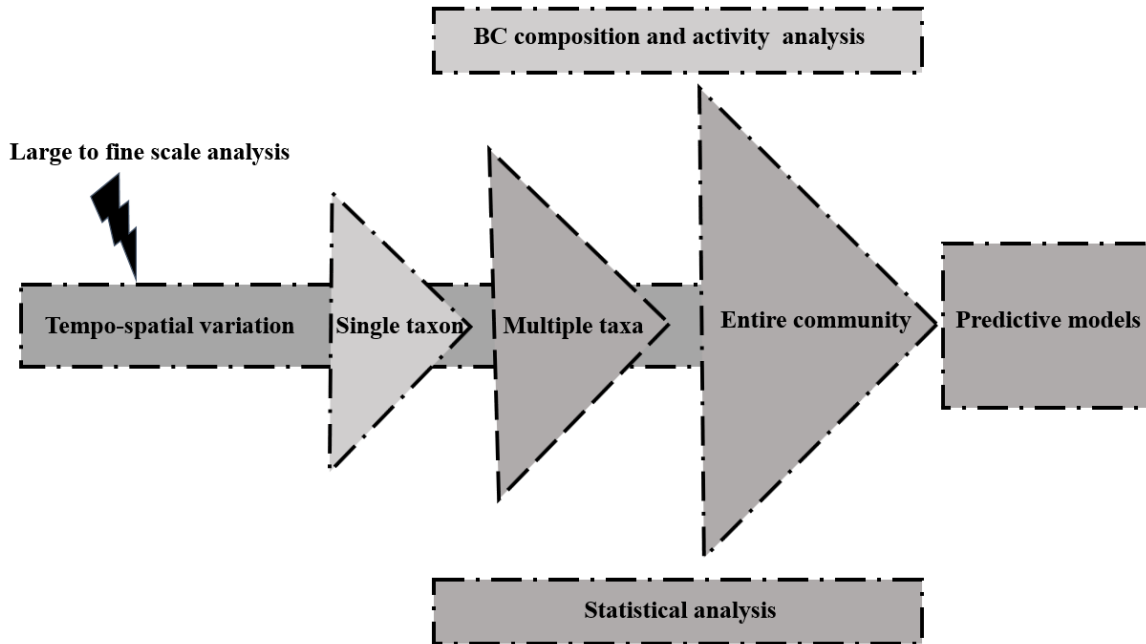
Figure 7.1. Characterizing the temporal and spatial variability of microbial community composition and function at different scales using diverse tools to improve our ability to predict future changes in patterns of BC composition and activity.

More specifically, temporal studies can identify taxa shared at different times, correlations with environmental conditions that affect the communities, and the relative contribution of different processes, including stochastic processes and priority effects, on the microbial community structure. Studies in different environments have shown that some communities exhibit cyclical temporal patterns (Gilbert et al., 2012), a monotonic trajectory (Koenig et al., 2011) or remain relatively stable over time (Sassoubre et al., 2015). In this thesis, three research chapters focused on temporal variation of the freshwater BC composition (chapter 2; fine-scale and chapter 3; broad-scale) and freshwater BC transcriptome (chapter 4; large scale). We captured a fine-scale (chapter 2) (Shahraki et al., 2020) and broad-scale (chapter 3) temporal variation in composition and gene transcription profile (chapter 4) of the freshwater BC.

We found a fine-scale temporal variation (bi-hourly and day/night variation) of the freshwater BC (Shahraki et al., 2020). To the best of our knowledge, our report is the first study to address the temporal variation of freshwater BC at a bi-hourly level in the diel cycle. This study revealed that the freshwater BC composition is changing at a

very fine temporal scale and the nature of that change would lead us to expect fine-scale variation at the metatranscriptome level. We found that some taxa such as *Actinobacteria* prefer daylight hours and some others such as *Enterobacteriaceae* prefer night hours (Shahraki et al., 2020). The presence of rhodopsins in *Actinobacteria* provides a potential mechanism for their supplemental energy generation through light-harvesting, and coupled with their UV stress resistance and potentially better contribution in biogeochemical cycling over day hours (Newton et al., 2011) might explain this pattern of dominance in daylight hours (Shahraki et al., 2020). However, further study is needed to characterize the mechanisms behind our observed fine-scale temporal variation. We also found fine-scale temporal variation in levels of *E. coli* (a FIB marker) which not only provides more evidence regarding the fine-scale temporal variation of the BC but also shows that fine-scale variation in the abundance of an indicator species should be a serious concern for human health risk assessment (Shahraki et al., 2020). For example, in many recreational settings; FIB levels use a predictor of water quality and are measured only one time (a single grab sample) (Boehm, 2007), usually between 8:00 to 16:00. Such sampling could result in sampling the lowest *E. coli* levels, based on our data, and hence mislead evening beachgoers on the safety of the water. It is highly recommended to replace current monitoring protocols with multiple sampling at different times of the day to obtain a high-resolution picture of water quality and human health risk (chapter 2; (Shahraki et al., 2020).

Our finding shows that changes in the composition over time (broad-scale; chapter 3) are likely correlated with changes in the gene transcription profile of the BC (broad-scale, chapter 4) and thus will influence BC activity and ecological services. More interestingly, we identified a composition shift from a community dominated by *Actinobacteria* (sensitive to nutrient overloading and low oxygen level) to one enriched by *Proteobacteria* (adapted to nutrient overloading) (Newton et al., 2011), *Bacteroidetes* (proficient in the degradation of complex biopolymers and dissolved organic matter) (Newton et al., 2011) and *Firmicutes* (diverse metabolic capabilities and resistant to oxygen limitation) (Martiny et al., 2006) over time (summer 2016-summer 2017). We also detected significant biodiversity loss over time, particularly

when we compared the BC of the two sampled summers (2016-2017) which did not show a cyclic pattern, as reported before (Gilbert et al., 2012; Ward et al., 2017). A recent microcosm study (Pradeep Ram et al., 2020), and our experimental study (chapter 5) have shown that nutrient overloading results in a loss of diversity in freshwater BCs. Thus, the dominance of OTUs belonging to *Proteobacteria, Bacteroidetes and Firmicutes* combined with a noticeable reduction in the abundance of *Actinobacteria* observed between the summers of 2016 and 2017 indicates that Lake Erie and Lake St. Clair likely received a higher load of nutrients in summer 2017 compared to 2016. Thus chapter 3 showed that large-scale temporal composition shifts (noncyclic), which could be attributed to changes in the eutrophication of Lake Erie and Lake St. Clair. These composition changes very likely resulted in significant changes in the gene transcription profile (ecological activity variation) of the freshwater BC, and hence the ecological services provided by the BC (although we did not test for transcriptome variation in chapter 3). We found nitrogen level as one of the drivers of temporal (broad-scale) variation of the freshwater BC gene transcription profile (chapter 4), which shows that temporal variation of the BC composition (chapter 3) and temporal gene transcription profile of the BC (chapter 4) can be regulated by nutrient level. Further study is needed to explore the regulatory impact of environmental variables such as nutrient level and eutrophication temporal variation on freshwater BC composition and its ecological activities. Our large-scale temporal analysis of the freshwater BC at two levels (composition and gene transcription levels) and fine-scale temporal variation (composition) of the whole community and specific indicator taxa shows that monitoring of BC at different scales in the ecosystems is an effective bio-monitoring tool to access ecosystem health and its associated human health implications (Figure 7.2).

Understanding how microbial communities vary at different spatial scales is important because diversity and ecological activity hotspots/deserts can be identified, correlations with environmental factors can be detected, and hypotheses about dispersal limitation or stochasticity of community assembly can be tested (Gonzalez et al., 2012). Spatial pattern analysis of BCs improves our ability to predict microbial diversity and ecological activity, and hence predetermine where microbes and their

associated functions are likely to be abundant in a given habitat. Overall spatial patterns in the BCs include subdividing analyses by taxon (from species to phylum-level) and can be used as inputs for niche modeling of the community (Gonzalez et al., 2012). Perhaps the most contested hypothesis in microbial ecology suggests that "everything is everywhere, but the environment selects" (O'Malley, 2007); that is, local environmental conditions, or 'filters', select for community assemblages that can best exploit local resources and thus survive. In the current thesis, three research chapters focused on spatial variation of the freshwater BC composition (chapter 2 and chapter 3) and metatranscriptome profile (chapter 4). The level of spatial variation in both community composition and gene transcription was significant but limited. Even in chapter 5, where we adapted the freshwater BC to different doses of stressor and then exposed the adapted BCs to a very high level of stress, we found very low BC source (spatial) effects in the response of adapted freshwater BCs (the original freshwater BC collected from two different sampling locations in Lake Erie and St. Clair). Although land use, spatial distance, climatic conditions and ecosystem physico-chemical properties are all suggested as drivers of spatial sorting of microbial communities among locations (Bru et al., 2011), we did not observe strong location effects. This may be due, in part, to the connection of the two lakes by the Detroit River (~45 km long) (Burniston et al., 2018). The Detroit River is the only natural outlet of water from Lake St. Clair to Lake Erie and has a very high discharge rate ~ 5,270 m^3/s (Madani et al., 2020). The eutrophication of Lake Erie (Watson et al., 2016) and St. Clair (Bocaniov et al., 2019) could have resulted in the formation of similar water column conditions, and consequently similar habitats which might have resulted in the spatial conservation of the composition of the BC of two lakes in our study. Moreover, the short distance among the sampling locations (maximum ~ 50 km), may not be enough to form distinct habitat with a different environmental conditions which ultimately resulted in the limited spatial variation of the freshwater BC composition (chapter 2 and 5) and gene transcription profiles (chapter 4). We detected limited spatial variation at both the lake and sampling point scales (chapters 2, 3 and 4), as well as significant temporal and spatial interaction effects (chapters 3 and 4), while our study was not designed to explore Great-Lakes wide spatial variation. We did indeed capture spatial

159

effects despite strong temporal variation in our analysis. Also, as we only collected samples from nearshore at beaches in the western basin of Lake Erie and the western shores of St. Clair our study likely under-estimates the spatial effect on BC.

Combining temporal and spatial data sets can reveal key features of dynamic ecosystems. As not all habitats change the same way over time, then combining temporal and spatial variation will provide a unique opportunity to explore the role of integration effects. Studies that track time series at multiple sites are essential for quantifying factors that affect BC dynamics, structure and ecological functions. We combined temporal and spatial analysis of the BC composition and metatranscriptome (chapter 2, 4 and 5). Describing different resolutions of BC spatial variability (from microns to continents), and temporal variability (from hours to decades) and ultimately the interaction effects in different ecosystems, will inform prediction of dynamics and responses to novel events particularly human introduced stressors. Although, studies highlighted temporal and spatial variation of bacterial community in aquatic ecosystems (Codling et al., 2018; Jones et al., 2012; Lear et al., 2014), however, to the best of our knowledge, the interaction effect of tempo-spatial variation on freshwater bacterial community in LGLs did not tested before.
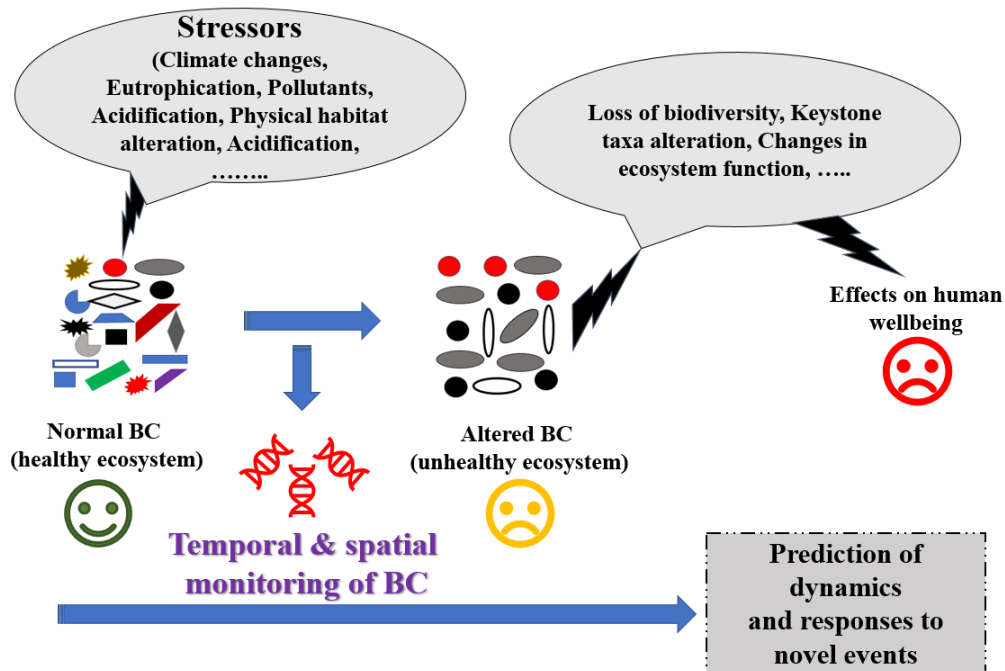
Figure 7.2. Conceptual links between stressors and the freshwater BC. BC composition can be used as a biomonitoring tool to identify changes in aquatic ecosystems resulting from perhaps unknown stressors. Such BC changes are very sensitive to ecosystem stress and may thus respond before the stress affects the rest of the food web.

Metabarcoding the 16S rRNA gene is a key tool for studies of the microbial community composition; however, it does not provide direct evidence of community functional capacity. Although advances in bioinformatics and computational approaches allow us to predict BC activity using 16S rRNA metabarcoding data (Langille et al., 2013), this only provides indirect evidence of the functional capabilities of the BC. Chapter 4 was designed to study the temporal and spatial variation of the gene transcription profile of the freshwater BC. Our metatranscriptome data showed, not surprisingly, that the freshwater BC is more ecologically active in summer and fall than winter. Butler et al, reported a high abundance of nitrogen-fixing bacteria involved in nitrification in the water column under ice-covered conditions (Butler et al., 2019). Out of our 52 selected physiologically and ecologically relevant genes, we identified some level of transcription of 60% of those genes in winter,

161

indicating that the BC is active even under low-temperature conditions. High transcription levels of cold-shock genes in winter highlights a potentially maladaptive shift in the metatranscriptome under low-temperature conditions (De Maayer et al., 2014). We identify transcripts of genes involved in nutrient cycling in winter (ice-free conditions), but their transcription levels were low. It has been suggested that ice duration drives the accumulation of nutrients such as nitrate in north temperate lakes (Powers et al., 2017), which could explain low transcription of nutrient cycling genes in our study (no ice covering in sampling time).

Environmental variables such as water temperature and daylight had a significant influence on a large-scale temporal variation on composition (chapter 3) and gene transcription profile (chapter 4) of the freshwater BC. On the other hand, we also found that water temperature did not affect the fine-scale temporal variation of BC composition (chapter 2) (Shahraki et al., 2020). These findings clearly show that environmental parameters (bottom-up variables; physical and chemical parameters) can have different effects on the composition and activity of the BC according to the scale of the study. Among the interacting forces structuring BCs, most attention has been on bottom-up effects, with lesser focus given to top-down (e.g., viral or protistan) effects on BC composition (Ram et al., 2016). We did not design our studies to test top-down effects on the BC composition, however, a recent microcosm study showed the maintenance of high diversity of freshwater BC in the presence of top-down factors (virus) and a loss of diversity of freshwater BC in the presence of bottom-up variables (nutrients) (Pradeep Ram et al., 2020). The relative importance of bottom-up versus top-down effects in aquatic ecosystems remains a longstanding and ongoing controversy due to the complexity of food web in the aquatic ecosystem and complex and critical interaction of microorganism (prokaryotic and eukaryotic) in their environment and aquatic food web. Our studies (chapters 2, 3 and 4) showed that environmental parameter effects (bottom-up variables) are dependent on the scale and study design, which could be true for top-down variables as well.

The occurrence and abundance of indicator microorganisms can serve as proxies for water quality (i.e., easily measured quantities that are correlated to often unknown agents that directly mediate waterborne risks such as pathogens, biotoxins and

chemicals) (Tan et al., 2015). Advances in molecular methods and next-generation sequencing (NGS) of DNA (metabarcoding and sequencing of 16S rRNA) have ushered in new opportunities for water quality assessment. NGS data is generally clustered into operational taxonomic units (OTUs) by identity thresholds (e.g., 95–99%) (Caporaso et al., 2010) or exact sequence variants (ESVs) by 100% sequence similarity (Porter and Hajibabaei, 2018). We were able to detect harmful bacteria in the metabarcoding data in chapters 2 and 3 (*Enterobacteriaceae*); however, we were not able to detect waterborne pathogens (at the species level) in either study. This was likely due to low resolution in OTU clustering resulting from a short 16S rRNA metabarcode sequence, coupled with a low abundance of waterborne pathogens. Although generating and analyzing metatranscriptome data still is challenging and expensive for water quality monitoring, we were able to detect 20 waterborne pathogens and 17 virulence genes associated with those pathogens in beach water (chapter 5). This elevated sensitivity is potentially due to high sequence read depth (~25 million reads/sample) combined with full gene transcript sequence. On the other hand, metatranscriptomic (RNA-based approach) only capture functional cells and their transcripts. These highlight the potential power of metatranscriptomic versus metabarcoding approaches for water quality assessment. Since metatranscriptomic is RNA-based approach, it. Our study, and a recent application of the study of the metatranscriptomic data analysis of beach sediment (VanMensel et al., 2020), provides more information regarding the high resolution of transcriptome analysis and its potential to capture pathogens and their associated health risks in aquatic ecosystems.

Eutrophication of large freshwater lakes, such as the LGLs, due to anthropogenic activity is one of the most challenging environmental problems facing water resources today (Bhagowati and Ahamad, 2019). Although one study showed, using microcosms (8 bacterial species), that nutrient overloading can cause negative interactions among species of the BC and could cause diversity loss (Ratzke et al., 2020), no information is available regarding the impact of pre-exposure to stress on BC response to very high levels of the stressor. In chapter 5, the freshwater BC was pre-exposed (adaptive phase) to low and high doses of a stressor (nutrients) and then challenged with a very high dose of the same stressor (challenge phase). *Proteobacteria* became the most dominant

taxa. More analysis showed a substantial difference in the adaptation response of different *Proteobacteria* and other taxa against challenge. This study showed that the high level of the stressor can provide better adaptation response against challenge by maintaining more diversity in taxa composition, keystone taxa and high horizontal gene transfer rate.

Culturing and counting *E. coli* (FIBs), remains the most common approach for monitoring pathogenic water pollution (Soller et al., 2016); however, it has several important limitations (time-consuming, ability to grow of some FIBs from aquatic habitats, poor correlation with the occurrence of pathogens and lack of host specificity) (Ishii and Sadowsky, 2008; Jang et al., 2017; Pattis et al., 2017). We designed a nanofluidic qRT-PCR (TaqMan® OpenArray®) chip and validated it by monitoring FIBs, MST markers and bacterial waterborne pathogens simultaneously (chapter 4) (Shahraki et al., 2019b) in artificial and natural samples. Our OpenArray chip successfully detected targeted genes in fecal and environmental samples. Previously designed molecular genetic methods for monitoring waterborne pathogens and FIBs (Ishii et al., 2013; Li et al., 2015) were not able to determine the source (species) of the contamination. Our OpenArray qRT-PCR chip, on the other hand, allows the simultaneous detection of waterborne pathogens, FIBs and MST markers with a less than 4 hour turn-around time - this includes sample collection to results (Shahraki et al., 2019a). MST markers in the designed OpenArray chip revealed that Canada goose and seagull are the leading source contamination at beaches in Windsor-Essex County (Shahraki et al., 2019b). All these results show that the designed OpenArray chip has the potential to use in water quality monitoring in large-scale in a different setting.

7.2   Significance

The results of this thesis highlighted the strong temporal variation with limited spatial variation along with minor interaction effects (sampling time and sampling location) in freshwater BC composition and gene transcription profile. As BC is the main part of the Great Lake ecosystem with direct or indirect effects on many ecosystem-level processes, our tempo-spatial variation in the community composition and transcription profiles in the different scales shows that BC variation could be used to predict the dynamic and impact of stressors on different ecosystems. On the other

hand, monitoring the freshwater BC community composition and its activity allows us to measure the influence of human-introduced stressors and could be used as biological tools for remediation purposes. Within this thesis, a new tool was designed and optimized to the monitoring of freshwater quality which can be used in other settings. Detection of a relatively high number of taxa and their virulence genes associated with water quality degradation in our metatranscriptomic data highlighted that water quality monitoring still is a significant challenge. Our transcriptome analyses provide a general picture regarding the application of transcriptome analysis to infer taxonomic profiles and the activity of harmful microbes in different ecosystems. Lastly, for the first time, our experimental design showed that high levels of stressor can cause irreversible changes in the BC composition, but pre-adaptation could save taxonomic diversity relative to original community against very high stressor challenge.

7.3 Future directions

We recorded a fine-scale temporal variation in the BC composition; however, we found that environmental variables (bottom-up variables) did not have much influence on this variation however, the mechanism(s) of this rapid change is not clear. Theoretically, many factors such as migration of taxa elsewhere, possibility of patchy physicochemical properties at fine-scale, interactions among coexisting species at fine-scale, water movement and turbulence in the water body, growth rate of different taxa (abundant and rare taxa) and etc. might have direct or indirect contribution in fine-scale temporal variation, however, more studies are needed to explore the mechanism(s) behind the fine-scale temporal variation of freshwater BC. Also, more studies are needed to understand the consequence of such fine-scale composition changes in the activity of the community. Although we found that *E. coli* levels are low in day hours relative to night, further study needs to evaluate the fine-scale temporal variation of harmful microbes including waterborne bacterial and viral agents to provide critical information for health risk assessment of recreational water.

We recorded temporal variation in both the composition and function of the freshwater BC. Our work clearly showed that environmental parameters (bottom-up variables; physical and chemical parameters) likely have strong influences on the composition and activity of BC, according to the scale of the study. Among the

interacting forces structuring BCs, much attention has been focused on bottom-up effects, with less attention given to top-down (viral or protistan) effects on BC composition (Ram et al., 2016). The relative importance of bottom-up versus top-down effects in aquatic ecosystems remains a longstanding and ongoing controversy due to the complexity of food web in the aquatic ecosystem and complex and critical interaction of microorganism (prokaryotic and eukaryotic) in their environment and aquatic food web. Study of the bottom-up and top-down effect on microbial community composition in different temporal scale would provide novel insight into the mechanism which involve in the ecology, evolution, dynamic and function of a freshwater ecosystem.

Sand Point (SP) beach is located in an urbanized area and is upstream (~1 km) of the entry point of the Little River Pollution Control Plant effluent to the Detroit River. We observed spatial variation in the transcriptome of SP beach relative to other beaches (chapter 4). As water flows from Lake St. Clair to Lake Erie and SP beach is located in upstream of the sewage effluent entry point, we believe that the detected spatial variation is probably related to urbanization rather than sewage effluent effects. This thesis was not designed to specifically study the sewage effluent effects on freshwater quality. More targeted studies are need address the sewage effluent effects on microbial community of fresh water and public beaches.

We also recorded substantial seasonal variation of the BC metatranscriptome, with the highest transcriptional activity in the summer. Further large-scale temporal studies of the freshwater BC metatranscriptome (dry year versus a year with high precipitation or year with ice cover versus a year with no ice) could provide more information regarding how climate change could influence the microbial metatranscriptome profiles and the ecological services of the Great Lakes. Our study provides field monitoring of harmful microbes and their virulence genes by metatranscriptomic data, but further study is needed to optimize the application of metatranscriptomic data analysis for water quality assessment.

We recorded the adapted BC response to the extreme challenge; however, we did not focus on functionality changes over adaptation and response phases. Further research is needed to understand how microbial taxa adapt to a different level of stress

by changing their gene expression profile and how they evolve genetically. Identification of the gene pools which are exchanging between BC under different stressor levels through horizontal gene transfer will provide further insight into microbial evolution under stressful environment.

