2022

Endogenous and exogenous factors driving bacterial community composition in aquatic ecosystems

Javad Sadeghi
University of Windsor

Endogenous and exogenous factors driving bacterial community composition in aquatic ecosystems

By

Javad Sadeghi

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

2022

Endogenous and exogenous factors driving bacterial community composition in aquatic ecosystems

by

Javad Sadeghi

APPROVED BY:

_____

K. Kidd, External Examiner

McMaster University

_____

P. Karpowicz

Department of Biomedical Sciences

_____

T. Pitcher

Great Lakes Institute for Environmental Research

_____

C. Semeniuk

Great Lakes Institute for Environmental Research

_____

S. Chaganti, Co-Advisor

University of Michigan

_____

D. Heath, Co-Advisor

Great Lakes Institute for Environmental Research

April 19, 2022

DECLARATION OF CO-AUTHORSHIP / PREVIOUS PUBLICATION

I. Co-Authorship

I hereby declare that this thesis incorporates material that is the result of joint research, as follows:

Chapters 2-4 were completed under the supervision of my supervisors, Dr. Daniel Heath and Dr. Subba Rao Chaganti. I primarily developed the key ideas, experimental designs, collecting the samples and laboratory work and completed the statistical analyses, interpretation and writing and I am the principle author on all chapters. Drs. Daniel Heath, and Subba Rao Chaganti are co-authors on chapters 2-4. The contribution of the co-authors was primarily through advice on the general research ideas, collaboration in collecting samples, review of results, consultation on statistical analyses, revising the manuscripts and finally in editing the presentation and thesis material. Abdolrazagh Hashemi Shahraki was only co-author for Chapter 2 and his contribution was in collecting samples.

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 one original paper that has been previously published in peer-reviewed journal, as follows:

Thesis Chapters	Authors	Title	Publication Status
2	Javad Sadeghi, Subba Rao Chaganti, Abdolrazagh Hashemi Shahraki, Daniel D Heath	Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes	Published in journal of Science of the Total Environment

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 bacterial community (BC) composition in various habitats, ranging from ecosystems to host anatomy, plays an important role in determining the nature and role of BC function in the ecosystem or host. However, the relative importance of host endogenous and environmental exogenous factors in determining the composition of the BC in aquatic habitats (e.g., freshwater lakes, fish hosts) remains poorly understood. To address this knowledge gap, this thesis makes several contributions to the estimation of the relative effects of endo-exogenous factors in driving the BC composition in aquatic ecosystem. To test the impact of biotic and abiotic factors on aquatic bacterial biodiversity, I collected water samples from sixty southern Ontario lakes and their BC and microbial eukaryotic community (MEC) compositions were determined using high throughput metabarcode sequencing of 16S and 18S rRNA gene fragments. Additionally, I sampled skin and gut BCs belonging to 17 fish species from 11 families (7 orders) at three distinct Laurentian Great Lakes (LGLs) habitats (Detroit River, Lake Erie, Lake Ontario) along with the associated aquatic BCs at those sites. These data allowed me to assess the extent to which host habitat and phylogeny predict gut and skin BC similarity. Finally, to address the effect of host microbiome on gene expression patterns, I manipulated the gut BC in Chinook salmon (Oncorhynchus tshawytscha) families using antibiotic and probiotic treatments (with healthy controls) and assessed host gene expression using transcriptome sequencing (RNA-Seq) on hindgut tissue samples to identify differentially expressed (DE) host genes.

Using a combination of parametric and non-parametric modelling, I showed deterministic processes (exogenous) prevail in shaping BC assembly in freshwater lakes, but that a combination of habitat-specific (e.g., microbial diversity associated with water) and species-specific (e.g., host ancestry, genotype, or diet) factors shape and promote divergence or convergence of the

microbiome BC across host fish species. Additionally, I showed that daily administration of antibiotics and probiotics resulted in significant and predictable changes in fish gut and the surrounding aquatic microbiota. Normal microbiota depletion by antibiotics generally led to downregulation of immune response gene and upregulation of apoptotic processes, while probiotic treatment affected post-translation modification and inflammatory response genes (over-expressed). While these effects were mostly due to microbiome-mediated mechanisms, host-related mechanisms were also detected (i.e., family effects).

In general, my thesis showed that BC composition in fish and lakes is regulated by assembly rules driven by exogenous abiotic and biotic factors (e.g., habitat, geography, microbial biodiversity, diet) and endogenous species-specific related factors (e.g., genetics, physiology, immunity). My work thus supports the deterministic view of BC composition variation across diverse habitats.

DEDICATION

To my family, especially my mother and father. Thank you for getting me here.

April 2022

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors, Dr. Daniel Heath and Dr. Subba Rao Chaganti for their expertise, guidance, assistance and mentorship during the course of my PhD. Their insights and leadership have sparked in me a true sense of excitement with respect to multidisciplinary research and exploratory science. They have provided me with invaluable knowledge and a skillset that goes beyond research itself, and for that I am truly grateful. Thanks Dr. Heath for inviting me to your mushroom farm! I had lots of fun apart from research going to BC for sampling. Thanks for making the environment friendly. Thank you Dr. Chaganti for both your academic guidance as well as guidance in my personal life.

I would like to thank my committee members, Dr. Trevor Pitcher, Dr. Christina Semeniuk, and Dr. Phillip Karpowicz for their wisdom and guidance. You were available whenever I needed help and played a key role in the shaping of this thesis.

Special thanks to my friends Zahra S Taboun, Keta Patel and Razagh Hashemi Shahraki, and Amirhossein Amirsoleymani for their help and support. I had lots of good times with you during my PhD that I will never forget.

The production of this thesis would not have been possible had it not been for the logistical and technical support of many talented individuals from Yellow Island, Quadra Island (British Columbia), and GLIER, Windsor (Ontario). I am greatly appreciative of all the knowledge I have gained from working in the field under the guidance of Drs. John and Ann Heath on Yellow Island. Additionally, I am incredibly thankful to the individuals who were instrumental to the success of my field and lab experiments: Elliott Haugen, Jane Drown, and Earl Heath.

I am grateful for the support of, Sara Jamieson; a person who created an incredibly efficient lab environment for the Heath Research Group. I would also like to thank Shelby Mackie and Jonathon Leblanc for their assistance with next-generation sequencing and OpenArray runs. I am also grateful to our lab manager, Sarah St. Louis. I also would like to thank past and present Heath lab members: Alex Kajtar, Chelsea Frank, Clare Venney, Colin Finerty, Edel Bai, Farwa Zaib, James Watkins, Kory Bertrand, Matt Charron, Mubarak Ziab, Nabeelah Lulat, and Shelby Wright. Finally, I am thankful to all the GLIER staff for their incredible patience and support over the years: Mary Lou Scratch, Christine Wiesner, Kendra Thompson-Kumar, Nia Khuong, and Tracie Coates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

can modulate the microbiome and the microbiome will affect the host's phenotype but note the indirect effect of the host's genome (chapters 3 & 4). C (green lines): another mechanism for three-way environment-host-microbiome interactions affecting host gene expression which results in changes in host phenotype (as shown in chapters 2,3,4). Probably C is the main way that the environment-host-microbiome interactions can change host gene expression patterns……

LIST OF APPENDICES

LIST OF ABBREVIATIONS/SYMBOLS

AIMD	Antibiotic-induced microbiome depletion
ASV	Amplicon Sequence Variant
BC	Bacterial communities
BCC	Bacterial community composition
BH FDR	Benjamini-Hochberg false-discovery rate
BLAST	The Basic Local Alignment Search Tool
CAP	Chloramphenicol
cDNA	Complementary DNA
CFU	Colony forming units
cox*1*	Cytochrome c oxidase I
Cq	Quantification cycle
CT	Cycle threshold
cytb	Cytochrome b
distLM	Distance-based linear model
dsDNA	Double-stranded deoxyribonucleic acid
eDNA	Environmental DNA
FC	Fold change
FDR	False discovery rate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GLM	Generalized liner model
GLIER	Great Lakes Institute for Environmental Research
GLMM	Generalised linear mixed model
HGT	Horizontal gene transfer
HTMS	Throughput metabarcode sequencing
KEGG	Kyoto Encyclopedia of Genes and Genomes
KW	Kruskal-Wallis
LCT	Lactose
LDA	Linear discriminant analysis
LGLs	Laurentian Great Lakes
LMM	Linear mixed model

MEC	Microbial eukaryotic community
N	Nitrogen
NGS	Next-generation sequencing
NJ	Neighbor-joining
NMDS	Non-metric multidimensional scaling
OMNRF	Ontario Ministry of Natural Resources and Forestry
OTC	Oxytetracycline
OTU	Operational taxonomic unit
PCoA	Principal coordinates analysis
PERMANOVA	Permutational multivariate analysis of variance
QIIME	Quantitative Insights into Microbial Ecology
qPCR	Quantitative polymerase chain r eaction
qRT-PCR	Quantitative real time PCR
RDP	Ribosomal database project
RIN	RNA integrity numbers
RNAseq	RNA sequencing
SAR	Stramenopiles, Alveolata, Rhizaria
SNP	Single-nucleotide polymorphism
SparCC	Sparse Correlations for Compositional
SPRI	Solid Phase Reversible Immobilization
SRA	Short Read Archive
SSU rRNA	Small subunit ribosomal ribonucleic acid
TP	Total phosphorus
WHO	World Health Organization

CHAPTER 1: GENERAL INTRODUCTION

1.1 Introduction

Ecosystems are the functional units composed of the components of biodiversity (animals, plants, fungi, and microorganisms) in a particular area interacting with each other and their abiotic environments (Levin, 2013). Generally, community ecology focuses on patterns of species diversity, abundance, and composition in ecosystems, and of the processes causing these relationships (Vellend, 2010). Biodiversity, a measure of the diversity of life in ecosystems, is a principal topic in ecology, because substantial loss (or change) of biodiversity could alter the function of the ecosystem and hence the habitat stability and benefits provided (Isbell et al., 2017). Yet, despite the diversity of mechanisms thought to shape patterns in ecological communities, all such mechanisms include only four processes: selection (difference in fitness among individuals of different species), drift (random changes in species abundances), speciation (the creation of new species), and dispersal (movement of organisms across space) (Hanson et al., 2012; Mallott and Amato, 2021; Vellend, 2010). While all ecosystems are affected by similar processes, the mechanisms driving microbial ecology are perhaps more challenging to describe due to logistics associated with their complexity and microscopic nature (Prosser et al., 2007). Deterministic (niche-base) and stochastic (neutral) processes are potential classes of mechanisms underpinning microbial community biogeography (Sadeghi et al., 2021). The impacts of abiotic (e.g., spatial effects, pH, nutrients) and biotic (e.g., cooperation, competition, predation) variables are generally defined as deterministic processes (Zhou and Ning, 2017). On the other hand, stochastic processes comprise the neutral theory, in which all taxa are expected to be functionally equivalent and not subject to strong environmental effects (Oliphant et al., 2019). In recent years, numerous facets of microbial biodiversity have been effectively studied across space, time, and ecological gradients

from various perspectives. Nonetheless, the mechanisms behind the complex and varied diversity patterns remain indistinct and controversial (Zhou and Ning, 2017).

Microorganisms are the most diverse group of life on our planet, living in almost every conceivable environment. However, different types of environments harbour strikingly different microbial communities. Studies have shown that salinity (Hou et al., 2017), pH (Banda et al., 2021), seasonality (Sun et al., 2017), and ecological interactions (Steele et al., 2011) are major factors determining the composition of microbial communities. Microorganisms not only exist in environmental habitats, but they are also found in micro-niches associated with individual eukaryotic hosts. A "microbiome" is defined as the microbial community occupying a reasonable well-defined habitat which has distinct physio-chemical properties (Berg et al., 2020). All existing organisms living today evolved from a single common ancestor which existed more than 1.2 billion years ago (Ley et al., 2008; Ros-Rocher et al., 2021). Consequently, this long history of interaction among the microbial communities and multicellular organisms shaped microbial communities as well as the evolution of vertebrates (Ley et al., 2008). This interaction resulted in specific, and sometimes obligate, associations with eukaryote hosts, ranging from insects (Moran et al., 2008) to primates (Yildirim et al., 2010). The majority of microbiome studies in vertebrates have focused on mammals, which encompass fewer than 10% of total vertebrate diversity (Sullam et al., 2012). Far fewer studies have focussed on fish, which consist of more than 32 000 species, originated over 600 million years ago and constitute almost half the total number of vertebrate species (Sullam et al., 2012; Zhang et al., 2014). Moreover, over three billion people every day depend on fish for at least 20% of their protein intake, with a global per capita consumption of almost 20 kg per year of fish (FAO, 2016). Given the evolutionary, ecological and social value of fish, it is

imperative to characterize the complex relationships between fish, the water they inhabit and the microbial communities around and within them.

1.1.1 Fish microbiome

Fish and their microbiomes exhibit a mutualistic relationship (Sehnal et al., 2021b) in which the microbial communities associated with the host have been shown to be involved in the host's metabolism and immunity, among other functions. In return, the host supports the colonization and nutritional needs of both surface and internal microbiota (Lescak and Milligan-Myhre, 2017a; Sehnal et al., 2021b). The microbiome has been referred to as a distinct organ because of its production of various vital molecules such as short-chain fatty acids (SCFAs) which can promote intestinal integrity in the host (Langlois et al, 2021). The metagenome of the human microbiome has even been named our 'second genome' (Zhu et al., 2010). These organisms affect genetic, metabolic, and immunologic functions that help normal host development and support general host health (Belkaid and Hand, 2014). The earliest study of fish-hosted microbial communities dates back to the late 1920s (Reed and Spence, 1929). Ever since that revolutionary paper, much effort has been devoted to characterizing the microbial communities associated with fish. Although an active research field, the structure and function of fish microbiome has not been characterized in depth, likely due to the complexity of the fish microbiomes. This limitation narrows the potential application of microbiome manipulation (and associated methods) in fish culture and commercial aquaculture (Merrifield and Ringo, 2014). However, with the advent of advanced molecular genetic techniques such as High Throughput Sequencing (HTS), we now have a better picture of the taxonomic diversity (structure) and dynamics of the microbial community in a range of fish species (Doane et al., 2020; Emie et al., 2021; Uren Webster et al., 2020; Xiao et al., 2021). In particular, metabarcoding using short fragment amplicon sequencing methods (e.g.

16S and 18S ribosomal RNA (rRNA) gene sequencing) has become the gold standard when evaluating the composition and diversity of microbiomes (Huse et al., 2012). Unlike culture-based methods, metabarcoding studies (16S rRNA) were able to show the true diversity of the fish microbiome; however, it was still found to be dominated by bacteria belonging to the Proteobacteria phylum (Ghanbari et al., 2015). Tissue-specific microbial communities have also been reported in fish (Legrand et al., 2020). Specifically, fish harbour different microbial communities across their anatomy (i.e. gut, gills and skin) (Krotman et al., 2020; Legrand et al., 2018). Although not yet reported, other body sites such as the eyes, buccal cavity, and urogenital opening no doubt have specialized microbiota; indeed, microbiomes associated urogenital region may provide a mechanism for vertical transmission of microbiome from dam to the eggs (Legrand et al., 2020).

Gut: Similar to mammals, fish have diverse intestinal microbiota that help in nutrient absorption, immune response, gut epithelial repair and development, and metabolism (Gomez et al., 2013; Lescak and Milligan-Myhre, 2017b; Maynard et al., 2012). Microbial colonisation of the fish gut mostly originates at the egg stage, with the environment (e.g. the surrounding water and the diet) (Egerton et al., 2018) and maternal transmission contributing (Sylvain and Derome, 2017). As soon as eggs emerge from the mother, the microbiota of the surrounding water as well as maternal microbiota come in contact with the eggs and subsequently have the chance to colonise the surface (Egerton et al., 2018; Gomez et al., 2013; Langlois et al., 2021; Llewellyn et al., 2014). Colonization of fish gut by surrounding water microbiota in combination with fish diet starts after hatching – likely the source of the first colonisers of the developing gastrointestinal tract (Colston and Jackson, 2016; Egerton et al., 2018; Gomez et al., 2013). As fish grow, the gut microbiota becomes further diversified and the fish gut microbiome achieves a complex assemblage of gut

associated microbes (Nayak, 2010). Over five hundred different bacteria species belonging to diverse phyla are reported to occupy the fish gut, mainly dominated by aerobes or facultative anaerobes (Romero and Navarrete, 2006; Talwar et al., 2018).

Skin: Skin surfaces of animals, especially in fish, provide an essential primary barrier against opportunistic pathogens while harbouring a diverse community of commensal microbes. Fish skin is covered in mucus that provides a barrier between the host and the surrounding water microbiota (Sehnal et al., 2021). Skin mucus is an important part of the fish immune system, and holds various immunoglobulins, antimicrobial peptides, mucins, and other mucosal products that protect the fish from pathogens (Gomez et al., 2013). Other biochemical products associated with fish skin include defensins, lysozymes and lectin-like agglutinins that also help in the innate immune response against pathogens (Guardiola et al., 2014). However, some co-evolved mutualistic and commensal microbes can use the mucus as an adhesion site and evade the primary defence mechanisms of the host (Ringo and Holzapfel, 2000). Unlike the gut microbiota, our knowledge of the even the composition of the skin microbiome in fish remains limited (Chiarello et al., 2018). Studies have shown that skin mucus can hosts a diverse community of commensal microorganisms, mostly bacteria (Krotman et al., 2020) but also fungi (Egerton et al., 2018). As the fish skin microbiota play a significant role in both adaptive and innate immunity, it is not surprizing that there are differences in both composition and diversity among host species and environments (Sehnal et al., 2021b). For example, a recent study showed that moving from fresh water to salt water shifted the skin-associated microbiota of Atlantic salmon (*Salmo salar*)(Lokesh and Kiron, 2016).

As sampling the skin microbiome is less invasive than gut microbiome sampling, requiring only a skin swab, the skin microbial community may be an important biomarker of fish health in

wild and/or protected species (Sehnal et al., 2021b). Moreover, studies have shown that the skin microbiota is more affected by the environment than the gut microbiota, perhaps making it an even better predictor of fish health (Sylvain et al., 2020). Taken together, each tissue has a unique microbial community which plays important roles in host development and tissue-specific physiology which is impacted by numerous factors including host diet, habitat use, and genetic background.

1.1.2 Microbial community of freshwater lakes

Bacteria as well other microorganisms play important roles in transforming nutrients and reintroducing them into the food web (Shahraki et al., 2021). Microorganisms are also involved in carbon cycling through microbial loops, and they can act as a food source for other organisms such as zooplankton (Buchan et al., 2014). Interactions between the bacteria and phytoplankton can be complex. For example, bacteria can support and promote the growth of phytoplankton via the recycling of nutrients, but at the same time, they also compete with phytoplankton and other organisms for essential nutrients (Buchan et al., 2014). Characterizing the factors that can change and govern bacterial community composition (BCC) can offer deeper insight into the processes and mechanisms operating in lake ecosystems and ultimately improve our basic knowledge of the microbial community and their interactions with other organisms at higher trophic level such as fish.

1.1.3 Factors affecting fish microbiome

Several drivers (Figure 1.1) of microbiome diversity have been reported for fish, including; host effects (e.g. genetics, gender, weight, age, vertical transmission of maternal microbiota, circadian rhythms, immunity, and intestinal motility) (Liu et al., 2021; Mallott and Amato, 2021; Xiao et al., 2021; Thaiss et al., 2016) environmental and dietary effects (e.g. water chemistry and

quality, diet) (Minich et al., 2020; Razak et al., 2019) and microbial characteristic effects (e.g. adhesion capacity, enzymes and metabolic capacity) (Prakash et al., 2011) and ecological interaction (He et al., 2018). While all of those factors contribute to the composition and diversity (and hence function) of the host's microbiome, this thesis focusses on two key factors: 1) environmental effects, and 2) host genomic effects.

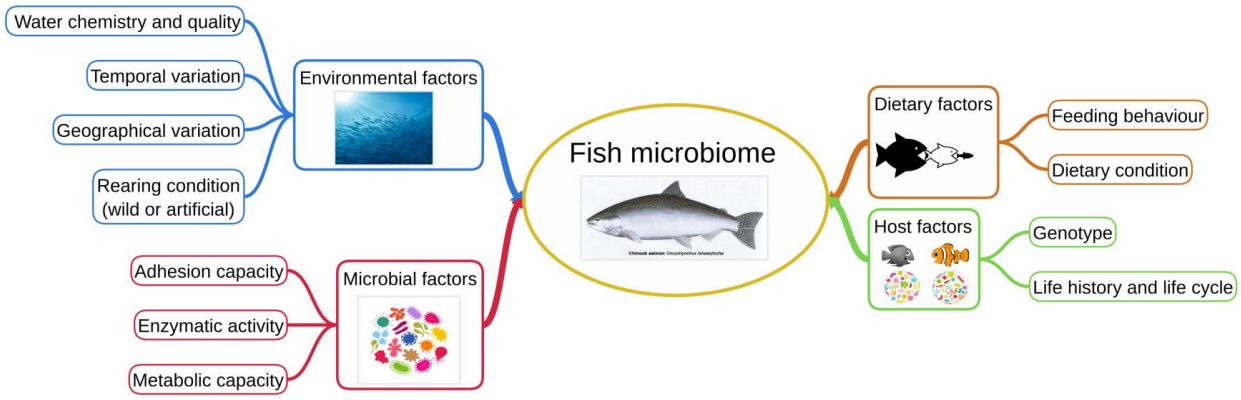


Figure 1.1. Schematic diagram showing biotic and abiotic factors affecting the fish microbiome

Environmental factors: Although microorganisms are found across all habitats, different types of environments hold remarkably distinct microbial communities. Environmental factors such as water chemistry and quality, salinity, season, and geospatial variation can deeply influence the composition of free-living and symbiotic microbial communities (Tarnecki et al., 2017). Studies have shown that water chemistry and quality have significant impacts on aquatic microbial communities (Bledsoe et al., 2016; Sadeghi et al., 2021). Specifically, nitrogen, phosphorus, dissolved and particulate organic matter, high ammonia concentrations, and suboptimal pH and salinity have all been shown to perturb the fish microbiota and may lead to compositional imbalances that pose a risk to fish health (Giatsis et al., 2015; Sylvain et al., 2016; Zhang et al., 2016). Temporal, and geographical factors can also affect the skin and gut microbial composition (Oh et al., 2016; Suzuki and Worobey, 2014; Woo et al., 2017). Skin microbiota have been shown

to be more sensitive to environmental factors than the gut microbiota, suggesting that the gut habitat is able to moderate community composition variation, despite environmental fluctuations (Dulski et al., 2020).