Our OpenArray chip could be useful for any water monitoring study; however, some minor modifications would increase the applicability and efficiency of the chip. It would be possible to use RNA as a target, instead of DNA, to distinguish dead or viable but dormant from live bacteria. Also incorporating waterborne viral pathogens (i.e. adenovirus, astrovirus, rotavirus) (Gall et al., 2015), cyanotoxin genes and modifying or adding more MST markers based on wildlife or domestic animals would increase the breadth of application of the chip. In our designed chip we only incorporated MST markers for a few animal sources, such as Canada goose, seagull, pig and dog markers; however, adding more MST markers such as cattle-specific markers and various wildlife animal markers could potentially increase the monitoring power of our OpenArray chip. Contaminants from domestic sewage could enter fresh water by both direct and indirect discharge, or by discharge of treated effluents from wastewater treatment plants. In recent years, various wastewater chemical markers, such as pharmaceutical care products and artificial sweeteners, have been proposed and utilized for identifying the wastewater contamination (Lim et al., 2017). The combination of our OpenArray chip with chemical tracers and stable isotopes could serve as promising tool for detecting sewage leakage to water bodies.

7.4   References

Bhagowati, B., Ahamad, K.U., 2019. A review on lake eutrophication dynamics and recent developments in lake modeling. Ecohydrology & Hydrobiology 19(1) 155-166.

Bocaniov, S.A., Van Cappellen, P., Scavia, D., 2019. On the role of a large shallow lake (Lake St. Clair, USA-Canada) in modulating phosphorus loads to Lake Erie. Water Resources Research.

Boehm, A.B., 2007. Enterococci Concentrations in Diverse Coastal Environments Exhibit Extreme Variability. Environmental Science & Technology 41(12) 8227-8232.

Bru, D., Ramette, A., Saby, N., Dequiedt, S., Ranjard, L., Jolivet, C., Arrouays, D., Philippot, L., 2011. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. The ISME journal 5(3) 532-542.

Burniston, D., Dove, A., Backus, S., Thompson, A., 2018. Nutrient concentrations and loadings in the St. Clair River–Detroit River Great Lakes Interconnecting Channel. Journal of Great Lakes Research 44(3) 398-411.

Butler, T.M., Wilhelm, A.-C., Dwyer, A.C., Webb, P.N., Baldwin, A.L., Techtmann, S.M., 2019. Microbial community dynamics during lake ice freezing. Scientific Reports 9(1) 1-11.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5) 335.

Caruso, G., Azzaro, M., Caroppo, C., Decembrini, F., Monticelli, L.S., Leonardi, M., Maimone, G., Zaccone, R., La Ferla, R., 2016. Microbial community and its potential as descriptor of environmental status. ICES Journal of Marine Science 73(9) 2174-2177.

Castello, L., Macedo, M.N., 2016. Large-scale degradation of Amazonian freshwater ecosystems. Global Change Biology 22(3) 990-1007.

Codling, G., Sturchio, N.C., Rockne, K.J., Li, A., Peng, H., Timothy, J.T., Jones, P.D., Giesy, J.P., 2018. Spatial and temporal trends in poly-and per-fluorinated compounds in the Laurentian Great Lakes Erie, Ontario and St. Clair. Environmental Pollution 237 396-405.

De Maayer, P., Anderson, D., Cary, C., Cowan, D.A., 2014. Some like it cold: understanding the survival strategies of psychrophiles. EMBO Reports 15(5) 508-517.

Fierro, P., Valdovinos, C., Arismendi, I., Díaz, G., De Gamboa, M.R., Arriagada, L., 2019. Assessment of anthropogenic threats to Chilean Mediterranean freshwater ecosystems: Literature review and expert opinions. Environmental Impact Assessment Review 77 114-121.

Fisher, J.C., Newton, R.J., Dila, D.K., McLellan, S.L., 2015. Urban microbial ecology of a freshwater estuary of Lake Michigan. Elementa (Washington, DC) 3.

Gall, A.M., Mariñas, B.J., Lu, Y., Shisler, J.L., 2015. Waterborne viruses: a barrier to safe drinking water. PLoS Pathogens 11(6) e1004867.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I., 2012. Defining seasonal marine microbial community dynamics. The ISME journal 6(2) 298-308.

Gonzalez, A., King, A., Robeson II, M.S., Song, S., Shade, A., Metcalf, J.L., Knight, R., 2012. Characterizing microbial communities through space and time. Current Opinion in Biotechnology 23(3) 431-436.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., 2008. A global map of human impact on marine ecosystems. Science 319(5865) 948-952.

Ishii, S., Sadowsky, M.J., 2008. Escherichia coli in the environment: implications for water quality and human health. Microbes and Environments 23(2) 101-108.

Ishii, S., Segawa, T., Okabe, S., 2013. Simultaneous quantification of multiple food and waterborne pathogens by use of microfluidic quantitative PCR. Applied and environmental microbiology AEM. 00205-00213.

Jang, J., Hur, H.G., Sadowsky, M.J., Byappanahalli, M., Yan, T., Ishii, S., 2017. Environmental Escherichia coli: ecology and public health implications—a review. Journal of Applied Microbiology 123(3) 570-581.

Jones, S.E., Cadkin, T.A., Newton, R.J., McMahon, K.D., 2012. Spatial and temporal scales of aquatic bacterial beta diversity. Frontiers in Microbiology 3 318.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., Ley, R.E., 2011. Succession of microbial consortia in the developing infant gut microbiome. Proceedings of the National Academy of Sciences 108(Supplement 1) 4578-4585.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R., 2013. Predictive functional profiling of

microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31(9) 814-821.

Lear, G., Bellamy, J., Case, B.S., Lee, J.E., Buckley, H.L., 2014. Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. The ISME Journal 8(8) 1715-1726.

Li, X., Harwood, V.J., Nayak, B., Staley, C., Sadowsky, M.J., Weidhaas, J., 2015. A novel microbial source tracking microarray for pathogen detection and fecal source identification in environmental systems. Environmental science & technology 49(12) 7319-7329.

Lim, F.Y., Ong, S.L., Hu, J., 2017. Recent advances in the use of chemical markers for tracing wastewater contamination in aquatic environment: a review. Water 9(2) 143.

Madani, M., Seth, R., Leon, L.F., Valipour, R., McCrimmon, C., 2020. Three dimensional modelling to assess contributions of major tributaries to fecal microbial pollution of lake St. Clair and Sandpoint Beach. Journal of Great Lakes Research 46(1) 159-179.

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4(2) 102-112.

Newton, R.J., Jones, S.E., Eiler, A., McMahon, K.D., Bertilsson, S., 2011. A guide to the natural history of freshwater lake bacteria. Microbiology and Molecular Biology Reviews 75(1) 14-49.

O'Malley, M.A., 2007. The nineteenth century roots of 'everything is everywhere'. Nature Reviews Microbiology 5(8) 647-651.

Pattis, I., Moriarty, E., Billington, C., Gilpin, B., Hodson, R., Ward, N., 2017. Concentrations of Campylobacter spp., Escherichia coli, Enterococci, and Yersinia spp. in the Feces of Farmed Red Deer in New Zealand. Journal of environmental quality 46(4) 819-827.

Porter, T.M., Hajibabaei, M., 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology 27(2) 313-338.

Powers, S.M., Labou, S.G., Baulch, H.M., Hunt, R.J., Lottig, N.R., Hampton, S.E., Stanley, E.H., 2017. Ice duration drives winter nitrate accumulation in north temperate lakes. Limnology and Oceanography Letters 2(5) 177-186.

Pradeep Ram, A., Keshri, J., Sime-Ngando, T., 2020. Differential impact of top-down and bottom-up forces in structuring freshwater bacterial communities. FEMS Microbiology Ecology 96(2) fiaa005.

Ram, A.P., Chaibi-Slouma, S., Keshri, J., Colombet, J., Sime-Ngando, T., 2016. Functional responses of bacterioplankton diversity and metabolism to experimental bottom-up and top-down forcings. Microbial Ecology 72(2) 347-358.

Ratzke, C., Barrere, J., Gore, J., 2020. Strength of species interactions determines biodiversity and stability in microbial communities. Nature Ecology & Evolution 4(3) 376-383.

Sassoubre, L.M., Yamahara, K.M., Boehm, A.B., 2015. Temporal stability of the microbial community in sewage-polluted seawater exposed to natural sunlight cycles and marine microbiota. Appl. Environ. Microbiol. 81(6) 2107-2116.

Shahraki, A.H., Chaganti, S.R., Heath, D., 2019a. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. Journal of Water and Health 17 37-49.

Shahraki, A.H., Chaganti, S.R., Heath, D.D., 2020. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microbial Ecology 1-13.

Shahraki, A.H., Heath, D., Chaganti, S.R., 2019b. Recreational water monitoring: Nanofluidic qRT-PCR chip for assessing beach water safety. Environmental DNA 1(4) 305-315.

Singh, B.K., Walker, A., 2006. Microbial degradation of organophosphorus compounds. FEMS Microbiology Reviews 30(3) 428-471.

Soller, J.A., Eftim, S., Wade, T.J., Ichida, A.M., Clancy, J.L., Johnson, T.B., Schwab, K., Ramirez-Toro, G., Nappier, S., Ravenscroft, J.E., 2016. Use of quantitative microbial risk assessment to improve interpretation of a recreational water epidemiological study. Microbial Risk Analysis 1 2-11.

Tan, B., Ng, C.M., Nshimyimana, J.P., Loh, L.-L., Gin, K.Y.-H., Thompson, J.R., 2015. Next-generation sequencing (NGS) for assessment of microbial water quality: current progress, challenges, and future opportunities. Frontiers in microbiology 6 1027.

VanMensel, D., Chaganti, S.R., Droppo, I.G., Weisener, C.G., 2020. Exploring bacterial pathogen community dynamics in freshwater beach sediments: A tale of two lakes. Environmental Microbiology 22(2) 568-583.

Ward, C.S., Yung, C.-M., Davis, K.M., Blinebry, S.K., Williams, T.C., Johnson, Z.I., Hunt, D.E., 2017. Annual community patterns are driven by seasonal switching between closely related marine bacteria. The ISME Journal 11(6) 1412-1422.

Watson, S.B., Miller, C., Arhonditsis, G., Boyer, G.L., Carmichael, W., Charlton, M.N., Confesor, R., Depew, D.C., Höök, T.O., Ludsin, S.A., 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. Harmful Algae 56 44-66.

1   **APPENDIX A; SUPPLEMENTARY INFORMATION OF CHAPTER 2**

2   **Table S2.1.** Mean water temperature, wind speed and solar radiation (bi-hourly data) for each

3   sampled diel cycle by day versus night hours for the four sampled beaches and three sample

4   months.

	Mean water temperature (˚C)					
	June		July		August	
	Day	Night	Day	Night	Day	Night
CH	23.5 (±1)	22.8 (±0.8)	27.6 (±0.8)	26.5 (±1.2)	24.9 (±0.9)	24.3 (±1.3)
HB	24.2 (±1.3)	23.2 (±1.3)	28.2 (±0.8)	26.5 (±1)	25.2 (±1)	24.8 (±0.5)
LP	26 (±0.4)	25.6 (±0.4)	26.5 (±0.7)	26.11 (±0.9)	24.5 (±0.6)	22.7 (±0.8)
SP	25.5 (±0.5)	23.6 (±1)	26.5 (±0.5)	26 (±0.5)	27.8 (±0.4)	27.2 (±0.5)
	Wind speed (Km/h)					
CH	5(±3)	7 (±3)	5 (±3)	7 (±3)	11 (±3)	7 (±3)
HB	5 (±3)	7 (±2)	6 (±1)	9 (±1)	6 (±2)	11 (±2)
LP	23 (±3)	13 (±3)	14 (±4)	8 (±3)	22 (±5)	17 (±4)
SP	16 (±3)	22 (±5)	14 (±3)	8 (±4)	22 (±2)	17 (±5)
	Solar radiation (W/m^2)					
CH	112-674	0	150-680	0	70-760	0
HB	112-674	0	150-680	0	70-760	0
LP	80-578	0	110-650	0	110-600	0
SP	80-578	0	110-650	0	110-600	0

5   **Table S2.2.** Distance based Liner Model (DistLM) analysis of environmental variables on the

6   BCC of each diel cycle.

		Variables	SS (trace)	Pseudo-F	P value	r^2	Proportion of explained variation
CH	June	Water temperature	1684.1	2.4287	0.002		0.03353
		Solar radiation	1792.8	2.5912	0.001	0.12	0.0357
		Wind speed	854.16	1.2111	0.109		0.01701
	July	Water temperature	4386.3	3.9674	0.002		0.05364
		Solar radiation	7657.5	7.2317	0.001	0.13	0.09364
		Wind speed	4140.9	3.7335	0.001		0.05064
	August	Water temperature	5294.8	6.0895	0.001		0.08003
		Solar radiation	3132.6	3.4791	0.002	0.11	0.04735
		Wind speed	1931.6	2.1051	0.009		0.0292

HB	June	Water temperature	9821.8	8.9237	0.001		0.11307
		Solar radiation	6453.1	5.6174	0.001	0.17	0.07429
		Wind speed	4119.7	3.4851	0.001		0.04743
	July	Water temperature	3621.6	5.4827	0.001		0.07264
		Solar radiation	2924.8	4.3621	0.001	0.11	0.05866
		Wind speed	1245.2	1.793	0.037		0.02498
	August	Water temperature	6389.8	6.8301	0.001		0.0889
		Solar radiation	2715.1	2.748	0.003	0.15	0.03778
		Wind speed	1895	1.8954	0.016		0.02636
LP	June	Water temperature	4613.6	3.8694	0.001		0.05238
		Solar radiation	3947.7	3.2847	0.001	0.11	0.04482
		Wind speed	4682.3	3.9303	0.001		0.05316
	July	Water temperature	10218	11.065	0.001		0.13649
		Solar radiation	6379.6	6.5211	0.001	0.21	0.08522
		Wind speed	4298.5	4.2642	0.001		0.05742
	August	Water temperature	6766.1	6.8409	0.001		0.08903
		Solar radiation	4643.6	4.5553	0.001	0.2	0.0611
		Wind speed	4197	4.9213	0.001		0.05653
LP	June	Water temperature	1915	2.2529	0.011		0.03118
		Solar radiation	1721.8	2.019	0.013	0.12	0.02804
		Wind speed	1293.3	1.5058	0.04		0.02106
	July	Water temperature	2556.8	4.2149	0.001		0.05679
		Solar radiation	1927.4	3.1309	0.003	0.12	0.04281
		Wind speed	1004.2	1.597	0.04		0.02231
	August	Water temperature	5267.7	4.7987	0.001		0.06416
		Solar radiation	3064	2.7135	0.004	0.14	0.03732
		Wind speed	2291.5	2.0096	0.027		0.02791

7   **Table S2.3.** Day/night variation of alpha diversity indexes

	June		July		August	
	Chao1	**Shannon**	**Chao1**	**Shannon**	**Chao1**	**Shannon**
CH	F=41.87, **p<0.0001** [a]	F=51.07, **p<0.0001** [a]	F=124.18, **p<0.0001** [a]	F=17.34, **p<0.0001** [a]	F=117.45, **p<0.0001** [a]	F=53.23, **p<0.0001** [a]
HB	F=54.23, **p<0.0001** [a]	F=0.19, p=0.65	F=14.34, **p=0.001** [b]	F=46.24, **p=0.008** [b]	F=77.21, **p<0.0001** [a]	F=74.55, **p<0.0001** [a]
LP	F=92.80, **p<0.0001** [a]	F=0.62, p=0.43	F=80.33, **p<0.0001** [a]	F=84.76, **p<0.0001** [a]	F=99.39, **p<0.0001** [a]	F=62.19, **p<0.0001** [a]
SP	F=14.78,	F=86.18,	F=2.1,	F=92.11,	F=97.11,	F=8.14,

	p=0.001 [a]	p=0.001 [a]	p=0.12	p<0.0001 [a]	p<0.0001 [a]	p=0.008 [b]

8   a; the mean value was significantly high in the night. b; the mean values were significantly high
9   in the day.

10

11 **Table S2.4.** Detail taxa (family level) profile of selected OTUs (mean abundance >0.1% in all sampling hours of the BCCs) from top

12 500 abundant OTUs of the BCC of diel cycles which showed a significant increase in their abundance either in the BCC of day or night

13 by LDA. We also provided the taxa profile of OTUs which showed a significant increase in their abundance at day community of some

14 diel cycles but showed a significant increase in their abundance at night community of other diel cycles.

Day specific OTUs

OTUs	Location (time) [a]	Phyla [b]	Class [c]	Order	Family
7	H (2)	Act	Acidi	Acidimicrobiales	C111
14	H (3), S (1,2)				
62	C (2), H (1,2), S (1)				
68	C (3)				
2	S (1)		Actino	Actinomycetales	ACK-M1
18	C (2), H (2)				
25	C (2)				
58	H (1)				
13	S (2)				
102	C (3), L (1)				
26	C (3), L (1,3)	Bac	Sapro	SaproSirales	Citinophagaceae
39	L (3)				
67	H (1), S (1)				
127	H (1), S (1)				
131	C (3), L (3)				
147	H (2)				
12	H (3)	Pro	Alpha	Rhodobacterales	Rhodobacteraceae
121	H (1)				

Night specific OTUs

OTUs	Location (time)	Phyla [d]	Class [e]	Order	Family
134	H (1), L (2)	Act	Acidi	Acidimicrobiales	C111
135	C (3)				
1	H (2)		Actino	Actinomycetales	ACK-M1
110	C (3), H (3)				
132	C (2), H (1)				Microbacteriaceae
41	L (1)				
49	H (2)				
33	C (3)	Bac	Flavo	Flavobacteriales	Cryomorphaceae
77	H (1)				
80	C (1)				
83	C (1)				
116	C (1,2)				
153	C (2, 3)				
99	H (1)				
44	C (1), S (2)		Sapro	Saprosirales	Citinophagaceae
118	H (2), S (2)				
150	C (2), S (2)				
43	C (1)	C y	C l	Stramenopiles	

OTUs presents in Day and Night

OTUs	Day/Night; Location (time) [f]	Phyla [g]	Class [h]	Order	Family
27	D; C (1), N; S (2)	Actino	Acidi	Acidimicrobiales	C111
98	D; H (3), N; H (1)				
32	D; H (1), N; L (1, 2)		Actin o	Actinomycetales	Microbacteriaceae
4	D; H (2), N; S (1)				
5	D; L (3), N; C (2), H (1, 2)	Bact	Cytoph	Cytophagales	Cyclobacteriaceae
37	D; C (1), H (3), N; L (3)				
91	D; H (1), N; C (3), S (2)				
40	D; C (3), N; H (1), L (1)		Flavo	Flavobacteriales	Cryomorphaceae
50	D; H (3), N; C (1, 2)				
15	D; L (2), N; C (1)				
9	D; H (3), N; C (2)		Shingo	Shingobacteriales	
57	D; C (3), N; C (1), H (1)				
86	D; L (3), N; C (2, 3), H (2)				
19	D; H (1), N; C (2, 3)	Plan	Planct	Pirellulales	Pirellulaceae
31	D; H (1), N; L (3)				
76	D; H (1), N; C (3), H (3)				
10	D; C (1), H (1), S (1), N; C (2)	Pro	Alpha	Caulobacterales	Caulobacteraceae
81	D; H (1), N; C (3)			Rhizobiales	

				Order	Family					Order	Family				Order	Family
59	H (3)			*Rickettsiales*	*Pelagibacteraceae*	89	H (1,2), S (2)					88	D; H (2, 3), N; C (2)	Beta	*Rhodobacterales*	*Hyphomonadaceae*
60	C (3)			*Shingomonadales*	*Shingomonadaceae*	103	C (1), H (2), S (1)					23	D; H (1)			*Rhodobacteraceae*
11	H (2), S (2)			*Shingomonadales*		161	C (2), S (1, 3)					8	D; C (1), L (2), N; L (3)		*Burkholderiales*	*Comamonadaceae*
52	H (2,3)					42	C (1), L (2)					45	D; H (1), N; S (1)			
63	H (1)					72	L (2), S (1)	Pro	Alpha	*Rhizobiales*	*Beijerinckiaceae*	65	D; H (1), N; C (3)			
28	C (1)		Beta	*Burkholderiales*	*Comamonadaceae*	53	H (3), L (2)					6	D; H (3), L (3), N; S (2)			
47	H (2)				*Oxalobacteraceae*	54	S (1)					24	D; H (1), N; S (2)			*Oxalobacteraceae*
21	C (3), H (3), L (2, 3)			*Methylophilales*	*Methylophilaceae*	117	C (3), S (1)					46	D; L (2), N; C (1,3), S (2)			
51	C (3)	Ver	Verru	*Verrucomicrobiales*	*Verrucomicrobiaceae*	22	C (1, 3)		Beta	*Burkholderiales*	*Comamonadaceae*	95	D; H (1), N; C (2, 3)	Gamma	*Aeromonadales*	*Aeromonadaceae*
87	H (2)					73	S (2)					29	D; C (1), N; L (3)		*Xanthomonadales*	*Xanthomonadaceae*
114	H (1)					113	C (3)									
						136	S (2)									
						30	S (2)									
						93	C (3)									

15 a: C; CH; H, L; L, S; SP, 1; June, 2; July, 3; August. b: Act; *Actinobacteria*, Bac, *Bacteroidetes*, Pro; *Proteobacteria*, Ver; *Verrucomicrobia*. c: Acidi;

16 *Acidimicrobiia*, Actino; *Actinobacteria*, Sapro; *Saprospirae*, Alpha; *Alphaproteobacteria*, Beta; *Betaproteobacteria*, Verru; *Verrucomicrobiae*, d: Cya;

17 *Cyanobacteria*. e: Flavo; *Flavobacteriia*, Cloro; *Chloroplast*. f: D; Day, N; Night. g: Plan; *Planctomycetes*., h: Cytoph; *Cytophagia*, Shingo; *Sphingobacteriia*,

18 Planct; *Planctomycetia*.

19

20 **Figure S2.1.** Line plots of Chao1 index over all four locations (CH, HB, LP and SP beaches) in

21 three different diel sampling events (June, July and August) based upon OTUs detected.