Trophic level and diet: Diet has been identified as a key factor affecting the diversity and community structure of not only fish gut but also skin microbiomes (Chiarello et al., 2018; Smith et al., 2015). One reason for this could be fishes having distinct diets might produce different surface mucus and this might affect their skin microbiome (Chiarello et al., 2018). Moreover, fasting in fish is known to substantially change intestinal microbiota composition (Wang et al., 2018; Xia et al., 2014). However, the effect of natural variation in diet, particularly when associated with host development (ontogenetic), on host-microbiota interactions is largely unknown (Leeming et al., 2019). Food additive such as antibiotics, prebiotics and probiotics have been used in aquaculture to improve grow rate and fish health (Hoseinifar et al., 2016; Navarrete et al., 2008; Yassir et al., 2002) and all are expected to affect microbiome composition. Prebiotics are non-living substrates that provide nutrients for resident microorganisms harboured by the host and can selectively increase the abundance of beneficial microbes in the gut. Moreover, probiotics are live bacteria that are applied used orally or through the water column and have quantifiable health benefits to the host through increase food digestibility by increasing different digestive enzymes such as proteases, alginate lyases, and amylases (Hoseinifar et al., 2018). Antibiotics have been used in aquaculture to prevent and treat bacterial diseases, however, there is a great chance that large proportion of the antibiotics to enter the environment. Moreover, using antibiotics frequently can raise antimicrobial resistance in fish farms (Miranda et al., 2018). Although prebiotics have many beneficial impacts on heath and performance as proved for several terrestrial animals, the use of prebiotics in aquaculture has been less investigated (Akhter et al., 2015).

Moreover, recent studies (Fan et al., 2020; Mallott and Amato, 2021) provided evidence that apart from environmental parameters other factors such as host-related genetic factors can change the microbiome structure.

The influence of host genome: Although the composition of both the gut and skin microbiomes in fish is affected by environmental and feeding-related factors, there is also strong evidence for a host genetic component in shaping microbial communities (Blekhman et al., 2015; He et al., 2018; Mallott and Amato, 2021; Uren Webster et al., 2018). In the vertebrate gut, microbial communities play vital physiological roles, influencing metabolic processes such as the breakdown of complex carbohydrates, the regulation of fat storage (i.e. production of short chain fatty acids by gut bacteria), and providing vital vitamin and amino acids to their host (Sullam et al., 2012). These key functions contribute to microbiomes being key factors in host adaptation, fitness and ultimately evolution (Xiong et al., 2019). Host genetic diversity (within and among populations) is believed to, in turn, impact microbial community composition (Tarnecki et al., 2017). Genetic differences among fish species, among populations within a species and among individuals within populations, can drive variation in immune response, metabolism, behaviour, and gene expression of the host, all of which are expected to change the structure of microbiome (Grieneisen et al., 2020; Nichols and Davenport, 2021; Sehnal et al., 2021b). Extensive research in teleosts as well as in humans has shown that, while the gastrointestinal (GI) microbiota is extremely variable from individual to individual (Boutin et al., 2014), family members (those that have similar genetic background) tend to have more similar microbiota than unrelated individuals (Spor et al., 2011; Steury et al., 2019). Indeed, the same microbiome can be shared between adult family members (Spor et al., 2011). However, the evidence for environmental effects on GI microbiome are substantially better characterized and documented than the effects of host genetic

background on the vital ‘microbial organ’, or microbiome (Dethlefsen et al., 2006; Ley et al., 2006).

1.1.4 Genomic approaches to studying the fish microbiome

Microbial community characterization is now undergoing a renaissance as high throughput sequencing techniques such as metabarcoding and metagenomics allow deep insight into hundreds or thousands of microbial taxa within a single sample (Feehery et al., 2013; Hamady and Knight, 2009). These studies are made possible by the finding that small fragments of ribosomal (16S rRNA and 18S rRNA) genes are satisfactory as a proxy for the full-length sequence for determining community composition (Janda and Abbott, 2007). 16S rRNA and 18S rRNA genes have conserved and variable regions, and the conserved regions can be used to design “universal” PCR primers while the amplified variable regions between the conserved primer sequences provide phylogenetic information allowing assignment to specific bacterial taxa (Ghanbari et al., 2015; Johny et al., 2021). Two widely used methods for taxonomic characterization of microbial communities are metagenomics and metabarcoding. Metagenomics is used to recreate whole organism genomes present in a sample, by employing both taxonomic and functional analytical methodologies (Zepeda Mendoza et al., 2015) On the other hand, DNA metabarcoding primarily focuses on identifying which species are present in a DNA sample. Metabarcoding is one of the most cost effective and efficient high-throughput sequencing methods that enable sequencing of multiple species present in individual environmental samples at one time. With the broad application of metabarcoding in microbial ecology and evolution, microbial ecologists can now characterize diverse and complex microbial community composition and infer putative function.

1.1.5 Host microbiome and gene expression interactions

Due to the contribution of microbiomes to host fitness, and the effects the host can have on the microbiome composition, the microbiome as a community has co-evolved with the host and *vice-versa* (Figure 1.2) (Foster et al., 2017). Thus, selection pressures on the host could, in theory, result in changes in the composition of the microbiota to alter their function to maximise benefits to the host (Foster et al., 2017), contributing to the host's adaptive response. If the microbial community function enhances host fitness, the result would assure the availability of host habitat for the microbiota over the longer term (Goodrich et al., 2017). Moreover, some host species display behavioral or other traits that ensure effective microbiome transfer to the next generation. Mechanisms for selecting, retaining, and transferring key elements of the microbiome are likely to be genetically encoded in the host, and the characterization of those genetic components will point to mechanisms underlying the evolution and function of host–microbe symbioses (Ley et al., 2006). All animals face the important challenge of building and maintaining diverse tissue function while maintaining sensitive and adaptive responses to their environment. This balance is most noticeable in the intestinal epithelium, which has significant roles in nutrient absorption / toxicant barrier function as well as immune response, while being continually exposed to complex microbial communities inside the intestinal lumen (Zheng et al., 2020). Although host-microbial interactions are widely recognized as valuable (Dabrowska and Witkiewicz, 2016), their regulatory molecular mechanisms are not well understood, especially in aquatic organismal hosts. Advanced molecular genetic-based technologies have permitted researchers to characterize tissue-specific host transcriptomes, which is important for determining not only the functional components of the host genome, but also how the host genome responds to environmental challenges. Transcriptomic studies also provide mechanistic insight into tissue development and regeneration as well as disease state (Ghanbari et al., 2015; Nichols and Davenport, 2021). The ability of intestinal

epithelial cells to maintain their physiological functions and respond properly to pathogens is enabled through regulation of gene expression, with transcription being the first, and rate limiting, step of that process. Genome-wide comparisons of transcript levels in intestinal epithelial cells from germ-free mice relative to mice harbouring a functional microbiome have shown hundreds of genes (e.g., genes involved in nucleotide metabolism and cell-cycle pathways) that have meaningfully differentially expressed mRNA levels (Thaiss et al., 2016). Interestingly, several of the mouse genes that appear to be transcriptionally controlled by the gut microbiome have zebrafish homologs, suggesting the existence of evolutionarily-conserved mechanisms (Davison et al., 2017).

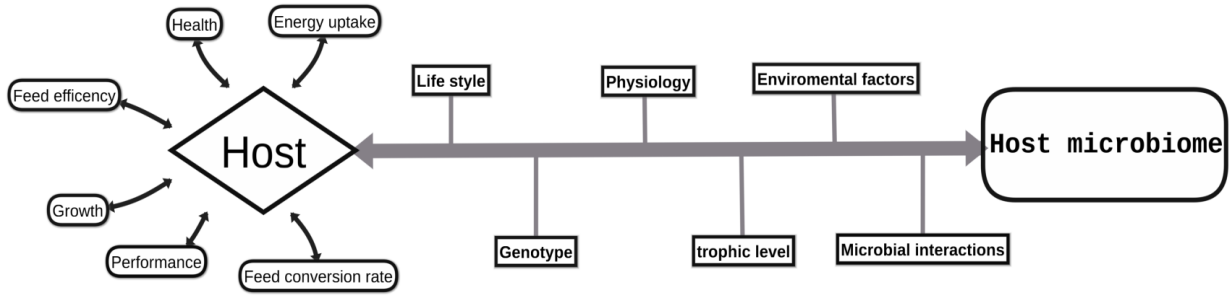


Figure 1.2. A schematic diagram showing the complex interactions between the host, it's environment and it's microbiome. A combination of biotic and abiotic factors such as genotype, fish physiological status (innate and adaptive immune systems), fish lifestyle (including diet), fish environment and microbial interactions affect the host microbiome composition, which will cause changes in host function and metabolic activities. These changes affect processes involved in growth, performance, energy storage and health in fish.

1.1.6 Thesis objectives

The main goal of this thesis is to characterize factors that shape and drive variation in natural microbiomes: both within the host (fish) and in it's aquatic environment. The specific objectives that address my main goal comprise three data chapters:

In Chapter 2, I investigate the bacterial and micro-eukaryote community composition and dynamics in southern Ontario lakes to determine which environmental factors shape the structure

and function of aquatic bacterial communities. More specifically, water samples as well as environmental parameters from sixty southern Ontario lakes were collected, and the bacterial and microbial eukaryotic community (BC and MEC) compositions were determined using high throughput metabarcode sequencing of 16S rRNA and 18S rRNA gene fragments. Moreover, I was interested in testing hypotheses about the relative contribution of deterministic versus stochastic processes as well as the contribution of biotic and abiotic factors in the assembly of bacterial communities of freshwater lakes. My hypotheses are deterministic biotic and abiotic factors prevail in shaping BC assembly in freshwater lakes, and within the deterministic factors, biotic factors will dominate abiotic factors in shaping the BCs. Lakes are excellent systems for investigating microorganism community dynamics because they have clear boundaries within lakes and strong environmental gradients among lakes. Understanding the drivers and controls of microbial (bacterial) communities will improve our knowledge of fundamental properties of bacterial communities and thus enhance our ability to predict community states and response to change.

In Chapter 3, I assess the diversity and taxonomic composition of skin, gut, and water microbial communities across 17 fish species sampled in three Great Lakes habitats. The relative importance of exogenous abiotic and biotic factors (e.g., habitat, geography, microbial biodiversity, diet) and endogenous host-related factors (e.g., genetics, physiology, immunity) in driving the composition of the fish microbiome remains poorly understood. I evaluate how environmental factors (water microbiome) as well as fish species identity can affect the gut and skin microbiomes across diverse fish species. Although it is well known that the intestinal microbiome can have an enormous impact on fish health, it is vital to determine how it varies among diverse hosts when variation in the aquatic environment is controlled for. This study will

improve our knowledge of variables that can affect fish microbiome hence fish fitness and health. Moreover, the inclusion of both the skin and gut microbiome in this study will further our knowledge of how skin microbiomes vary in natural fish populations. Ultimately, evaluating fish microbiome variation as a fish health biomarker using skin swabs may help fish conservation and management by providing a rapid, on-invasive method to evaluate fish stock status and health.

In Chapter 4, I treated Chinook salmon (*Oncorhynchus tshawytscha*) fry with probiotics and antibiotics to measure the range of effects an altered gut microbiome would have on the host. I characterized host response to microbiome manipulation through gut tissue gene expression evaluated using transcriptome sequencing (RNA-Seq) and quantitative real-time PCR at 50 candidate genes. I also performed 16S rRNA sequencing to characterize the microbiome of the gut as well as the water that fish were reared in. My hypothesis was that by altering the fish gut microbiome, host gene expression, especially the expression of key immune responses in the GI tract, will change and these changes will have both beneficial (probiotic) or deleterious (antibiotic). The results from this chapter will shed light on how microbial communities can alter host function through changes in gene expression patterns and thus characterize the mechanisms of microbiome effects on host performance generally.

In all data chapters, I apply genetic, ecological and evolutionary theory to characterize the nature and extent of the forces driving aquatic as well as host-associated microbiome composition using advanced molecular genetic techniques (transcriptomics, metabarcoding, high-throughput qRT-PCR). By testing for abiotic and abiotic factors driving bacterial community composition, I determined that deterministic and stochastic processes are shaping the BC composition, and hence function, across a broad array of temperate freshwater lakes as well as fish species. Moreover, I also demonstrated that habitat-specific factors (e.g., aquatic microbial communities, geographical

variation) along with species-specific (e.g., host ancestry, genotype, or diet) promotes divergence or convergence of fish microbiome. Finally, I was able to explore how the microbiome co-evolved with their host, and how this bidirectional interaction is contributing to host phenotype, and ultimately health and fitness.

1.2 References

Akhter N, Wu B, Memon AM, Mohsin M. Probiotics and prebiotics associated with aquaculture: A review. Fish Shellfish Immunol 2015; 45: 733-41.

Banda JF, Zhang Q, Ma L, Pei L, Du Z, Hao C, et al. Both pH and salinity shape the microbial communities of the lakes in Badain Jaran Desert, NW China. Sci Total Environ 2021; 791: 148108.

Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell 2014; 157: 121-41.

Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome 2020; 8: 103.

Bledsoe JW, Peterson BC, Swanson KS, Small BC. Ontogenetic Characterization of the Intestinal Microbiota of Channel Catfish through 16S rRNA Gene Sequencing Reveals Insights on Temporal Shifts and the Influence of Environmental Microbes. PLoS One 2016; 11: e0166379.

Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. Genome Biol 2015; 16: 191.

Boutin S, Sauvage C, Bernatchez L, Audet C, Derome N. Inter individual variations of the fish skin microbiota: host genetics basis of mutualism? PLoS One 2014; 9: e102649.

Buchan A, LeCleir GR, Gulvik CA, Gonzalez JM. Master recyclers: features and functions of bacteria associated with phytoplankton blooms. Nat Rev Microbiol 2014; 12: 686-98.

Chiarello M, Auguet JC, Bettarel Y, Bouvier C, Claverie T, Graham NAJ, et al. Skin microbiome of coral reef fish is highly variable and driven by host phylogeny and diet. Microbiome 2018; 6: 147.

Colston TJ, Jackson CR. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. Molecular ecology 2016; 25: 3776-3800.

Dabrowska K, Witkiewicz W. Correlations of Host Genetics and Gut Microbiome Composition. Front Microbiol 2016; 7: 1357.

Davison JM, Lickwar CR, Song L, Breton G, Crawford GE, Rawls JF. Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor Hepatocyte nuclear factor 4 alpha. Genome Res 2017; 27: 1195-1206.

Dethlefsen L, Eckburg PB, Bik EM, Relman DA. Assembly of the human intestinal microbiota. Trends Ecol Evol 2006; 21: 517-23.

Doane MP, Morris MM, Papudeshi B, Allen L, Pande D, Haggerty JM, et al. The skin microbiome of elasmobranchs follows phylosymbiosis, but in teleost fishes, the microbiomes converge. Microbiome 2020; 8: 93.

Dulski T, Kozlowski K, Ciesielski S. Habitat and seasonality shape the structure of tench (Tinca tinca L.) gut microbiome. Sci Rep 2020; 10: 4460.

Egerton S, Culloty S, Whooley J, Stanton C, Ross RP. The gut microbiota of marine fish. Frontiers in microbiology 2018; 9: 873.

Emie AG, Francois-Etienne S, Sidki B, Nicolas D. Microbiomes of clownfish and their symbiotic host anemone converge before their first physical contact. Microbiome 2021; 9: 109.

Fan P, Bian B, Teng L, Nelson CD, Driver J, Elzo MA, et al. Host genetic effects upon the early gut microbiota in a bovine model with graduated spectrum of genetic variation. ISME J 2020; 14: 302-317.

FAO. Contributing to Food Security and Nutrition for All,(The State of World Fisheries and Aquaculture). Food and Agricultural Organisation of the United Nations Rome, 2016.

Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. PLoS One 2013; 8: e76096.

Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature 2017; 548: 43-51.

Ghanbari M, Kneifel W, Domig KJ. A new view of the fish gut microbiome: advances from next-generation sequencing. Aquaculture 2015; 448: 464-475.

Giatsis C, Sipkema D, Smidt H, Heilig H, Benvenuti G, Verreth J, et al. The impact of rearing environment on the development of gut microbiota in tilapia larvae. Sci Rep 2015; 5: 18206.

Gomez D, Sunyer JO, Salinas I. The mucosal immune system of fish: the evolution of tolerating commensals while fighting pathogens. Fish & shellfish immunology 2013; 35: 1729-1739.

Goodrich JK, Davenport ER, Clark AG, Ley RE. The Relationship Between the Human Genome and Microbiome Comes into View. Annu Rev Genet 2017; 51: 413-433.

Grieneisen L, Muehlbauer AL, Blekhman R. Microbial control of host gene regulation and the evolution of host-microbiome interactions in primates. Philos Trans R Soc Lond B Biol Sci 2020; 375: 20190598.

Guardiola FA, Cuesta A, Abellán E, Meseguer J, Esteban MA. Comparative analysis of the humoral immunity of skin mucus from several marine teleost fish. Fish & shellfish immunology 2014; 40: 24-31.

Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. Genome Res 2009; 19: 1141-52.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 2012; 10: 497-506.

He X, Chaganti SR, Heath DD. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microb Ecol 2018; 75: 140-151.

Hoseinifar SH, Eshaghzadeh H, Vahabzadeh H, Peykaran Mana N. Modulation of growth performances, survival, digestive enzyme activities and intestinal microbiota in common carp (C yprinus carpio) larvae using short chain fructooligosaccharide. Aquaculture research 2016; 47: 3246-3253.

Hoseinifar SH, Sun YZ, Wang A, Zhou Z. Probiotics as Means of Diseases Control in Aquaculture, a Review of Current Knowledge and Future Perspectives. Front Microbiol 2018; 9: 2429.

Hou D, Huang Z, Zeng S, Liu J, Wei D, Deng X, et al. Environmental Factors Shape Water Microbial Community Structure and Function in Shrimp Cultural Enclosure Ecosystems. Front Microbiol 2017; 8: 2359.

Huse SM, Ye Y, Zhou Y, Fodor AA. A core human microbiome as viewed through 16S rRNA sequence clusters. PLoS One 2012; 7: e34242.

Isbell F, Gonzalez A, Loreau M, Cowles J, Diaz S, Hector A, et al. Linking the influence and dependence of people on biodiversity across scales. Nature 2017; 546: 65-72.

Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol 2007; 45: 2761-4.

Johny TK, Puthusseri RM, Bhat SG. A primer on metagenomics and next-generation sequencing in fish gut microbiome research. Aquaculture Research 2021.

Krotman Y, Yergaliyev TM, Shani RA, Avrahami Y, Szitenberg A. Dissecting the factors shaping fish skin microbiomes in a heterogeneous inland water system. Microbiome 2020; 8: 1-15.

Langlois L, Akhtar N, Tam KC, Dixon B, Reid G. Fishing for the right probiotic: Host-microbe interactions at the interface of effective aquaculture strategies. FEMS Microbiology Reviews 2021.

Leeming ER, Johnson AJ, Spector TD, Le Roy CI. Effect of Diet on the Gut Microbiota: Rethinking Intervention Duration. Nutrients 2019; 11.

Legrand TP, Catalano SR, Wos-Oxley ML, Stephens F, Landos M, Bansemer MS, et al. The inner workings of the outer surface: skin and gill microbiota as indicators of changing gut health in yellowtail kingfish. Frontiers in microbiology 2018; 8: 2664.

Legrand TP, Wynne JW, Weyrich LS, Oxley AP. A microbial sea of possibilities: current knowledge and prospects for an improved understanding of the fish microbiome. Reviews in Aquaculture 2020; 12: 1101-1134.

Lescak EA, Milligan-Myhre KC. Teleosts as Model Organisms To Understand Host-Microbe Interactions. J Bacteriol 2017a; 199.

Lescak EA, Milligan-Myhre KC. Teleosts as model organisms to understand host-microbe interactions. Journal of bacteriology 2017b; 199: e00868-16.

Levin SA. Encyclopedia of biodiversity: Elsevier Inc., 2013.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 2008; 6: 776-88.

Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 2006; 124: 837-48.

Liu B, Song C, Gao Q, Liu B, Zhou Q, Sun C, et al. Maternal and environmental microbes dominate offspring microbial colonization in the giant freshwater prawn Macrobrachium rosenbergii. Sci Total Environ 2021; 790: 148062.

Llewellyn MS, Boutin S, Hoseinifar SH, Derome N. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Frontiers in microbiology 2014; 5: 207.

Lokesh J, Kiron V. Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. Sci Rep 2016; 6: 19707.

Mallott EK, Amato KR. Host specificity of the gut microbiome. Nat Rev Microbiol 2021; 19: 639-653.

Maynard CL, Elson CO, Hatton RD, Weaver CT. Reciprocal interactions of the intestinal microbiota and immune system. Nature 2012; 489: 231-241.

Merrifield DL, Ringo E. Aquaculture nutrition: gut health, probiotics and prebiotics: John Wiley & Sons, 2014.

Minich JJ, Petrus S, Michael JD, Michael TP, Knight R, Allen EE. Temporal, Environmental, and Biological Drivers of the Mucosal Microbiome in a Wild Marine Fish, Scomber japonicus. mSphere 2020; 5.

Miranda CD, Godoy FA, Lee MR. Current Status of the Use of Antibiotics and the Antimicrobial Resistance in the Chilean Salmon Farms. Front Microbiol 2018; 9: 1284.

Moran NA, McCutcheon JP, Nakabachi A. Genomics and evolution of heritable bacterial symbionts. Annu Rev Genet 2008; 42: 165-90.

Navarrete P, Mardones P, Opazo R, Espejo R, Romero J. Oxytetracycline treatment reduces bacterial diversity of intestinal microbiota of Atlantic salmon. J Aquat Anim Health 2008; 20: 177-83.

Nayak SK. Role of gastrointestinal microbiota in fish. Aquaculture Research 2010; 41: 1553-1573.

Nichols RG, Davenport ER. The relationship between the gut microbiome and host gene expression: a review. Hum Genet 2021; 140: 747-760.

Oh J, Byrd AL, Park M, Program NCS, Kong HH, Segre JA. Temporal Stability of the Human Skin Microbiome. Cell 2016; 165: 854-66.

Oliphant K, Parreira VR, Cochrane K, Allen-Vercoe E. Drivers of human gut microbial community assembly: coadaptation, determinism and stochasticity. ISME J 2019; 13: 3080-3092.

Prakash S, Rodes L, Coussa-Charley M, Tomaro-Duchesneau C. Gut microbiota: next frontier in understanding human health and development of biotherapeutics. Biologics 2011; 5: 71-86.

Prosser JI, Bohannan BJ, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, et al. The role of ecological theory in microbial ecology. Nat Rev Microbiol 2007; 5: 384-92.

Razak SA, Griffin M, Mischke C, Bosworth B, Waldbieser G, Wise D, et al. Biotic and abiotic factors influencing channel catfish egg and gut microbiome dynamics during early life stages. Aquaculture 2019; 498: 556-567.

Reed GB, Spence CM. The intestinal and slime flora of the haddock: a preliminary report. Contributions to Canadian Biology and Fisheries 1929; 4: 257-264.

Ringo E, Holzapfel W. Identification and characterization of carnobacteria associated with the gills of Atlantic salmon (Salmo salar L.). Syst Appl Microbiol 2000; 23: 523-7.

Romero J, Navarrete P. 16S rDNA-based analysis of dominant bacterial populations associated with early life stages of coho salmon (Oncorhynchus kisutch). Microbial ecology 2006; 51: 422-430.

Ros-Rocher N, Perez-Posada A, Leger MM, Ruiz-Trillo I. The origin of animals: an ancestral reconstruction of the unicellular-to-multicellular transition. Open Biol 2021; 11: 200359.

Sadeghi J, Chaganti SR, Shahraki AH, Heath DD. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. Sci Total Environ 2021; 781: 146771.

Sehnal L, Brammer-Robbins E, Wormington AM, Blaha L, Bisesi J, Larkin I, et al. Microbiome composition and function in aquatic vertebrates: small organisms making big impacts on aquatic animal health. Frontiers in microbiology 2021; 12: 358.