22
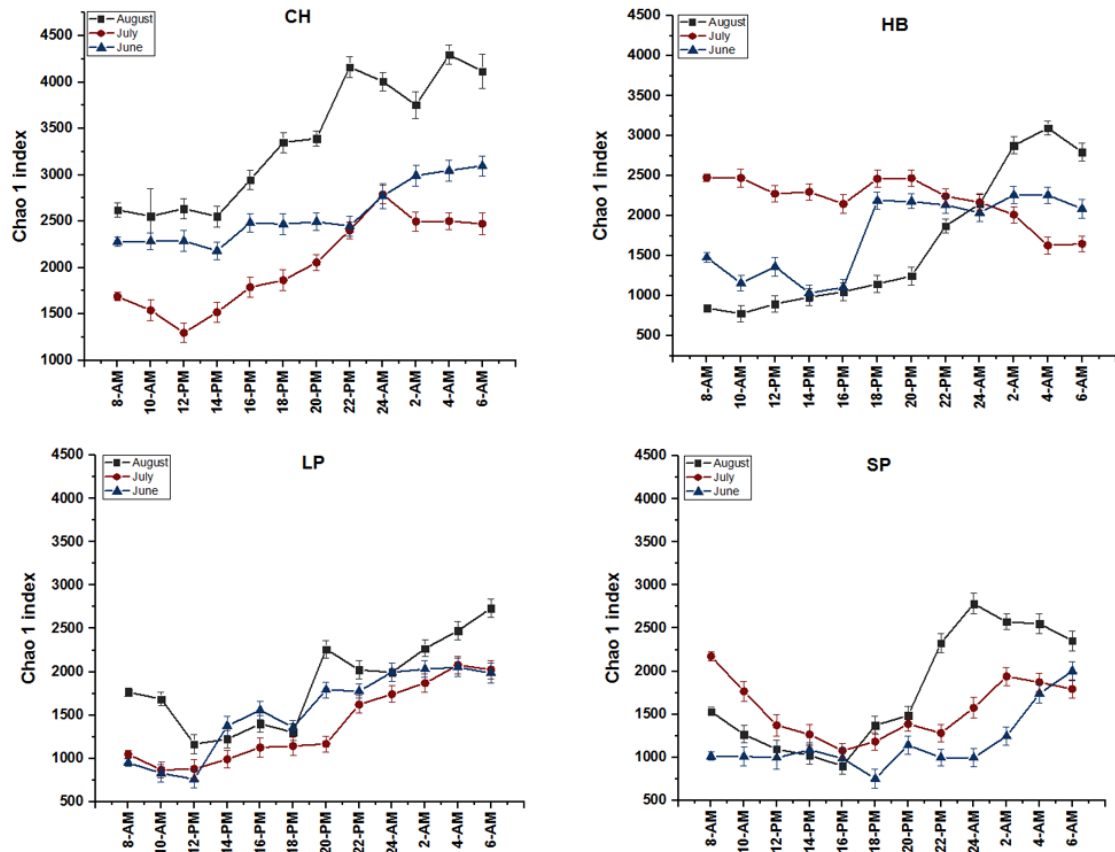
23
24

25   **Figure S2.2.** Line plots of Shannon index over all four locations (CH, HB, LP and SP beaches) in

26   three different diel sampling events (June, July and August) based upon OTUs detected.
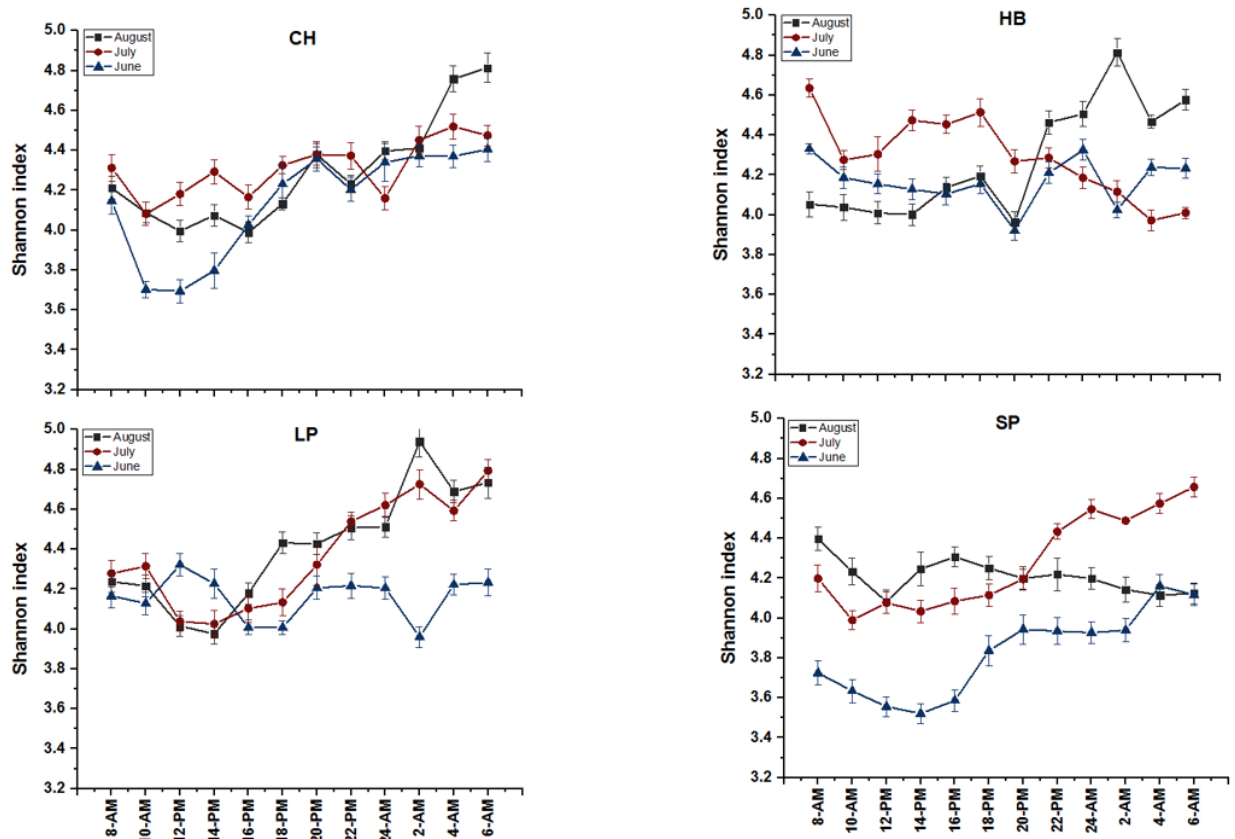
27

28   **Figure S2.3.** Relative abundance of bacterial taxa at phyla/class level for each beach (CH, HB, LP

29   and SP) over each diel cycle in June, July and August. X axis presenting sampling hours which

30   started from 8:00 AM and finished next day at 6:00 AM with 2-hours intervals. Each plot divided

31   into two-part, day hours (8:00 AM, 10:00 AM, 12:00 PM, 14:00 PM, 16:00 PM and 18:00 PM)

32   and night hours (20:00 PM, 22:00 PM and 24:00 PM, 2:00 AM, 4:00 AM and 6:00 AM). The bars

33   show phyla/class with at least 0.1% relative abundance in combined water samples over a diel

34   cycle. The "others" category includes the sum of all phyla that occurred at less than 0.1% relative

35   abundance.

36
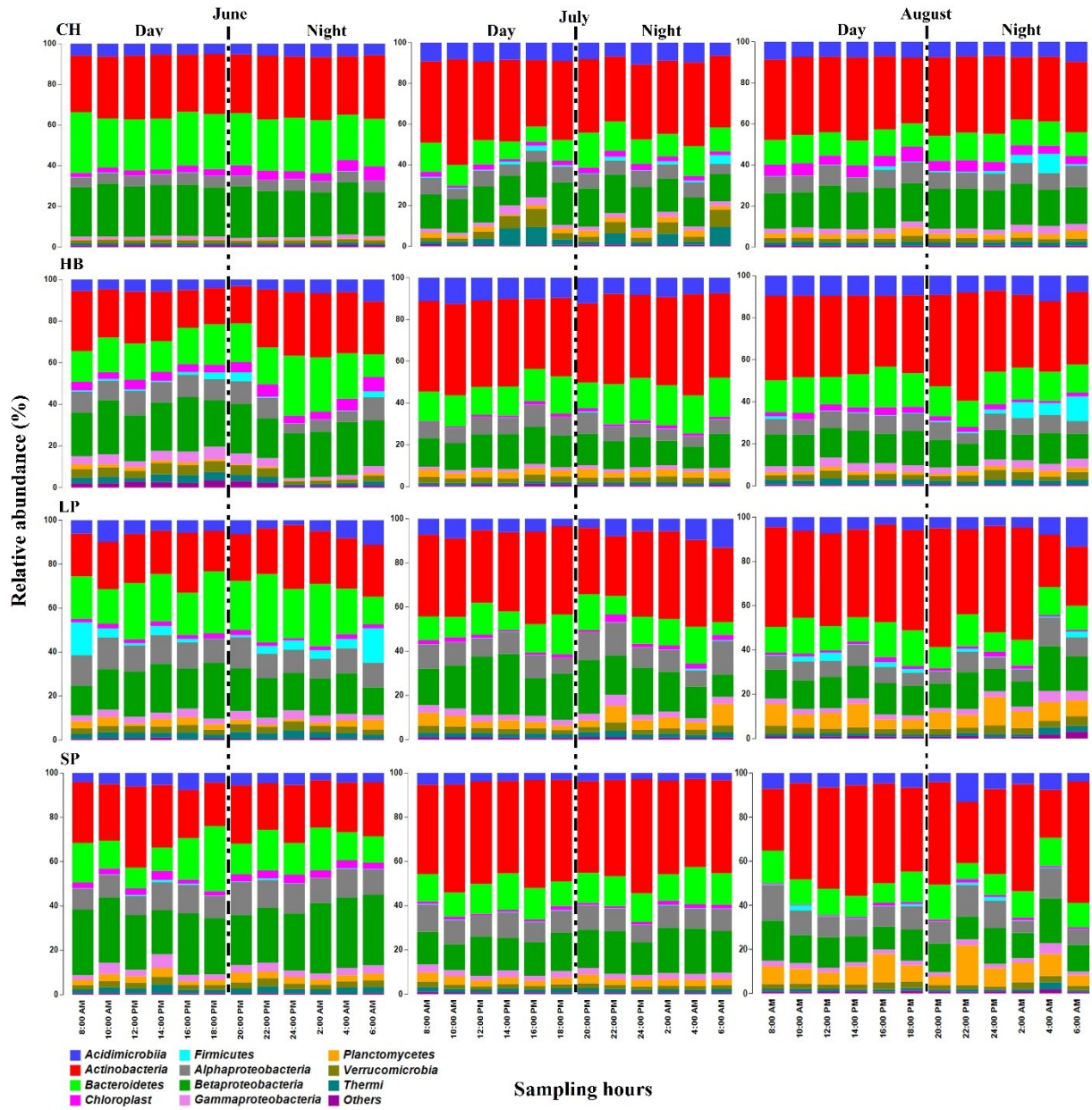
Figure S2.4. Graphics of LDA effect size (LEfSe) for the day versus night community of each diel. Horizontal bars represent the effect size for each taxon (family). The length of the bar represents the $\log^{10}$ transformed LDA score, indicated by vertical dotted lines. The day community is indicated by red, and the night community by green. The family of bacteria with statistically significant change ($p < 0.05$) in the relative abundance is written alongside the horizontal lines.

42

LDA SCORE (log 10)
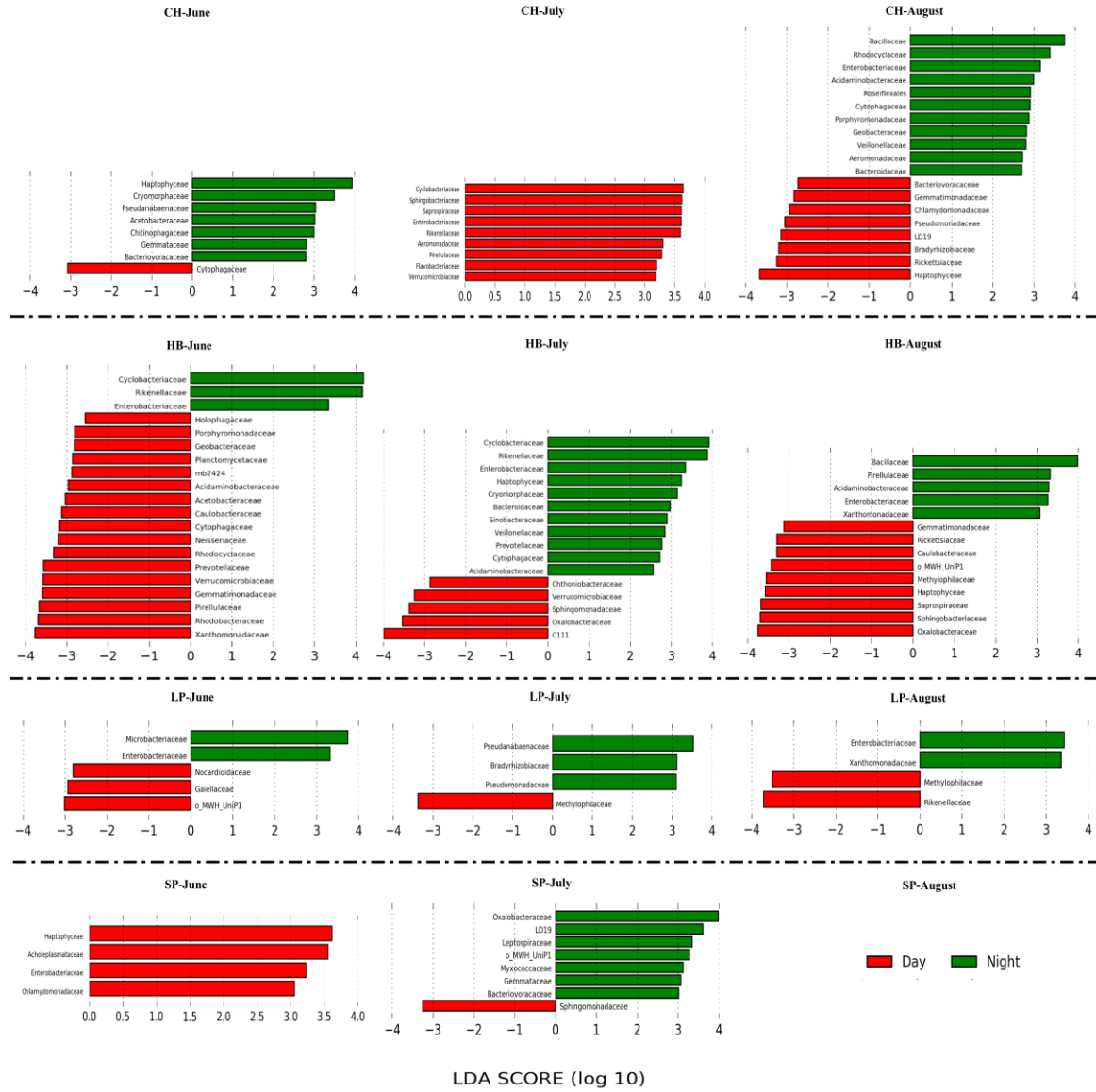
43 **APPENDIX B; SUPPLEMENTARY INFORMATION OF CHAPTER 3**

44 **Table S3.1.** Taxa which shown significant spatial variation (p<0.05) in the six sampling locations

45 in Lakes Erie and St Clair.

Phylum.Class	Locations (Pairwise)	LDA effect	p value
	CB-CH		
Firmicutes.Bacilli	CH	3.43	0.002
Gemmatimonadetes.Gemmatimonadetes	CB	2.36	0.003
Proteobacteria.Gammaproteobacteria	CB	3.0	0.001
Proteobacteria.Gammaproteobacteria	CB	3.95	0.014
Verrucomicrobia.Verrucomicrobiae	CB	2.4	0.026
	CB-HB		
Acidobacteria.Solibacteres	CB	2.55	0.037
Cyanobacteria.Chloroplast	CB	2.54	0.047
Cyanobacteria.Nostocophycideae	CB	2.17	0.016
Firmicutes.Bacilli	HB	3.23	0.001
Proteobacteria.Gammaproteobacteria	CB	3.01	0.002
	CB-LP		
Actinobacteria.Acidimicrobiia	CB	3.77	0.023
Actinobacteria.Thermoleophilia	CB	3.42	0.029
Bacteroidetes.Saprospirae	CB	3.48	0.033
Cyanobacteria.Chloroplast	CB	3.2	0.011
Firmicutes.Bacilli	LP	3.11	0.001
	CB-PP		
Cyanobacteria.Chloroplast	CB	3.32	0.028
Firmicutes.Bacilli	PP	3.33	0.003
	CB-SP		
Actinobacteria.Thermoleophilia	CB	3.39	0.002
Bacteroidetes.Flavobacteriia	CB	3.05	0.012
Bacteroidetes.Saprospirae	CB	3.47	0.010
Cyanobacteria.Chloroplast	CB	2.52	0.024
Firmicutes.Bacilli	CB	2.43	0.001
	CH-HB		
Cyanobacteria.Chloroplast	CH	3.07	0.032
	CH-LP		
Actinobacteria.Acidimicrobiia	CH	3.85	0.006
Planctomycetes.Planctomycetia	LP	4.12	0.014
Proteobacteria.Gammaproteobacteria	LP	3.43	0.002
Proteobacteria.Gammaproteobacteria	LP	4.01	0.024
	CH-SP		

Actinobacteria.Acidimicrobiia	CH	3.75	0.005
Actinobacteria.Thermoleophilia	CH	3.29	0.047
Bacteroidetes.Flavobacteriia	CH	3.37	0.008
Cyanobacteria.Chloroplast	CH	3.14	0.049
Proteobacteria.Gammaproteobacteria	SP	3.54	0.003
HB-LP			
Bacteroidetes.Sphingobacteriia	LP	3.32	0.026
Proteobacteria.Gammaproteobacteria	LP	2.45	0.003
HB-SP			
Acidobacteria.Acidobacteria.6	HB	4.01	0.009
Bacteroidetes.Flavobacteriia	HB	3.57	0.004
Gemmatimonadetes.Gemmatimonadetes	HB	2.61	0.042
Proteobacteria.Gammaproteobacteria	SP	3.45	0.003
Verrucomicrobia.Pedosphaerae	HB	2.4	0.030
PP-LP			
Actinobacteria.Acidimicrobiia	PP	4.0	0.020
Proteobacteria.Gammaproteobacteria	LP	2.43	0.002
PP-SP			
Acidobacteria.Acidobacteria.6	PP	3.87	0.025
Actinobacteria.Acidimicrobiia	PP	3.70	0.017
Bacteroidetes.Flavobacteriia	PP	3.66	0.006
Proteobacteria.Gammaproteobacteria	SP	2.22	0.004
LP-SP			
Bacteroidetes.Bacteroidia	LP	2.13	0.015
Bacteroidetes.Flavobacteriia	LP	3.63	0.018
Planctomycetes.Phycisphaerae	LP	2.14	0.013
Planctomycetes.Planctomycetia	LP	3.04	0.011
Verrucomicrobia.Verrucomicrobiae	LP	2.26	0.037

46

47   **Table S3.2.** SIMPER results (above the diagonal) and pairwise PERMANOVA probabilities
48   (below the diagonal) of 5 broad clusters of the BCCs. p values were adjusted using a sequential
49   Bonferroni correction for multiple comparisons.

Clusters	1	2	3	4	5
1		54.12	55.75	49.61	47
2	**0.0001**		50.21	57.74	48.12
3	**0.0001**	**0.0001**		51.72	49.55
4	**0.0001**	**0.0001**	**0.0001**		49.36
5	**0.0001**	**0.0001**	**0.0001**	**0.0001**	

50   **Table S3.3.** Pairwise comparison of diversity indexes between the BCCs of 5 clusters.

Clusters		Chao1	Shannon	PCo1	PCo2
1	**2**	**0.000**	**0.014**	**0.000**	**0.000**
	3	**0.000**	**0.000**	**0.000**	**0.000**
	4	**0.000**	**0.000**	**0.009**	0.881
	5	**0.000**	**0.000**	**0.000**	0.776
2	**1**	**0.000**	**0.014**	**0.000**	**0.000**
	3	**0.003**	**0.000**	**0.017**	**0.001**
	4	0.995	0.092	0.4	**0.000**
	5	0.998	**0.004**	0.83	**0.000**
3	**1**	**0.000**	**0.000**	**0.000**	**0.000**
	2	**0.003**	**0.000**	**0.017**	**0.001**
	4	**0.016**	0.053	**0.000**	**0.000**
	5	**0.000**	0.239	0.156	**0.000**
4	**1**	**0.000**	**0.000**	**0.009**	0.881
	2	0.995	0.092	0.400	**0.000**
	3	**0.016**	**0.053**	**0.000**	**0.000**
	5	0.946	0.914	**0.027**	0.283
5	**1**	**0.000**	**0.000**	**0.000**	0.776
	2	0.998	**0.004**	0.833	**0.000**
	3	**0.000**	0.239	0.156	**0.000**
	4	0.946	0.914	**0.027**	0.283

51   **Table S3.4.** Pairwise comparison of temporal variation of taxa (class level) between the BCCs of

52   5 clusters.

Phylum.Class	Clusters	LDA effect	p value
Cluster 1 versus Cluster-2			
Acidobacteria.Solibacteres	1	4.042	**0.0025**
Actinobacteria.Acidimicrobiia	1	4.180	**0.0022**
Actinobacteria.Actinobacteria	1	4.650	**0.0026**
Actinobacteria.Thermoleophilia	1	3.763	**0.002**
Bacteroidetes.Cytophagia	1	3.861	**0.0028**
Bacteroidetes.Saprospirae	1	3.867	**0.0024**
Cyanobacteria.Chloroplast	3	3.399	**0.0021**
Cyanobacteria.Synechococcophycideae	3	3.583	**0.0018**
Firmicutes. Bacilli	3	4.457	**0.0023**

Gemmatimonadetes.Gemmatimonadetes	1	3.641	**0.0025**
Planctomycetes.Phycisphaerae	1	3.282	**0.002**
Planctomycetes.Planctomycetia	1	3.148	**0.0024**
Proteobacteria.Betaproteobacteria	3	4.088	**0.0027**
Proteobacteria.Gammaproteobacteria	3	4.204	**0.0024**
Verrucomicrobia.Pedosphaerae	1	3.201	**0.0025**
Cluster 1 versus Cluster 3			
Acidobacteria.Acidobacteria_6	1	3.028	**0.0018**
Acidobacteria.Solibacteres	1	3.704	**0.0023**
Actinobacteria.Acidimicrobiia	1	4.072	**0.0025**
Actinobacteria.Actinobacteria	1	4.674	**0.002**
Actinobacteria.Thermoleophilia	1	3.740	**0.0024**
Bacteroidetes.Cytophagia	1	3.856	**0.0027**
Bacteroidetes.Saprospirae	1	3.822	**0.0018**
Chloroflexi.Chloroflexi	1	2.895	**0.0023**
Cyanobacteria.Chloroplast	4	2.829	**0.0025**
Cyanobacteria.Nostocophycideae	1	3.313	**0.0015**
Firmicutes.Bacilli	4	4.830	**0.0025**
Gemmatimonadetes.Gemmatimonadetes	1	3.065	**0.0095**
Planctomycetes.Phycisphaerae	1	2.845	**0.007**
Proteobacteria.Gammaproteobacteria	4	4.138	**0.006**
Verrucomicrobia.Pedosphaerae	1	2.928	**0.008**
Verrucomicrobia.Verrucomicrobiae	1	2.575	**0.002**
Cluster 1 versus Cluster 4			
Acidobacteria.Acidobacteria_6	1	2.869	**0.0025**
Acidobacteria.Solibacteres	1	3.359	**0.0022**
Actinobacteria.Actinobacteria	1	4.573	**0.0026**
Actinobacteria.Thermoleophilia	1	3.534	**0.002**
Bacteroidetes.Cytophagia	1	3.894	**0.0028**
Bacteroidetes.Flavobacteriia	1	3.566	**0.0024**
Bacteroidetes.Saprospirae	1	3.745	**0.0018**
Bacteroidetes.Sphingobacteriia	1	2.889	**0.0023**
Chloroflexi.Chloroflexi	1	2.665	**0.0025**
Cyanobacteria.Nostocophycideae	1	3.066	**0.002**
Cyanobacteria.Synechococcophycideae	5	2.784	**0.0024**
Firmicutes.Bacilli	5	4.773	**0.0027**
Firmicutes.Clostridia	1	3.914	**0.0018**
Gemmatimonadetes.Gemmatimonadetes	1	2.742	**0.0023**
Nitrospirae.Nitrospira	1	3.637	**0.0025**
Planctomycetes.Phycisphaerae	1	3.008	**0.0015**
Proteobacteria.Alphaproteobacteria	1	3.269	**0.0025**

Proteobacteria.Betaproteobacteria	1	4.056	**0.0095**
Proteobacteria.Deltaproteobacteria	1	3.121	**0.0018**
Thermi.Deinococci	1	2.216	**0.0023**
Verrucomicrobia.Pedosphaerae	1	2.844	**0.0046**
Verrucomicrobia.Verrucomicrobiae	1	2.494	**0.0034**
Cluster 1 versus Cluster 5			
Actinobacteria.Acidimicrobiia	1	4.061	**0.0025**
Actinobacteria.Actinobacteria	1	4.561	**0.002**
Actinobacteria.Thermoleophilia	1	3.732	**0.0024**
Bacteroidetes.Cytophagia	1	3.894	**0.0027**
Bacteroidetes.Flavobacteriia	1	3.671	**0.0018**
Bacteroidetes.Saprospirae	1	3.828	**0.0023**
Bacteroidetes.Sphingobacteriia	1	4.077	**0.0025**
Firmicutes.Bacilli	2	4.743	**0.0052**
Planctomycetes.Phycisphaerae	1	3.674	**0.0032**
Verrucomicrobia.Pedosphaerae	1	3.742	**0.002**
Cluster 2 versus Cluster 3			
Acidobacteria.Solibacteres	3	3.229	**0.004**
Chloroflexi.Chloroflexi	3	2.668	**0.0094**
Cyanobacteria.Chloroplast	3	2.671	**0.002**
Cyanobacteria.Nostocophycideae	3	3.041	**0.0052**
Cyanobacteria.Synechococcophycideae	3	3.500	**0.0032**
Firmicutes.Bacilli	4	4.615	**0.002**
Planctomycetes.Phycisphaerae	3	2.771	**0.005**
Proteobacteria	3	4.480	**0.008**
Cluster 2 versus Cluster 4			
Acidobacteria.Acidobacteria	3	2.319	**0.0085**
Acidobacteria.Solibacteres	3	2.438	**0.0072**
Actinobacteria.Acidimicrobiia	5	3.992	**0.008**
Actinobacteria.Thermoleophilia	5	3.128	**0.007**
Bacteroidetes.Flavobacteriia	3	3.866	**0.0028**
Bacteroidetes.Saprospirae	5	3.314	**0.0024**
Bacteroidetes.Sphingobacteriia	3	3.494	**0.0032**
Chloroflexi.Chloroflexi	3	2.565	**0.0018**
Cyanobacteria.Chloroplast	3	2.936	**0.0023**
Cyanobacteria.Nostocophycideae	3	3.102	**0.0025**
Cyanobacteria.Synechococcophycideae	3	3.430	**0.002**
Firmicutes.Bacilli	5	4.459	**0.0024**
Firmicutes.Clostridia	3	2.612	**0.0027**
Nitrospirae.Nitrospira	3	2.733	**0.0024**
Planctomycetes.Planctomycetia	5	3.449	**0.0025**

Proteobacteria.Alphaproteobacteria	3	3.564	**0.0022**
Proteobacteria.Betaproteobacteria	3	4.323	**0.0026**
Proteobacteria.Deltaproteobacteria	3	4.078	**0.002**
Cluster 2 versus Cluster 5			
Actinobacteria.Acidimicrobiia	2	3.778	**0.0028**
Bacteroidetes.Flavobacteriia	3	3.891	**0.0024**
Bacteroidetes.Saprospirae	2	3.565	**0.0032**
Cyanobacteria.Nostocophycideae	2	4.261	**0.0018**
Cyanobacteria.Synechococcophycideae	3	3.630	**0.0023**
Firmicutes.Bacilli	2	4.517	**0.0025**
Planctomycetes.Planctomycetia	2	3.425	**0.002**
Proteobacteria.Betaproteobacteria	3	4.251	**0.0024**
Verrucomicrobia.Pedosphaerae	2	3.795	**0.0028**
Cluster 3 versus Cluster 4			
Actinobacteria.Acidimicrobiia	5	3.935	**0.0034**
Actinobacteria.Actinobacteria	5	4.115	**0.0024**
Actinobacteria.Thermoleophilia	5	3.473	**0.009**
Bacteroidetes.Saprospirae	5	3.569	**0.007**
Cyanobacteria.Chloroplast	4	3.842	**0.0043**
Cyanobacteria.Synechococcophycideae	5	3.663	**0.0053**
Planctomycetes.Planctomycetia	5	3.648	**0.0097**
Cluster 3 versus Cluster 5			
Actinobacteria.Actinobacteria	2	4.208	**0.0023**
Cyanobacteria.Nostocophycideae	2	4.016	**0.0025**
Proteobacteria	4	4.087	**0.002**
Verrucomicrobia.Pedosphaerae	2	3.612	**0.0024**
Cluster 4 versus Cluster 5			
Actinobacteria.Acidimicrobiia	5	3.708	**0.002**
Cyanobacteria.Nostocophycideae	2	2.522	**0.0052**
Proteobacteria.Alphaproteobacteria	2	3.269	**0.0032**
Proteobacteria.Betaproteobacteria	2	3.804	**0.002**
Proteobacteria.Deltaproteobacteria	2	3.020	**0.005**
Verrucomicrobia.Pedosphaerae	2	2.458	**0.002**

53

54   **Table S3.5.** Pairwise comparison of alpha diversity and Bray–Curtis dissimilarity PCo1 & 2

55   indexes among 15 months of sampling across 6 different locations

	Months/years	Chao1	Shannon	PCo1	PCo2
6/16	7/16	1.000	1.000	1.000	1.000
	8/16	1.000	1.000	1.000	.886
	9/16	.163	**.000**	**.022**	.183
	10/16	.198	**.000**	.248	.992
	11/16	1.000	.275	.174	1.000
	12/16	1.000	.986	.859	1.000
	1/17	**.008**	.458	.057	.921
	2/17	.104	**.006**	**.020**	.981
	3/17	**.000**	**.000**	**.001**	1.000
	4/17	**.000**	**.000**	.075	1.000
	5/17	**.000**	**.000**	**.000**	1.000
	6/17	**.011**	**.000**	.152	1.000
	7/17	.554	**.024**	.431	.995
	8/17	.288	**.030**	.520	.977
7/16	6/16	1.000	1.000	1.000	1.000
	8/16	1.000	1.000	1.000	.999
	9/16	.094	**.000**	**.014**	.675
	10/16	.117	**.000**	.180	1.000
	11/16	.998	.052	.122	.998
	12/16	.998	.755	.777	.995
	1/17	**.003**	.115	**.037**	.439
	2/17	.057	**.000**	**.012**	.637
	3/17	**.000**	**.000**	**.001**	.997
	4/17	**.000**	**.000**	.050	1.000
	5/17	**.000**	**.000**	**.000**	1.000
	6/17	**.005**	**.000**	.105	1.000
	7/17	.399	**.002**	.334	1.000
	8/17	.180	**.003**	.415	1.000
8/16	6/16	1.000	1.000	1.000	.886
	7/16	1.000	1.000	1.000	.999
	9/16	**.022**	**.000**	**.014**	.997
	10/16	**.029**	**.001**	.179	1.000
	11/16	.954	.362	.122	.696

	12/16	.947	.995	.777	.625
	1/17	**.000**	.562	**.037**	**.047**
	2/17	**.012**	**.011**	**.012**	.100
	3/17	**.000**	**.000**	**.001**	.682
	4/17	**.000**	**.000**	.050	.932
	5/17	**.000**	**.000**	**.000**	.965
	6/17	**.001**	**.001**	.105	.994
	7/17	.145	**.038**	.334	1.000
	8/17	**.049**	**.047**	.415	1.000
9/16	6/16	.163	**.000**	**.022**	.183
	7/16	.094	**.000**	**.014**	.675
	8/16	**.022**	**.000**	**.014**	.997
	10/16	1.000	1.000	1.000	.931
	11/16	.735	.339	1.000	.080
	12/16	.754	**.007**	.835	**.061**
	1/17	1.000	.188	1.000	**.001**
	2/17	1.000	.989	1.000	**.003**
	3/17	.858	1.000	1.000	.076
	4/17	.901	1.000	1.000	.239
	5/17	**.019**	.982	.904	.312
	6/17	1.000	1.000	1.000	.492
	7/17	1.000	.915	.994	.912
	8/17	1.000	.888	.986	.970
10/16	6/16	.198	**.000**	.248	.992
	7/16	.117	**.000**	.180	1.000
	8/16	**.029**	**.001**	.179	1.000
	9/16	1.000	1.000	1.000	.931
	11/16	.788	.779	1.000	.942
	12/16	.805	.058	1.000	.911
	1/17	.999	.583	1.000	.171
	2/17	1.000	1.000	1.000	.304
	3/17	.815	.982	.899	.937
	4/17	.866	1.000	1.000	.997
	5/17	**.015**	.753	.318	.999
	6/17	1.000	1.000	1.000	1.000
	7/17	1.000	.999	1.000	1.000
	8/17	1.000	.997	1.000	1.000

	6/16	1.000	.275	.174	1.000
	7/16	.998	.052	.122	.998
	8/16	.954	.362	.122	.696
	9/16	.735	.339	1.000	.080
	10/16	.788	.779	1.000	.942
	12/16	1.000	.989	.998	1.000
11/16	1/17	.126	1.000	1.000	.987
	2/17	.610	.993	1.000	.999
	3/17	**.009**	**.046**	.951	1.000
	4/17	**.012**	.498	1.000	1.000
	5/17	**.000**	**.006**	.423	1.000
	6/17	.166	.802	1.000	1.000
	7/17	.982	1.000	1.000	.956
	8/17	.879	1.000	1.000	.884
	6/16	1.000	.986	.859	1.000
	7/16	.998	.755	.777	.995
	8/16	.947	.995	.777	.625
	9/16	.754	**.007**	.835	.061
	10/16	.805	.058	1.000	.911
	11/16	1.000	.989	.998	1.000
12/16	1/17	.136	.999	.956	.994
	2/17	.631	.339	.819	1.000
	3/17	**.010**	**.000**	.296	1.000
	4/17	**.013**	**.017**	.975	1.000
	5/17	**.000**	**.000**	**.030**	1.000
	6/17	.178	.066	.997	1.000
	7/17	.985	.609	1.000	.930
	8/17	.892	.661	1.000	.837
	6/16	**.008**	.458	.057	.921
	7/16	**.003**	.115	**.037**	.439
	8/16	**.000**	.562	**.037**	**.047**
	9/16	1.000	.188	1.000	**.001**
1/17	10/16	.999	.583	1.000	.171
	11/16	.126	1.000	1.000	.987
	12/16	.136	.999	.956	.994
	2/17	1.000	.960	1.000	1.000
	3/17	1.000	**.019**	.997	.989

	4/17	1.000	.307	1.000	.872
	5/17	.295	**.002**	.732	.802
	6/17	1.000	.612	1.000	.621
	7/17	.941	.997	1.000	.194
	8/17	.995	.998	.999	.115
2/17	6/16	.104	**.006**	.020	.981
	7/16	.057	**.000**	.012	.637
	8/16	**.012**	**.011**	.012	.100
	9/16	1.000	.989	1.000	.003
	10/16	1.000	1.000	1.000	.304
	11/16	.610	.993	1.000	.999
	12/16	.631	.339	.819	1.000
	1/17	1.000	.960	1.000	1.000
	3/17	.929	.685	1.000	.999
	4/17	.956	.998	1.000	.962
	5/17	**.034**	.258	.915	.927
	6/17	1.000	1.000	1.000	.803
	7/17	1.000	1.000	.993	.338
	8/17	1.000	1.000	.983	.217
3/17	6/16	**.000**	**.000**	.001	1.000
	7/16	**.000**	**.000**	.001	.997
	8/16	**.000**	**.000**	.001	.682
	9/16	.858	1.000	1.000	.076
	10/16	.815	.982	.899	.937
	11/16	**.009**	**.046**	.951	1.000
	12/16	**.010**	**.000**	.296	1.000
	1/17	1.000	**.019**	.997	.989
	2/17	.929	.685	1.000	.999
	4/17	1.000	.999	.994	1.000
	5/17	.868	1.000	1.000	1.000
	6/17	1.000	.977	.963	1.000
	7/17	.422	.408	.736	.952
	8/17	.705	.360	.650	.875
4/17	6/16	**.000**	**.000**	.075	1.000
	7/16	**.000**	**.000**	.050	1.000
	8/16	**.000**	**.000**	.050	.932
	9/16	.901	1.000	1.000	.239

	10/16	.866	1.000	1.000	.997
	11/16	**.012**	.498	1.000	1.000
	12/16	**.013**	**.017**	.975	1.000
	1/17	1.000	.307	1.000	.872
	2/17	.956	.998	1.000	.962
	3/17	1.000	.999	.994	1.000
	5/17	.818	.939	.662	1.000
	6/17	1.000	1.000	1.000	1.000
	7/17	.492	.972	1.000	.998
	8/17	.770	.959	1.000	.990
5/17	6/16	**.000**	**.000**	.000	1.000
	7/16	**.000**	**.000**	.000	1.000
	8/16	**.000**	**.000**	.000	.965
	9/16	**.019**	.982	.904	.312
	10/16	**.015**	.753	.318	.999
	11/16	**.000**	**.006**	.423	1.000
	12/16	**.000**	**.000**	.030	1.000
	1/17	.295	**.002**	.732	.802
	2/17	**.034**	.258	.915	.927
	3/17	.868	1.000	1.000	1.000
	4/17	.818	.939	.662	1.000
	6/17	.234	.727	.463	1.000
	7/17	**.002**	.104	.169	1.000
	8/17	**.008**	.086	.125	.996
6/17	6/16	**.011**	**.000**	.152	1.000
	7/16	**.005**	**.000**	.105	1.000
	8/16	**.001**	**.001**	.105	.994
	9/16	1.000	1.000	1.000	.492
	10/16	1.000	1.000	1.000	1.000
	11/16	.166	.802	1.000	1.000
	12/16	.178	.066	.997	1.000
	1/17	1.000	.612	1.000	.621
	2/17	1.000	1.000	1.000	.803
	3/17	1.000	.977	.963	1.000
	4/17	1.000	1.000	1.000	1.000
	5/17	.234	.727	.463	1.000
	7/17	.966	.999	1.000	1.000

	8/17	.998	.998	1.000	1.000
7/17	6/16	.554	**.024**	.431	.995
	7/16	.399	**.002**	.334	1.000
	8/16	.145	**.038**	.334	1.000
	9/16	1.000	.915	.994	.912
	10/16	1.000	.999	1.000	1.000
	11/16	.982	1.000	1.000	.956
	12/16	.985	.609	1.000	.930
	1/17	.941	.997	1.000	.194
	2/17	1.000	1.000	.993	.338
	3/17	.422	.408	.736	.952
	4/17	.492	.972	1.000	.998
	5/17	**.002**	.104	.169	1.000
	6/17	.966	.999	1.000	1.000
	8/17	1.000	1.000	1.000	1.000
8/17	6/16	.288	**.030**	.520	.977
	7/16	.180	**.003**	.415	1.000
	8/16	**.049**	**.047**	.415	1.000
	9/16	1.000	.888	.986	.970
	10/16	1.000	.997	1.000	1.000
	11/16	.879	1.000	1.000	.884
	12/16	.892	.661	1.000	.837
	1/17	.995	.998	.999	.115
	2/17	1.000	1.000	.983	.217
	3/17	.705	.360	.650	.875
	4/17	.770	.959	1.000	.990
	5/17	**.008**	.086	.125	.996
	6/17	.998	.998	1.000	1.000
	7/17	1.000	1.000	1.000	1.000

56
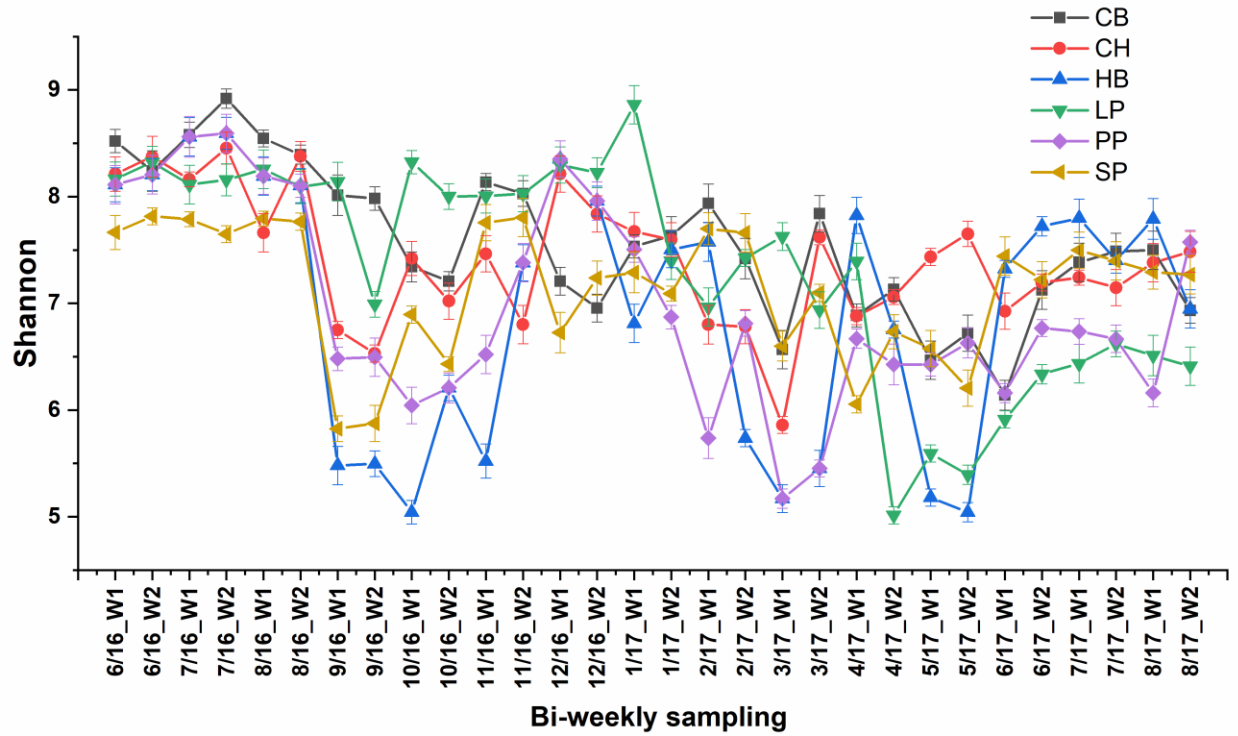
Figure S3.1. Bi-weekly variation in the Shannon index for the 6 different sampling locations (CB, CH, HB, LP, PP and SP) over 15 months of sampling (June 2016 - August 2017). The X-axis shows the time of sampling (bi-weekly sampling).

57
58
59

60

61

62
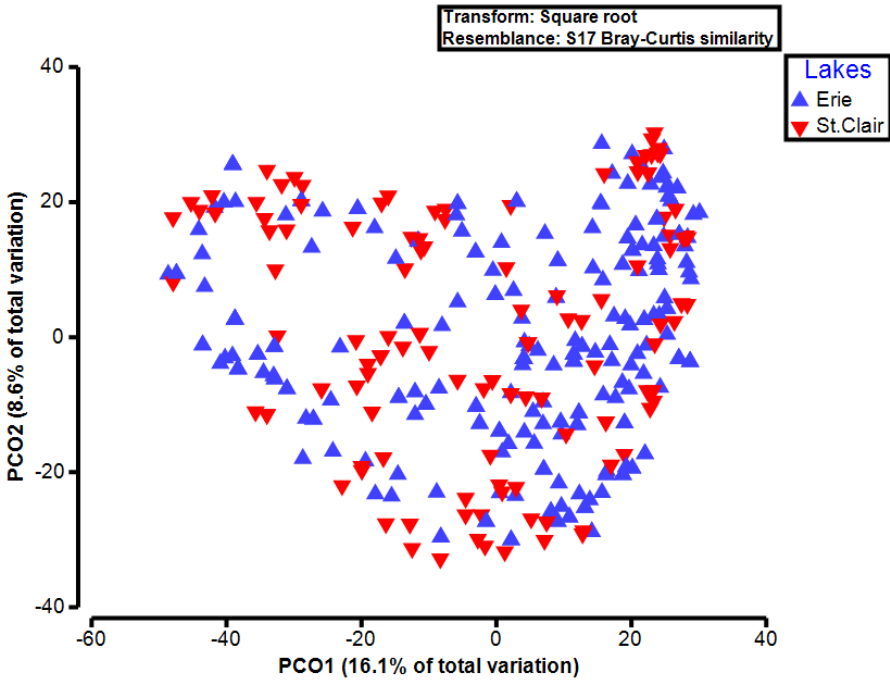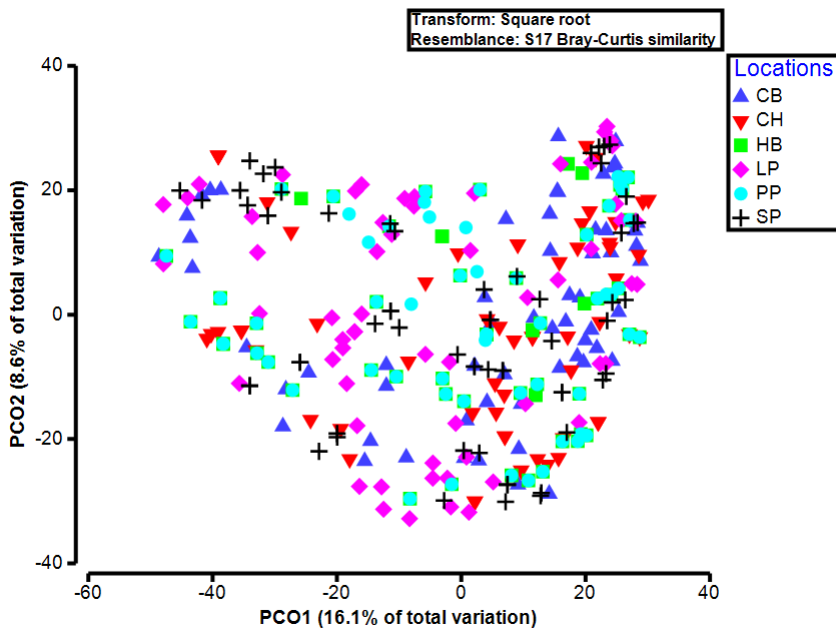
Figure S3.2. Multivariant principle coordinate analysis (PCoA) plot of the Bray–Curtis similarity matrix of the BCCs of Lake Erie and Lake St. Clair (panel A) and six different locations (CB, CH, HB, LP, PP and SP) (panel B) across 15 months of sampling.

66
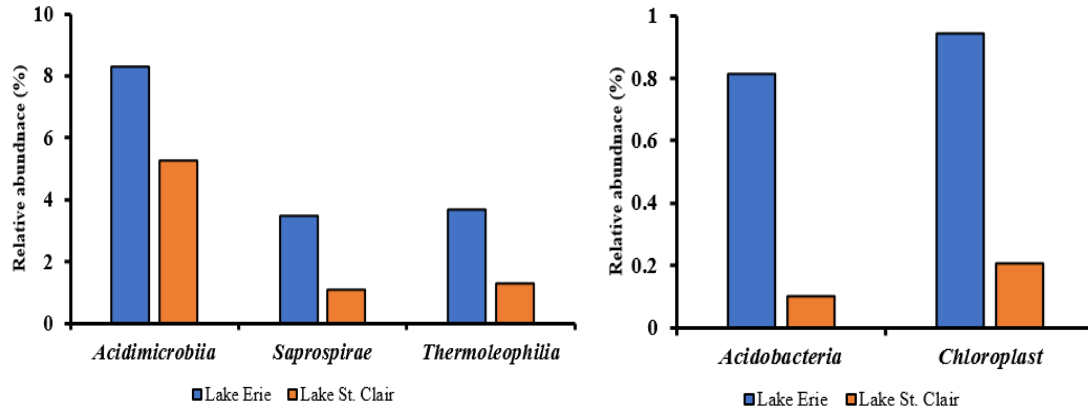
Figure S3.3. Taxa with significant spatial variation in their relative abundance among two lakes (Lake Erie and St. Clair). The relative abundance of all 5 classes were significantly higher (p<0.05) in Lake Eire relative to Lake St. Clair.