Sehnal L, Brammer-Robbins E, Wormington AM, Blaha L, Bisesi J, Larkin I, et al. Microbiome Composition and Function in Aquatic Vertebrates: Small Organisms Making Big Impacts on Aquatic Animal Health. Front Microbiol 2021b; 12: 567408.

Shahraki AH, Chaganti SR, Heath D. Spatio-temporal dynamics of bacterial communities in the shoreline of Laurentian great Lake Erie and Lake St. Clair's large freshwater ecosystems. BMC Microbiol 2021; 21: 253.

Smith CC, Snowberg LK, Gregory Caporaso J, Knight R, Bolnick DI. Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. ISME J 2015; 9: 2515-26.

Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol 2011; 9: 279-90.

Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J 2011; 5: 1414-25.

Steury RA, Currey MC, Cresko WA, Bohannan BJM. Population Genetic Divergence and Environment Influence the Gut Microbiome in Oregon Threespine Stickleback. Genes (Basel) 2019; 10.

Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight R, et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Mol Ecol 2012; 21: 3363-78.

Sun W, Xia C, Xu M, Guo J, Sun G. Seasonality Affects the Diversity and Composition of Bacterioplankton Communities in Dongjiang River, a Drinking Water Source of Hong Kong. Front Microbiol 2017; 8: 1644.

Suzuki TA, Worobey M. Geographical variation of human gut microbial composition. Biol Lett 2014; 10: 20131037.

Sylvain F-É, Cheaib B, Llewellyn M, Correia TG, Fagundes DB, Val AL, et al. pH drop impacts differentially skin and gut microbiota of the Amazonian fish tambaqui (Colossoma macropomum). Scientific reports 2016; 6: 1-10.

Sylvain FE, Derome N. Vertically and horizontally transmitted microbial symbionts shape the gut microbiota ontogenesis of a skin-mucus feeding discus fish progeny. Sci Rep 2017; 7: 5263.

Sylvain FE, Holland A, Bouslama S, Audet-Gilbert E, Lavoie C, Val AL, et al. Fish Skin and Gut Microbiomes Show Contrasting Signatures of Host Species and Habitat. Appl Environ Microbiol 2020; 86.

Talwar C, Nagar S, Lal R, Negi RK. Fish gut microbiome: current approaches and future perspectives. Indian journal of microbiology 2018; 58: 397-414.

Tarnecki AM, Burgos FA, Ray CL, Arias CR. Fish intestinal microbiome: diversity and symbiosis unravelled by metagenomics. J Appl Microbiol 2017; 123: 2-17.

Thaiss CA, Levy M, Korem T, Dohnalova L, Shapiro H, Jaitin DA, et al. Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. Cell 2016; 167: 1495-1510 e12.

Uren Webster TM, Consuegra S, Hitchings M, Garcia de Leaniz C. Interpopulation Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity. Appl Environ Microbiol 2018; 84.

Uren Webster TM, Rodriguez-Barreto D, Castaldo G, Gough P, Consuegra S, Garcia de Leaniz C. Environmental plasticity and colonisation history in the Atlantic salmon microbiome: A translocation experiment. Mol Ecol 2020; 29: 886-898.

Vellend M. Conceptual synthesis in community ecology. Q Rev Biol 2010; 85: 183-206.

Wang AR, Ran C, Ringø E, Zhou ZG. Progress in fish gastrointestinal microbiota research. Reviews in Aquaculture 2018; 10: 626-640.

Woo S, Yang SH, Chen HJ, Tseng YF, Hwang SJ, De Palmas S, et al. Geographical variations in bacterial communities associated with soft coral Scleronephthya gracillimum. PLoS One 2017; 12: e0183663.

Xia JH, Lin G, Fu GH, Wan ZY, Lee M, Wang L, et al. The intestinal microbiome of fish under starvation. BMC Genomics 2014; 15: 266.

Xiao F, Zhu W, Yu Y, He Z, Wu B, Wang C, et al. Host development overwhelms environmental dispersal in governing the ecological succession of zebrafish gut microbiota. NPJ Biofilms Microbiomes 2021; 7: 5.

Xiong JB, Nie L, Chen J. Current understanding on the roles of gut microbiota in fish disease and immunity. Zool Res 2019; 40: 70-76.

Yassir R, Adel M, Azze A. Use of probiotic bacteria as growth promoters, antibacterial and the effect on physiological parameters of Orechromis niloticus. J of Fish Dis 2002; 22: 633-642.

Yildirim S, Yeoman CJ, Sipos M, Torralba M, Wilson BA, Goldberg TL, et al. Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. PLoS One 2010; 5: e13963.

Zepeda Mendoza ML, Sicheritz-Ponten T, Gilbert MT. Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Brief Bioinform 2015; 16: 745-58.

Zhang M, Sun Y, Liu Y, Qiao F, Chen L, Liu W-T, et al. Response of gut microbiota to salinity change in two euryhaline aquatic animals with reverse salinity preference. Aquaculture 2016; 454: 72-80.

Zhang Z, Chen J, Li L, Tao M, Zhang C, Qin Q, et al. Research advances in animal distant hybridization. Sci China Life Sci 2014; 57: 889-902.

Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res 2020; 30: 492-506.

Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev 2017; 81.

Zhu B, Wang X, Li L. Human gut microbiome: the second genome of human body. Protein Cell 2010; 1: 718-25.

CHAPTER2: MICROBIAL COMMUNITY AND ABIOTIC EFFECTS ON AQUATIC BACTERIAL COMMUNITIES IN NORTH TEMPERATE LAKES

2.1 Introduction

An ecosystem can be defined simply as the collection of living organisms and their interactions (biotic) coupled with non-living factors (abiotic) that occur in a specific locale. Part of what sustains an ecosystem is the constant interchange of energy between its biotic (plants, animals, and microorganisms) and abiotic components (habitat, water, soil, light, etc.) (Gurung et al., 2001). Unicellular organisms, or microorganisms, consist of a wide range of microscopic taxa, representing prokaryotes (i.e. bacteria and archaea) and unicellular microbial eukaryotes (i.e. ciliates and amoebae) (Thorp and Covich, 2009). The microbial community is comprised of diverse groups of species that live and interact within a defined habitat or space (often defined as a "niche") (Berg et al., 2020). However, the processes driving patterns of microbial diversity among ecosystems remain poorly characterized and highly controversial (Nemergut et al., 2013). A framework has been established that postulates deterministic (niche-base) and stochastic (neutral) factors as potential mechanisms underpinning microbial biogeography (Liu et al., 2019; Liu et al., 2020a). Based on niche theory (deterministic processes), abiotic (e.g., nutrients) and biotic (e.g., competition, cooperation and predation) factors regulate community structure (Zhou and Ning, 2017). On the other hand, the stochastic process hypothesis proposes that all taxa are functionally equivalent, and not subject to strong environmental effects and community assembly is driven primarily by ecological drift coupled with variation in dispersal (Oliphant et al., 2019). This means that spatial distance would result in increasing divergence in microbial community composition, hence community composition should exhibit a distance-decay relationship (Zhang et al., 2019a). The controversy over the factors that drive microbial community composition

reflects a critical knowledge gap in the field of microbial ecology: the nature of the proximate and ultimate ecological forces that structure these communities (Oliphant et al., 2019).

The complexity of microbiomes encourages a shift from reductionist approaches that focus on individual taxa in isolation, to more holistic approaches that emphasize interactions among members of the community and their environments (Layeghifard et al., 2017). For example, co-occurrence patterns in microbial community have been used to assess community assembly rules (Fuhrman, 2009). Numerous ecological processes hypothetically contribute to co-occurrence patterns, including both deterministic and stochastic processes. Network analyses using high-throughput sequencing data may allow the nature of the interactions among organisms with the bacterial community (BC) to be characterized through comparisons of BC composition across environmental clines. As BCs exhibit diverse interactions with other taxa (e.g., protists), all the interacting communities (and abiotic factors) should ultimately be included in the analysis (Berry and Widder, 2014; Fuhrman, 2009).

Microbial communities associated with fresh water form the foundation of freshwater food webs and are the primary biogeochemical agents involved in nutrient cycling; yet the factors affecting community composition remain poorly characterized (Percent et al., 2008). To address the nature of the interactions among the aquatic microbial community in combination with biotic and abiotic factors affecting that community, sixty lakes located on the edge of the Precambrian Shield in central Ontario, Canada were sampled for microbial metabarcoding. These lakes have been surveyed regularly by the Ontario Ministry of the Environment, Conservation and Parks' Dorset Environmental Science Centre since the late 1970s (Nelligan et al., 2019), providing valuable metadata to test hypotheses about environmental factors affecting microbial community composition. As the microbial communities within the Precambrian Shield

have yet to be fundamentally described, simple baseline data on microbial community composition within these systems are valuable; however, network and correlational analyses facilitate the determination of how BC interact with other organisms and their environment. Here, we use high throughput sequencing and metabarcoding to characterize the factors that structure the BCs of small temperate freshwater lakes. We tested the following hypotheses: (1) deterministic biotic and abiotic factors prevail in shaping BC assembly in freshwater lakes, (2) within the deterministic factors, biotic factors will dominate abiotic factors in shaping the BCs, and (3) spatial factors will play a minor role (after correcting for abiotic variation) in driving bacterial assemblage in freshwater ecosystems. Our data first provide valuable baseline data on the composition and variation in microbial communities among small temperate lakes. More importantly, our results define the relative influence of biotics and abiotic factors on freshwater lake BC composition. These results can be used to better monitor, predict and respond to changes in the heath and stability of lake ecosystems. Our work also provides information on a neglected component of the freshwater microbial community, the micro-eukaryotic communities (MECs), and their interactions with their associated BCs.

2.2 Materials and Methods:

2.2.1 Sample collection and physicochemical analysis

Sixty lakes were sampled in southern Ontario over the course of three days (7-9, October 2017). Within each lake, we sampled either one location (for 50 lakes), two locations (for 8 lakes) or three locations (for 2 lakes), making a total of 72 locations across all 60 lakes. The number of locations sampled was based primarily on the availability of public access and lake surface area (sampling sites with their coordinated are provided in Supplementary Table S2.1). Sampling sites with less disturbance and human activity at the site were selected. At each sampling location, two (n= 43 lakes) or three (n= 17 lakes) 500 mL water samples of lake water from the top meter (0–1.0 m) were collected from the shore without disturbing the sediment (Supplementary Table S2.1). In total, 162 bottles of lake water were collected across all 60 lakes. Samples were collected in sterile 500mL Nalgene™ HDPE bottles that were rinsed with sample site water prior to collection. Field negative controls were collected at the beginning and in the middle of each sampling day to ensure that the sampling and transportation did not cause sample contamination. Field negative controls were generated by transferring double distilled water from one sterile bottle to another, these were stored with the other samples. All water samples were stored in a cooler with ice and filtered within two hours of collection using 0.22-micron pore size, 47 mm diameter polycarbonate filters (Isopore™, Millipore, MA). After filtration, each filter was cut in half and each half was placed in a 2 mL sterile tube with 50-100 g glass beads (0.1 mm diameter, Bio-Spec Products, Bartlesville, US) and stored on dry ice in the field. In total, 324 half-filter samples were processed (Supplementary Table S2.1), shipped to the lab (on dry ice) and stored at -20 ^{o}C until DNA extraction.

We collected metadata for each sampled lake (11 environmental variables), those variables can be grouped into three categories: (i) physical or geographical characteristics that are constant or do not vary at an annual scale (lake depth, surface area, volume, shoreline length, altitude), (ii) specific physicochemical variables were selected for their known importance for microbial communities in lake systems (pH, total phosphorus (TP), calcium, Secchi depth), and (iii) spatial position variables (latitude, longitude). Physicochemical data were obtained from Ministry of the Environment, Conservation and Parks (Sutey et al., 2019). Environmental data were collected during 2017 and the samples for chemical analysis were collected from offshore. Geographical (altitude) and spatial data were collected using cell phone application (Google Earth, Google LLC) during sample collection for each lake.

2.2.2 Environmental DNA (eDNA) extraction

eDNA was extracted using a published sucrose lysis buffer protocol (Shahraki et al., 2019). In brief, the filters were placed in 2-mL tubes with 400 µL of sucrose lysis buffer (400 mM NaCl, 750 mM sucrose, 20 mM ethylenediaminetetraacetic acid, 50 mM Tris-HCl pH 9.0). The samples were homogenized using a Mini-beadbeater-16 (Lab Services BV, Nederland) for 1 min, three times. Then, each sample was treated with 20 µL lysozyme (10 $mg \cdot mL^{-1}$), and 60 µL of SDS (1%) (Sigma-Aldrich, USA); followed by a one-hour incubation at 37 °C. Following that incubation, 2 µL proteinase K (20 $mg \cdot mL^{-1}$) (Thermo Scientific, USA) was added and the solution was incubated, rocking, overnight at room temperature (25 °C). At the end of incubation, the Proteinase K was inactivated for one hour at 60 °C. Finally, 100 µL of the lysate was used to extract DNA using a "Bead-Robotic" purification protocol with solid-phase reversible immobilization (SPRI) paramagnetic beads on an automated liquid handling platform (Tecan Freedom Evo150 Liquid Handling Platform, Perkin Elmer, USA) (Shahraki et al., 2019).

For each DNA extraction plate (96 wells) one negative control (sterile membrane filter) was included to test for lab-based contamination during DNA extraction. The extracted eDNA was stored at -20 °C until PCR amplification of the 16S rRNA and 18S rRNA gene fragments was performed.

2.2.3 DNA Library Construction and Sequencing

Bacterial metabarcoding was performed using the variable V5-V6 region of the bacterial 16S rRNA gene. We chose to use the V5-V6 region due to amplicon length restrictions, and the need to maximize the sequencing read depth (Liu et al., 2007). The V5-V6 region primers (787F = acctgcctgccg-ATTAGATACCCNGGTAG; 1046R = acgccaccgagc-CGACAGCCATGCANCACCT) amplified a 260 bp fragment providing substantial taxonomic resolution. The primer sequences had 12 base extensions on the 5' end (lower case base codes on the primer sequence) to facilitate library preparation for high through-put metabarcode sequencing (HTMS). The V5–V6 region of the 16S rRNA gene was amplified in a 25 μL PCR, consisting of 2.5 μL 10X Taq reaction buffer, 0.5 μL each of 10 μM forward and reverse primers, 0.1 μL of Taq polymerase (5 U/ μL), 1.0 μL of 10 μM dNTPs, 3.5 μL of 20 mM $MgSO_4$, and 2.0 μL of extracted eDNA (ultrapure water for negative PCR controls). The thermal cycling protocol for the first round PCR consisted of: 95 °C for 3 minutes followed by 28 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 m, and a final elongation at 72 °C for 10 m. First-round PCR amplification was verified by visualizing amplicons on an agarose gel. If an appropriate band was observed, the remaining PCR solution was purified using Sera-Mag Magnetic Beads (GE, Healthcare Life Science, UK).

The purified PCR products were used as a template for a second, short-cycle, PCR to ligate the adaptor and barcode sequences necessary for sample identification and HTMS. The

second PCR was conducted in a total volume of 25 µL, consisting of 2.5 µL of 10X Taq reaction buffer, 25 mM $MgSO_4$, 0.2 mM of each dNTP, 0.4 µL 10 µM forward primer (UniA, CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXGATacctgcctgccg), 0.4 µL 10 µM reverse primer (UniB, CCTCTCTATGGGCAGTCGGTGATacgccaccgagc), 0.1 µL Taq polymerase 5U/µL and 10 µL of the cleaned first-round PCR product. The string of ten to twelve X's in the primer sequence represents the sample ID barcode sequence. The thermal cycle protocol for the second ligation PCR was 94 °C for 3 min, then 8 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min, and final elongation at 72 °C for 7 min. The second PCR product was visualized on an agarose gel and all field samples produced bands, whereas negative controls did not (field, DNA extraction, and PCR negative controls). The (positive) amplified samples were pooled roughly based on their band intensity (between 1- 5 µL for samples with strong to faint bands respectively) and the combined PCR products were cleaned and purified using the QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). Six field sampling-negative controls, four eDNA extraction-negative controls (one for each 96 well PCR plate (PROGENE®)), and eight PCR amplification-negative controls (four for BC and four for MEC samples) were also included in our amplicon pool for the sequencing library, despite there being no visible band on the agarose gels. The concentration of the purified PCR product mix (library) was measured on an Agilent 2100 Bioanalyzer with a High Sensitivity DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library was then diluted to 60 $pmol \cdot \mu L^{-1}$ and sequenced on an Ion PGM™ System using the Ion PGM™ Sequencing 400 bp chemistry and an Ion 318™ Chip (Thermo Fisher Scientific, Burlington, ON, Canada).

Micro-eukaryote metabarcoding using 18S rRNA: V9 region of 18S rRNA gene (PCR primers: F = acctgcctgccg GTACACACCGCCCGTC'; R = acgccaccgagcTGATCCTT

CTGCAGGTTCACCTAC) was used as the metabarcoding target region, following a previously described protocol (Petri et al., 2019). The PCR reagents were same as for the 16S rRNA PCR (above). The thermocycling protocol for the first PCR amplification of the V9 region of the 18S rRNA gene was an initial denaturing stage at 94 °C for 2 min, followed by 28 cycles of: denaturing at 94 °C for 15 s, annealing at 55 °C for 15 s, extension at 72 °C for 30 s, followed by a final elongation step at 72 °C for 7 min (Petri et al., 2019). The second short-cycle ligation PCR to create the barcoded library was the same as for the 16S metabarcoding ligation second PCR (above). The 18S PCR products were cleaned, barcoded, and sequenced following the same protocols as for 16s rRNA.

2.2.4 Sequence Data Processing

16SrRNA Sequences: After sequencing was complete, sequence reads were filtered using Ion PGM™ software to remove unwanted polyclonal and low-quality sequences. De-multiplexing and quality filtering of the 16S sequences were done in QIIME 1.9.0. Chimera detection was performed using Usearch quality filtering in QIIME (Edgar et al., 2011; Kuczynski et al., 2011). Operational taxonomic units (OTUs) were defined at 97% sequence similarity using the UCLUST algorithm. Representative sequences were aligned with the PyNast algorithm and taxonomy assignment used the default method (RDP). Singleton OTUs were discarded (n=2) before further analysis to avoid possible biases. Rarefaction plots were generated in QIIME as described previously (Kuczynski et al., 2011). Briefly, rarefied OTU tables were generated using the *multiple_rarefactions.py* script, then alpha diversity indices (Chao1 and number of observed OTUs) were computed for each rarefied OTU table with *alpha_diversity.py* and collated using the *collate_alpha.py* script. Finally, alpha diversity rarefaction plots were generated by *make_rarefaction_plots.py* script. The rarefaction cut-off was set to 2000 reads per sample

because most of the sample rarefaction curves plateaued at 2000 reads. Samples that had sequence read numbers below the rarefaction cut-off (2000) were removed from further analysis. As a result, 270 samples covering 59 lakes (98%) remained. All negative controls (field sampling negative controls, DNA extraction negative controls, and PCR negative controls) contained fewer than 30 sequence reads and were excluded from the rest of the analysis. Sequences were deposited in the NCBI Short Read Archive (SRA) under accession number SRR12080477.

18SrRNA Sequences: Bioinformatic analyses were conducted using QIIME 1.9.0. For demultiplexing and quality filtering of the raw sequence reads we used the default settings, except the minimum sequence length was changed to 130bp (Kuczynski et al., 2011). OTUs were defined at 97% sequence similarity using the UCLUST algorithm. Representative sequences were aligned with the PyNast algorithm and taxonomy assignment was done using blast against the SLIVA 104 reference database (http://qiime.org/1.4.0/tutorials/processing_18S_data.html). Singleton OTUs were removed (n=2) from further analyses. Alpha diversity indices were computed as for 16s rRNA data. The rarefaction cut-off was set to 2000 reads per sample and samples that had sequence read numbers below the rarefaction cut-off were removed from further analyses. 18S PCR negative controls contained fewer than 10 sequence reads and were excluded at this step. As a result, 237 samples from 54 lakes (90%) remained. For the OTU table, OTUs assigned to fungi or vertebrate animals were removed and only OTUs related to freshwater invertebrates (protozoa and micrometazoans (less than 2mm)) were retained. OTUs classified as “Uncultured_stramenopile”, “Uncultured_alveolate”, “Uncultured_cercozoan”, and “Uncultured_freshwater_cercozoan” (without class or order classifications) were combined and renamed as uSAR (uncultured Stramenopiles, Alveolata, and Rhizaria). All invertebrate fauna classified by SILVA were

double-checked with the latest scientific nomenclature in the World Register of Marine Species (WORMS) (World Register of Marine Species (WoRMS), 2020). The sequences were deposited in the NCBI SRA under accession number SRR12080473.

2.2.5 Statistical analysis

Combining replicates: We included location (multiple sites within a lake), biological (multiple water samples at a site) and technical (DNA extracted from the two halves of the filter) replicates in our study design. To test whether the replicates contributed significantly to the variation in MEC and BC composition, we used a nested ANOVA in the R package "lme4" across all lakes with technical replicates (filter halves) nested within biological replicates (multiple samples at a site) nested within location replicates (multiple locations within a lake) nested within lakes. Our dependant variables were alpha diversity indices (Chao1 and number of observed OTUs) and beta diversity. Beta diversity was measured using principal coordinate analysis (PCoA) conducted in PRIMER (version 7.0.13) based on the Bray-Curtis similarity distance. PCoA1, PCoA2, PCoA3 were included as dependant variables based on % variance explained and the eigenvalue (e). For BC, PCoA1, PCoA2, and PCoA3 explained 14% (e=13.3), 8% (e=7.7), and 7% (e=6.8) of the observed variation among samples. Other PCoA axes had eigenvalues < 5.0 and explained less than 5% of the variance. For MEC, PCoA1, PCoA2, and PCoA3 explained 12.7% (e=2.2), 9.4% (e=1.6), and 7.8% (e=1.3) of the observed variation among samples. Other PCoA axes for MEC had eigenvalues < 1.0 and explained less than 1% of the variance. Moreover, we also tested the replicate effects on BCs using nested (as above) permutational multivariate analysis of variance (PERMANOVA) with the *adonis* function of the vegan R package with the Bray Curtis similarity index applied across all sample data (Team, 2013a). The nested ANOVA and PERMANOVA tests were not significant across alpha and beta

diversity measures for location, sample, and filter replicates (Supplementary Tables S2.2–S2.3). As a result, we combined all replicate data within each lake; to do this we summed the number of sequence reads across replicates for each OTU within the 59 (BC) and 54 (MEC) lakes. We used the *biom convert* command (McDonald et al., 2012) to convert the resulting combined OTU tables to BIOM format. Alpha and rarefied BIOM tables were generated following the same protocols for the 18S and 16S rRNA sequencing data. The rarefaction cut-off was set to 2000 reads per sample and a total of 59, and 54 samples (lakes) remained for the BCs and MECs, respectively. All subsequent analyses are based on the replicate-combined samples.

Diversity analyses: Microbial diversity (both for bacteria (16S) and micro-eukaryotes (18S)) was measured using a series of OTU-based estimates of alpha- and beta-diversity (as described above). OTUs with a read frequency greater than 0.001 percent of the total sequence reads were selected for further analysis as some of the low abundance OTUs may have derived from sequencing errors or other artifacts. Moreover, removing low read OTUs will dramatically reduce the computational workload (Unno, 2015; Wallace et al., 2018). Alpha diversity and taxonomic summary analyses were conducted to determine variation in microbial community composition among all the lakes. Stacked barplots of the relative abundance of the BCs at the order level, and the MECs at the class level were generated using the ggplot package in R (version 4.0.) (Team, 2013a).