70

71
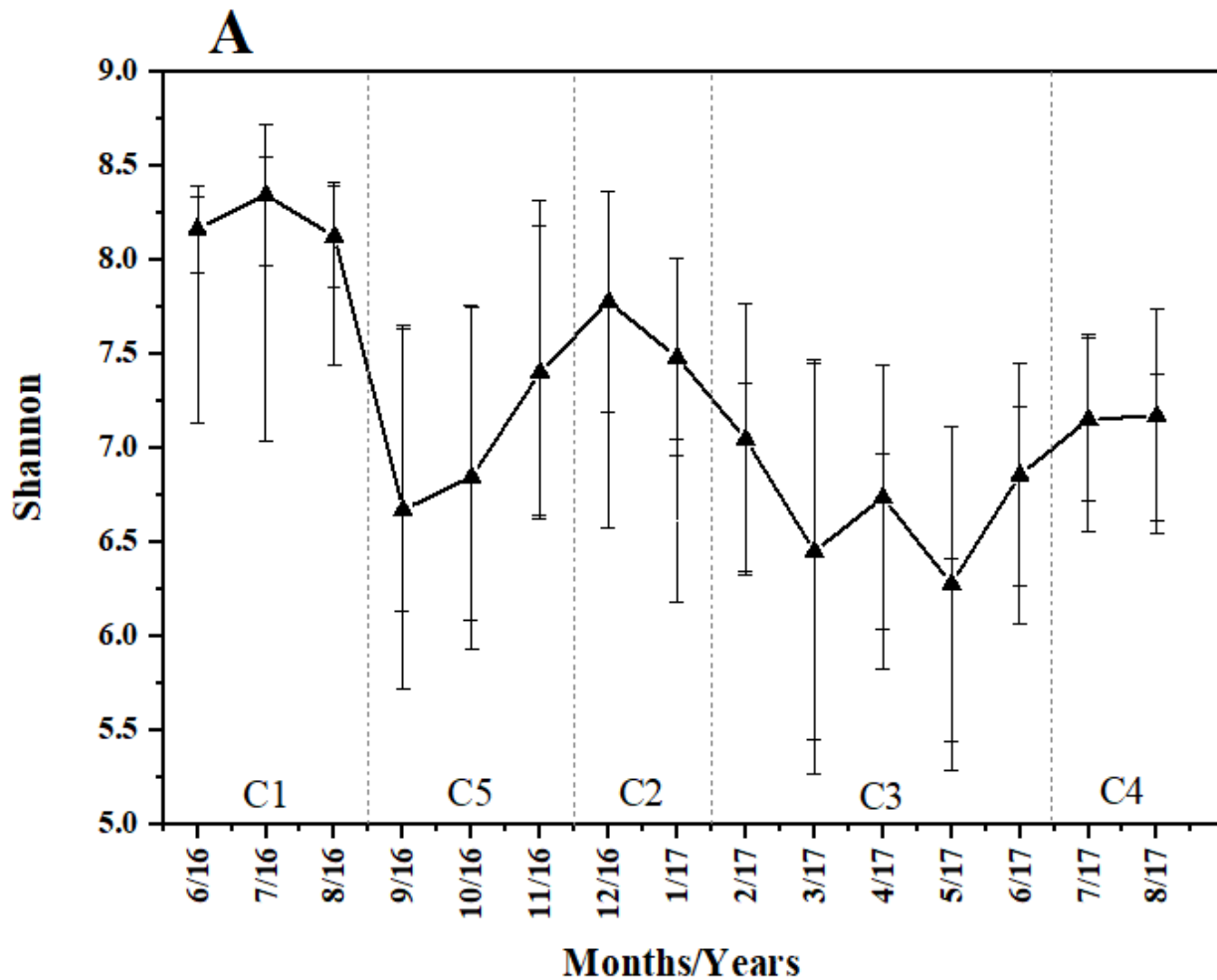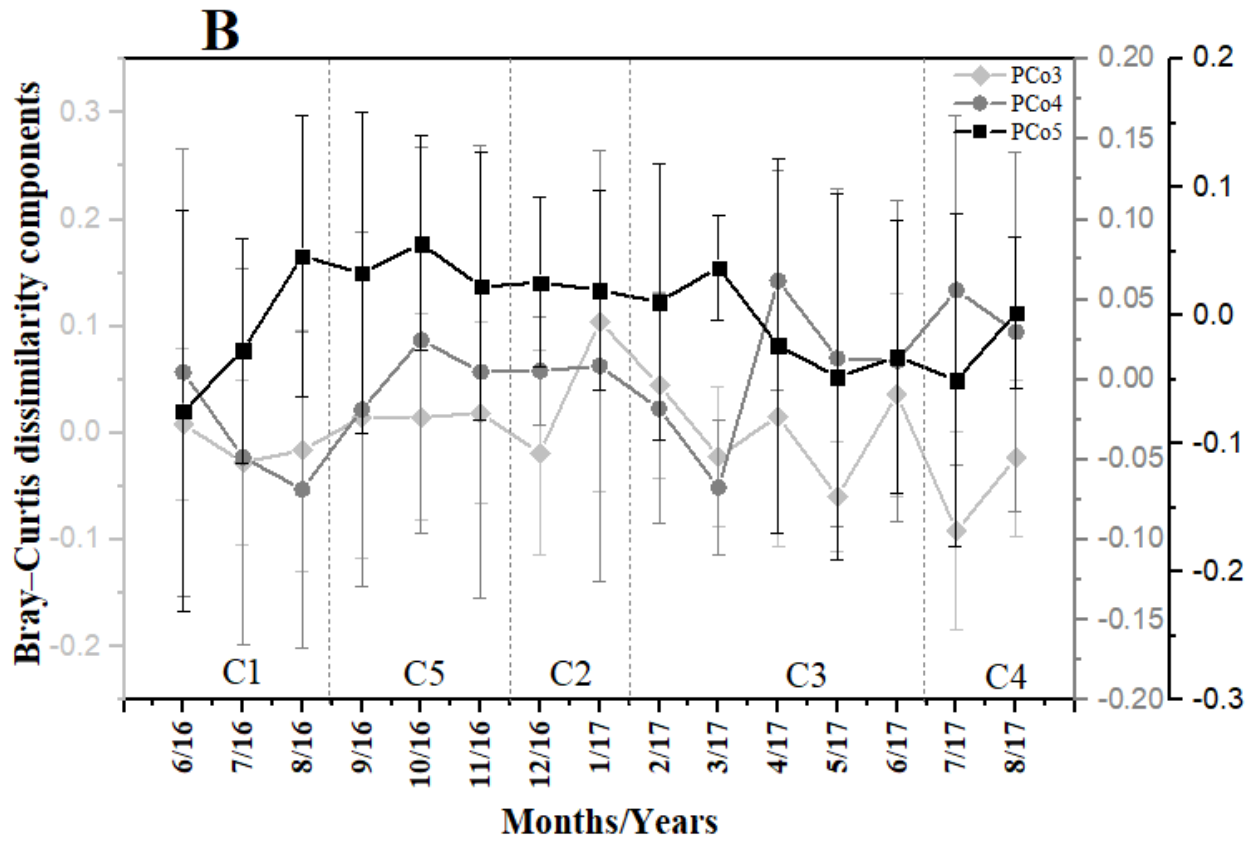
Figure S3.4. Line plots of monthly changes of Shannon (top panel) and Bray–Curtis dissimilarity
components; PCo3-5 (bottom panel) of 6 different locations over 15 months of sampling. C1-5:
Cluster 1-5 based on figure 4.
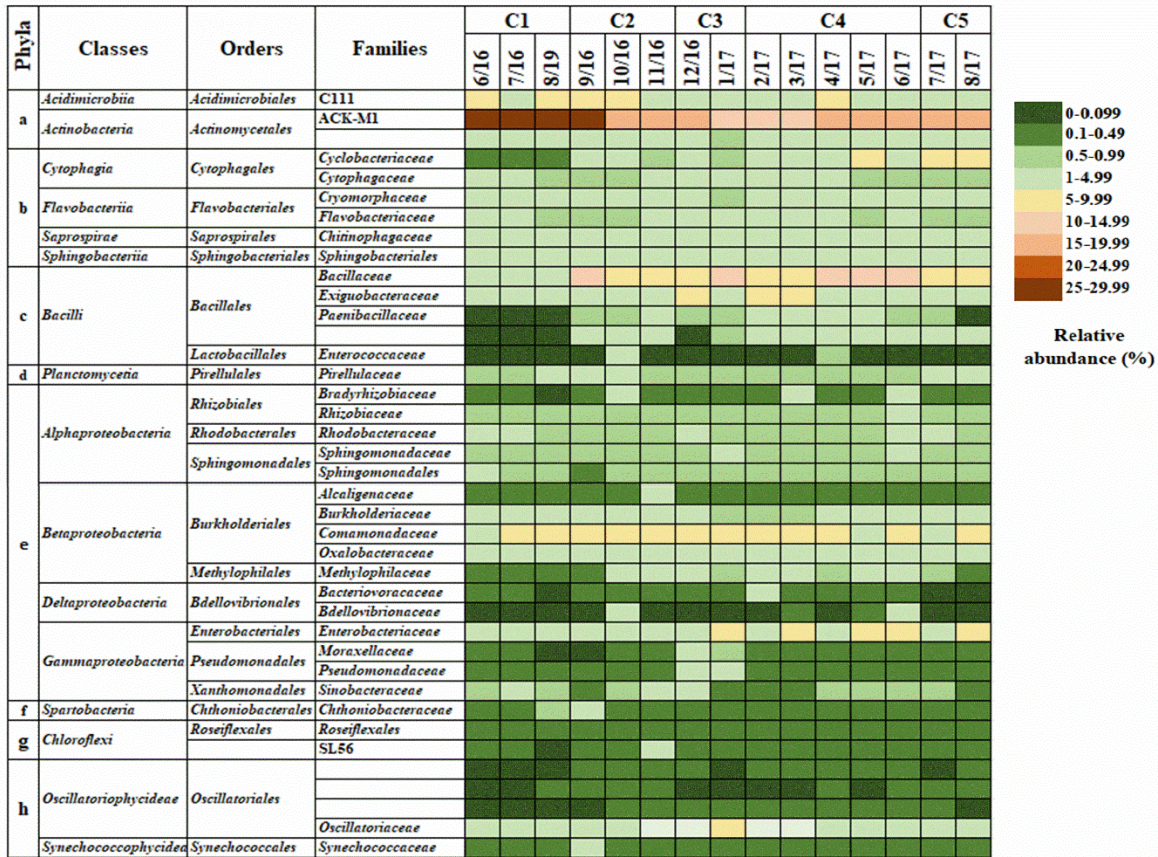
Phyla; a: *Actinobacteria*, b: *Bacteroidetes*, c: *Firmicutes*, d: *Planctomycetes*, e: *Proteobacteria*, f: *Verrucomicrobia*, g: *Chloroflexi*, h: *Cyanobacteria*

Figure S3.5. Phylogenetic affiliations of the top heterotrophic bacterial (panel a) and phototrophic (panel b) OTU groups from the 15 months of sampling in Lake Erie and St. Clair over 2016 and 2017. Due to space constraints, only taxa which had relative abundance more than 1% in at least one sampling month are presented for heterotrophic bacterial (panel a).

196

80
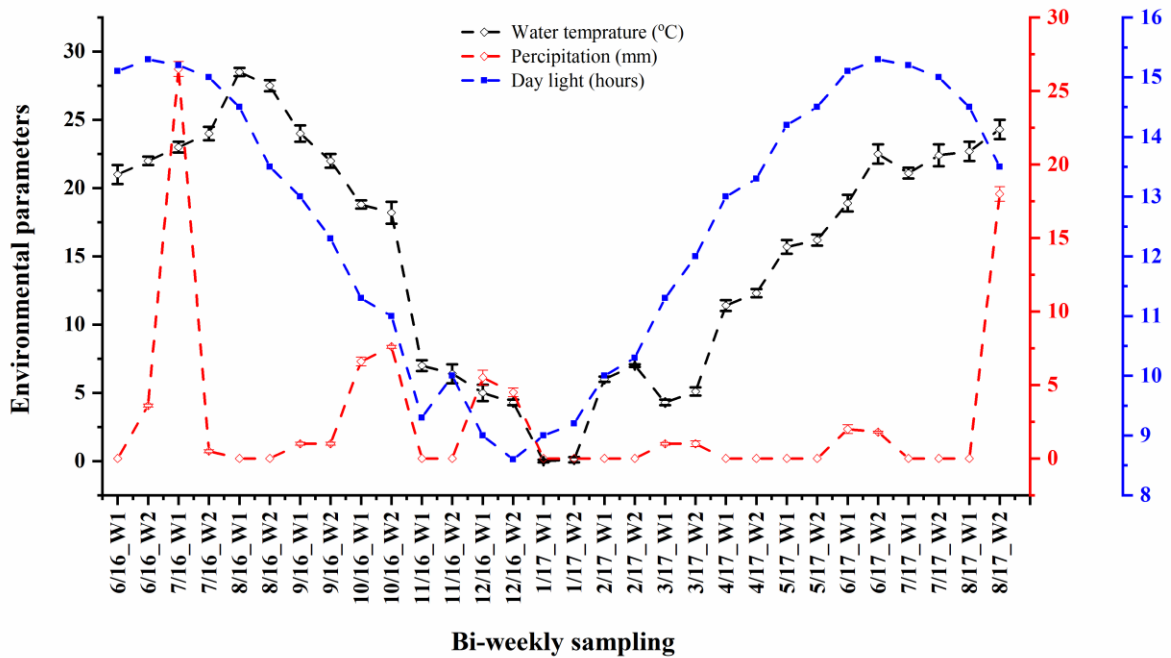
Figure S3.6. Pattern of environmental parameter variation (water temperature, precipitation and daylight duration) for the 6 sampling locations over 15 months of the sampling. As air and water temperature both had same pattern spatially and temporally (no significant variation) we only plotted water temperature. For each month, 2 different weeks were sampled (week 1 and 2). Sampling was started June 2016 and ended August 2017. Error bars, showing the standard deviation of water temperature and precipitation among 6 sampling locations.

87

Distance based Liner Model (DistLM) analysis of environmental variables on the BCC of 15 months

Variables	SS (trace)	Pseudo-F	P value	Proportion of explained variation
Day light hours	16759	9.44	0.001	0.096
Precipitation	8532.2	4.56	0.001	0.049
Water temperature	27499	16.64	0.001	0.159

DistLM sequential tests (specified solutions)

R²	RSS	Number of variables	Selection of variables
0.21	22894	3	Day light hours, precipitation, water temperature

88

89 **Figure S3.7.** Distance based Redundancy Analysis (dbRDA) of freshwater microbiota. Relative
90 position of water samples in the biplot is based on Bray Curtis similarity of square root transformed
91 relative abundance at the OTU level. Vectors indicate the weight and direction of the
92 environmental variables that were best predictors of the BCCs of different months as suggested by
93 the results of the distance-based linear model (distLM). The dbRDA axes describe the percentage
94 of the fitted or total variation explained by each axis while being constrained to account for group
95 differences. Sample IDs indicate the sampling months.

APPENDIX C; SUPPLEMENTARY INFORMATION OF CHAPTER 4

Table S4.1. Sequencing statistics of the metatranscriptomic profiles obtained from the Illumina HiSeq 4000 run for all samples are summarized in the

Samples ID	Number of reads	Failed in QC (%)	Processed high-quality sequences	High-quality sequences		
				rRNA (%)	% (reads) of high-quality transcripts with known function	% high quality reads with unknown function
CH1	29,798,479	83	5,056, 742	81	4.7 (237,666)	14.3
CH1C	34,468,587	84	5,515,120	80	5.2 (286,786)	14.8
CH18	30,489,974	83	5,183,290	79	4.2 (217,698)	16.8
CH19	34,223,787	83	5,818,004	76	5.2 (302,536)	18.2
CH21	32,577,928	81	6,189,806	76	6.5 (402,337)	17.5
CH22	30,713,668	84	4,914,186	77	6 (294,851)	17
SL1	30,180,206	85	4,527,030	78	5.3 (239,932)	16.7
SL1C	31,185,883	85	4,667,882	80	4.7 (219,390)	15.3
SL18	34,738,192	88	4,168,583	81	5.2 (216,766)	13.8
SL19	33,112,182	86	4,635,705	81	4.5 (208,606)	14.5
SL21	33,732,219	87	4,384,188	81	4.8 (210,441)	14.2
SL22	34,534,774	82	6,216,259	81	4 (248,650)	15
SP1	31,089,357	85	4,663,403	80	4.8 (223,843)	15.2
SP1C	35,001,268	88	4,200,152	82	5.1 (214,207)	12.9
SP18	32,132,331	84	5,141,172	81	4 (205,646)	15
SP19	33,130,782	87	4,307,001	81	4.8 (206,736)	14.2
SP21	32,252,781	84	5,160,444	79	9 (464,439)	12
SP22	33,243,171	84	5,318,907	80	8.2 (436,143)	11.8

Table S4.2. Spatial variation of the transcripts of the biological/ecological relevant genes detected in the transcriptomic data of three sampling locations. We only presented the genes which had known function and were part of physiological and ecological pathways and expressed by at least two different species of bacteria.

Pathways	Function	Gene(s)	Sampling locations		
			CH	**SL**	**SP**
Aminoacyl-tRNA biosynthesis	D-aminoacyl-tRNA deacylase	*dtd*	+	+	+[a]
Fatty acid biosynthesis	Malonyl CoA-acyl carrier protein transacylase	*fabD*	+	+	+[a]
	3-oxoacyl-[acyl-carrier-protein] synthase 2	*fabF*	+	+	+[a]
Glycolysis	Pyruvate dehydrogenase	*pdhA*	+	+	+[a]
	Pyruvate kinase	*pyk*	+	+	+[a]
Phototrophy	Photosystem I	*psaA*	+	+	+[a]
	Photosystem II	*psbA, psbA1, psbA2, psbD1, psb27, psbO*	+	+	+[a]
Porphyrin metabolism	Uroporphyrinogen-III synthase	*hemD*	+	+	+[a]
Purine metabolism	dITP/XTP pyrophosphatase	*rdgB*	+	+	+[a]
RNA polymerase	DNA-directed RNA polymerase subunit beta	*rpoB*	+[b]	+	+
Two-component system	Transcriptional regulatory protein PhoP	*phoP*	+[b]	+	+
	Transcriptional regulatory protein RstA	*rstA*	+[b]	+	+

'+' indicates occurrence in three sampling locations. Letters indicates significantly high transcript frequency in; SP versus CH and SL (a) and CH versus SP (b).

Table S4.3. Environmental variables measured in winter, summer and fall (2019) at three different locations.

Environmental variables	Winter			Summer			Fall		
	CH	**SL**	**SP**	**CH**	**SL**	**SP**	**CH**	**SL**	**SP**
Water temperature (°C)	1.2	1.5	1.3	23	23.5	24	12	12.5	13
Daylight (hour)	9	9	9	15	15	15	12	12	12
TN (NO3+NO2) (mg/L)	0.23	0.26	0.24	0.71	0.58	1.38	1.21	1.7	2.8
TP (mg/L)	0.02	0.02	0.08	0.01	0.01	0.25	0.09	0.08	0.13

Table S4.4. Distance-based Liner Model (DistLM) analysis of environmental variables effects on the metatranscriptomic data.

Variable	SS (trace)	Pseudo-F	P value	Proportion of explained variation
Day light	9906.9	11.465	**0.001**	0.41745
Water temperature	11801	15.826	**0.001**	0.49727
TN	6394.9	5.9018	**0.004**	0.26947
TP	1911.3	1.4015	0.222	0.080539
	$r^2=0.82311$			
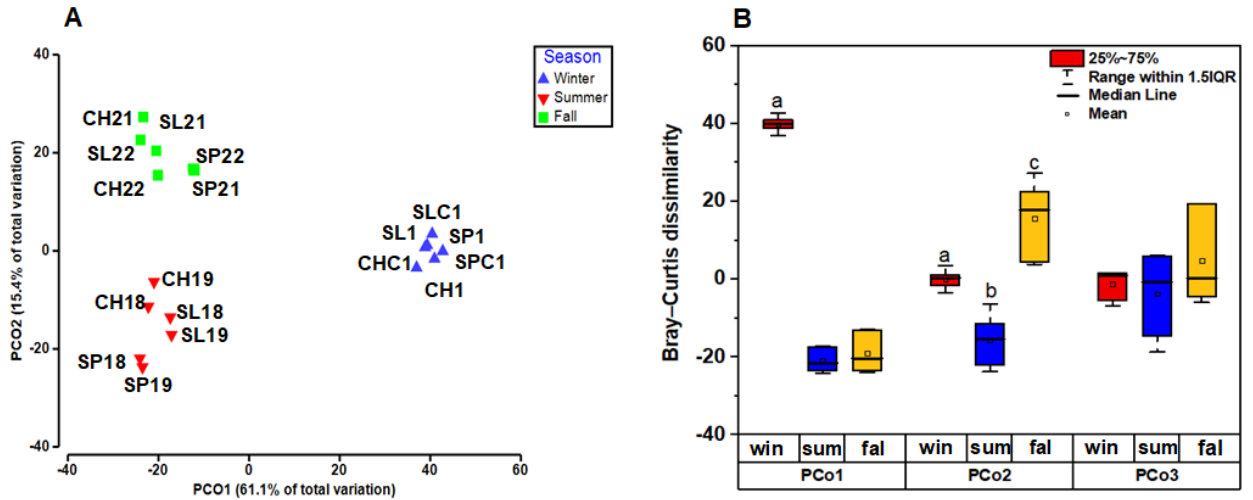
Figure S4.1. Patterns of similarity among freshwater bacterial meta-transcripts at three locations over three seasons. Panel A: Principal coordinates analysis (PCoA) based on the Bray–Curtis dissimilarity matrix generated using metatranscriptomic data (transcript reads) of the samples collected from Lake Eire (2 locations) and Lake St. Clair (1 location) over winter (win), summer (sum) and fall (fall). Panel B: A box plot showing the mean and variation of PCo1, PCo2 and PCo3 over winter (win), summer (sum) and fall (fall). The letters on top of boxes indicate significant differences based on one-way ANOVA.
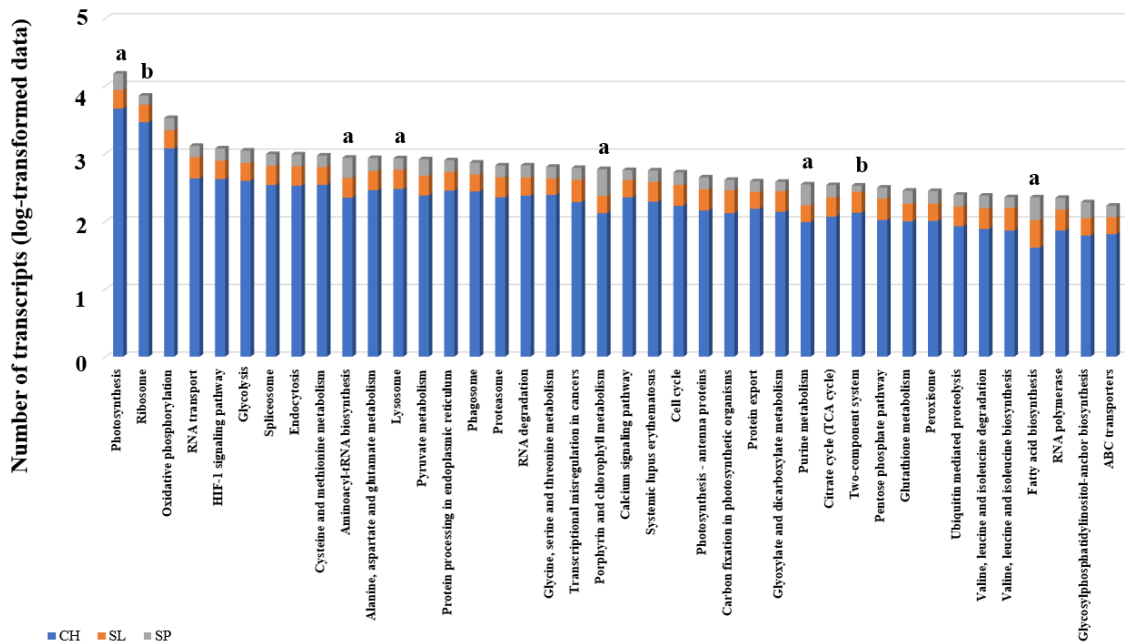
Figure S4.2. Out of the top 50 abundant KEGG pathways, the 40 most abundant KEGG pathways in the metatranscriptomic data of CH (blue), SL (green) and SP (red) are displayed. The pathways marked with the letters were highly expressed (p<0.05) at; SP versus CH and SL (a) and CH versus SP (b).
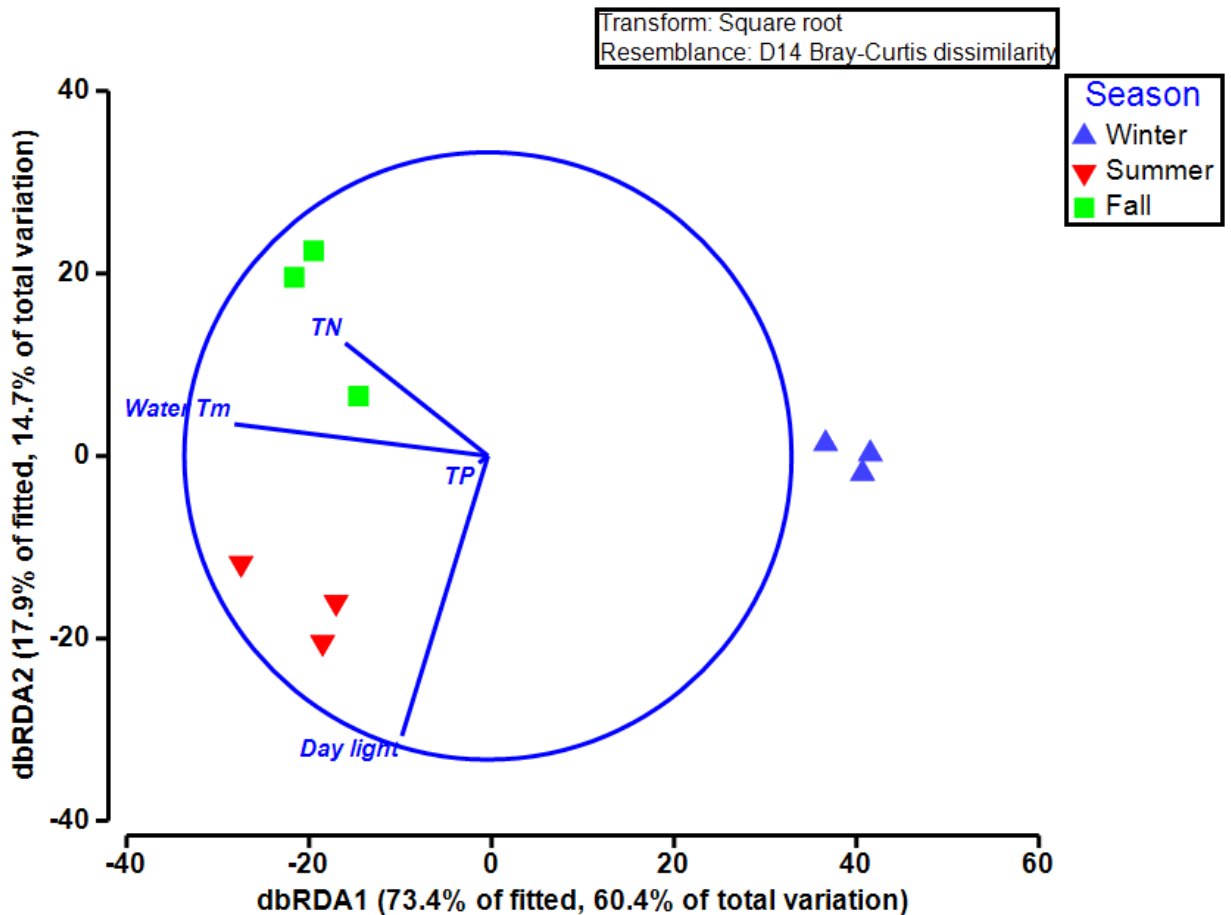
Figure S4.3. Distance-based redundancy analysis (dbRDA) plot showing the influence of water temperature (Water Tm), daylight, total phosphate (TP) and total nitrogen (TN) on metatranscriptomic data of freshwater BC in winter, summer and fall (2019).