Clustering analyses: OTU tables were square-root transformed prior to calculating the Bray–Curtis similarity matrices to reduce skew. To visualize patterns among the sampled lakes, Bray–Curtis similarity matrices were used in Paleontological Statistics Software Package for Education and Data Analysis (PAST, version 4.03) (Hammer et al., 2001) to create a neighbor-joining (NJ) tree. To test for significant differences among the identified NJ clusters, we ran a

global PERMANOVA test (for among cluster differences), and in cases of significant cluster effects, we also performed pairwise PERMANOVA to determine post-hoc specific cluster differences. We present the NJ trees with a sample map to aid in interpretations of spatial patterns using QGIS (Version 3.16.0) with OpenLayers plugin and Stamen Toner/OSM layer.

Correlation analysis: The biogeochemical and physical parameters were log-transformed (ln (value + 1)) to achieve normal distribution. Pearson correlation analyses were used to test for correlations between the bacterial alpha and beta diversity indices with environmental and biotic factors (18S alpha and beta diversity) using the *cor.test* function in R with "pearson" for correlation coefficient, the *p*-value was adjusted using the Bonferroni correction (Team, 2013a). DistLM (distance-based multivariate multiple regression based on a linear model) (McArdle and Anderson, 2001) conducted in PRIMER (version 7.0.13) was used to examine the relationships between the biological communities (18s Chao1, 18s PCOA1, 2, and 3) and multivariate environmental data (lake depth, volume, surface area, shoreline length, TP, pH, calcium, Secchi depth, altitude, latitude, and longitude) (Legendre and Anderson, 1999). As the physical variables (depth, volume, surface area, shoreline length) were highly correlated, a Principal Component Analysis was performed using PAST (version 4.03) software. The first and second Principal Components (PC1, and PC2) were selected based on eigenvalue > 1.0 and subsequently were incorporated in our DistLM model. Moreover, we chose to include only Chao1 as our one alpha diversity measure as including multiple alpha diversity indices might cause redundancy in our analysis. The BEST DistLM model building procedure with 9999 permutations was used to determine the combination of individual factors that accounted for the greatest proportion of variation in BC composition, wherein factor addition was evaluated stepwise and was based on threshold improvement in the model's adjusted R^2. As DistLM does not accept missing data,

only lakes with all physicochemical data shared between BC and MEC were included (40 lakes). Individual variables (marginal tests) and groups of variables (sequential tests (conditional on relationships among variables with the community data)) were tested against the null hypotheses of no relationship between the biological distribution and environmental data.

2.2.6 Co-occurrence Network Analysis

Given the expected complexity of biotic interactions between and within the BCs and MECs among the sampled lakes, we used SparCC (Sparse Correlations for Compositional data) in Python to determine patterns of co-occurrence (positive) and co-exclusion (negative) relationships using OTU abundance data (Friedman and Alm, 2012). In brief, because microbial community data are sparse (with high numbers of OTUs with '0' counts in some lakes), the rarefied OTU table was filtered to remove any OTUs with a total sequence abundance of less than 0.01% across all samples. Removing less abundant taxa would help to better present (and interpret) the interactions among high abundant taxa; these abundant taxa have disproportionally important roles with other taxa relative to rare taxa. The resulting data set consisted of 400 and 602 OTUs for BC and MEC, respectively for the network analysis. To minimize the effects of compositional bias, we used the SparCC method. This method evaluates the variance of the log-ratio for transformed data, rather than the relative abundance, to infer pairwise relations (Friedman and Alm, 2012). The two-side pseudo p-values were calculated using python scripts based on bootstrapping with 1000 repetitions. A network plot showing correlation values higher than 0.55 or less than -0.55 and pseudo *P*-value less than <0.05 was generated. Cytoscape (v. 3.6.1) was employed to visualize the resulting networks (Shannon et al., 2003). We used yFiles Organic Layout to draw the network. For better visualization and to improve taxonomic relationship resolution, taxa are presented at the phylum level and uSAR and SAR were

combined and presented as SAR. Nodes (OTUs) with a high degree (>10), high closeness centrality (>0.24), and low betweenness centrality (<0.07) were identified as "keystone" taxa (Dai et al., 2020; Huber et al., 2020; Xue et al., 2020). The topology of the network reflects interactions among microorganisms. For example, the degree value describes the level of connectedness between OTUs, and the betweenness centrality provides information on how critical an OTU is to the connectedness of a network (Mondav et al., 2017).

2.3 Results

2.3.1 Combining replicate samples

Nested ANOVA for alpha (Chao1 and Observed OTUs) and beta diversity (PCoA1 and PCoA2, and PCoA3) indexes, as well as the nested PERMANOVA, showed no significant differences between the various types of replicate samples (locations (within a lake), biological (two or three bottles per site) and technical (extraction from filter halves)); however, there were highly significant lake effects (Supplementary Tables S2.2–S2.3). We thus combined sequence data across location, biological and technical replicates to maximize sequence read depth, resulting in a total of 59, and 54 samples (lakes) for BC and MEC (respectively) remaining for further analysis.

2.3.2 Structure of Microbial Communities

BC composition: Alpha diversity varied considerably (Chao1 ranged from 919 to 3824 (Supplementary Figure S2.1) and observed numbers of OTUs ranged from 539-1464 among lakes) and showed clear differences among BCs in southern Ontario lakes. After removing low abundance OTUs (less than 0.001% of reads), 10,012 OTUs remained across all the 59 lakes. The taxonomic analysis of the 16S rRNA gene sequences revealed the dominance of three bacterial orders: Actinomycetales (33%), Burkholderiales (12%), Enterobacteriales (11%) across all sampled lakes (Figure 2.1). Lakes varied considerable in their BC; for example, Enterobacteriales dominated L1-L26, however, Actinomycetales and Burkholderiales were more abundant in L27 to L59. Individual lakes showed high variation in taxon abundance, for example Lake Gananoque (L5) had the lowest abundance (1.6%) of Actinomycetales, while Big Bald

Lake (L31) had the highest (65%). Approximately 78% of the total bacteria taxonomic abundance was classified into six bacterial orders.

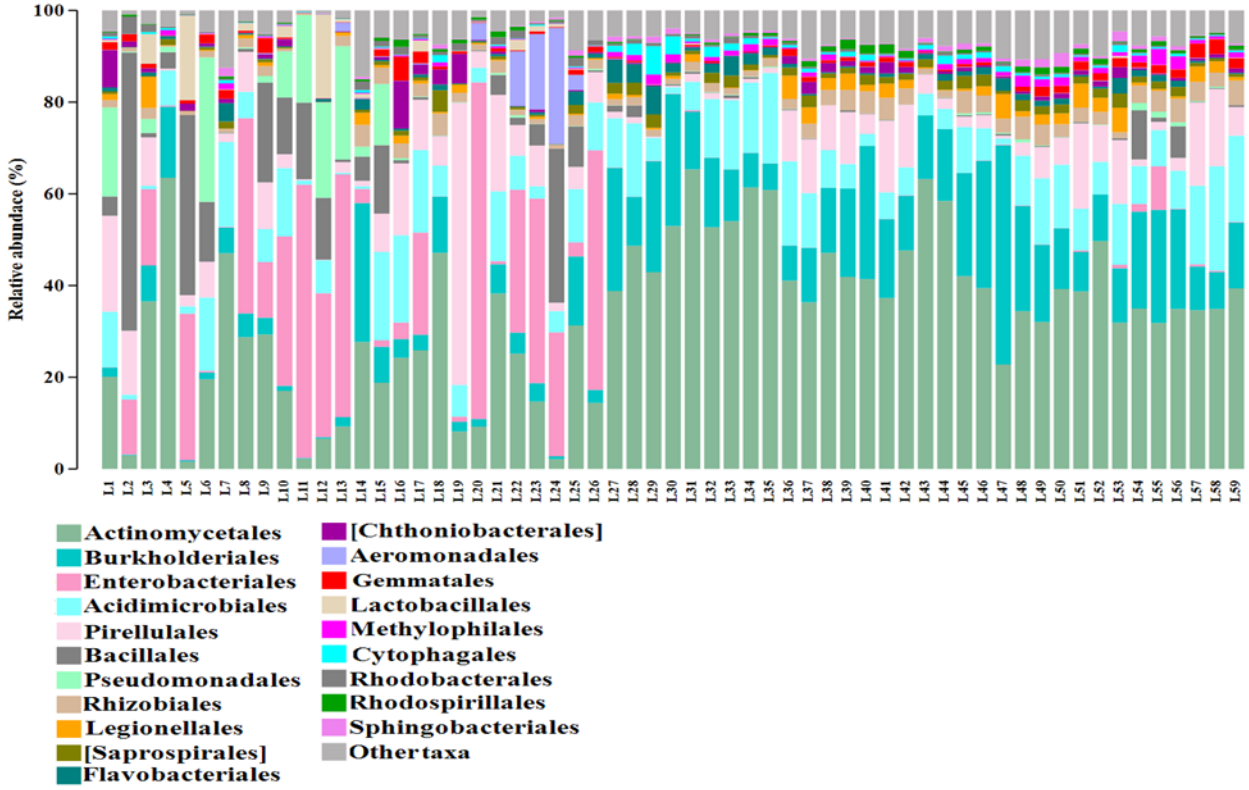


Figure 2.1. Relative abundance (greater than 0.01%) of lake bacterial community composition presented at the order level for the 59 sampled lakes. The 'other taxa' category includes the sum of all bacterial orders that occurred at less than 0.1% relative abundance.

MEC composition: MEC alpha diversity (Chao1 ranged from 317 to 1283 (Supplementary Figure S2.1), and observed numbers of OTUs ranged from 166-721 among the sampled lakes) showed clear differences among lakes. However, overall, the BCs (Chao1 mean= 2722 ± 700) were more diverse than the MECs (Chao1 mean= 690 ± 177). For the MECs, after removing low abundance OTUs (< 0.001%), a total of 3380 OTUs remained. Of these 3380 OTUs, 412 OTUs were assigned to fungi and vertebrates and were removed. Taxonomic analysis of the MECs revealed that approximately 62% of the total micro-eukaryote taxonomic abundance was classified into 6 major classes; uSAR (23%), Maxillopoda (10%), Spirotrichea

(9%), Cryptophyceae (8%), Chrysophyceae (4%), Prymnesiophyceae (4%), and Synurophyceae (4%). The SAR supergroup (uSAR, Stramenopiles (Chrysophyceae, Synurophyceae), Alveolata (Spirotrichea), and Rhizaria (Cryptophyceae)) were the most abundant taxa (Figure 2.2).

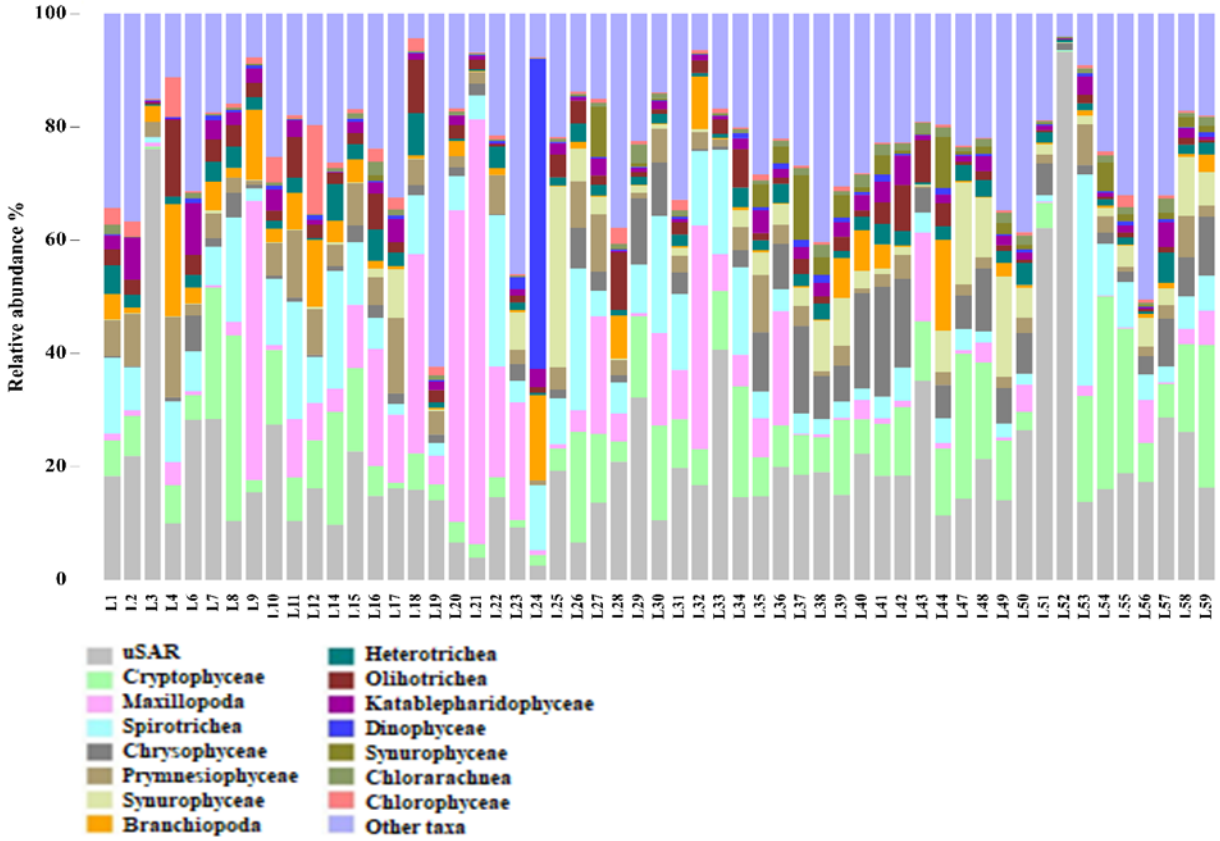


Figure 2.2. Relative abundance (greater than 0.01%) of lake MEC composition presented at the class level for the 54 sampled lakes. The 'other taxa' category includes the sum of all ME orders that occurred at less than 0.1% relative abundance.

2.3.3 Clustering analysis

NJ clustering of the BCs using the Bray-Curtis distance matrix resulted in six distinct clusters or clades, with most of the lake BCs falling into clusters I (n=28, 47%) and II (n=11, 19%) (Figure 2.3). The geographical distribution of the clusters reflects a clear spatial effect (Figure 2.3). PERMANOVA confirmed the statistical significance of the NJ cluster effect

(F=3.3, P=0.001). However, the subsequent post-hoc pairwise PERMANOVA analyses revealed that some of the individual clusters were not significantly different from others (Supplementary Table S2.4). The NJ clustering analysis for the MEC resulted in five clusters, with most of the lakes falling into Cluster I (n=18, 33%) and Cluster V (n=17, 31%) (Figure 2.3). The PERMANOVA analysis revealed a significant MEC NJ cluster effect (F=3.4, P=0.001). The post-hoc pairwise PERMANOVA analyses showed that all clusters were statistically significant from each other (Supplementary Table S2.4).

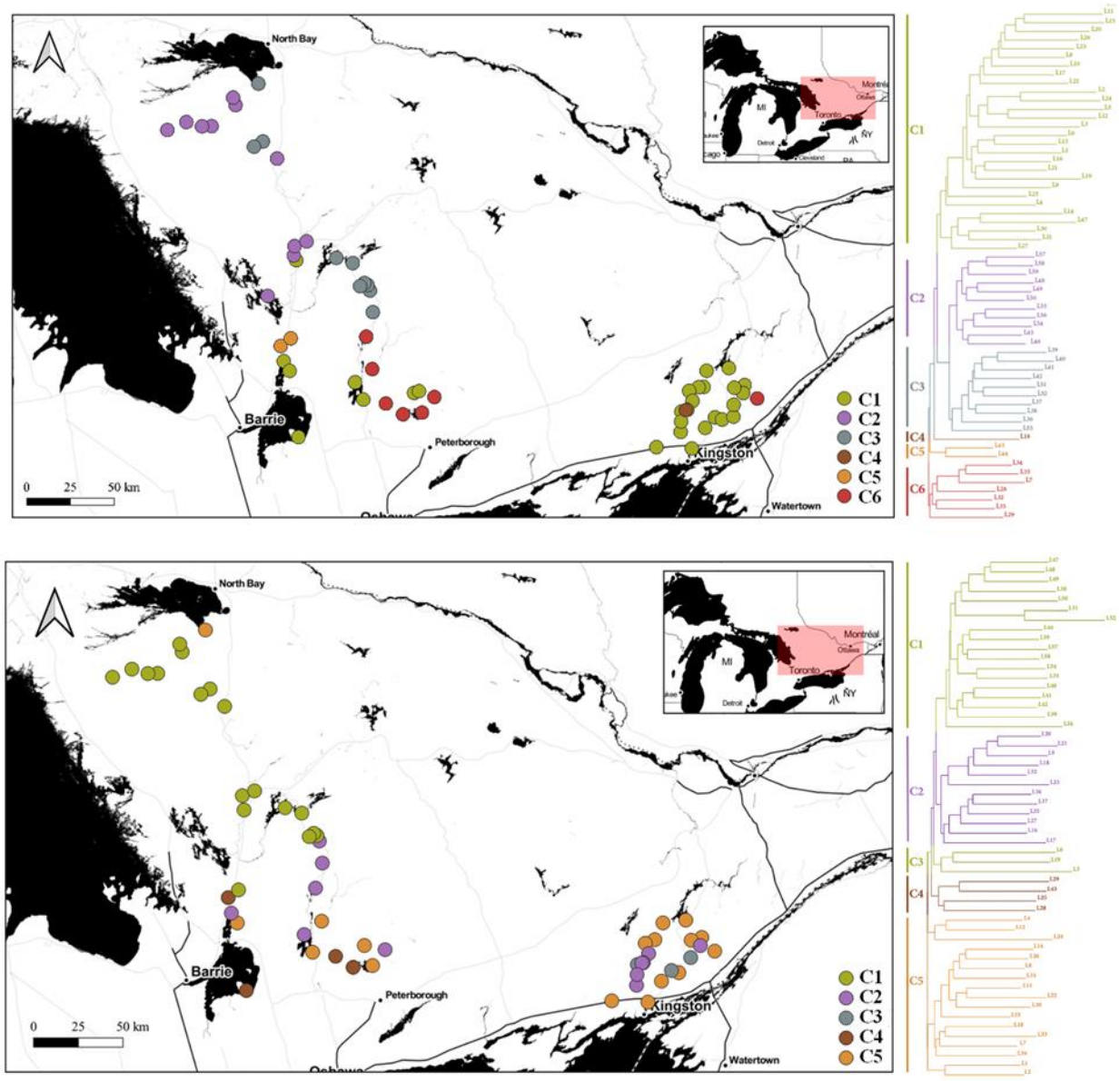


Figure 2.3. Map of the selected lakes for BC (top) and MEC (bottom). Lakes are color-coded based on NJ cluster assignment. Lake names with their coordinates are listed in supplementary Table S2.1.

2.3.4 Geochemistry of the studied lakes

There was considerable variation in lake depth (3 to 92 m), surface area (18 to 88,052 ha), volume (0.40 to 11000 m^3 x10^6), shoreline length (2.8-1013 km), altitude (78 to 480 m), Secchi depth (0.40 to 8.0 m), TP (0.001 to 0.09 mg/L), calcium (1.5 to 44.0 mg/L) and pH (6.4 to 8.5) among the sampled lakes (Supplementary Table S2.5). BC alpha and beta diversity indices were tested for correlation with environmental variables (PC1 and PC2 of physical variables

(lake depth, surface area, volume, shoreline length), TP, Secchi depth, calcium, pH, altitude), spatial position variables (latitude and longitude) and biotic variables (i.e., MEC composition (Chao1, Observed OTUs, and PCoA1 and PCoA2)). Pearson correlational analyses showed alpha and beta diversity indices were significantly and positively or negatively correlated with lake's spatial, geographical, physical, chemical, or biological variables (Figure 2.4).

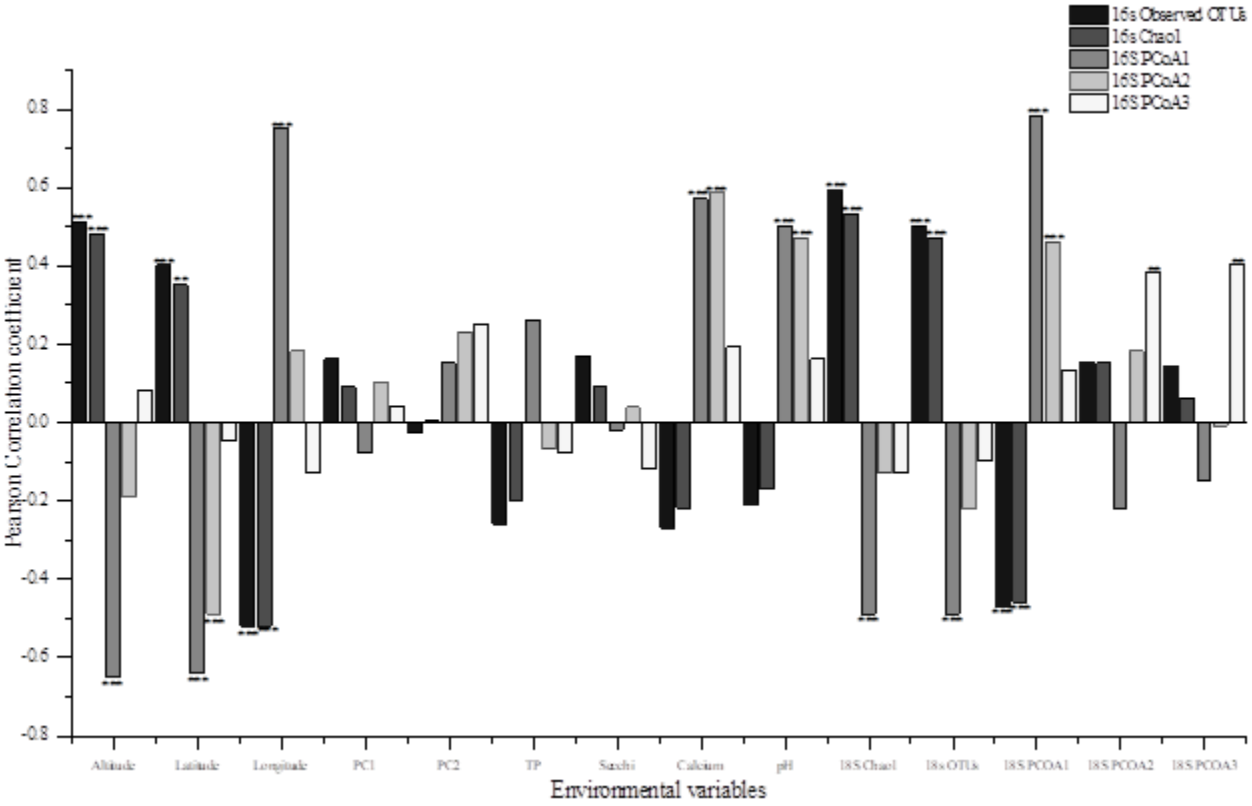


Figure 2.4. Pearson correlation coefficients between bacteria alpha (Chao1, observed OTUs) and beta diversity (PCoA1, PCoA2, PCoA3) indices with lake physicochemical, geographical, spatial, and biological variables. Asterisks show significant correlation (*** = $p<0.001$; ** = $p<0.01$; * = $p<0.0$). Detailed results are available in the supplementary Table S2.6.

DistLM was used to determine the relative contribution of the biotic and abiotic variables to variation in the BC composition among the lakes. Marginal DistLM values (Table 2.1) show the proportion of variation in the BC composition each variable explains when corrected for all other variables. Environmental variables (physical, chemical, and lake altitude) accounted for a

substantial component of the BC diversity variation (31%; Table 2.1). Spatial variables (Lat & Long) were the second most important set of variables and accounted for 23% of the BC diversity variation (Table 2.1). Biological variables (i.e., MEC effects) accounted for 23% of the diversity variation (Table 2.1). Among the chemical variables, calcium and pH were statistically significant and accounted for 13 percent of the BC diversity variation, while the total BC diversity variation explained by all chemical variables was 18% (Table 2.1). The geographical variable (lake altitude) was statistically significant and accounted for 7 percent of the BC diversity variation (Table 2.1). None of the lake physical characteristics variables contributed to the model significantly, but taken together, they accounted for 5% percent of the BC diversity variation (Table 2.1). The DistLM BEST explained up to 43% of the BC diversity variation (Supplementary Table S2.7). MEC PCoA1 was the single BEST predictive factor (Supplementary Table S2.7). In addition to MEC PCoA1, longitude (two-factor model) and latitude (three-factor model) emerged as the BEST predictor variables.

Table 2.1. Marginal DistLM values show the proportion of bacterial community variation explained by lake spatial, physical, chemical, or biological variables

Factor	Pseudo-F	Proportion
Spatial variables		
Latitude	4.18	0.099***[a]
Longitude	5.13	0.118***
Environmental variables		
Physical characteristics[b]		
PC1	0.96	0.024
PC2	1.21	0.03

Chemical variables		
TP	1.17	0.030
Calcium	2.7	0.067***
pH	2.4	0.060**
Secchi depth	0.97	0.025
Geographical variable		
Altitude	3.02	0.073***
Biological variables		
18S PCoA1	5.32	0.122***
18S PCoA2	1.02	0.026
18S PCoA3	1.08	0.027
18S Chao1	2.19	0.054**

a *** = p<0.001; ** = p<0.01; * = p<0.0

b Based on PCA with lake depth, volume, surface area, volume, shoreline length.

2.3.5 Network analysis

Our SparCC analyses identified 121 unique OTU nodes (83 BC nodes and 38 MEC nodes), linked with 224 interactions (edges). We identified 34 interactions that were co-exclusion, or negative associations (red lines; Figure 2.5), and among those 34 negative interactions, 18 edges (53%) were between two BC nodes (a red line connecting circle with a circle), 7 (21%) were between a BC and an MEC node (red line connecting circle with triangle), and 8 (26%) were between two MEC nodes (red line connecting triangular with triangle). Of the 190-remaining co-occurrence, or positive associations (black lines; Figure 2.5), 148 edges (77 %) were between two BC nodes (black line connected circle with circle), 8 (4%) were between a BC and an MEC node (black line connected circle with triangular), and 34 (18%) were between two MEC nodes. Overall, there were more interactions within the BCs and MECs than between

them. Generally, Actinobacteria (42 nodes, 35%), Proteobacteria (26 nodes, 21%), and SAR (24 nodes, 20%) dominated in terms of the number of representative nodes (Figure 2.5). We performed a "general network analysis" in this co-occurrence network to define the basic topologies of the microbial communities. The network analysis showed that only BC harbor keystone taxa (taxa that are of high degree, high closeness, and low betweenness scores and thus have a disproportional influence within the network). All BC keystone taxa were from the Actinobacteria (family ACK-M1), and Proteobacteria (family Enterobacteriaceae) phyla. Not surprisingly, less abundant taxa (node size) generally had fewer connections than highly abundant taxa. Overall, the BC network (which included the only keystone taxa) was much more complex than the MEC network.

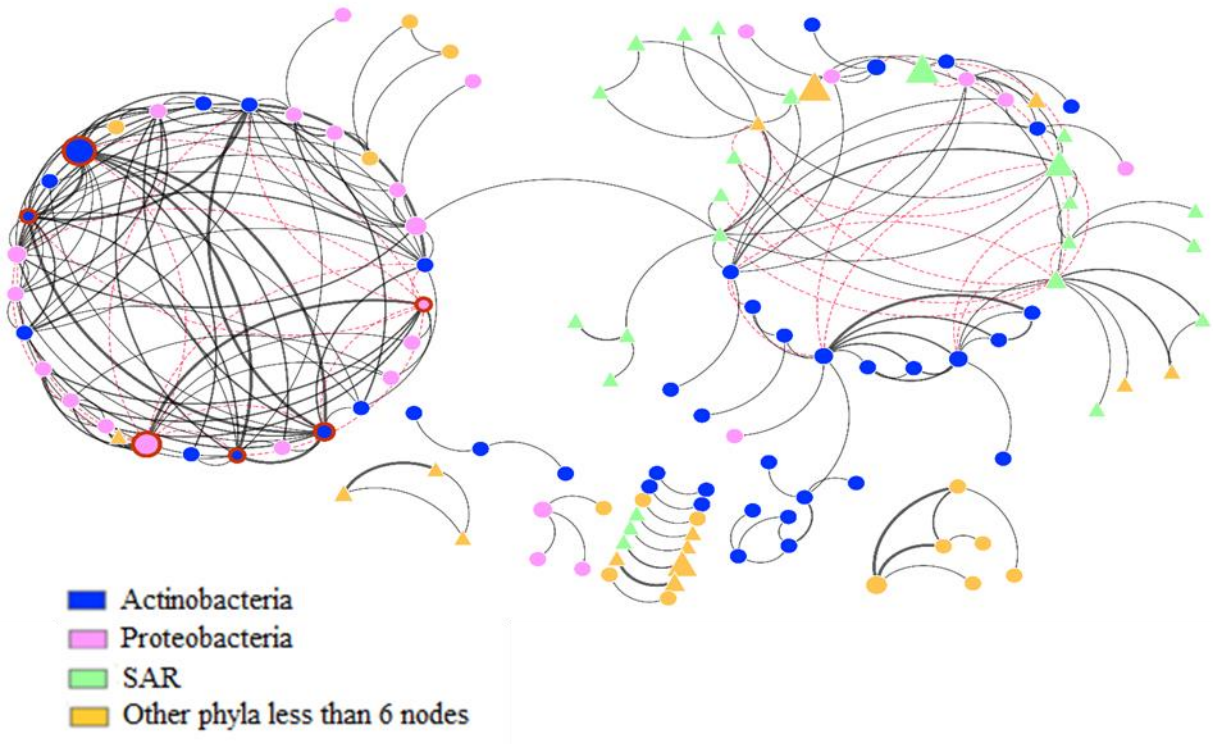


Figure 2.5. Association networks showing the highest correlations between microbial OTUs (correlation ≥0.55 or ≤-0.55). Node shapes correspond to bacteria (circle) or MEC (triangular) OTUs. Node colors are based on their phylum (phyla with fewer than 6 nodes are colored coded the same). Black lines represent positive correlations (co-abundance interactions) and red lines, negative correlations (co-exclusion interactions). Node symbol size and edge thickness are based on OTU read number, and strength of correlation, respectively. Keystone species are nodes with a red border (N = 6).

2.4 Discussion

Our study characterized temperate lake microbial communities and identified the prevailing processes of community assemblage and major BC drivers in a collection of ~60 southern Ontario lakes. The BCs found in all lakes were generally consistent with previous reports of bacterial phyla found in other North American lakes (Morrison et al., 2017; Mou et al., 2013). Members of *Actinomycetales*, *Burkholderiales* (*Betaproteobacteria*), and *Enterobacteriales* (*Gammaproteobacteria*) orders dominated the BCs in our selected lakes. Previous studies showed that Actinobacteria and Proteobacteria phyla are common resident of freshwater lakes (Liu et al., 2021; Mateus-Barros et al., 2021; Newton et al., 2011). Several factors such as supplemental mode of energy generation, size and composition of the bacterial cell wall, and growth rate, could be the reasons for their broad success (Kiersztyn et al., 2019; Newton et al., 2011; Shade et al., 2007). Moreover, in this study, members of *Enterobacteriales* were more abundant in lakes near populated areas (L1-L26) relative to more isolated lakes. One reason for this could be lakes around populated areas are experiencing increased nutrient loading, which would select against Actinobacteria (Haukka et al., 2006). Bacteria belonging to *Enterobacteriales*, such as *Escherichia coli*, are found in freshwater lakes (Shahraki et al., 2021); however, they are widely considered transient members that originate from anthropogenic or zoonotic sources (Newton et al., 2011). Micro-eukaryotic diversity in freshwater lakes is not as well documented as bacterial diversity; however, the MECs characterized in our study were similar to those reported in other freshwater ecosystems (Bjorbækmo et al., 2020; Gendron et al., 2019; Mikhailov et al., 2019). In this study, members of SAR dominated the MEC in the sampled lakes. This is consistent with other studies indicating that the SAR supergroups comprise a range of parasites and bacterial grazers (Boaventura et al., 2018). Moreover, we found that biotic and abiotic variables were major factors shaping freshwater BCs, indicating that

deterministic processes are likely more of a driver than stochastic processes in freshwater lakes, at least for the lakes we sampled. However, microbial communities in lakes that were geographically close to each other were more similar than those further apart (Mantel test using Bray-Curtis dissimilarity matrix versus geographic distance; R= 0.37, P = 0.001; data not shown). The similarity between these communities may be due to dispersal and ecological drift coupled with similar environmental conditions at spatially close sampling sites. Such "isolation by distance" patterns of diversity are consistent with stochastic assembly processes. Likely the freshwater BC composition is determined by a combination of deterministic and stochastic processes; however, our analyses indicate the deterministic process are dominant in the systems we sampled. Our co-occurrence network analyses provided a broad picture of the organization and interactions of the lake microbial communities, clearly indicating fundamental differences in the complexity and connectivity of the BCs and MECs. Specifically, the co-occurrence networks showed that BCs are more connected than MECs, with Actinobacteria, Proteobacteria and SAR dominating the network nodes. These results generally support our hypothesis that deterministic biotic and abiotic factors prevail in shaping BC assembly in freshwater lakes. However, within the deterministic factors, biotic factors did not dominate the abiotic factors in shaping the BCs among our sampled lakes, not supporting our hypothesis. Finally, stochastic processes (dispersal limitation and ecological drift) are likely contributing significantly to BC composition among our lakes, as spatial factors were found to have a substantial contribution, again not in support of our prediction.

2.4.1 Factors driving community structure

Many studies have characterized how aquatic BC diversity distribution patterns reflect environmental factors (Bordez et al., 2016; Freedman and Zak, 2015). However, previous such

studies of BCs have been limited by focusing on only abiotic factors (Zhang et al., 2019b) or by studying only one location (lake) (Gendron et al., 2019). Using multiple diversity indices as well as a variety of biotic and abiotic driving factors, we characterized variation in BCs and tested whether biotic and abiotic factors were significantly and directly related to different aspects of bacterial diversity – in addition to being a driver of composition/structure. Recently, it has been accepted by many microbial ecologists that niche-based (deterministic) and neutral-based (stochastic) processes together shape local community structure (Aguilar and Sommaruga, 2020; Vellend, 2010; Zhou and Ning, 2017). However, the relative importance of deterministic and stochastic processes in controlling community assembly, succession, and biogeography in different environments is still a central debate (Stegen et al., 2015; Zhou and Ning, 2017).

Deterministic processes are the result of the selection imposed by environmental conditions (e.g., pH, TP, salt, etc.) and biotic interactions (i.e. predator-prey, mutualism, competition, etc.) which govern community structure, leading to more similar or dissimilar structure among communities (Aguilar and Sommaruga, 2020; Tripathi et al., 2018). In this study, biotic and abiotic factors explained 31 and 23 percent of BC composition variation respectively, which shows 54 percent of the variation is explained by deterministic processes. As a result, deterministic factors were more important in determining the community assembly patterns in the selected freshwater lakes than stochastic process. Similar to our study, deterministic process has been reported to be the major driver shaping BC composition in freshwater lakes (Aguilar and Sommaruga, 2020; Hanson et al., 2012; Huber et al., 2020). For example, Aguilar and Sommaruga (2020) studied one alpine oligotrophic and one subalpine mesotrophic lake. In both lakes, deterministic process was the main assembly process and explained 67% of the BC turnover (Aguilar and Sommaruga, 2020). Our analysis showed that

BC composition was significantly correlated with pH and calcium. Similar results were found in other studies showing environmental conditions (such as pH and calcium) govern community structure (Yang et al., 2019; Zhou and Ning, 2017). We also found a significant relationship between BC diversity and altitude, with high-altitude lakes having higher microbial diversity. A similar relationship was also reported in other studies that postulated that altitude reflects important changes both biotic and abiotic characteristics (Ortiz-Alvarez and Casamayor, 2016; Siles and Margesin, 2016). Moreover, the lake MEC composition explained 20% of BC beta diversity, indicating interactions between bacteria and micro-eukaryotes (i.e. predator-prey, mutualism, competition, etc.) are structuring the composition of the BC. To the best of our knowledge, only a few recent studies have included biotic factors into their analysis of BC composition and diversity, and, similar to our results, those studies also reported that biotic interactions contribute significantly to community assembly in the lake ecosystem (Gendron et al., 2019; Mikhailov et al., 2019).

In contrast, stochastic process assumes that microbial diversity (e.g., bacteria) is controlled by non-selective processes, such as ecological drift (change in the relative abundance of taxa in a location due to chance demographic fluctuations) mediated by dispersal (movement of taxa across space) (Hanson et al., 2012; Zhang et al., 2019a). This means that increasing spatial distance would result in increasing divergence in BC composition, hence BC should exhibit a distance-decay relationship indicating that community similarity decreases with increasing geographic distance (Huber et al., 2020). In our study, spatial factors explained 23% of BC composition variation (i.e., stochastic process). Moreover, our Mantel test also showed a significant a distance-decay relationship (R= 0.37, P = 0.001; see above) among the sampled lakes where lakes close to each other had more similar community patterns than those further

away (Figure 3.3 and Table 3.1). Several studies have shown that restricted dispersal among BCs can lead to distance-decay relationship, while high dispersal rates undermine the relationship (Hanson et al., 2012; Liu et al., 2020b; Zhou and Ning, 2017). Overall, spatial, chemical, geographical and biological variables all contributed to BC community variation in the selected lakes, supporting a growing body of evidence that a combination of niche-based (deterministic) and neutral processes (stochastic) affect BC composition (Huber et al., 2020; Zhou and Ning, 2017).

Although environmental and spatial factors described some of the BC composition variation, still considerable variation was left unexplained. There are two possible explanations for this; first, we may have not measured some key environmental factors that drive variation in microbial composition (for example nitrogen and carbon). Secondly, as suggested by Evans *et al*. (Evans et al., 2017), complex interactions among microbial dispersal, drift and selection are likely to alter or obscure linear relationships between environmental variables and microbial community composition. Moreover, we collected our water samples from the shore and at one depth (0-1.0 m), so while we collected from multiple locations for the large lakes, our design did not capture within-lake variation in microbial community composition. It is likely that different depth or offshore microbial communities might have considerably different community structure (e.g., (Meyerhof et al., 2016)) which may have contributed to the unexplained variation in community composition.

2.4.2 Co-occurrence networks of bacteria and microbial eukaryotes

A key component in the community assembly of complex aquatic microbial communities is how the high-level ecological processes impact the relationships and interactions among the organisms in lakes, either among domains (e.g., bacteria vs protists) or even within one domain

(different bacterial taxa). Network analyses provide insights into the patterns of interaction within and among communities, insights not possible with diversity estimates, community structure descriptions and correlation-based analyses (Ziegler et al., 2018). Our network analysis showed interactions within the BCs and MECs are mostly positive and more prevalent than between those domains. Elevated levels of positive interactions in our network coupled with strong deterministic assembly processes indicate the coexistence of complex and diverse taxonomic assemblages may be facilitated by specialized niche use (Faust and Raes, 2012; Fuhrman et al., 2015). Furthermore, we found the BC was more inter-connected relative to the MEC, a pattern previously reported (Nelligan et al., 2019; Zancarini et al., 2017). Since bacteria are also known to exhibit metabolic syntrophism (bacterial cooperation to fulfill biochemical pathways needed to access environmental nutrients), our observation of a dominance of positive network interactions among members of the BCs makes functional sense (Stubbendieck et al., 2016). While positive interactions between the MEC and BC were rare in our study, they have been reported between members of Bacteroidetes (Cytophagales and Flavobacteriales) and various phytoplankton taxa such as Haptophyceae, and Cryptophyta (Mikhailov et al., 2019), consistent with our study. Those positive interactions are generated by members of Bacteroidetes utilizing organic matter produced by the phytoplankton (Mikhailov et al., 2019). Our co-occurrence analyses showed that the microbes within a few BC phyla (e.g., Actinobacteria) exhibited abundant, significant, and predominantly positive interactions among their taxa. This pattern of interactions may indicate that, in our mature microbial communities, niche separation and absence of competition among those groups resulted in reduced competition, and hence few negative interactions. Positive interactions in microbial network analyses could be due to common favored conditions, or perhaps cooperative phenomenon such as syntrophic (eating

together). Alternatively, negative network interactions may reflect competition for limited resources, or predator–prey relationships (Fuhrman, 2009). In our network analysis, there were a few important negative interactions between members of Actinobacteria or Proteobacteria and the SAR groups. One explanation for those (rare) negative interactions may be competition for nutrients - BC and MEC growth is often limited by the availability of macronutrients such as nitrogen and phosphorus leading to possible competition scenarios (Amin et al., 2012).

The "keystone" taxa identified in this study (i.e., highly connected taxa or network hubs (Steele et al., 2011)) have disproportionately important roles in maintaining network structure through high connectivity or critical links to other network node taxa (Berry and Widder, 2014). In our study, members of Actinobacteria and Proteobacteria were identified as keystone taxa. Similar BC keystone taxa were identified in other studies, indicating they may fulfill key ecological functions such as nitrogen fixation or ammonia oxidation (Mikhailov et al., 2019; Xun et al., 2019) and hence are conserved across freshwater ecosystems. Among the microbial eukaryotes, members of the SAR supergroup (uSAR and SAR) had the highest values of betweenness and centrality, which underscores their important role in the MECs; however, they did not reach the threshold of keystone taxa. In general, targeted removal of keystone species results in rapid community change and may cause networks to disassemble, and thus, keystone taxa play an important role in providing ecosystem stability (Bissett et al., 2013).

2.5 Conclusion

A critical knowledge gap in the field of microbial ecology is the relative role of diverse environmental factors in shaping microbial composition and function. In the present study, we found important differences in the microbial (both bacterial and micro-eukaryotic) communities among our selected freshwater lakes. Using environmental (chemical, physical, and geographical

(altitude)), spatial, and biological community data for the sampled lakes, we found that chemical, geographical (altitude), spatial and biological factors were major factors associated with BC composition variation among the lakes. Interestingly, we found that both deterministic and stochastic processes together shape the BC; however, deterministic process prevail in our selected lakes. Thus, some form of environmental filtering (biotic or abiotic) generally plays a significant role in driving the assembly of bacterial communities. On the other hand, spatial factors (stochastic process) did explain a significant proportion of the variation in BC composition. Our microbial community network analyses showed that communities of bacteria were more connected relative to MEC. Moreover, BC and MEC communities were mostly isolated and interactions between BC and MEC were more rare and weaker. Highly connected taxa (keystone taxa) were identified within the bacterial networks (Actinobacteria, Proteobacteria), and while the MEC showed strongly connected node taxa, they did not constitute “keystone” status. Partitioning the relative biotic and abiotic effects on bacterial communities across lakes in a gradient of latitude, longitude and is crucial to predicting how communities will adapt to environmental changes such as climate change and determining how ecosystem dynamics may change with those shifts.

2.6 Reference

Aguilar P, Sommaruga R. The balance between deterministic and stochastic processes in structuring lake bacterioplankton community over time. Mol Ecol 2020; 29: 3117-3130.

Amin SA, Parker MS, Armbrust EV. Interactions between Diatoms and Bacteria. Microbiology and Molecular Biology Reviews 2012; 76: 667-+.

Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome 2020; 8: 103.

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

Bissett A, Brown MV, Siciliano SD, Thrall PH. Microbial community responses to anthropogenically induced environmental change: towards a systems approach. Ecology Letters 2013; 16: 128-139.

Bjorbækmo MFM, Evenstad A, Røsæg LL, Krabberød AK, Logares R. The planktonic protist interactome: where do we stand after a century of research? The ISME journal 2020; 14: 544-559.

Boaventura CM, Coelho FJRC, Martins PT, Pires ACC, Duarte LN, Uetanabaro APT, et al. Micro-eukaryotic plankton diversity in an intensive aquaculture system for production of Scophthalmus maximus and Solea senegalensis. Aquaculture 2018; 490: 321-328.

Bordez L, Jourand P, Ducousso M, Carriconde F, Cavaloc Y, Santini S, et al. Distribution patterns of microbial communities in ultramafic landscape: a metagenetic approach highlights the strong relationships between diversity and environmental traits. Molecular Ecology 2016; 25: 2258-2272.

Dai WF, Sheng ZL, Chen J, Xiong JB. Shrimp disease progression increases the gut bacterial network complexity and abundances of keystone taxa. Aquaculture 2020; 517.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011; 27: 2194-200.

Evans S, Martiny JB, Allison SD. Effects of dispersal and selection on stochastic assembly in microbial communities. ISME J 2017; 11: 176-185.

Faust K, Raes J. Microbial interactions: from networks to models. Nat Rev Microbiol 2012; 10: 538-50.

Freedman Z, Zak DR. Soil bacterial communities are shaped by temporal and environmental filtering: evidence from a long-term chronosequence. Environmental Microbiology 2015; 17: 3208-3218.

Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol 2012; 8: e1002687.

Fuhrman JA. Microbial community structure and its functional implications. Nature 2009; 459: 193-199.

Fuhrman JA, Cram JA, Needham DM. Marine microbial community dynamics and their ecological interpretation. Nat Rev Microbiol 2015; 13: 133-46.

Gendron EMS, Darcy JL, Hell K, Schmidt SK. Structure of bacterial and eukaryote communities reflect in situ controls on community assembly in a high-alpine lake. Journal of Microbiology 2019; 57: 852-864.

Gurung TB, Kagami M, Yoshida T, Urabe J. Relative importance of biotic and abiotic factors affecting bacterial abundance in Lake Biwa: an empirical analysis. Limnology 2001; 2: 19-28.

Hammer O, Harper DA, Ryan PD. Palaeontological statistics software package for education and data analysis. Palaeontologia Electronica 2001; 4.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol 2012; 10: 497-506.

Haukka K, Kolmonen E, Hyder R, Hietala J, Vakkilainen K, Kairesalo T, et al. Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. Microbial Ecology 2006; 51: 137-146.

Huber P, Metz S, Unrein F, Mayora G, Sarmento H, Devercelli M. Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. ISME J 2020; 14: 2951-2966.

Kiersztyn B, Chrost R, Kalinski T, Siuda W, Bukowska A, Kowalczyk G, et al. Structural and functional microbial diversity along a eutrophication gradient of interconnected lakes undergoing anthropopressure. Sci Rep 2019; 9: 11144.