APPENDIX D; SUPPLEMENTARY INFORMATION OF CHAPTER 5

Table S5.1. Networks characteristics of different treatment

Characteristics	P2-Co	C-LND	R-LND	C-HND	R-HND
Clustering coefficient	0.45	0.49	0.5	0.48	0.39
Network centralization	0.31	0.4	0.4	0.34	0.25
Characteristic path length	2.3	2.2	2.2	2.33	2.5
Average number of neighbors	42.5	29	49.7	27.5	17
Number of nodes	253	181	222	199	127

Table S5.2. Keystone taxa profile of each the BC of control (P2-Co), challenge communities (C-LND and C-HND) and relaxed communities (R-LND and R-HND). Data are presented at the phyla and family levels.

Treatments	Families	Phyla	OTUs
P2-Co	*Solibacterales*	*Acidobacteria*	250
	C111		370
	Microbacteriaceae		13
	ACK-M1	*Actinobacteria*	63,311, 367
	Cytophagaceae	*Bacteroidetes*	790
	Chitinophagaceae		1020
	Cryomorphaceae		639
	Sphingobacteriaceae		5, 64, 131, 320, 396, 1007
	Chloroflexi	*Chloroflexi*	1567
	Roseiflexales		94
	Bacillaceae	*Firmicutes*	342
	Staphylococcaceae		1159, 1241
	Planococcaceae		24, 840, 1073
	Phycisphaerales	*Planctomycetes*	125, 286
	Isosphaeraceae		59
	Gemmataceae		164, 454, 1178
	Pirellulaceae		20, 1136
	Planctomycetaceae		170, 297
	Sinobacteraceae	*Proteobacteria*	318, 460
	Acetobacteraceae		985
	Rhodospirillaceae		105, 523

	Caulobacteraceae		27
	Comamonadaceae		9, 12, 83, 281, 421
	Ellin6067		156
	Hyphomonadaceae		56, 249
	Oxalobacteraceae		559
	Pseudomonadaceae		18
	Rhodocyclaceae		725
C-LND	ACK-M1	*Actinobacteria*	357
	C111		116
	Cytophagaceae	*Bacteroidetes*	31, 73
	Cyanobacteria	*Cyanobacteria*	77
	Synechococcaceae		42
	Pseudomonadaceae	*Proteobacteria*	189
	Alphaproteobacteria		401
	Comamonadaceae		2,4,9,26
	Erythrobacteraceae		425, 449
	Burkholderiales		1490
	Rhizobiaceae		134
	Acetobacteraceae		127, 284
	Bdellovibrionaceae		174, 317
	Sinobacteraceae		3843
	Xanthomonadaceae		1432
	Sphingomonadaceae		130, 726
	Rhodobacteraceae		182, 199
C-HND	C111	*Actinobacteria*	124, 251,1584
	Saprospiraceae	*Bacteroidetes*	1517
	Cytophagaceae		73, 2131
	Sphingobacteriaceae		175
	Synechococcaceae	*Cyanobacteria*	42
	WD2101	*Planctomycetes*	273
	Gemmataceae		2126, 2212
	Rhodobacteraceae	*Proteobacteria*	351, 442, 509, 806, 1670
	Caulobacteraceae		16, 823, 836
	BD7-3		312
	Acetobacteraceae		560, 792
	Sphingomonadaceae		475, 580
	Xanthomonadaceae		138, 384
	Oxalobacteraceae		8
	Methylophilaceae		417, 684, 912,
	Rhizobiaceae		277

	Comamonadaceae		21,115
	Chthoniobacteraceae	*Verrucomicrobia*	17
R-LND	PK29	*Acidobacteria*	45
	Actinomycetales	*Actinobacteria*	38
	Microbacteriaceae		13
	Gaiellales		207
	ACK-M1		63311
	Cytophagaceae	*Bacteroidetes*	76, 790
	Sphingobacteriaceae		5, 64, 131, 320, 396
	Chitinophagaceae		265, 477,1020
	Cryomorphaceae		639
	Bacillaceae	*Firmicutes*	311
	Phycisphaerales	*Planctomycetes*	125
	Gemmataceae		164, 454
	Caulobacteraceae	*Proteobacteria*	11
	Comamonadaceae		400, 526,1347,1776
	Hyphomonadaceae		23
	Oxalobacteraceae		8, 86
	Erythrobacteraceae		28,118. 1098, 2987
	Alcaligenaceae		121
	Alphaproteobacteria		551
	Sphingomonadaceae		5,15,34,61,667
	Verrucomicrobia	*Verrucomicrobia*	34
R-HND	*Microbacteriaceae*	*Actinobacteria*	13
	ACK-M1		3, 334
	Synechococcaceae	*Cyanobacteria*	42
	Streptophyta		622
	Planococcaceae	*Firmicutes*	676
	Streptococcaceae		142
	Bacillaceae		228
	Longimicrobiaceae	*Gemmatimonadetes*	322, 600
	Isosphaeraceae	*Planctomycetes*	59
	Pirellulaceae		20
	Methylophilaceae	*Proteobacteria*	684, 912
	Comamonadaceae		4,12,21
	Mycoplasmataceae		1794
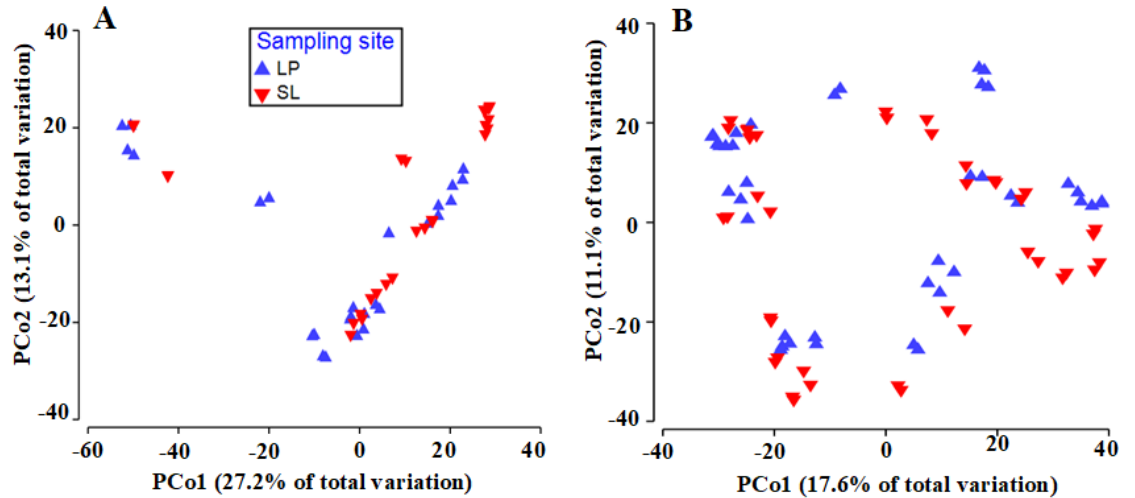
Figure S5.1. PCoA analysis of Barry-Curtis dissimilarity matrix of the BCs (weeks 1-4) of phase 1 (panel A) and phase 2 (panel B) of two sampling sites; SL and LP.
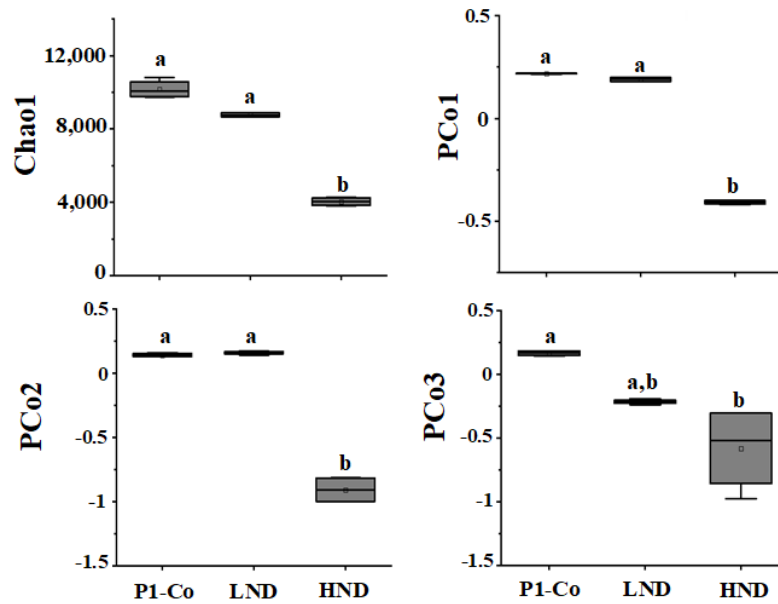
Figure S5.2. Box plots showing the variation of Chao1, PCo1, PCo2 and PCo3 of the adaptation (phase 1) experiment (week 4). Letters show significant differences between the BC of P1-Co and two adapted BCs; LND and HND. The thick bar is median, upper and lower quartiles represent 75% and 25% of the data respectively. Whiskers are used to indicate variability outside the upper and lower quartiles.
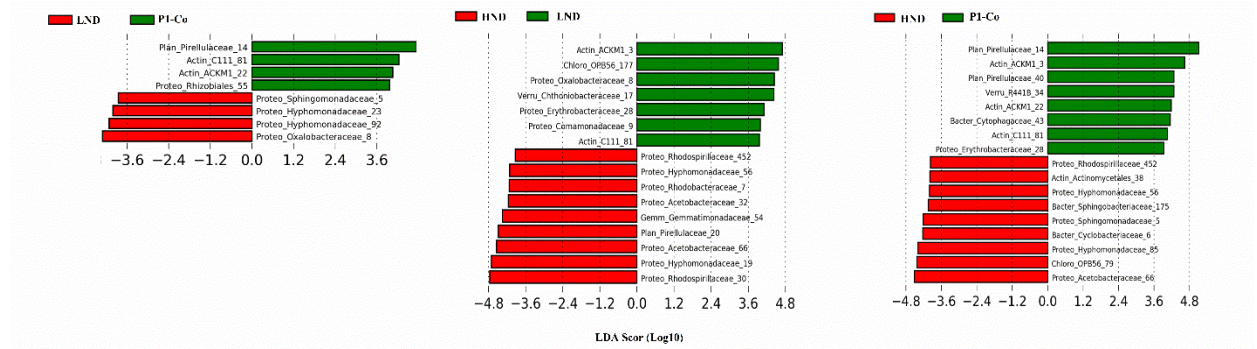
Figure S5.3. Graphic showing the LDA effect size (LEfSe) of highly abundant OTUs of the BCs of P1-Co, LND and HND. Horizontal bars represent the effect size for each taxon. The length of the bar represents the log^{10} transformed LDA score, indicated by vertical dotted lines. OTUs (numbers) and their taxa information with statistically significant change (p < 0.05) in the relative abundance is written alongside the horizontal lines. Taxonomic abbreviations: Actin; *Actinobacteria*, Bacter; *Bacteroidetes*, Chloro; *Chloroflexi*, Gemm; *Gemmatimonadetes*, Plan; *Planctomycetes*, Proteo; *Proteobacteria* and Verru; *Verrucomicrobia.*
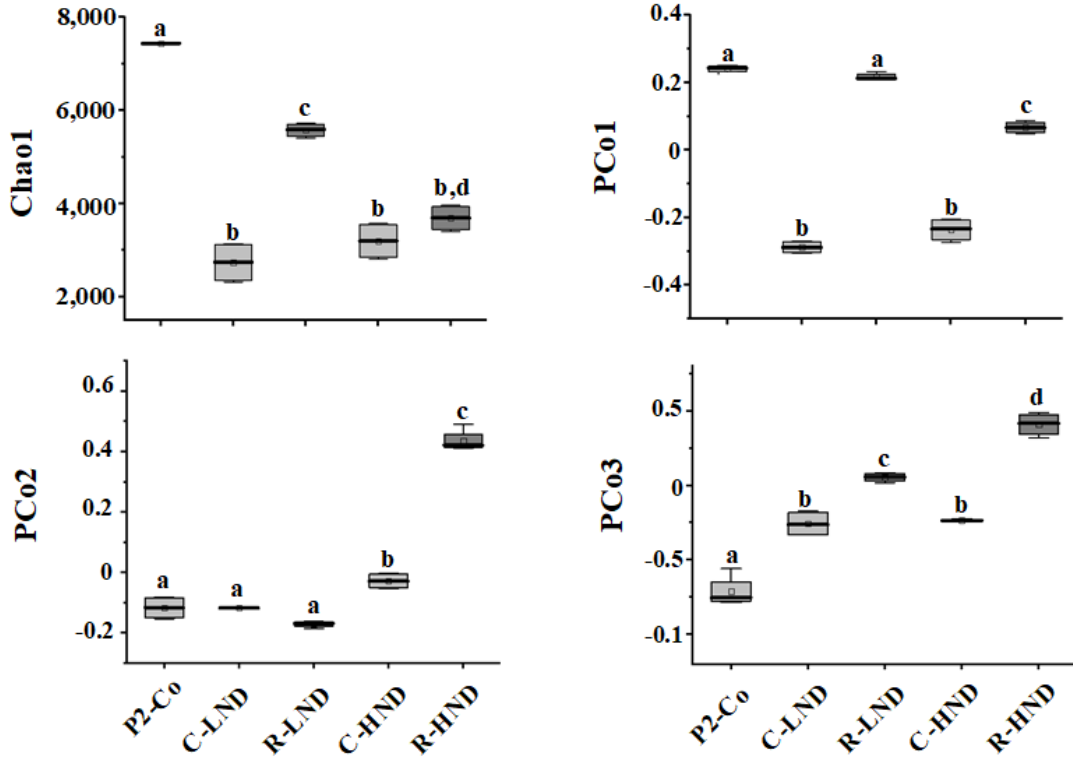
Figure S5.4. Box plots showing the variation of Chao1, PCo1, PCo2 and PCo3 of the BCs in phase 2 (week 4). Significant variation was highlighted by letters between the BCs of P2-Co, C-LND, C-HND, R-LND and R-HND. The thick bar is median, upper and lower quartiles represent 75% and 25% of the data respectively. Whiskers are used to indicate variability outside the upper and lower quartiles.

Figure S5.5. Graphics of LDA effect size (LEfSe) for high abundant OTUs of the BCs among different treatments of challenge phase (week 4). Panel A; P2-Co, C-LND and R-LND, panel B; P2-Co, C-HND and R-HND, panel C; challenge and relaxed communities' comparison. Horizontal bars represent the effect size for each taxon. The length of the bar represents the log^{10} transformed LDA score, indicated by vertical dotted lines. OTUs (numbers) and their taxa information with statistically significant change (p < 0.05) in the relative abundance is written alongside the horizontal lines. Taxa abbreviation: Acido;
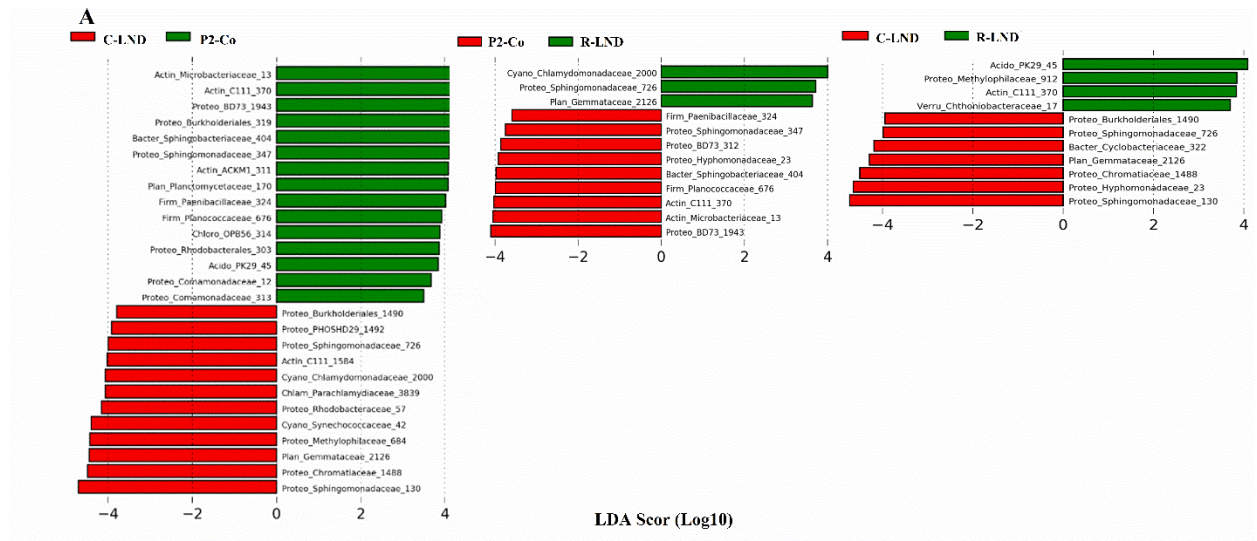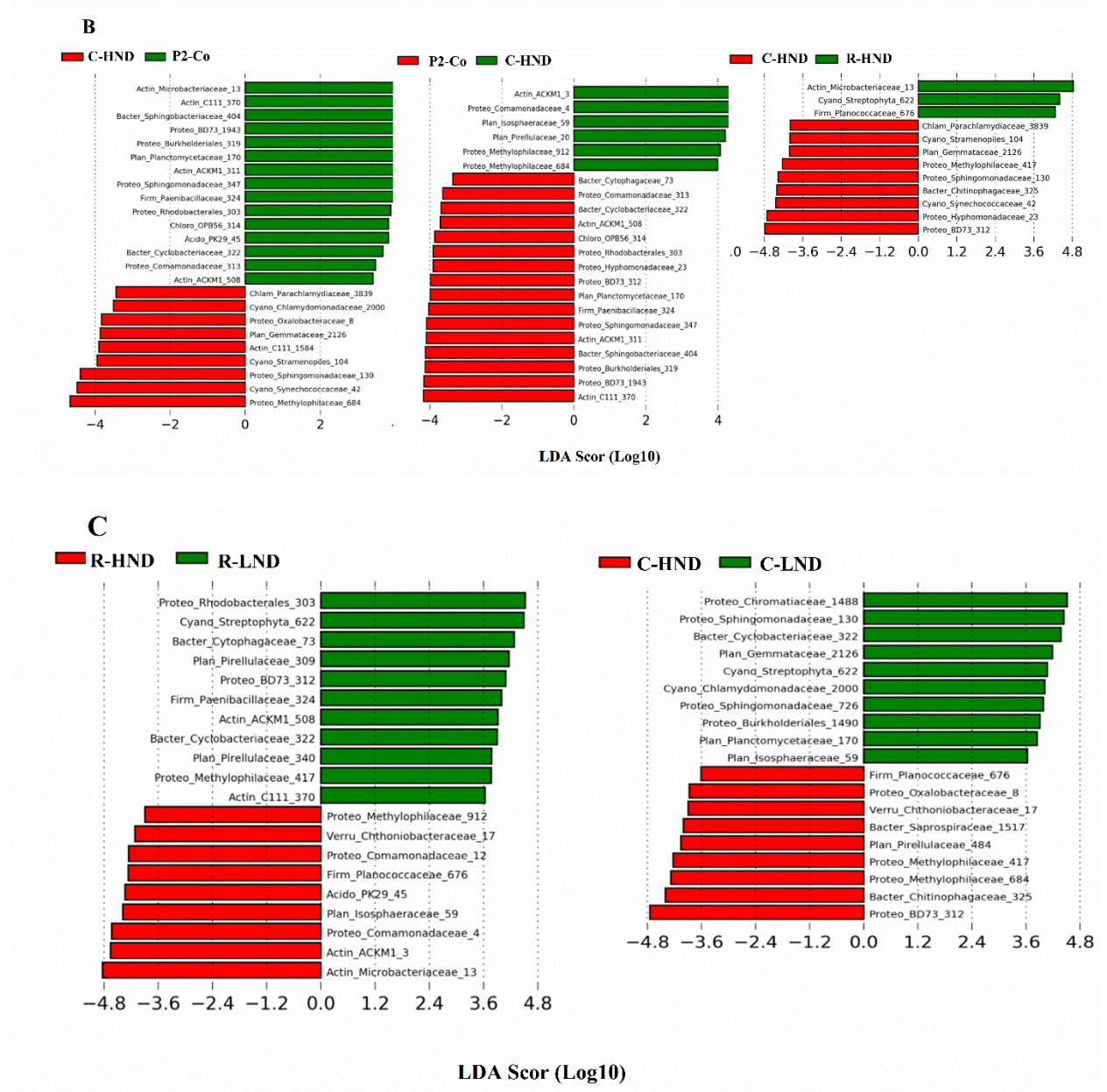
Acidiobacteria, Actin; *Actinobacteria*, Bacter; *Bacteroidetes*, Chlam; *Chlamydiae*, Chloro; *Chloroflexi*, Cyano; *Cyanobacteria*, Firm; *Firmicutes*, Gemm; *Gemmatimonadetes*, Plan; *Planctomycetes*, Proteo; *Proteobacteria* and Verru; *Verrucomicrobia.*
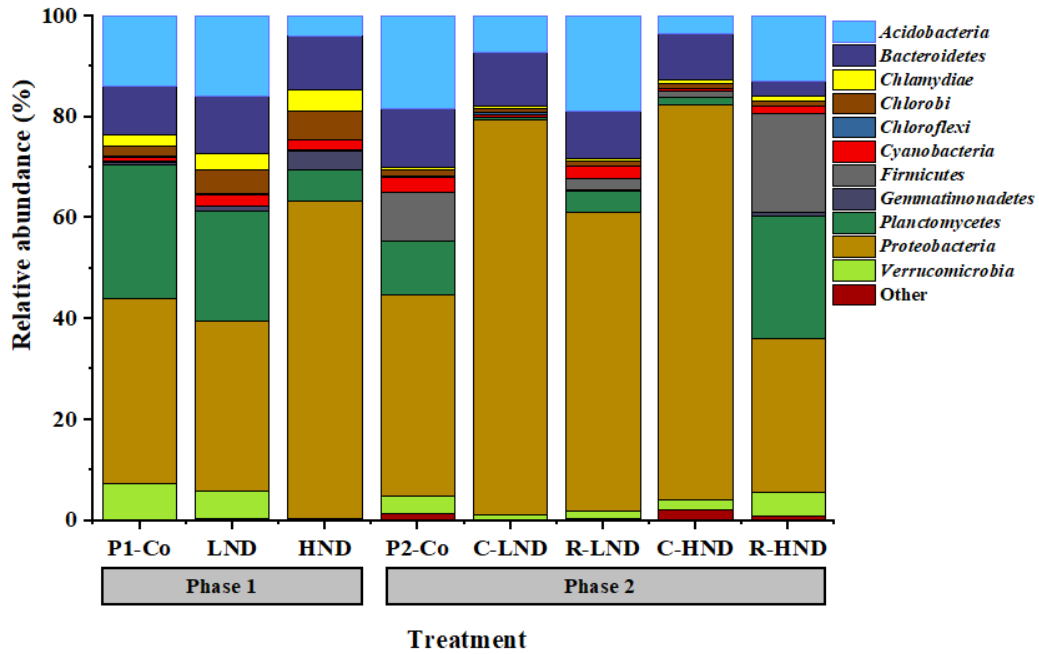
Figure S5.6. Taxa composition (phyla level) of the BCs of different treatments at phase 1 (adaptation phase) and 2 (challenge phase) over week 4.