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Bioinformatics 2011; Chapter 10: Unit 10 7.

Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: A Network Perspective. Trends Microbiol 2017; 25: 217-228.

Legendre P, Anderson MJ. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments (vol 69, pg 1, 1999). Ecological Monographs 1999; 69: 512-512.

Liu J, Meng Z, Liu X, Zhang X-H. Microbial assembly, interaction, functioning, activity and diversification: a review derived from community compositional data. Marine Life Science & Technology 2019: 1-17.

Liu J, Zhu S, Liu X, Yao P, Ge T, Zhang XH. Spatiotemporal dynamics of the archaeal community in coastal sediments: assembly process and co-occurrence relationship. ISME J 2020a.

Liu K, Liu Y, Hu A, Wang F, Chen Y, Gu Z, et al. Different community assembly mechanisms underlie similar biogeography of bacteria and microeukaryotes in Tibetan lakes. FEMS Microbiol Ecol 2020b; 96.

Liu K, Yao T, Pearce DA, Jiao N, Zeng Y, Guo B, et al. Bacteria in the lakes of the Tibetan Plateau and polar regions. Sci Total Environ 2021; 754: 142248.

Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res 2007; 35: e120.

Mateus-Barros E, de Melo ML, Bagatini IL, Caliman A, Sarmento H. Local and Geographic Factors Shape the Occupancy-Frequency Distribution of Freshwater Bacteria. Microb Ecol 2021; 81: 26-35.

McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology 2001; 82: 290-297.

McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, et al. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. Gigascience 2012; 1: 7.

Meyerhof MS, Wilson JM, Dawson MN, Michael Beman J. Microbial community diversity, structure and assembly across oxygen gradients in meromictic marine lakes, Palau. Environ Microbiol 2016; 18: 4907-4919.

Mikhailov IS, Zakharova YR, Bukin YS, Galachyants YP, Petrova DP, Sakirko MV, et al. Co-occurrence Networks Among Bacteria and Microbial Eukaryotes of Lake Baikal During a Spring Phytoplankton Bloom. Microbial Ecology 2019; 77: 96-109.

Mondav R, McCalley CK, Hodgkins SB, Frolking S, Saleska SR, Rich VI, et al. Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. Environmental Microbiology 2017; 19: 3201-3218.

Morrison JM, Baker KD, Zamor RM, Nikolai S, Elshahed MS, Youssef NH. Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). Plos One 2017; 12.

Mou XZ, Jacob J, Lu XX, Robbins S, Sun SL, Ortiz JD. 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 2013; 39: 352-357.

Nelligan C, Jeziorski A, Ruhland KM, Paterson AM, Smol JP. Long-term trends in hypolimnetic volumes and dissolved oxygen concentrations in Boreal Shield lakes of south-central Ontario, Canada. Canadian Journal of Fisheries and Aquatic Sciences 2019; 76: 2315-2325.

Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and Processes of Microbial Community Assembly. Microbiology and Molecular Biology Reviews 2013; 77: 342-356.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev 2011; 75: 14-49.

Oliphant K, Parreira VR, Cochrane K, Allen-Vercoe E. Drivers of human gut microbial community assembly: coadaptation, determinism and stochasticity. ISME J 2019; 13: 3080-3092.

Ortiz-Alvarez R, Casamayor EO. High occurrence of Pacearchaeota and Woesearchaeota (Archaea superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. Environmental Microbiology Reports 2016; 8: 210-217.

Percent SF, Frischer ME, Vescio PA, Duffy EB, Milano V, McLellan M, et al. Bacterial community structure of acid-impacted lakes: what controls diversity? Appl Environ Microbiol 2008; 74: 1856-68.

Petri B, Chaganti SR, Chan PS, Heath D. Phytoplankton growth characterization in short term MPN culture assays using 18S metabarcoding and qRT-PCR. Water Res 2019; 164: 114941.

Shade A, Kent AD, Jones SE, Newton RJ, Triplett EW, McMahon KD. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnology and Oceanography 2007; 52: 487-494.

Shahraki AH, Chaganti SR, Heath D. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. J Water Health 2019; 17: 37-49.

Shahraki AH, Chaganti SR, Heath DD. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microb Ecol 2021; 81: 1-13.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-504.

Siles JA, Margesin R. Abundance and Diversity of Bacterial, Archaeal, and Fungal Communities Along an Altitudinal Gradient in Alpine Forest Soils: What Are the Driving Factors? Microbial Ecology 2016; 72: 207-220.

Steele JA, Countway PD, Xia L, Vigil PD, Beman JM, Kim DY, et al. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. Isme Journal 2011; 5: 1414-1425.

Stegen JC, Lin X, Fredrickson JK, Konopka AE. Estimating and mapping ecological processes influencing microbial community assembly. Front Microbiol 2015; 6: 370.

Stubbendieck RM, Vargas-Bautista C, Straight PD. Bacterial Communities: Interactions to Scale. Front Microbiol 2016; 7: 1234.

Sutey P, Anderson J, Xu R, Rusak J, Pereira C, Evans J, et al. User Guide for the Submission of Water Samples. Ministry of Environment Conservation and Parks, Chemistry Laboratory Dorset, Ontario 2019; 2019.

Team RC. R: A language and environment for statistical computing. Vienna, Austria, 2013.

Thorp JH, Covich AP. Ecology and classification of North American freshwater invertebrates: Academic press, 2009.

Tripathi BM, Stegen JC, Kim M, Dong K, Adams JM, Lee YK. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. ISME J 2018; 12: 1072-1083.

Unno T. Bioinformatic Suggestions on MiSeq-Based Microbial Community Analysis. J Microbiol Biotechnol 2015; 25: 765-70.

Vellend M. Conceptual Synthesis in Community Ecology. Quarterly Review of Biology 2010; 85: 183-206.

Wallace JG, Kremling KA, Kovar LL, Buckler ES. Quantitative Genetics of the Maize Leaf Microbiome. Phytobiomes Journal 2018; 2: 208-224.

World Register of Marine Species (WoRMS). 2020.

Xue M, Guo Z, Gu X, Gao H, Weng S, Zhou J, et al. Rare rather than abundant microbial communities drive the effects of long-term greenhouse cultivation on ecosystem functions in subtropical agricultural soils. Sci Total Environ 2020; 706: 136004.

Xun W, Li W, Xiong W, Ren Y, Liu Y, Miao Y, et al. Diversity-triggered deterministic bacterial assembly constrains community functions. Nat Commun 2019; 10: 3833.

Yang J, Jiang H, Dong H, Liu Y. A comprehensive census of lake microbial diversity on a global scale. Sci China Life Sci 2019; 62: 1320-1331.

Zancarini A, Echenique-Subiabre I, Debroas D, Taib N, Quiblier C, Humbert JF. Deciphering biodiversity and interactions between bacteria and microeukaryotes within epilithic biofilms from the Loue River, France. Sci Rep 2017; 7: 4344.

Zhang CJ, Pan J, Duan CH, Wang YM, Liu Y, Sun J, et al. Prokaryotic Diversity in Mangrove Sediments across Southeastern China Fundamentally Differs from That in Other Biomes. Msystems 2019a; 4.

Zhang L, Zhao TT, Wang Q, Li L, Shen TT, Gao G. Bacterial community composition in aquatic and sediment samples with spatiotemporal dynamics in large, shallow, eutrophic Lake Chaohu, China. Journal of Freshwater Ecology 2019b; 34: 575-589.

Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev 2017; 81.

Ziegler M, Eguiluz VM, Duarte CM, Voolstra CR. Rare symbionts may contribute to the resilience of coral-algal assemblages. Isme Journal 2018; 12: 161-172.

CHAPTER 3: HOST SPECIES AND HABITAT SHAPE FISH BACTERIAL COMMUNITIES: PHYLOSYMBIOSIS BETWEEN FISH AND THEIR MICROBIOME

3.1 Introduction

Host-associated microbiomes, specifically the bacterial community (BC) present inside and on host surfaces, influence a broad range of host immunological, evolutionary, and ecological functions (Bordenstein and Theis, 2015; Woodhams et al., 2020). The microbiome definition refers to the entire microbial ecosystem, including the microorganisms (prokaryotes and eukaryotes), their genomes, and their surrounding habitats (Marchesi and Ravel, 2015) and is hypothesized to have co-evolved with its host (Doane et al., 2020). Significant research effort has focused on the importance of exogenous abiotic and biotic factors (e.g., habitat, geography, microbial biodiversity, diet) and endogenous host-related factors (e.g., genetics, physiology, immunity) in driving the composition of the microbiome (Minich et al., 2020a; Minich et al., 2020b; Riiser et al., 2020). Deterministic (endogenous and exogenous) factors are thought to be the dominant forces shaping species composition of BCs (Chiarello et al., 2019; Rothschild et al., 2018; Zhou and Ning, 2017). On the other hand, stochastic process-driven microbiome assembly (population growth, colonization, extinction, and speciation) assumes that BCs species are neutral, and community assembly is the result of stochastic dispersal and drift through which organisms are randomly lost and replaced (Heys et al., 2020; Sadeghi et al., 2021). Recent studies showed that both stochastic and deterministic processes are shaping the microbiome BC, but still a central question is the extent to which these two processes influence the host, as well as all of their associated microbes (Kohl et al., 2018; Zhou and Ning, 2017). Despite the known effects of habitat and host-specific factors on the microbiome across host taxa, very little is known about the degree of variation in fish-associated BC diversity and composition that occurs within and among species (Uren Webster et al., 2018). Moreover, to know how host-associated BCs may influence host

phenotype, and hence contribute to evolutionary processes, quantifying the degree and nature of among-host taxa microbiome variation and the systematic drivers behind it seems crucial.

We expect that hosts and their microbiomes are linked eco-evolutionarily and the microbiome composition will recapitulate the phylogeny of their host, the basis of phylosymbiosis theory (Brooks et al., 2016). Essentially, this predicts that hosts that are phylogenetically similar will have microbiomes that are more similar and *vise-versa* (Mallott and Amato, 2021). Phylosymbiosis may occur through stochastic and/or deterministic processes (Lim and Bordenstein, 2020), mirroring genetic evolution by drift and/or selection. However, patterns of phylosymbiosis are expected to be affected by exposure to habitat microbiome(s) (Lutz et al., 2019). Although most studies in which phylosymbiosis has been identified have focused on microbes inhabiting internal organs, such as the gastrointestinal tract (Brooks et al., 2016), recent work suggests that external host microbiomes (e.g., skin) can also exhibit a phylosymbiosis signal (Ross et al., 2018). In expanding the range of studied vertebrate microbiomes, questions about the range of environmental, ecological and evolutionary factors that shape gut and skin microbiomes (or more specifically, BCs), and the functions of those communities, still remained challenging. For example, similar gut BCs are found among phylogenetically related mammals but also among unrelated mammal species with similar diets (Ley et al., 2008a; Ley et al., 2008b; Muegge et al., 2011).

Most host-microbial interaction studies are mainly focused on mammalian species, with few studies focussed on other species (Pascoe et al., 2017). Teleosts encompass over one half of vertebrate diversity (Nelson et al., 2016) and are one of the most successful groups of vertebrates on Earth (Colston and Jackson, 2016). Teleosts are represented by more than 32 000 species, originated over 600 million years ago and exhibit a variety of physiologies, natural histories and

ecologies (Li et al., 2018). However, their success and species radiation perhaps would not have been possible without the help of their microbiome (Ghanbari et al., 2015). Additionally, their long history of co-evolution and symbiotic relationships with microbes (compared to mammals which evolved 160 million years ago (Ley et al., 2008a)) make them good candidates to study host-microbial interactions. However, while most published studies of fish microbiomes include the gut microbiome, few studies have included other key microbiome habitats such as skin (Krotman et al., 2020; Sylvain et al., 2020). Thus, the role the skin microbiome likely has on host health is under-studied, and, importantly for this study, the skin microbiome may be under separate selective pressures (Uren Webster et al., 2018).

The five Laurentian Great Lakes (LGLs) in North America form the largest freshwater ecosystem on the planet and have provided valuable ecosystem services for humans for centuries (Ozersky et al., 2021). The LGLs and their associated drainages include diverse ecosystems, complex trophic interactions, and mixed habitats including forests, wetlands, agricultural and urban areas, thus establishing a powerful natural ecological laboratory to study important questions about microbial ecology, host microbial interactions and co-evolution. While much work has been done on characterizing microbial communities in the LGLs (Shahraki et al., 2021; Paver et al., 2020), there is a critical need to examine the roles of host species and habitat on the microbiome of fish at the community level. Fish and microbial communities of LGLs provide a powerful study system to study the intra- and interspecific divergence of host-associated intestinal and skin BCs among members of the fish communities.

The aim of this study was to characterize fish microbiome BCs among a variety of host species across three locations to: (i) identify patterns of fish-associated bacterial communities, (ii) quantify the connectivity between water BCs and those of the fishes, and (iii) determine the extent to which

BCs among fish follow a pattern of phylosymbiosis, while correcting for the effects of location. Our analyses will allow us to test three main hypotheses, including host-related deterministic processes (host-based selection pressures on BC composition) are the main drivers of BC composition variation, and hence most BC composition variation will be found among species. We hypothesize that due to the host endogenous effects, the water BC will be distinct from the host-associated BCs due to host-based selection pressures driving BCC; however the skin BC is expected to be less controlled by host-related factors than the gut BC. Furthermore, we predict that the gut microbiome will be more strongly associated with host phylogeny than the skin microbiome, as the skin BC will be affected by the local environmental microbiome. Characterizing intra-specific variation in fish microbiomes will help managers, particularly in aquaculture settings, to effectively manipulate gut and skin BCs to promote animal health and well being. On the other hand, such information may provide wild fish population managers tools to assess fish stock status and health. Perhaps more exciting however, is investigating possible existence of phylosymbiosis in fish as more research on this can shed light on rules governing the microbial community associated with fish and their co-evolutionary dynamics with teleosts in general.

3.2 Material methods

3.2.1 Study sites

Samples were collected from three locations (Detroit River, Lake Erie and Lake Ontario)

Detroit River: The Detroit River of the LGLs is a 51 km channel that comprises the lower portion of the Huron-Erie Corridor, connecting Lake St. Clair to Lake Erie (Lapointe, 2014). Samples were collected around Fighting Island (from 42°10'56.5"N 83°06'39.9"W up to 42°14'04.2"N 83°06'32.1"W) from July 17 to August 29, 2018.

Lake Erie: Samples came from the western basin of Lake Erie, a shallow (mean 6m), warm, and productive basin (area = 26 km^2) fed by the Detroit River and several smaller tributaries draining agricultural watersheds. Samples were collected on October 18, 2018

Lake Ontario: Fish were collected from the Bay of Quinte and the eastern basin of Lake Ontario from July 31 to August 1, 2018. The Bay of Quinte is a large (254 km2), Z-shaped embayment with a history of nutrient and anthropogenic stress that feeds into the eastern basin and the upper St Lawrence River. The eastern basin is a mildly eutrophic, bathymetrically complex outlet basin of Lake Ontario with depths averaging ~20m.

3.2.2 Sample collection

Detroit River: Fish were captured using a single anode boat electrofisher (Smith-Root 5.0 GPP) set to use pulsed DC current at 60 Hz using between 30%– 60% of the range to maintain a current of 6–8 A (Klinard et al., 2018). All fish were euthanized with an overdose of tricaine methanesulfonate (MS222) following the protocols of the Ontario Ministry of Natural Resources and Forestry (OMNRF) and with Canadian Council on Animal Care policies. The skin swabs for all fishes were immediately taken by gently rubbing a sterile cotton swab (VWR™, cat: 470019-172) on almost 50% of the total surface on the right side of each fish. Dorsal, ventral, and pectoral

fin areas were swabbed for larger fish. Swab samples were placed into 2mL tubes and were stored on ice with the whole fish and transported to the Great Lakes Institute for Environmental Research (GLIER) at the University of Windsor and immediately frozen at -20°C. Within 2-4 hours after capture the fish were dissected, and gut content samples were collected and stored frozen at -20°C. Water samples (500 mL) were collected at the sampling sites and transported to GLIER for filtration and further analysis. Water samples were filtered using 0.22-µm pore size, 47 mm diameter polycarbonate membranes (Isopore™, Millipore, MA) and stored at -20ºC until DNA extraction.

Lake Erie: Fish were captured using bottom-set, graded mesh monofilament gillnets. After capture, all fish were euthanized in compliance with the protocols of the OMNRF and with Canadian Council on Animal Care policies. The skin, fish and water sampling, as well as the transportation and storage of the samples were same as for Detroit River.

Lake Ontario: Fish were captured using bottom-set, graded mesh (38-152 mm) monofilament gillnets (Yuille et al., 2015). Upon capture, all fish were euthanized following the protocol described for Detroit River and Lake Erie. Skin swab samples and gut tissue (foregut, midgut, and hindgut) with contents were taken within 2-4 hours of fish capture at the Glenora Fisheries Station, Ontario, Canada. Samples were stored in a 50 mL falcon tubes filled with 45 mL of a high salt solution (700 g/L Ammonium Sulfate, 25 mM Sodium Citrate, 20 mM Ethylenediaminetetraacetic acid, pH 5.2) for 48 hours to let the salts penetrate the samples (swabs, and gut content), and then stored at -20 ºC until DNA extraction. Water samples (500 mL) were collected and filtered at the Glenora Fisheries Station by using 0.22-µm filters pore size, 47 mm diameter and the filters kept in high salt solution until delivered to GLIER. All samples were stored at -20°C until DNA extraction.

3.2.3 Sample information

A total of 334 fish from Detroit River (n= 98, 29%), Lake Erie (n= 90, 27%), and Lake Ontario (n=146, 44%) belonging to 17 fish species were collected for microbiome characterization. The fish species included herbivores, invertivores, invertivores/carnivores, invertivores/detritivores, invertivores/herbivores, and planktivores (Table 3.1 and Supplementary Table S3.1) (Eakins, 2020; Heuvel et al., 2019). For each of the 334 fish, samples of the skin mucus (swab), and gut content (3-5 gram of gut content (hindgut) as well as gut tissue) and were collected and fork length and total weight were recorded. Gut samples were taken after carefully dissecting the fish with a new razor blade or a sterilize scissors to isolate a section comprising hindgut with both tissue and gut content. For Detroit River, Lake Erie, and Lake Ontario (based on sampling locations) a total of 4, 3, and 5 water samples were collected at the site of capture, respectively.

3.2.4 DNA Extraction, Library Construction and Sequencing

DNA was extracted from swabs as well as whole fish hindgut samples, which included content as well as the surrounding tissue (~3-5 gram) using a sucrose lysis buffer protocol as previously described (Shahraki et al., 2019). The V5 (787 F-acctgcctgccg-ATTAGATACCCNGGTAG) and V6 (1046 R-acgccaccgagc-CGACAGCCATGCANCACCT) variable regions of the 16S rRNA were selected for PCR amplification. The first and second PCR conditions were same as our previously published methods (Shahraki et al., 2019). Briefly, The PCR protocol for the first round PCR consisted of: 95 °C for 3 min followed by 28 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 m, and a final step at 72 °C for 10 m. For each 96 well PCR plate, one negative control corresponded to PCR mix with ultra-pure water instead of DNA template, was used. After first-round PCR amplification, the results were verified by

visualizing amplicons on an agarose gel. After checking and verifying first-round PCR products, the PCR solution was purified using Sera-Mag Magnetic Beads (GE, Healthcare Life Science, UK). A second round of PCR was conducted on purified PCR products to ligate the adaptor and barcode (10 -12 bp) sequences necessary for sample identification and sequencing. The second PCR was set at 94 °C for 3 min, then 8 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min, and final extension at 72 °C for 7 min. The second round PCR amplifications were visualized on an agarose gel and combined in proportions based on their estimated concentration (between 1- 5 μL for samples with strong clear to faint bands). Subsequently, the combined samples were gel extracted from an agarose gel and cleaned and purified using QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). In total 299 (89%), 330 (99%), and 10 (83%) samples for gut, skin swab and water were amplified successfully for first and second PCR. We also included eight PCR blanks (one for each 96 PCR plate) in our library. The concentration of purified PCR product mix (library) was measured on an Agilent 2100 Bioanalyzer with a High Sensitivity DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library concentration was then diluted to 60 pmol·μL^{-1} and sequenced on an Ion S5™ sequencing system using the Ion S5™ sequencing reagents and an Ion 530™ Chip (Thermo Fisher Scientific, ON, Canada).

3.2.5 Processing of 16S sequences

Two FASTQ files (one for gut and water samples and one for swab samples) were analyzed using the Quantitative Insights Into Microbial Ecology (QIIME2-2020.11) platform (Bolyen et al., 2019). The FASTQ sequence files were demultiplexed using the *cutadapt demux-single* command to remove sample barcode and primer sequences. Additionally, *cutadapt trim-single* was used to identify and remove the sequencing adapters for the demultiplexed data (Martin, 2011). The

DADA2 pipeline (*dada2 denoise-pyro*) was used to denoise single-end sequences, dereplicate and filter chimeras, followed by Amplicon Sequence Variant (ASV) picking (Callahan et al., 2016). Chimeric sequences were removed using the *removeBimeraDenovo* function with the "consensus" method with default values used, except the read truncation length was set to 270 (*p-trunc-len 270*). The two ASV tables and representative sequences were merged using *feature-table merge* and *feature-table merge-seqs*, respectively. Taxonomic classification was performed using the *feature-classifier* plugin (Bokulich et al., 2018) and the SILVA 138-99 reference database (Quast et al., 2013). This plugin supports taxonomic classification of features using the Naive Bayes method. All ASVs were aligned with mafft (Katoh et al., 2002) (via *phylogeny align-to-tree-mafft-fasttree* command) and used to construct a phylogeny with fasttree (Price et al., 2010). After quality control, chimera removal and combining the two feature tables, the table was summarized with the *feature-table summarize* command. A total of 18 147 574 sequences were obtained for 647 samples (299 gut samples, 10 water samples, 330 skin samples plus eight negative controls). The eight negative controls had zero to 10 sequence reads with no consistent taxa present. The *decontam* (Davis et al., 2018) (version 1.8.0) in the R package were used to identify the blank sample ASVs as possible contaminants in our sample BCs; however, none of the blank sample ASVs were identified as a contaminant. Thus, the negative controls were excluded from the rest of the study, and we assumed contamination was not an issue for our data set.

We used the *taxa filter-table*, to remove ASVs related to mitochondria, chloroplast and eukaryote sequences (3% of total sequence). We also removed "Unassigned" ASVs (13%), as well as Bacteria and Archaea ASVs without phylum assignment (1%), resulting in a total of 14 990 847 sequences remaining. The ASV table was rarefied to 3000 reads per sample for the alpha and beta diversity estimation because most of the rarefaction curves plateaued at ~3000 reads as variation

in read depth can introduce bias in both alpha and beta diversity indices. Samples with fewer than 3000 reads were deleted. These deleted samples included three gut and six swab samples (Lake Trout (1 gut, 2 swab samples), White Perch (1 gut), White Bass (1 gut), Blacknose shiner (1 swab), Pumpkinseed (1 swab), White Sucker (1 swab), Freshwater Drum 1 swab)). This decreased the total number of samples to 630 samples (296 gut, 324 swab and 10 water samples) (Table 1). ASVs were retained for further analysis only if they had at least 10 sequence reads in at least two samples. So, the final analysis included13 884 500 reads (93% of reads were retained) and 6 597 ASVs (15% of total ASVs (43 763)).

3.2.6 BC Alpha and Beta Diversity:

BC alpha diversity indices were calculated for each sample using QIIME alpha diversity alpha command. The calculated alpha diversity indices were Chao1 (a metric for species richness), and Faith's phylogenetic diversity (PD) (a metric that incorporates both species richness and species evenness). We estimated beta diversity as the Bray–Curtis dissimilarity matrix among all samples. The rarefied ASV table was used for the rest of analyses unless stated. Raw data sets are available at the Sequence Read Archive of NCBI with a PRJNA701818 BioProject accession number.

Table 3.1. Summary of fish species in the Great Lakes sampled for gut and skin, and water microbiome. We provide a description of their taxonomic information, normal diet of the host species and where the fish were sampled. The number of samples included in our analyses are shown for both skin and gut samples (and total number for each fish species) after quality filtering and rarefaction is shown.

Order	Family	Genus	Species	Fish species (common name)	Feeding	Detroit River		Lake Erie		Lake Ontario		T*
						G*	S*	G	S	G	S	
Acanthuriformes	*Sciaenidae*	*Aplodinotus*	*grunniens*	Freshwater Drum	Invertivore/ carnivore	1	5	17	16	10	10	59
Atheriniformes	*Atherinopsidae*	*Labidesthes*	*sicculus*	Brook silverside fish	Planktivore	7	7	-	-	-	-	14
Clupeiformes	*Clupeidae*	*Alosa*	*pseudoharengus*	Alewife	Planktivore	-	-	-	-	20	21	41

Clupeiformes	*Clupeidae*	*Dorosoma*	*cepedianum*	Americangizzard Shad	Herbivore	9	10	-	-	2	2	23
Cypriniformes	*Leuciscidae*	*Notropis*	*atherinoides*	Emerald shiner	Planktivore	5	15		-	-	-	20
Cypriniformes	*Leuciscidae*	*Notropis*	*heterolepis*	Blacknose shiner	Invertivore/ herbivore	11	8	-	-	-	-	19
Cypriniformes	*Leuciscidae*	*Cyprinella*	*spiloptera*	Spotfin shiner	Invertivore/ herbivores	12	11	-	-	-	-	23
Cypriniformes	*Catostomidae*	*Catostomus*	*commersonii*	White sucker	Invertivore/ detritivore	5	9	-	-	2	2	18
Gobiiformes	*Gobiidae*	*Neogobius*	*melanostomus*	Round goby	Invertivore	-		8	8	16	16	48
Perciformes	*Percidae*	*Sander*	*vitreus*	Walleye	Invertivore/ carnivores	-	-	19	19	6	7	51
Perciformes	*Moronidae*	*Morone*	*chrysops*	White bass	Invertivore/ carnivores	-	-	10	10	10	11	41
Perciformes	*Moronidae*	*Morone*	*americana*	White perch	Invertivore/ carnivores	-	-	21	21	11	12	65
Perciformes	*Centrarchidae*	*Lepomis*	*gibbosus*	Pumpkinseed	Invertivore/ carnivores	5	10	-	-	5	5	25
Perciformes	*Centrarchidae*	*Ambloplites*	*rupestris*	Rock Bass	Invertivore/ carnivores	3	7	-	-	3	3	16
Perciformes	*Percidae*	*Perca*	*flavescens*	Yellow perch	Invertivore/ carnivores	10	10	15	15	15	16	81
Salmoniformes	*Salmonidae*	*Salmo*	*trutta*	Brown trout	Invertivore/ carnivore	-	-	-	-	16	16	32
Salmoniformes	*Salmonidae*	*Salvelinus*	*namaycush*	Lake trout	Invertivore/ carnivores	-	-	-	-	22	22	44
-	-	-	-	Water samples	-	3	-	2	-	5	-	10

* Abbreviation: G (Gut), S (Skin), T (Total number of fish)

3.2.7 Statistical analysis of sequence variants

Fish versus Environmental (Water) BCs: To test for the effect of the environmental (water) BCs on gut and skin samples, taxonomical compositions of the BCs from the different sample types (gut, skin and water) were visualized using stacked barplots of the relative abundance of the bacteria at the phylum and family level with the online tool MicrobiomeAnalyst (Chong et al.,

2020). Moreover, differences in alpha diversity (species richness and evenness (Chao1, PD)) among the different sample types (gut vs water, as well as skin vs water samples) in all locations combined as well as within each location were statistically tested using the Kruskal-Wallis (KW) rank test, followed by a posthoc Dunn test with Bonferroni corrected *P* values in SPSS (IBM SPSS 25). Subsequently, Non-metric multidimensional scaling (NMDS) plot using the Bray Curtis distance matrix was performed using vegan (v.2.5-7) (Oksanen et al., 2013) and ggplot2 (v.3.3.5) (Wickham, 2016) packages in R (v.4.1.0) (Team, 2013) to visualize sample type clustering among the sample types (skin, gut, and water). We then used permutational analyses of variance (PERMANOVAs) using *adonis2* in the vegan (v.2.5-7) (Oksanen et al., 2013) package in R (Team, 2013) to test for significant differences in beta diversity among the sample types (gut vs water, as well as skin vs water samples).

Fish BCs: Taxonomic composition of the BCs of gut and skin samples from seventeen fish species sampled at three locations (Detroit River, Lake Erie, and Lake Ontario) were visualized using stacked barplots of the relative abundance of the bacteria at the family level using phyloseq (McMurdie and Holmes, 2013) and ggplot2 (Wickham, 2016) packages built on R (Team, 2013). Bacteria families with relative abundance less than 10% (ranged between 0-9.99%) in all samples were combined and presented as "Family <10 percent" in barplots. To test for the differential abundance of bacterial taxa between the gut and skin microbiomes, the ASV table data were aggregated to the family level in R (v.4.1.0) using phyloseq package (McMurdie and Holmes, 2013). Subsequently, the non-normalized ASV table was used for the negative binomial Wald test in DESeq2 (Love et al., 2014). Default DESeq2 setting with negative binomial generalized linear model (GLM) fitting with Wald significance tests were performed. *P*-values were adjusted for multiple testing using Benjamini Hochberg false discovery rate correction ($FDR < 0.05$)

(Benjamini and Hochberg, 1995). Bacterial taxa showing differential abundance were defined with the thresholds of FDR < 0.05 and $|\log_2$ Fold Change (FC)| > 2. Differences in alpha diversity (Chao1, PD) between gut and skin samples was statistically tested using the KW rank test (as described in Fish versus Environmental (Water) BCs section). To test for significant beta diversity differences between gut and skin samples PERMANOVAs analysis using *adonis2* in vegan (v.2.5-7) (Oksanen et al., 2013) package in R (Team, 2013).

Endogenous and exogenous factor analyses: Gut and skin samples clustering based on location and fish species were visualized using NMDS plot. Observed differences were assessed for significance with PERMANOVAs analyses using *adonis2* in vegan (v.2.5-7) (Oksanen et al., 2013) package in R (Team, 2013). To avoid type one error, fish species that were captured at only one location were excluded from the PERMANOVA. To examine the effect of biological (fish species identity, and weight) and environmental (location) factors on alpha (Chao1, PD) and beta (PCoA 1-5 (gut samples: selected axes had at least eigenvalue >2, percent of variation > 6%; skin samples: selected axes had at least eigenvalue >5, percent of variation > 3%)) diversity indices for gut and skin samples we built an independent linear mixed model (LMM) in R (v.4.1.0) using lme4 (Bates et al., 2014) package including location, and fish species as fixed factors, and fish weight as a random factor. We also $\log_{10}$ transformed weight, Chao1, and PD to meet the normality and heteroscedasticity assumptions of LMMs.

3.2.8 Host-microbiome phylosymbiosis

To test if skin and gut microbiome composition were linked with host phylogeny, cytochrome c oxidase I (*COX1*) and cytochrome b (*cytb*) sequences for the host fish species were downloaded from NCBI website (https://www.ncbi.nlm.nih.gov/gene) and combined to create an artificial sequence. The artificial sequences were aligned using MUSCLE (Edgar, 2004) on the

CIPRES Science Gateway v.3.1 (Miller et al., 2011). Subsequently, pair-wise phylogenetic distances between the species were calculated using the Kimura two-parameter model of substitution in MEGA-X (version 10.2.6) (Tamura et al., 2007). Finally, to evaluate the effect of host phylogeny on microbiome dissimilarity (phylosymbiosis), we performed Mantel tests (999 permutations) in R (version 4.0.) using the vegan (V2.5-7) package (Hammer et al., 2001) to compare the bacterial community Bray-Curtis dissimilarity matrixes (skin and gut) and fish species phylogenetic distance matrix. To do so, Bray-Curtis dissimilarity data were combined among samples within fish species.

3.3 Results

Fish versus Water BCs: Taxonomic compositions of the skin, gut and water BCs were distinct; however, the water BC was most divergent. Across all fish species, the fish microbiome was dominated at the phylum level by Proteobacteria (78 % (skin), 56% (gut)), Fusobacteriota (6% (skin), 19% (gut)), and Firmicutes (7% (skin), 18 % (gut)). However, the water sample taxa showed a different set of common taxa, with only Proteobacteria in common among the dominate taxa (Proteobacteria (38%), Actinobacteriota (27%), Bacteroidota (18%)) (Supplementary Figure S3.1). At the family level, the fish microbiome was dominated by members of *Aeromonadaceae* (gut (21%), skin (19%)), *Fusobacteriaceae* (gut (19%), skin (6%)), *Enterobacteriaceae* (gut (17%), skin (17%)), and Moraxellaceae (gut (1%), skin (20%)) (Supplementary Figure S3.2). However, at the family level, the water microbiome was dominated by *Comamonadaceae* (23%), *Sporichthyaceae* (19%), and *Chitinophagaceae* (5%) (Supplementary Figure S3.2).

Measures of alpha diversity (Chao1, PD) showed higher diversity in the water than in the fish gut and skin samples (Chao1: KW 65, P <0.0001; PD: KW 74, P <0.0001) (Supplementary Figure S3.3, Table 3.2). The post-hoc pairwise comparisons revealed that when water samples were compared against gut and skin samples, gut samples were more different (Chao1: test statistic -292, adj P <2.00E-06; PD: test statistic -316, adj P < 1.97E-07) compared to skin vs water (Chao1: test statistic -192, adj P <2.00E-03; PD: test statistic -208, adj P < 1.00E-03) samples (Supplementary Figure S3.3, Table 3.2). We also tested the differences between gut vs water, and skin vs water within each location. The differences between gut vs water compared to skin vs water were more pronounced when we performed the analysis within each location (Table 3.2).

The NMDS plot also showed clear separation between the fish microbiome BCs and the water BCs (Figure 3.1). The gut and skin BCs showed considerable overlap in the NMDS; however, the water BC was separated from the fish BCs and grouped together, indicative of different community

composition. PERMANOVA analyses confirmed the statistical significance of the NMDS clusters (PERMANOVA pseudo-F: 12.8, *P* value <0.001). Subsequently, pairwise PERMANOVA showed significant effects for the separate analyses of gut (t-value: 3.02; p-value < 0.001) and skin (F-value: 3.01; p-value < 0.001) compared to water BCs.

Table 3.2. Results of the Kruskal-Wallis H test followed by a *post hoc* Dunn test testing differences in alpha diversity indices (Chao1, and Faith's phylogenetic diversity (PD)) for gut, skin, and water samples among all the locations followed by separate tests within each location.

Variables	Diversity index	Post-hoc Pairwise Comparisons	Test Statistic	Std. Error	Adj. Sig.[a]	d	Kruskal-Wallis H
All locations	Chao1	gut-skin	-99	14.6	2.94E-11	2	63 ***[b]
		gut-water	-292	58.5	2.00E-06		
		skin-water	-192	58.4	2.00E-03		
	PD	gut-skin	-108	14.6	4.69E-13	2	74 ***
		gut-water	-316	58.5	1.97E-07		
		skin-water	-208	58.4	1.00E-03		
Lake Ontario	Chao1	gut-skin	-36	9.8	7.00E-04	2	23 ***
		gut-water	-134	37.7	1.00E-03		
		skin-water	-98	37.7	0.02		
	PD	gut-skin	-62	9.8	7.14E-10	2	52 ***
		gut-water	-160	37.7	5.00E-05		
		skin-water	-98	37.7	0.02		
Lake Erie	Chao1	gut-skin	-67	7	0.00E+00	2	76 ***
		gut-water	-85	37	0.06		
		skin-water	-18.	37	1		
	PD	gut-skin	-72	7	0.00E+00	2	88 ***
		gut-water	-97	37	0.02		
		skin-water	-24	37	1		
Detroit River	Chao1	gut-skin	7	7	1	2	8*
		gut-water	-76	27	0.01		
		skin-water	-69	27	0.03		
	PD	gut-skin	17	7	0.04	2	13**
		gut-water	-84	27	7.00E-03		
		skin-water	-66	27	0.01		

a. Significance values have been adjusted by the Bonferroni correction for multiple tests.
b. Significance codes: $0.01 < P \leq 0.05$*, $0.001 < P \leq 0.01$**, $P \leq 0.001$***

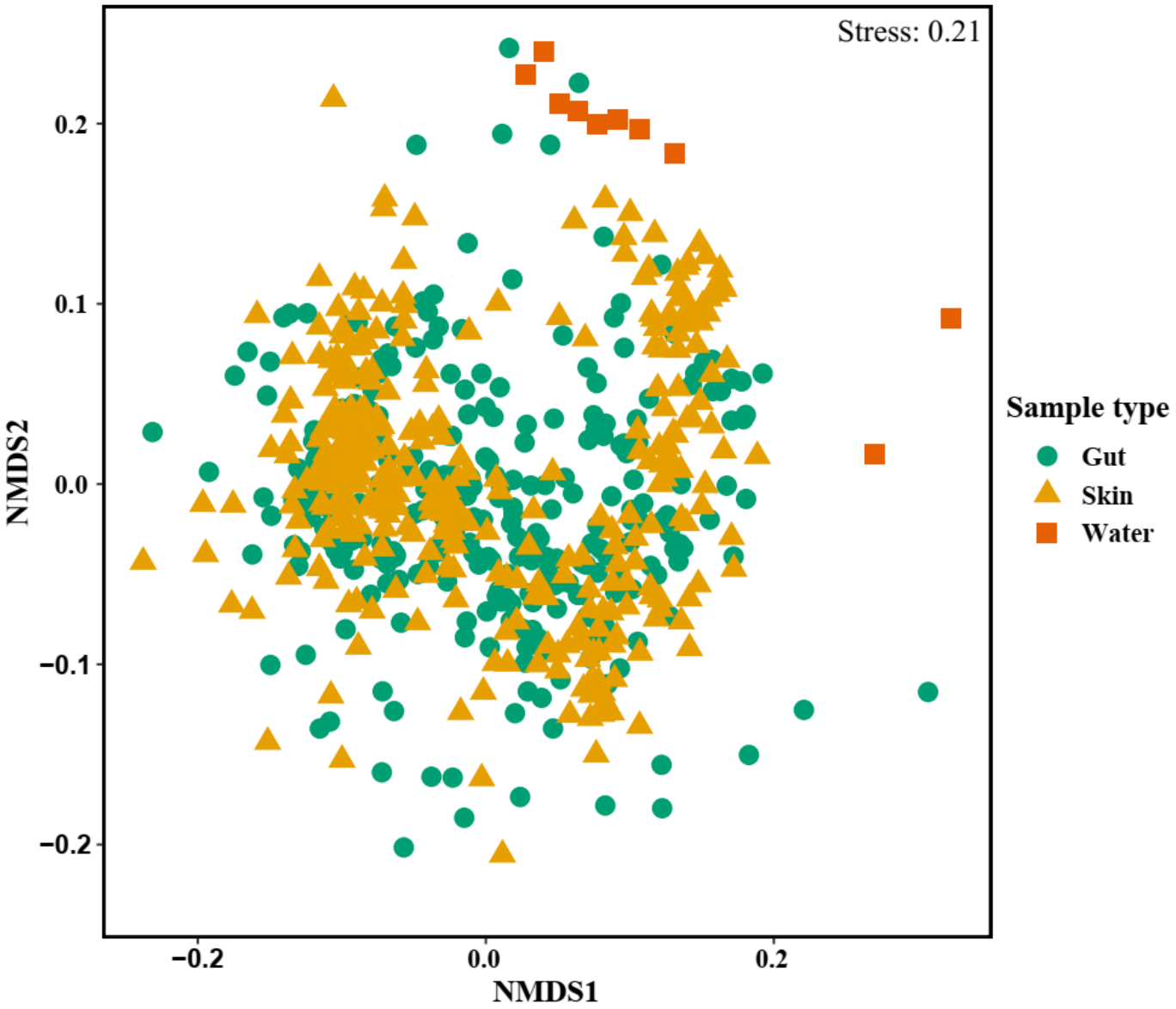


Figure 3.1. NMDS plot of BCs associated with fish gut, skin, and water samples based on Bray-Curtis dissimilarity distance. Shapes and colours are based on sample types.

Fish Skin vs Gut BCs: Taxonomic analysis of the BCs of skin and gut showed that Proteobacteria (Supplementary Figure S3.1), and specifically members of *Aeromonadaceae*, and *Enterobacteriaceae* families dominate the fish microbiomes (Figure 3.2, 3.3, and Supplementary

Figure S3.2). However, gut and skin samples exhibited different BCs (Figure 3.2, 3.3). Differential abundance analysis at the family level was used to acquire more specific insight into differences in microbiome BC composition between the gut and skin microbiomes. Comparing the family abundances between skin vs gut identified 37 bacteria families that had statistically significant differences (Supplementary Table S3.2). For example, while members of *Deinococcaceae*, *Exiguobacteraceae*, and *Moraxellaceae* family were at high abundance in skin samples, it was rare for gut samples. On the other hand, *Microbacteriaceae* and *Lachnospiraceae* were higher in gut samples (Supplementary Table S3.2).

Figure 3.2. Bar plots shoring relative abundance of gut bacterial community composition presented at the family level for all fish species at the three sample locations (DR; Detroit River, LE; Lake Erie, LO; Lake Ontario). Each bar is representative of an individual fish within that species.

Figure 3.3. Bar plot showing relative abundance of skin bacterial community composition presented at the family level for all fish species at the three sample locations (DR; Detroit River, LE; Lake Erie, LO; Lake Ontario). Each bar is representative of an individual fish within that species.

Moreover, the alpha diversity comparison between gut and skin samples showed that skin samples had higher diversity relative to gut samples (Chao1: test statistic -99, adj P <2.94E-11; PD: test statistic -108, adj P < 4.69E-13) (Table 3.2, Supplementary Figure S3.3). Although gut and skin samples showed considerable overlap in the NMDS plot (Figure 3.1), our PERMANOVA analysis showed that gut and skin samples had significantly different BCs (t-value: 4.08; p-value < 0.001). Factors affecting fish microbiome: The BC composition in the gut and skin samples of different fish species sampled at different locations often showed high levels of variation for both skin and

gut microbiomes among host species, within host species and among locations, highlighting the potential effect of location, and fish species on fish BCs (Figure 3.2, 3.3). Furthermore, NMDS analysis showed that BC composition clustered based on the host fish location (Lake Erie, Lake Ontario, Detroit River) (Figure 3.4A-B), as well as host fish species within the location for both gut and skin BCs (Figure 3.4C-D). PERMANOVA analyses supported those clustering patterns and showed highly significant effects of location, host fish species and their interaction on both the gut and skin BCs (Table 3.3).

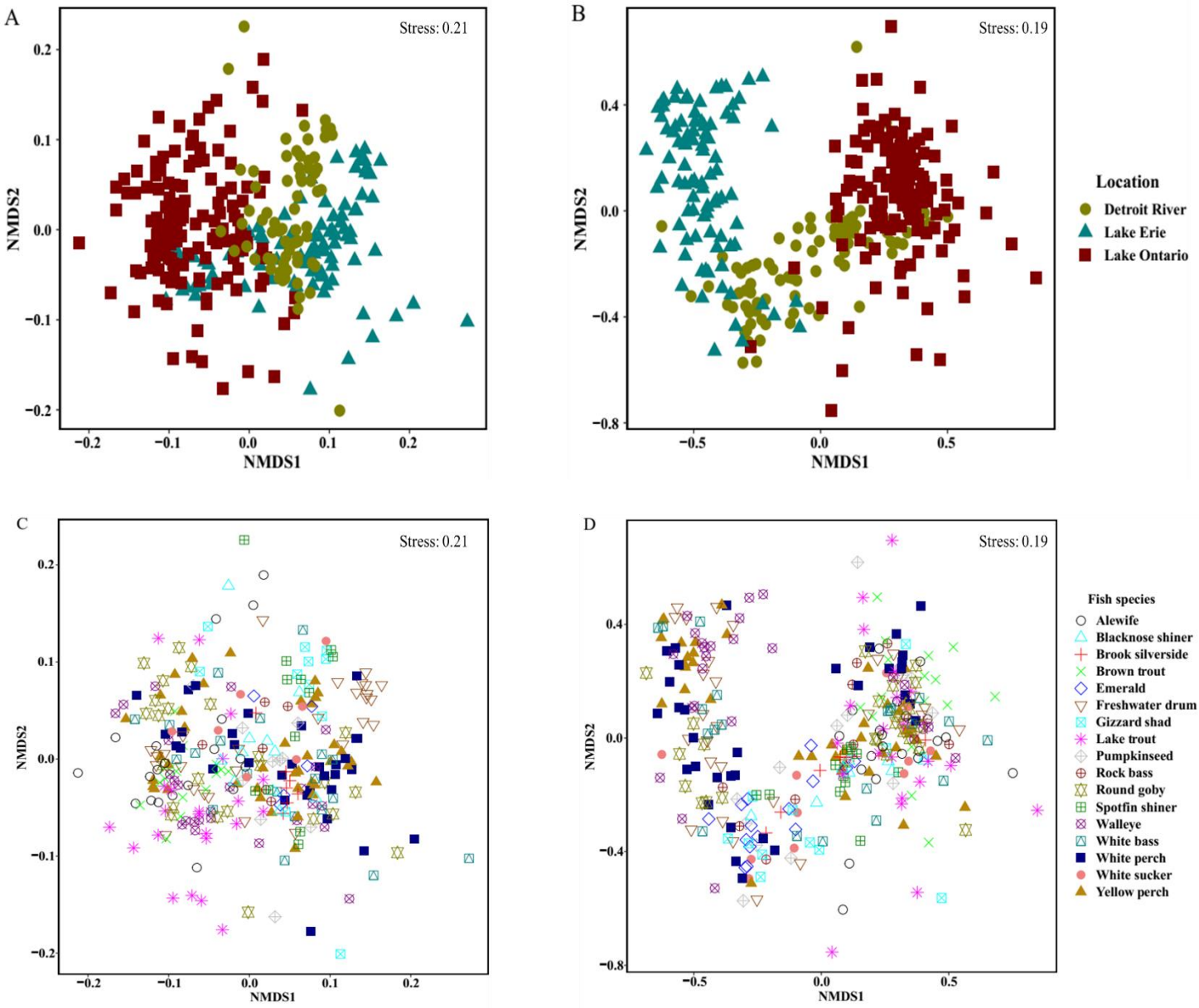


Figure 3.4. NMDS plots of skin and gut microbiome bacterial community composition based on Bray-Curtis distance matrices for 17 species of fish sampled at three locations (Detroit River, Lake Erie and Lake Ontario). The left panels A and C) show the NMDS for the gut microbiome bacterial community compositions, while the right panels (B and D) show the NMDS for the skin microbiome bacterial communities in the same fish. The top NMDS plots (A and B) are coded to show the location of capture for each fish, while the bottom panels (C and D) are coded to show the species of each fish sampled.