APPENDIX E; SUPPLEMENTARY INFORMATION OF CHAPTER 6

Table S6.1. List and the sequences of the primers and probes designed in this study for quantitative qPCR assays. Primers were designed for 2 FIBs, 7 MST markers and 15 waterborne pathogens

Groups	Target	Genes	Size (bp)	FAM probe sequence (5'-3')	Forward Primer Sequence (5'-3')	Reverse Primer Sequence (5'-3')
A	1	23 rRNA	71	ACCGTAAACCGACACAGGTA	GTGAGAAAACAACTGCCCGT	CGCTCGCTCACCTTAGGATA
A	2	*UidA*	70	ACCGATACCATCAGCGATCT	GTAATGTTCTGCGACGCTCA	AATAACGGTTCAGGCACAGC
B	3	16S rRNA	90	TTGGAACTGAGACACGGTCCA	TAGGGGTTCTGAGAGGAAGGTC	CATTGACCAATATTCCTCACTGCT
B	4	16S rRNA	91	AGATTAAAGAATTTCGGTA	CGATGGTATCTTAAGCGCACATGCAATT	CTAGGTGGGCCGTTACCCCGCCTACTAC
B	5	16S rRNA	70	GACCTTCCCTTTACTCAGGGATAGCCT	GTCAGCTTGCTGACTCTGATGGCGA	GCAGGGAAACAAAGGCATCAGGTA
B	6	16S rRNA	81	GTCGAATTAATGCCTGATGT	CCGGGCTATTCCTGACCATGGGGAAGGT	TTGTTCCACATCAGATGCCGTCCGTG
B	7	16S rRNA	83	GGAAACAGGTGCTAATACCGC	CATCAGAGGGGGACAACACT	AAGCGCCTTTCAACAAAGAA
B	8	MT-ND2	81	TTCAATGAATCTGAGGAGGCT	TATCCGCCATCCCATACATT	GTGTGAGGGTGGGACTGTCT
B	9	*NifH*	90	TTGTGCGGTAGAAGAATCCC	CGAAAGCTGATACAACACGTACA	CCTCTTCACTGGGTTTTCTGTT
C	10	*GltA*	90	AGGCGCATTATCTGCTTTCT	TATGGCAATCATGGTTGGTG	GTAATTTCGCGGTGGTTGAT
C	11	*Lip*	73	ATGCTGGCTGGCTGTGGTGG	TAATTTACGCCGCGGTTGTC	GTGTCGCCTTTGTTATCGTCAC
C	12	*GlyA*	89	CCCACAAAACCTTACGTGGT	GGTGAGCATCCAAGTCCATT	CATGATGATACCGCCTCTTGGACCAC
C	13	*HipO*	83	CCTTTGCAAGAATGCACAAA	TTCGTGCAGATATGGATGCT	CGCAAGCATGCATTACATTT
C	14	*manC*	76	TGACCAAGGTGAGCGATACA	CATGGGGAAAATTTGATTCG	CCCTCACCAGGTTTCACAAT
C	15	*manC*	79	TAAGCCAAAAGGATTGCCAT	GGCTGGAATGACGTGGGTTCTTGGTCA	TTGAGCACATCCCCATGGCACACATTAC
C	16	*manC*	66	CAGGGCGAACCTTTAGGTTT	GCCAGGCGTGACCATTATGAACGTG	CGGGTCGTGCACACAAAATGGAGTGG
C	17	*PhoE*	78	GATTAACGACGACCTGACCG	TCGGTATTAAAGGCGAAACG	CGGTTTTGTTACCGCAGAAT
C	18	*MipA*	72	TGAAAGACGTTCTTAACAAGTTTCA	GGCTTTAACCGAACAGCAAA	GCAGTACGTTTTGCCATCAA
C	19	*Hly*	83	CAAGGATTGGATTACAATAAAAACAA	ACGCGGATGAAATCGATAAG	GTCACTGCATCTCCGTGGTA
C	20	*RegA*	81	CGACGACGTGGACAAGCT	ACTATCACGGCATTCCTTCG	AACAGAAACTGCCGATGACC
C	21	*InvA*	88	AAAGGTTCAGAACGTGTCGC	TGTCACCGTGGTCCAGTTTA	GGGCATACCATCCAGAGAAA

C	22	*IpaH*	64	AAGTAAATCTGCGGGCCGT	CTGTTGTCCGTCACTCATGG	ATTAACCTCTTCGCCGGACT
C	23	*gyrA*)	67	TCCCTAACTTGTTAGCCAATGG	GCCGTCAGTCTTACCTGCTC	ACCTACCGCGATACCTGATG
C	24	*CtxA*	77	GGAGCATAGAGCTTGGAGGG	CAGCAGCAGATGGTTATGGA	ATGATGAATCCACGGCTCTT

A; FIBs markers, B; MST markers, C; waterborne pathogens. 1; *Enterococcus* spp., 2; *Escherichia coli*, 3; *Bacteroides-Prevotella* spp, 4; *Bacteroides* spp. (Dog marker), 5; *Bacteroides* spp. (Canada goose marker), 6; *Bacteriodes* spp. (Pig marker), 7; *Catellicoccus marimammalium* (Seagull), 8; C40 mitochondria (Human marker), 9; *Methanobrevibacter smithii* (Human marker), 10; *Acinetobacter baumannii*, 11; *Aeromonas hydrophila*, 12; *Campylobacter coli*, 13; *Campylobacter jejuni*, 14; *E. coil* O157:H7, 15; *E. coli* O111, 16; *E. coli* O26; 17; *Klebsiella pneumoniae*, 18; *Legionella pneumophila*, 19; *Listeria monocytogenes*, 20*; Pseudomonas aeruginosa*, 21; *Salmonella typhimurium*, 22; *Shigella* spp., 23; *Staphylococcus aureus*, 24; *Vibrio cholerae*

Table S6.2. Quality and concentration of DNA extracted for samples used in this study

Samples and their source		Sample's IDs	260/280 ratio	260/230 ratio	Concentration (ng/µL)
Environmental samples	Beach water	CI	1.99	2.17	66
		HB	1.84	2.1	58
		LP	1.88	1.98	88
		PP	1.91	1.99	56
		SL	1.78	2.15	59
		SP	1.93	2.2	76
	Pore water	CI	1.76	2.11	160
		HB	2.1	2.22	156
		LP	1.88	2.05	172
		PP	1.78	1.98	69
		SL	1.79	2.14	110
		SP	1.87	1.95	112
	Sand	CI	1.68	1.89	61
		HB	1.83	2.11	57
		LP	1.79	2.15	59
		PP	1.75	1.98	81
		SL	1.92	2.15	59
		SP	1.97	2.18	69
	Stream water	HB	1.83	2.08	88
		LP	1.77	2.05	109
		PP	1.89	2.02	132
	Pond water	1	2.05	2.11	122
		2	2.1	2.17	156
		3	1.93	2.15	101
Fecal samples	Dog	1	1.83	2.12	270
		2	1.78	2.05	130
		3	1.98	1.98	232
		4	1.81	1.88	178
	Canada goose	1	1.89	1.91	249
		2	1.99	1.96	257
		3	2.1	2.05	290
		4	2.05	2.23	238
	Pig	1	1.79	2.21	211
		2	1.88	2.11	189
		3	1.67	2.05	156
		4	1.69	1.99	197
	Seagull	1	1.87	1.95	234
		2	1.92	2.21	256
		3	1.89	2.07	279

		4	1.69	2.11	266
	Human	1	1.84	2.17	302
		2	1.87	2.06	319
	Sewage	1	1.89	1.98	412
		2	1.88	1.95	456
		3	1.71	2.1	480

Table S6.3. Mean of C_T values for all markers obtained by SYBR green qPCR (top) and TaqMan® OpenArray® plate (bottom) for a known concentration of each marker.
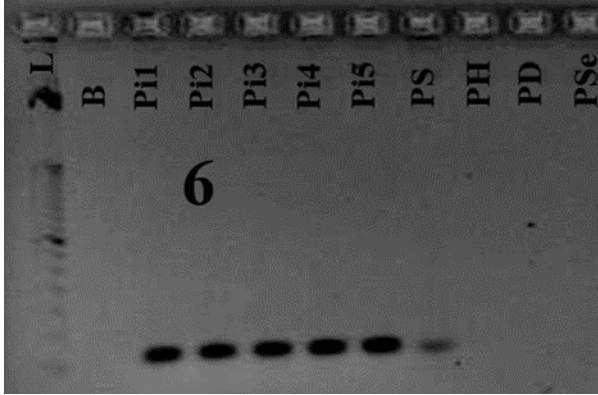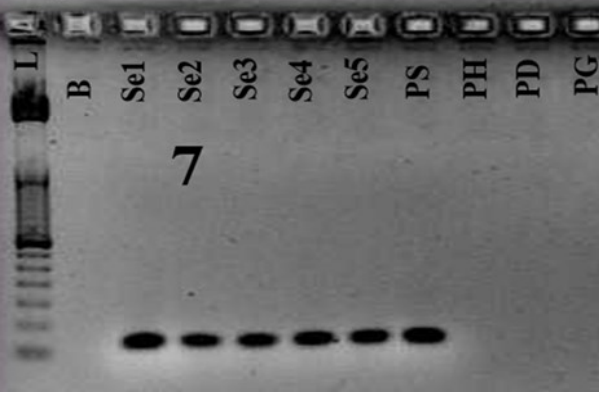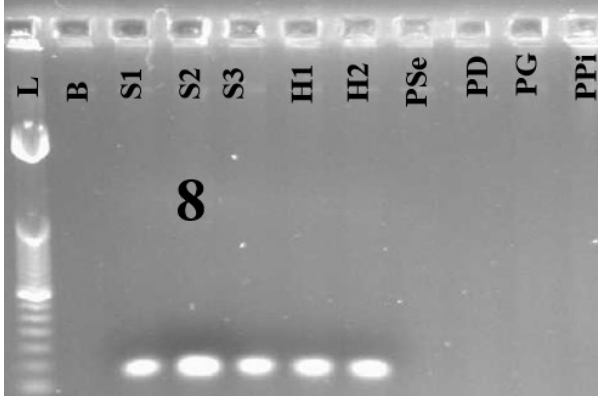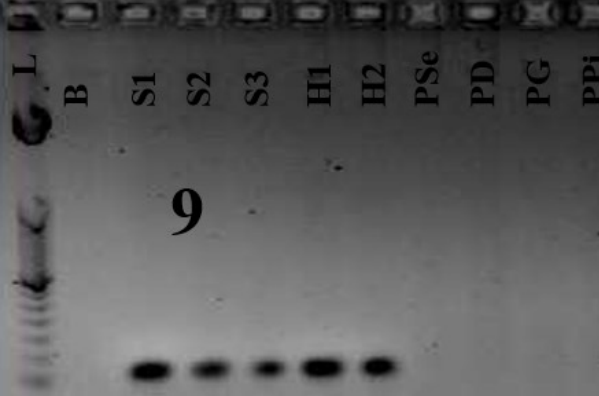
Markers/species	SYBR green qPCR Mean C_T value (copies/reaction)							
	2,000,000	200,000	20,000	2,000	200	20	2	1
Acinetobacter baumannii	18.41	21.27	24.94	27.68	30.43	33.55	ND	ND
Aeromonas hydrophila	17.48	20.80	24.63	27.53	30.53	33.12	ND	ND
Bacteroides spp. (General marker)	17.54	20.98	24.54	27.55	30.14	33.71	ND	ND
Bacteroides spp. clone CGOF52 (Goose marker)	17.55	21.05	24.28	27.34	29.72	33.56	ND	ND
Bacteroides A3 (Dog marker)	17.16	20.72	24.52	27.69	30.11	33.28	ND	ND
Bacteriodes Cluster 1, PigA4 (Pig marker)	17.40	21.61	24.78	27.60	30.22	34.06	37.88	ND
Campylobacter coli	16.56	21.96	24.81	27.49	30.03	33.26	37.87	ND
Campylobacter jejuni	18.01	21.16	24.56	28.49	30.02	32.46	ND	ND
Catellicoccus marimammalium (Seagull marker)	17.42	21.12	24.04	27.33	30.55	34.64	ND	ND
Enterococcus spp.	17.17	21.06	23.67	27.97	30.22	34.03	ND	ND
Escherichia coli (23S rRNA)	17.42	20.62	23.74	27.25	30.24	33.74	37.18	ND
Escherichia coli (*uidA*)	17.50	20.56	23.20	27.47	31.06	34.50	ND	ND
Escherichia coil O157:H7	18.09	21.14	24.27	27.23	31.54	34.23	ND	ND
Escherichia coli O111	17.40	20.52	24.04	27.81	31.05	33.84	ND	ND
Escherichia coli O26	17.74	21.98	23.55	26.07	30.90	33.11	37.87	ND
Human C40 mitochondria (Human marker)	17.06	21.50	23.41	26.57	30.51	33.89	37.28	ND
Klebsiella pneumoniae	17.58	21.63	24.05	27.60	30.44	32.89	ND	ND
Legionella pneumophila	17.29	20.62	23.37	27.82	31.24	34.25	ND	ND
Listeria monocytogenes	17.39	20.66	23.77	28.03	31.15	34.67	ND	ND
Methanobrevibacter smithii (Human marker)	17.39	20.66	24.77	27.53	31.35	34.72	ND	ND
Pseudomonas aeruginosa	17.08	21.50	24.05	27.07	30.20	33.95	37.60	ND
Salmonella typhimurium	17.09	21.66	24.77	27.37	31.45	34.20	ND	ND
Shigella spp.	17.40	21.18	23.30	27.81	29.41	33.58	ND	ND
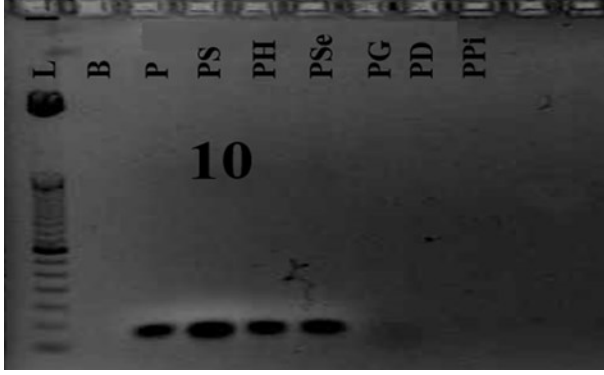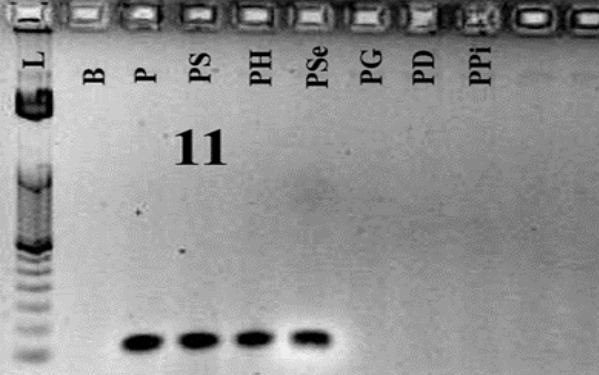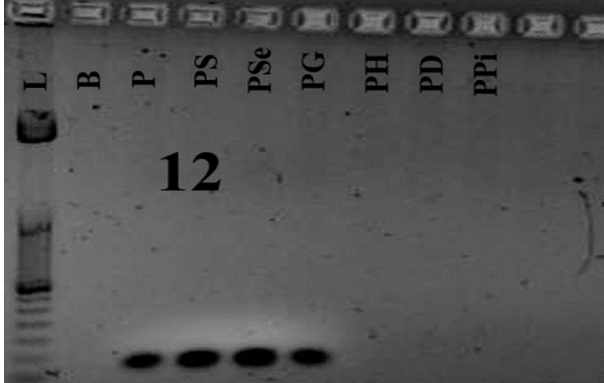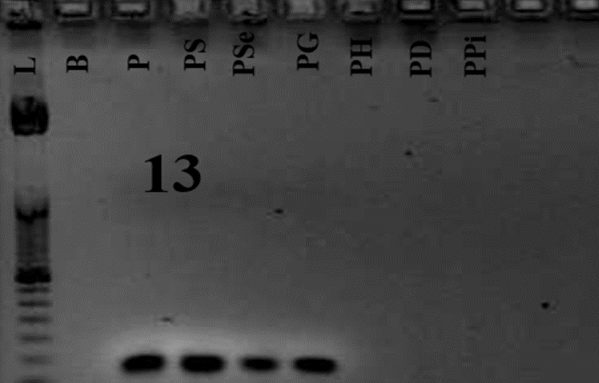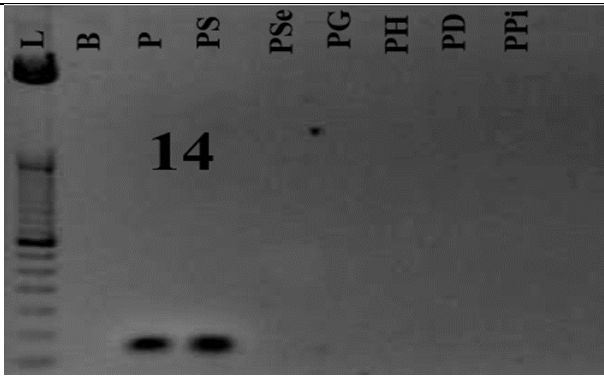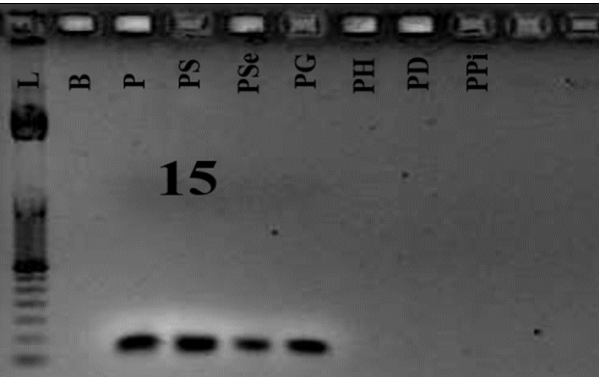
Markers/species	TaqMan® OpenArray® plate Mean C$_T$ value (copies/hole)							
	2,000,000	200,000	20,000	2,000	200	20	2	1
Staphylococcus aureus	17.25	21.77	24.77	27.10	31.05	33.67	ND	ND
Vibrio cholerae	17.19	21.55	23.41	26.57	31.54	34.23	ND	ND
Acinetobacter baumannii	17.81	20.28	23.40	27.16	29.94	33.94	36.10	ND
Aeromonas hydrophila	16.91	21.42	23.75	26.17	29.10	33.17	36.58	37.11
Bacteroides spp. (General marker)	16.91	21.42	23.75	26.17	29.10	33.17	35.58	ND
Bacteroides spp. clone CGOF52 (Goose marker)	16.78	19.09	22.86	26.05	29.07	32.30	35.46	ND
Bacteroides A3 (Dog marker)	16.17	20.73	24.29	27.01	29.11	33.39	37.28	ND
Bacteriodes Cluster 1, PigA4 (Pig marker)	16.46	19.19	22.32	25.22	28.96	31.16	34.20	37.34
Campylobacter coli	16.64	20.50	24.28	27.15	29.21	33.52	36.08	ND
Campylobacter jejuni	17.60	19.72	23.80	27.10	29.14	33.44	36.07	ND
Catellicoccus marimammalium (Seagull marker)	16.96	19.78	22.25	25.12	28.30	32.44	35.58	ND
Enterococcus spp.	16.13	19.73	22.04	25.84	29.10	33.64	37.20	ND
Escherichia coli (23S rRNA)	17.52	19.18	22.11	26.09	29.64	33.02	36.81	ND
Escherichia coli (*uidA*)	16.65	19.61	22.21	25.16	28.63	31.59	34.55	36.23
Escherichia coil O157:H7	17.21	20.44	23.12	26.11	29.65	32.92	35.47	ND
Escherichia coli O111	16.03	19.82	22.95	26.11	29.20	32.95	35.09	ND
Escherichia coli O26	16.19	19.79	22.62	25.13	28.11	31.01	34.79	ND
Human C40 mitochondria (Human marker)	16.22	19.14	23.00	25.10	28.10	32.64	35.70	ND
Klebsiella pneumoniae	16.78	19.09	22.86	26.05	29.07	32.30	35.46	ND
Legionella pneumophila	16.25	19.54	23.60	26.12	29.21	32.47	34.21	37.18
Listeria monocytogenes	16.20	19.89	22.81	26.12	29.10	33.90	36.11	ND
Methanobrevibacter smithii (Human marker)	16.76	19.29	22.75	25.07	28.14	32.66	35.26	ND
Pseudomonas aeruginosa	16.19	20.13	22.99	26.08	29.53	32.06	35.23	37.52
Salmonella typhimurium	16.09	19.78	22.26	26.35	29.81	32.79	35.51	ND
Shigella spp.	16.67	19.58	22.57	25.07	28.81	32.33	35.81	ND
Staphylococcus aureus	16.29	19.68	22.56	25.16	29.55	32.17	35.16	37.88
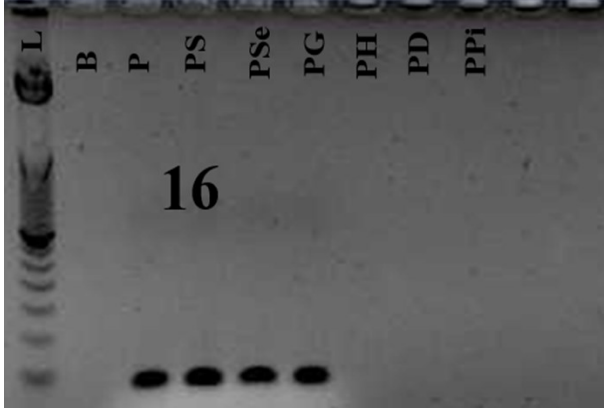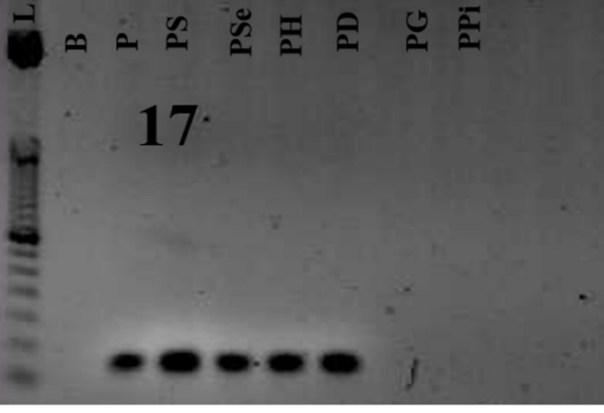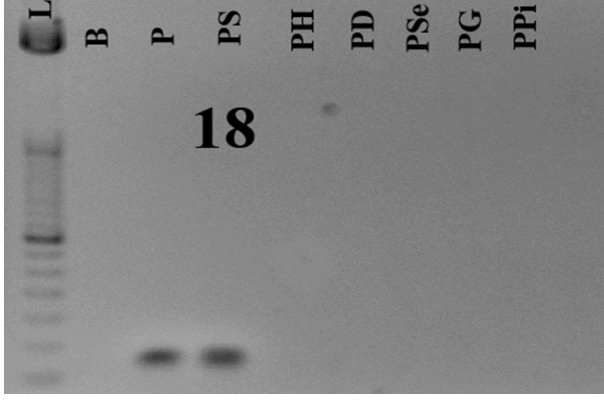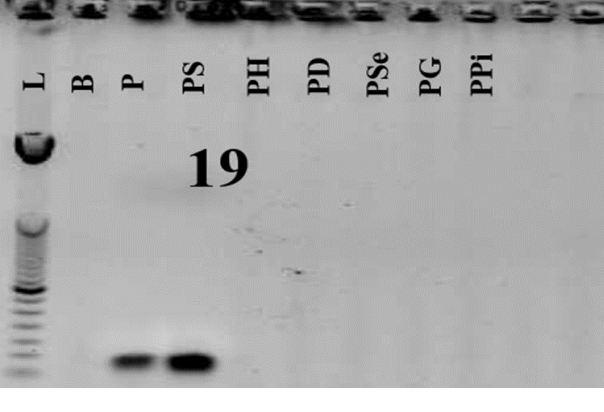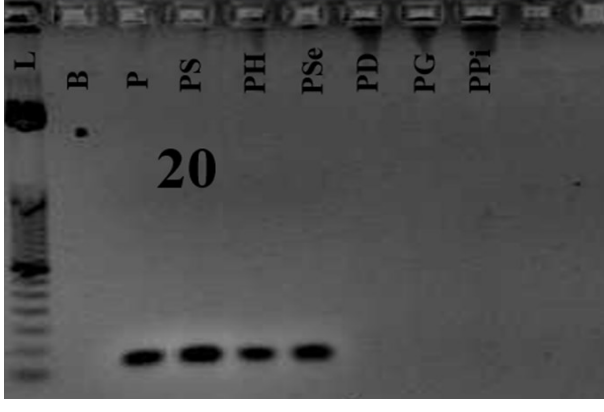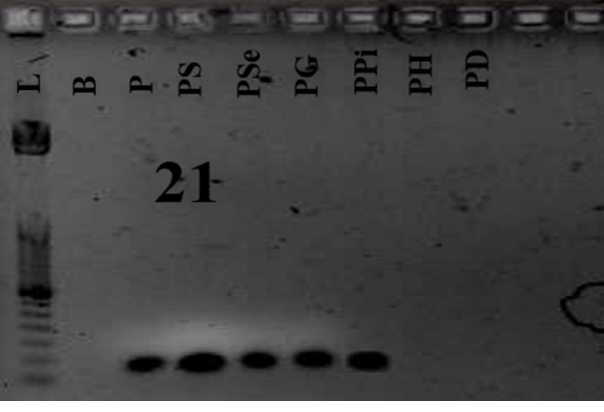
| *Vibrio cholerae* | 16.42 | 19.64 | 22.24 | 25.29 | 29.15 | 32.06 | 35.32 | ND |
| ND; not determined | | | | | | | | |

Species/markers	*gltA*	*lip*	16S rRNA	16S rRNA	*glyA*	*hipo*	16S rRNA	16S rRNA	23 rRNA	*manC*	*uidA*	23 rRNA	*manC*	*manC*	MT-ND2	*phoE*	*mipA*	*hly*	*nifH*	16S rRNA	*regA*	*invA*	*ipaH*	*gyrA*	*ctxA*
Acinetobacter baumannii																									
Aeromonas hydrophila																									
Bacteroides spp.																									
Canada goose *bacteroides* spp.																									
Campylobacter coli																									
Campylobacter jejuni																									
Catellicoccus marimammalium																									
Dog *bacteroides* A3																									
Enterococcus spp.																									
Escherichia coil O157:H7																									
Escherichia coli																									
Escherichia coli O111																									
Escherichia coli O26																									
Human C40 mitochondria																									
Klebsiella pneumoniae																									
Legionella pneumophila																									
Listeria monocytogenes																									
Methanobrevibacter smithii																									
Pig *bacteroides* Cluster 1																									
Pseudomonas aeruginosa																									
Salmonella typhimurium																									

Shigella spp.		
Staphylococcus aureus		
Vibrio cholerae		

Figure S6.1. Summary of the specificity of the designed primers for qPCR assays for targeted genes (X-axis) in this study was verified using conventional PCR for FIBs, MST markers and waterborne pathogens (Y-axis). To evaluate the specificity, genomic DNA from pure bacterial culture (FIBs and waterborne pathogens) and DNA extracted from host-specific fecal material (MST markers) were used. The detection of *Shigella* spp. by *uidA*-targeted qPCR (specific for *E. coli*) was considered as true positive. Filled boxes are negative and white boxes showing target-specific amplification.

Gel 1: 23 rRNA gene amplification of *Enterococcus* spp., P: *Enterococcus faecalis* DSM 20478

Gel 2: *uidA* gene amplification of *Escherichia coli*., P: *E coli* K-12, a: *E. coli* O157:H7 ATCC 43895, b: *E. coli* O26:H11, c: *E. coli* O111, d: *Shigella sonnei* CCUG 68726

Gel 3: 16S rRNA gene amplification of *Bacteroides-Prevotella* spp. a: sewage samples, b: human fecal samples, c: Canada goose fecal samples, d: seagull fecal samples, e: pig fecal samples and f: dog fecal samples

Gel 4: 16S rRNA gene amplification of *Bacteroides* A3; dog marker. D1-5: dog fecal samples	Gel 5: 16S rRNA gene amplification of *Bacteroides* CGOF52; Canada goose marker. G1-5: Canada goose fecal samples
Gel 6: 16S rRNA gene amplification of *Bacteroides* PigA4; pig marker. Pi1-5: pig fecal samples	Gel 7: 16S rRNA gene amplification of *Catellicoccus marimammalium*; seagull marker. Se1-5: seagull fecal samples
Gel 8: C40 mitochondria gene amplification; human marker. S1-3: sewage samples and H: human fecal samples	Gel 9: *nifH* gene amplification of *Methanobrevibacter smithii;* human marker; S1-3: sewage samples and H: human fecal samples

Gel 10: *gltA* gene amplification of *Acinetobacter baumannii*; P: *A. baumannii* DSM 30007

Gel 11: *lip* gene amplification of *Aeromonas hydrophila*; P: *A. hydrophila* DSM 30187

Gel 12: *glyA* gene amplification of *Campylobacter coli*; P: *C. coli* DSM 4689

Gel 13: *hipO* gene amplification of *Campylobacter jejuni*; P: *C. jejuni* DSM 4688

Gel 14: *manC* gene amplification of *E. coil* O157:H7; P: *E. coli* O157:H7 ATCC 43895

Gel 15: *manC* gene amplification of *E. coil* O111; P: *E. coli* O111

Gel 16: *manC* gene amplification of *E. coil* O26; P: *E. coli* O26:H11

Gel 17: *phoE* gene amplification of *Klebsiella pneumoniae*; P: *K. pneumoniae* DSM16358

Gel 18: *mipA* gene amplification of *Legionella pneumophila*; P: *L. pneumophila* DSM 7513

Gel 19: *hly* gene amplification of *Listeria monocytogenes*; P: *L. monocytogenes* DSM 20600
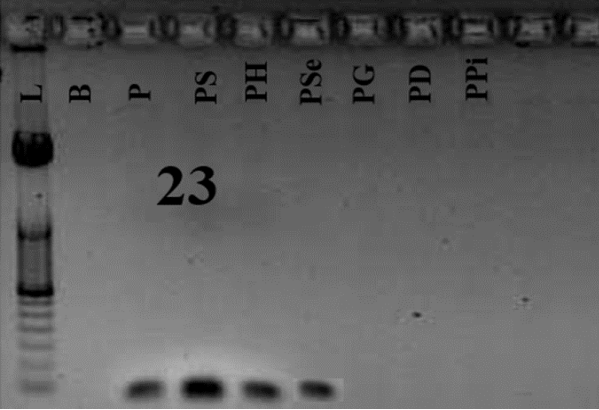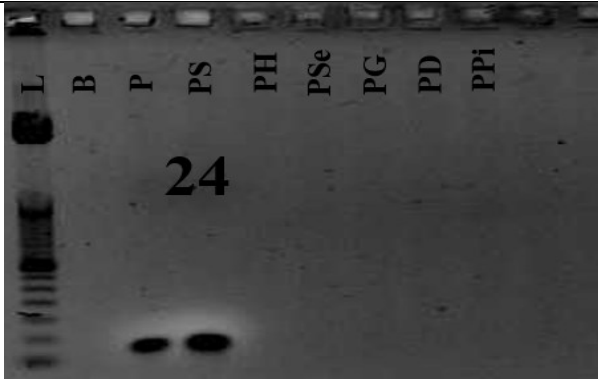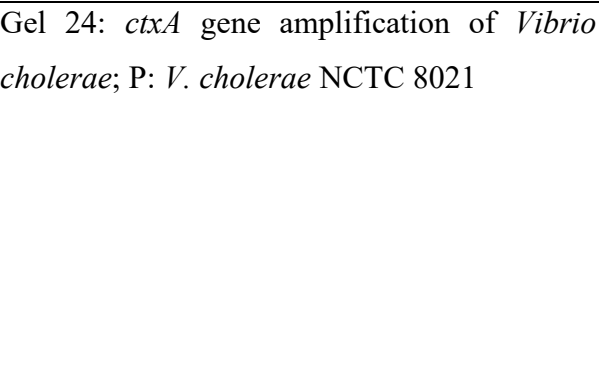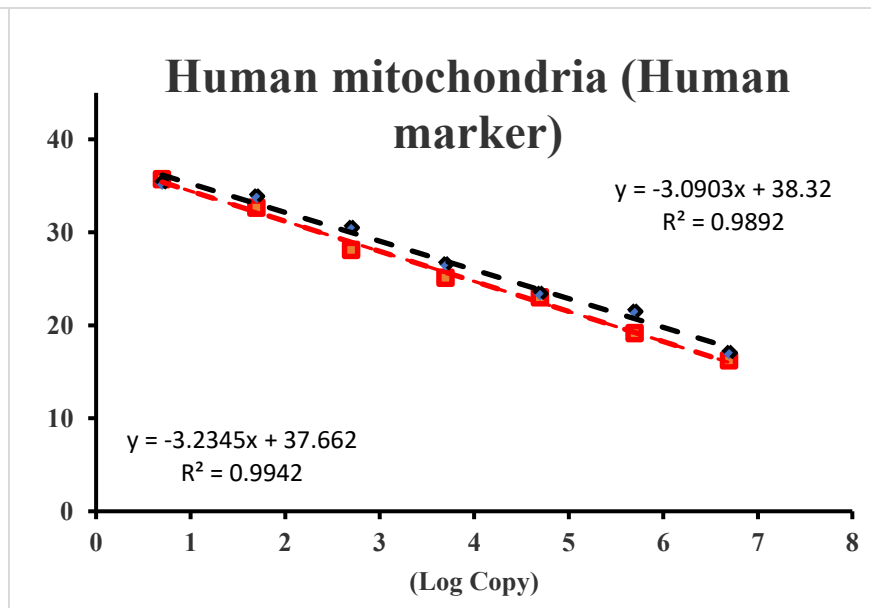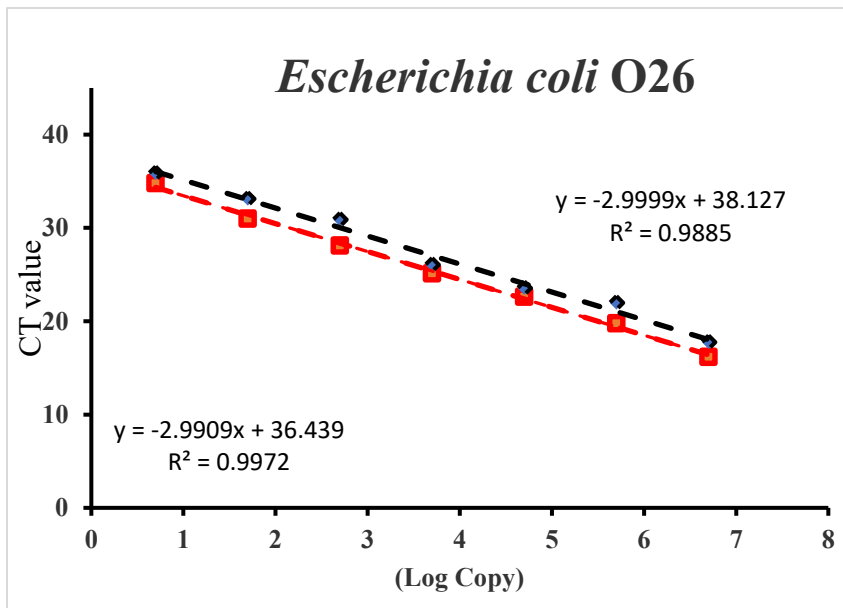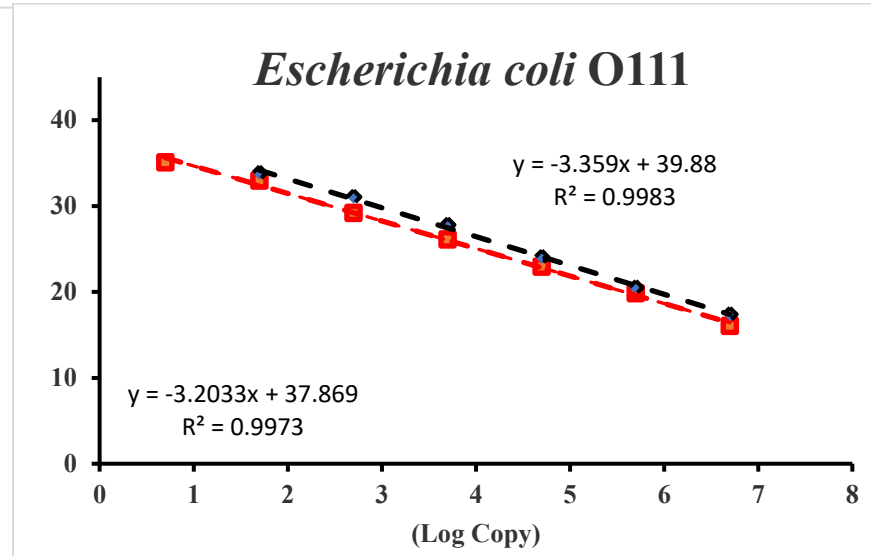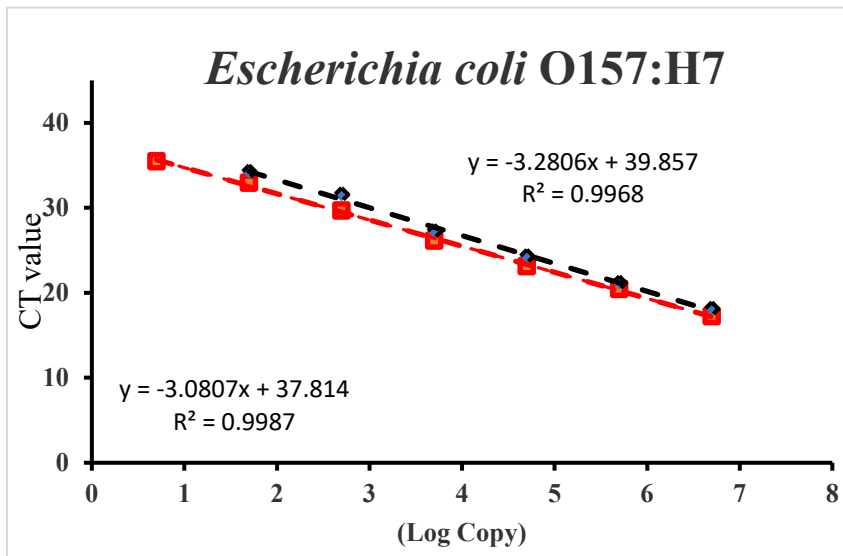
Gel 20: *regA* gene amplification of *Pseudomonas aeruginosa*; P: *P. aeruginosa* DSM 50071	Gel 21: *invA* gene amplification of *Salmonella typhimurium*; P: *S. typhimurium* DSM 17085
Gel 22: *ipaH* gene amplification of *Shigella* spp.; P: *Shigella sonnei* CCUG 68726	Gel 23: *gyrA* gene amplification of *Staphylococcus aureus*; P: *S. aureus* DSM 20231
	Gel 24: *ctxA* gene amplification of *Vibrio cholerae*; P: *V. cholerae* NCTC 8021

Figure S6.2. Specificity assessment of 24 sets of designed primers in this study was examined by conventional PCR trials. The assessment was performed on the pure culture of different bacterial species as well as individual and pooled fecal samples. L; 50 bp Ladder, B; blank (negative control), P; positive control, PS; pooled sewage samples, PH; pooled human fecal samples; PD; pooled dog fecal samples, PC; pooled Canada goose fecal samples, PSe; pooled seagull fecal samples and PPi; pooled pig fecal samples.

Acinetobacter baumannii

y = -3.0267x + 38.755
R² = 0.9984

y = -3.1687x + 38.668
R² = 0.9964

Aeromonas hydrophila

y = -3.1509x + 38.91
R² = 0.9961

y = -3.3741x + 39.039
R² = 0.9937

Bacteroides spp. (General marker)

y = -3.182x + 39.104
R² = 0.9978

y = -3.0304x + 37.794
R² = 0.9927

Bacteroides spp. (Canada goose marker)

y = -3.1183x + 38.679
R² = 0.997

y = -3.1662x + 37.656
R² = 0.9986

Bacteroides spp. (Dog marker)

y = -3.1981x + 39.006
R² = 0.9951

y = -3.3384x + 39.202
R² = 0.9913

Bacteroides spp. (Pig marker)

y = -3.0631x + 38.694
R² = 0.993

y = -3.0957x + 36.925
R² = 0.996

Campylobacter coli

y = -3.0625x + 38.468
R² = 0.986

y = -3.2852x + 39.033
R² = 0.9935

Campylobacter jejuni

y = -2.9362x + 38.115
R² = 0.985

y = -3.1499x + 38.345
R² = 0.9946

Catellicoccus marimammalium (Seagull marker)

y = -3.3625x + 39.968
R² = 0.9979

y = -3.1158x + 37.302
R² = 0.9938

Enterococcus spp.

y = -3.3171x + 39.615
R² = 0.9954

y = -3.5028x + 39.193
R² = 0.9947

Escherichia coli (23S rRNA)

y = -3.1791x + 38.785
R² = 0.9988

y = -3.3234x + 38.631
R² = 0.9913

Escherichia coli (uidA)

y = -3.4497x + 40.199
R² = 0.996

y = -3.0039x + 36.596
R² = 0.9993

Klebsiella pneumoniae

y = -3.0449x + 38.484
R² = 0.9946

y = -3.1662x + 37.656
R² = 0.9986

Legionella pneumophila

y = -3.4601x + 40.293
R² = 0.9967

y = -3.1186x + 37.533
R² = 0.9935

Pseudomonas aeruginosa

y = -3.0936x + 38.507
R² = 0.9928

y = -3.1515x + 37.713
R² = 0.9984

Salmonella typhimurium

y = -3.3565x + 40.184
R² = 0.9939

y = -3.2793x + 38.214
R² = 0.9972

Shigella spp.

y = -3.1458x + 38.655
R² = 0.9895

y = -3.1852x + 37.617
R² = 0.9965

Staphylococcus aureus

y = -3.2081x + 39.406
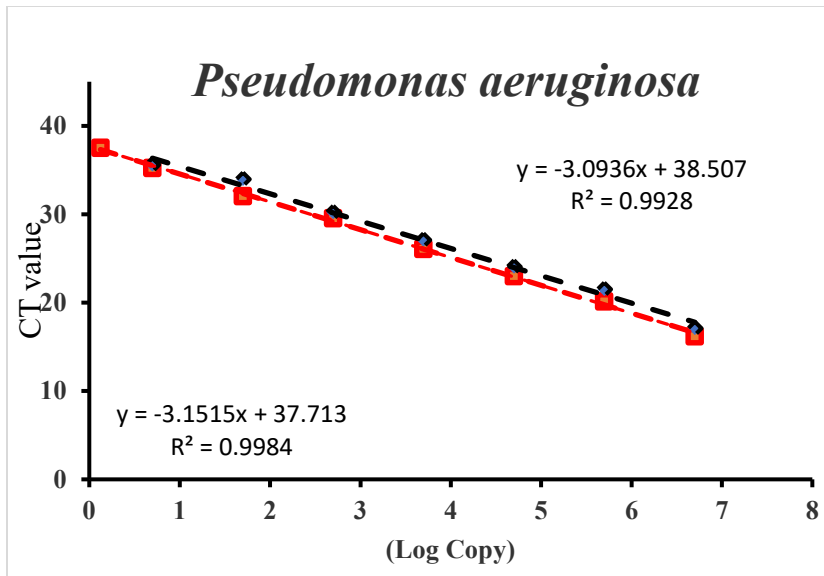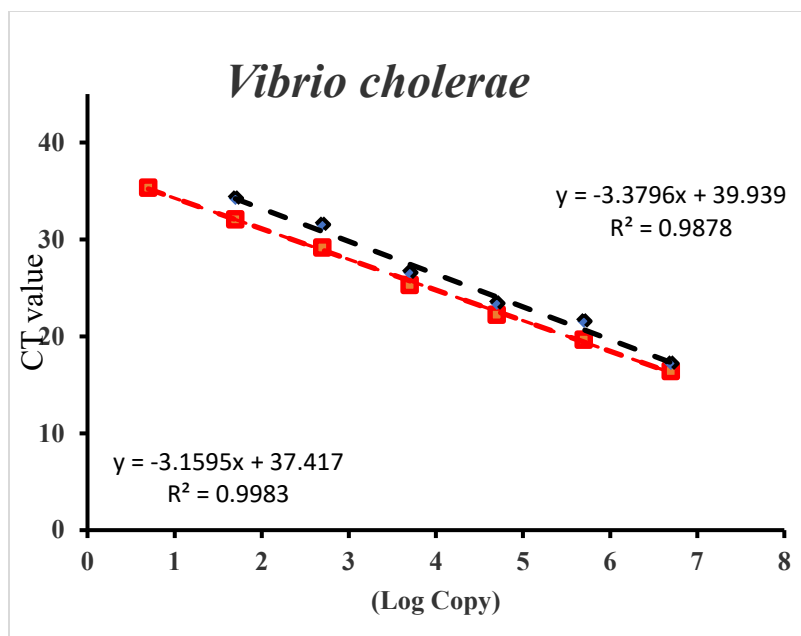R² = 0.9922

y = -2.9909x + 36.439
R² = 0.9972

Figure S6.3. Standard curves were obtained for each target by qPCR (black dotted line) and TaqMan® OpenArray® plate (red dashed line). The standard curves were generated for each targeted waterborne pathogen, FIBs and MST markers by linear regression analysis of C_T mean values obtained by qPCR and TaqMan® OpenArray® vs. the amounts of known copy number of template DNA (log copies/reaction for SYBR green qPCR and log copies/hole for OpenArray® plate). The linear regression equations and goodness-of-fit (R^2) values are also shown for each assay.

VITA AUCTORIS

NAME:	Abdolrazagh Hashemi Shahraki
PLACE OF BIRTH:	Shahrak, Chaharmahal and Bakhtiari Province, Iran
YEAR OF BIRTH:	1979
EDUCATION:	Uromia university, Uromia, Iran, B.Sc. in Laboratory of Sciences, 2002
	Ahvaz Joundishapur University of Medical Science, Ahvaz, Iran, M.Sc. in Medical Microbiology, 2005