Table 3.3. PERMANOVA results for bacteria community beta diversity (Bray-Curtis dissimilarity matrix) testing for the effects of sample location, fish taxonomy and their interaction for both skin and gut samples. Only fish species captured at two or more locations were included in this analysis.

Variables	df	SS	R^2	F value
Gut				
Location	2	10.5	0.13	17***
Fish species	9	6.6	0.08	2.4***
Fish species* location	10	6.3	0.07	2***
Res	181	55.7	0.70	
Total	202	79.2	1.00	
Skin				
Location	2	15.2	0.17	26***
Fish species	9	5.9	0.06	2.2***
Fish species * location	10	6.3	0.07	2.1***
Res	202	59	0.68	
Total	223	86.6	1.00	

Significance codes: $P \leq 0.001$***

To examine the effect of location, fish species, and weight on the alpha (Chao1, PD) and beta diversity for BCs from different anatomical sites (skin and gut), a LMM was used. Weight was not significant for alpha and beta diversity indices and was dropped from the model. LMM results showed that all tested variables had significant effects on the fish microbiome (Table 3.4). However, the effect of location on alpha diversity was more pronounced for skin samples compared to gut samples (Table 3.4). On the other hand, fish species identity had more significant effects (F value as well as significance level) on gut bacterial diversity and richness than on skin BC diversity (Table 3.4). Moreover, location, fish species and their interaction had significant effects on beta diversity indices (PCs) of gut and skin samples (Table 3.4).

Table 3.4. LMM testing the effect of fish taxonomy, locations, interaction between locations with fish species

Gut				Skin			
Chao1				**Chao1**			
Variables	df	Sum Sq	F value	Variables	df	Sum Sq	F value
Location	2	0.1	1.4	Location	2	5.5	46.8***
Fish species	16	2.9	2.7***[a]	Fish species	16	1.3	1.3
Location * fish species	9	0.9	1.5	Location * fish species	2	2.1	3.9***
PD				**PD**			
Location	2	0.1	1.5	Location	2	4.3	55***
Fish species	16	2.1	2.85***	Fish species	16	1.1	1.8*
Location * fish species	2	0.7	1.7	Location * fish species	2	1	2.9**
PCoA axis1				**PCoA axis1**			
Location	2	41287	91.9***	Location	2	91172	215***
Fish species	16	6649	1.8*	Fish species	16	12747	3.7***
Location * fish species	2	4978	1.2*	Location * fish species	2	4034	2.1*
PCoA axis2				**PCoA axis2**			
Location	2	9441	23.5***	Location	2	7674	32.8***
Fish species	16	16052	4.9***	Fish species	16	9789	5.2***
Location * fish species	2	6030	3.3***	Location * fish species	2	4591	4.3***
PCoA axis3				**PCoA axis3**			
Location	2	6765	26.8***	Location	2	10839	49***
Fish species	16	10217	5***	Fish species	16	7167	4***
Location * fish species	2	5774	5***	Location * fish species	2	6225	6.2***
PCoA axis4				**PCoA axis4**			
Location	2	482	3.3*	Location	2	2515	18.7***
Fish species	16	10029	8.6***	Fish species	16	13735	12.7***
Location * fish species	2	820	1.2	Location * fish species	2	2897	4.7***
PCoA axis5				**PCoA axis5**			
Location	2	2991	17.5***	Location	2	462	2.6
Fish species	16	3176	2.3**	Fish species	16	6551	4.7***
Location * fish species	2	1959	2.5**	Location * fish species	2	3860	4.9***

a. Significance codes: $0.01 < P \leq 0.05$*, $0.001 < P \leq 0.01$**, $P \leq 0.001$***

3.3.1 Phylosymbiosis

Mantel tests of pairwise correlations between host phylogenetic distances and Bray Curtis BC dissimilarity values revealed a significant increase in BC dissimilarity with increasing host evolutionary distance for both gut (r= 0.18, $P < 0.05$) skin samples (r= 0.26, $P < 0.01$) supporting phylosymbiosis for fish gut and skin samples. (Figure 3.5).

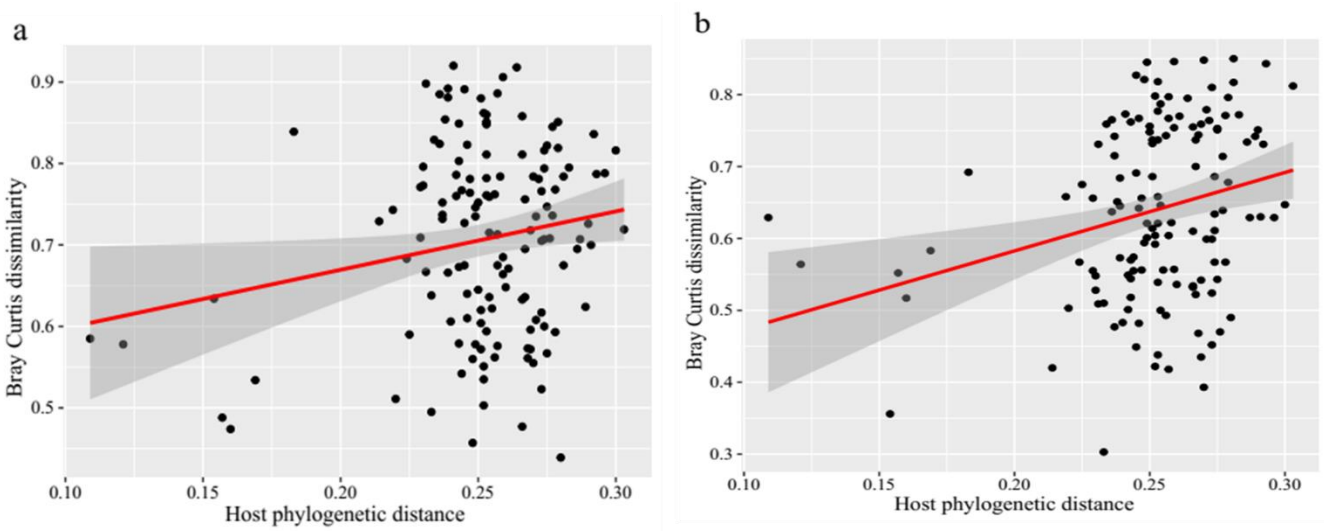


Figure 3.5. Scatterplot of host phylogenetic distance vs Bray Curtis dissimilarity matrix for both gut (a) and skin (b) samples. Samples are combined based on species identity. Host phylogeny is based on combining the sequence of *COX-1* and cytb genes.

3.4 Discussion

Diverse exogenous and endogenous factors have been reported that can drive the composition of fish microbiomes (Minich et al., 2020a; Sylvain et al., 2020; Uren Webster et al., 2018). However, the relative contributions of those factors in determining the composition of teleost fish microbiome remains poorly understood (Riiser et al., 2020; Sylvain et al., 2020). Our study evaluated how the bacterial component of the gut and skin microbiomes in 17 wild freshwater fish species sampled at three locations is shaped according to the aquatic environment, host species, and host phylogeny. While the fish microbiomes (skin and gut) were very different from the surrounding environmental microbiome, sample location (Detroit River, Lake Erie, Lake Ontario) had a strong effect on the composition of both the gut and skin BCs across all species of fish. Moreover, host species as well as host species-by-location interactions were significant but did not explain as much variation in the fish BC composition as the location effect. However, the fish species effect was phylogenetically consistent, indicative of long-term co-evolutionary relationships. We also observed significant but weak relationships between host phylogenetic distance and microbiome dissimilarity, consistent with phylosymbiosis. Overall, our results indicate that the host's habitat has a considerably greater role than host-specific selection in the assembly and composition of host-associated microbiome and must be accounted for in any assessment of host-specific effects on the microbiome.

The aquatic microbial communities are thought to be the main source for the bacterial component of the fish microbiomes, including the gut and skin (Galbraith et al., 2018). Nevertheless, even with the on-going and constant exposure to the bacteria in the surrounding aquatic environment, studies indicate that fish harbor microbiomes that are distinct from the water microbiome (Chiarello et al., 2018; Reinhart et al., 2019). Our work supported those previous findings; we showed that the BCs in the fish gut and skin microbiomes were different than in the

surrounding water. Previous studies showed that Proteobacteria, Fusobacteria, Firmicutes, and Bacteroidetes often comprise up to 95% of the fish microbiota (Li et al., 2017; Minich et al., 2020b; Sylvain et al., 2020; Uren Webster et al., 2020), and our results were consistent with those studies. This pattern is expected as Proteobacteria play an important role in the growth of fishes through nutrient cycling and the mineralization of organic compounds (Kirchman, 2002), while Firmicutes and Fusobacteria have roles in fatty acid absorption, lipid metabolism, fermentative process, and degradation of oligosaccharides in fish (Ghanbari et al., 2015). The dominant bacterial phyla in the fish microbiome in our study had some overlap with the water BC, where Proteobacteria, Actinobacteriota, and Bacteroidota were most abundant phyla in the aquatic environment (which agrees with previous studies (Krotman et al., 2020; Sylvain et al., 2020; Uren Webster et al., 2020)).

While broad comparisons at the phylum level are valuable, a more detailed differential abundance analysis at the family level provides more specific insights into BC variation between gut and skin microbiomes. Our differential abundance analysis showed fundamental differences between the BCs of skin and gut across a diverse array of host fish species. For example, *Deinococcaceae* and *Exiguobacteraceae* were generally more abundant in the skin mucus microbiome, relative to the gut. On the other hand, members of *Microbacteriaceae* and *Lachnospiraceae* were more common in the gut microbiome. Bacteria in the family *Deinococcaceae* are obligate aerobes, and have a high resistance to ionizing radiation (gamma- and/or ultra-violet (UV) radiation) (Slade and Radman, 2011). This makes it unsurprising that *Deinococcaceae* was more abundant in the skin microbiome, as skin experiences more exposure to solar radiation that the gut. Their resistance to ionizing radiation may also contribute to *Deinococcaceae* being reported in habitats associated with high levels of solar radiation, including

fish skin (Hu et al., 2021), hot springs (Ferreira et al., 1997) and rivers (Lee et al., 2017). In our study, we found a general pattern of elevated levels of *Deinococcaceae* in pelagic species (such as yellow perch, walleye) while lower levels were observed in deep water or benthic species (such as brown trout, lake trout, and round goby) (Figure 3.3).The prevalence of *Exiguobacterium* in the skin microbiome in this study is consistent with it's reported occurrence under a large range of environmental conditions (Dastager et al., 2015), including fresh water (White et al., 2018) and skin in humans (Tena et al., 2014). So perhaps one reason for the elevated abundance of *Exiguobacterium* in fish skin samples relative to the gut samples in our study is the ability of *Exiguobacterium* to thrive under high variable conditions (Vishnivetskaya et al., 2009), an important consideration for fish skin microbiomes that are exposed to considerable environmental variation relative to the gut habitat. Our detection of *Microbacteriaceae* at higher levels in the gut microbiome is consistent with previous work that reported them in various terrestrial and aquatic ecosystems (Evtushenko and Takeuchi, 2006), as well as associated with fish at various life stages (Jami et al., 2015; Liu et al., 2014; Nguyen et al., 2020). The association of *Lachnospiraceae* with the fish gut microbiomes likely reflects their important functional role as intestinal symbionts of vertebrates (Arroyo et al., 2019), acting as butyrate producers residing in the intestinal microbiome. Members of *Lachnospiraceae* have been reported in the fish gut (Escalas et al., 2021; Ricaud et al., 2018), and indeed have been shown to have a symbiotic link to their host in surgeonfish (Arroyo et al., 2019). Finally, we found substantially elevated levels of *Enterobacteriaceae* in the fish skin and gut samples, relative to the water samples (Supplementary Figure S3.1). Dominance of *Enterobacteriaceae* taxa in freshwater fish species microbiomes have been widely reported (Egerton et al., 2018; Gajardo et al., 2016; Khurana et al., 2020) likely

reflecting the importance of these bacteria for fish health and homeostasis (Huang et al., 2020; Krotman et al., 2020).

A large volume of published work shows that both endogenous and exogenous factors contribute to the teleost microbiome composition (Chiarello et al., 2019; Doane et al., 2020; Minich et al., 2020a; Riiser et al., 2020). Endogenous factors can act at the individual level to drive variation in the microbiome (e.g., via life history or health status, and host genetics), the population level (e.g., via adaptation to local selection pressures) or the species-level (e.g., via genomic variation and ancestry) (Wong and Rawls, 2012). On the other hand, abiotic (climate, water chemistry, geography, etc.) and biotic (such as water microbiome) exogenous factors can contribute to variation in microbiome composition as well (Llewellyn et al., 2016). Our analyses of alpha and beta diversity indices for both the skin and gut BCs showed that habitat (exogenous) and host fish species (endogenous) had significant effects on the microbiome BCs. A variety of studies have reported both environmental and host species effects on the gut and skin microbiome composition in fishes (Fu et al., 2020; Kim et al., 2021; Minich et al., 2020a; Sylvain et al., 2020). Generally, the skin microbiota is reported to be more affected by environmental factors than the gut microbiome (Chiarello et al., 2019; Sylvain et al., 2020), which is not surprizing, given the close contact between the aquatic environment and fish skin habitat (Guivier et al., 2020). Surrounding environments such as water and sediment are thought to be major sources of skin and gut microbiome bacteria (Wu et al., 2012; Xing et al., 2013). In this study, sample location dominated host species effects, with, as expected, a stronger effect for the skin BC (R^2 = 0.17) than for the gut BC (R^2 = 0.13). This pattern of skin versus gut effects was despite the strong divergence between the fish and water microbiomes. However, our Kruskal-Wallis also showed that skin BCs were more similar to that found in the water microbiome than the gut BCs. As the

skin is in constant direct contact with the surrounding water microbiome the skin BCs would be expected to reflect at least part of the bacteriological composition of the surrounding water (Steiner et al., 2021). By contrast, the gut microbiome is known to be strongly influenced by host-related factors and diet (Xiao et al., 2021). Our study showed approximately equal effects of host species on the gut and skin BCs, this may be due to the inclusion of multiple host species that utilized the aquatic habitat quite differently. Given that the gut microbiome habitat is highly controlled by the host's physiology, only bacterial species adapted to that environment would be expected to thrive in the gut, hence the host should have considerable effect on the gut BC composition (Sylvain et al., 2020).

Past work has shown that the gut and skin microbiome among diverse taxa are affected by host endogenous factors such as genome composition, ancestry, and diet (Boutin et al., 2014; Miyake et al., 2015; Wu et al., 2012). Our observed host fish species effects on both gut and skin microbiome BCs agree with previous research showing interindividual, population, and species variation for microbial community composition (Boutin et al., 2014), all of which were interpreted as due to host endogenous factors. Indeed, in this study we included 17 different fish species sampled at three different locations and while we found a strong host species effect, the large location effect detected may actually reflect local dietary variation among the study species (Kim et al., 2021; Sevellec et al., 2019). Thus, our reported exogenous factor (location), may actually include a component of endogenous effects. This is further supported by the significant species-by-location interaction effect; such an effect reflects species-specific sample location effects, which would logically be due to differences in host diet, at least for the gut microbiome BC. The specific mechanisms driving variation in the fish BC are still unclear, despite substantial research (including this work), likely due to complex interactions among possible mechanisms. Different

fish capture methods (electrofishing and trawling) were used in the collection gut and skin samples, therefore it is possible that capture methods may have an effect on microbial composition of skin and gut samples; however, those effects will be confounded with the location effect.

Based on phylosymbiosis, BC similarity is predicted to decrease with the increasing evolutionary divergence of host organisms. Moreover, Phylosymbiotic associations between the host and its associated microbiome can impact host fitness and support the assumption that hosts are adapted to their indigenous microbiomes. For example, hybridization between two different host species can initiate mismatches in this mutualistic relationship between the host and its associated microbiome as creating a hybrid species by combining independently evolved host genotypes may cause a breakdown in either microbial colonization and create dysbiosis or host control of the microbiome (Lim and Bordenstein, 2020). Phylosymbiosis can be driven by numerous factors, including phenotypic divergence between fishes that are phylogenetically distant (Brooks et al., 2016), co-evolution between the individual bacteria in the microbiome and the host (Miyake et al., 2016). Additionally, evolutionary processes such as selection, and drift can also shape the BC and patterns of phylosymbiosis (Mallott and Amato, 2021; O'Brien et al., 2020). Numerous studies have documented an effect of the host fish species on microbiome BCs (Doane et al., 2020; Fu et al., 2020; Huang et al., 2020). Our results showed correlations between BC composition and host fish taxonomy, and this was the same for gut and skin mucus. Previous studies showed that host-specific microbiomes are a widespread pattern in nature, occurring in many host organisms (Mallott and Amato, 2021), including within the class of mammalia (Ross et al., 2018) and fish (Doane et al., 2020; Pollock et al., 2018). Although some microbial lineages may still co-diversify with hosts, phylosymbiosis itself is not an indicator of host–microorganism co-adaptation or co-evolution. We observed strong host species-level effects on bacterial alpha

and beta diversity as well as significant correlation between phylogeny and Bray Curtis bacterial dissimilarities (phylosymbiosis) for gut and skin samples. Evidence for phylosymbiosis in non-mammalian vertebrate animal including amphibians and fish are inconsistent (Mallott and Amato, 2021). For example, some studies showed the presence of phylosymbiosis (e.g., in fish (Doane et al., 2020)), whereas others report mixed or weak evidence (Riiser et al., 2020; Sylvain et al., 2020). Moreover, our results support our hypothesis, that skin mucus was more affected by environmental condition compared to gut mucus samples and recent studies on Northern Pike from southwestern Quebec, Canada and Amazonian 6 fish populations aligns with our results (Reinhart et al., 2019; Sylvain et al., 2020) and strengthens our hypothesis. We also found phylosymbiosis was more pronounced in fish skin mucus compared to gut. This was not expected as skin microbial communities are highly connected with environmental physicochemical parameters (Sylvain et al., 2020). One possibility is that there might be some unmeasured confounding factors that are driving the correlations. However, given that our fish species ranged across 17 fish species, it is probably true phylosymbiosis between fish and gut and skin BC composition. Identifying the mechanisms that contribute to patterns of phylosymbiosis between hosts and their microbiomes, as well as the factors that reinforce or undermine it, are critical directions for a more robust evaluation of the role of co-evolution between microbiome communities and host organisms.

3.5 Conclusion

In conclusion, our findings contribute to the characterization of the modulators of microbiome composition and diversity across fish taxa. While many studies have characterized fish microbiome BCs, few of those studies included multiple species sampled in the wild. We analyzed organismal surface and internal microbiomes, as well as the aquatic environmental microbiome, across 17 freshwater fish species sampled at three locations in the Great Lakes ecosystem. Our

study design provided a robust test of the relative effects of habitat, host species and their interaction on the BCs of two key microbiomes associated with fish health and fitness. Not surprisingly, we found that the fish microbiome BCs are distinct from the aquatic environmental BC, but that sampling location had a strong effect on BC composition, nevertheless. Curiously, we found strong host species-by-location interaction effects for both skin and gut microbiome BCs, indicating that the species effects varied among the three sampled locations, possibly due to local fish diet and/or habitat-use differences. As expected, we also found a significant (based on F value), but less strong effect of host fish species on both the gut and skin microbiome BCs. Based on the host fish species effect, we tested for phylosymbiosis between the host phylogeny and both the gut and skin microbiome BC and found weak but significant correlations between host phylogenetic distance and microbial dissimilarity in the skin microbiome. This suggests that both the gut and skin BCs co-evolved with their host species, although ecological covariation also contributes substantially. Investigations of the nature of fish-microbe associations, and whether they are sustained, functional relationships or transient effects of fish and habitat associations are critical to further our understanding of the potential beneficial interactions between hosts and their microbiomes.

3.6 Reference

Arroyo FA, Pawlowska TE, Choat JH, Clements KD, Angert ER. Recombination contributes to population diversification in the polyploid intestinal symbiont Epulopiscium sp. type B. ISME J 2019; 13: 1084-1097.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 1995; 57: 289-300.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 2018; 6: 90.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019; 37: 852-857.

Bordenstein SR, Theis KR. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. PLoS Biol 2015; 13: e1002226.

Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. PLoS Biol 2016; 14: e2000225.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016; 13: 581-3.

Chiarello M, Paz-Vinas I, Veyssiere C, Santoul F, Loot G, Ferriol J, et al. Environmental conditions and neutral processes shape the skin microbiome of European catfish (Silurus glanis) populations of Southwestern France. Environ Microbiol Rep 2019; 11: 605-614.

Choi SG, Jeon SH, Lee JB, Joo ES, Lim S, Jung HY, et al. Deinococcus rubellus sp. nov., bacteria isolated from the muscle of antarctic fish. J Microbiol 2016; 54: 796-801.

Colston TJ, Jackson CR. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known and unknown. Mol Ecol 2016; 25: 3776-800.

Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 2012; 9: 772.

Dastager SG, Mawlankar R, Sonalkar VV, Thorat MN, Mual P, Verma A, et al. Exiguobacterium enclense sp. nov., isolated from sediment. Int J Syst Evol Microbiol 2015; 65: 1611-1616.

Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 2018; 6: 226.

Doane MP, Morris MM, Papudeshi B, Allen L, Pande D, Haggerty JM, et al. The skin microbiome of elasmobranchs follows phylosymbiosis, but in teleost fishes, the microbiomes converge. Microbiome 2020; 8: 93.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 2012; 29: 1969-73.

Eakins RJ. Ontario Freshwater Fishes Life History Database. 2020; Version 5.04.

Egerton S, Culloty S, Whooley J, Stanton C, Ross RP. The Gut Microbiota of Marine Fish. Front Microbiol 2018; 9: 873.

Engels C, Ruscheweyh HJ, Beerenwinkel N, Lacroix C, Schwab C. The Common Gut Microbe Eubacterium hallii also Contributes to Intestinal Propionate Formation. Front Microbiol 2016; 7: 713.

Escalas A, Auguet J-C, Avouac A, Seguin R, Gradel A, Borrossi L, et al. Ecological Specialization Within a Carnivorous Fish Family Is Supported by a Herbivorous Microbiome Shaped by a Combination of Gut Traits and Specific Diet. Frontiers in Marine Science 2021; 8: 91.

Evtushenko LI, Takeuchi M. The family microbacteriaceae. The prokaryotes 2006; 3: 1020-1098.

Ferreira AC, Nobre MF, Rainey FA, Silva MT, Wait R, Burghardt J, et al. Deinococcus geothermalis sp. nov. and Deinococcus murrayi sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. Int J Syst Bacteriol 1997; 47: 939-47.

Fu H, Zhang L, Fan C, Liu C, Li W, Cheng Q, et al. Environment and host species identity shape gut microbiota diversity in sympatric herbivorous mammals. Microb Biotechnol 2020.

Gajardo K, Rodiles A, Kortner TM, Krogdahl A, Bakke AM, Merrifield DL, et al. A high-resolution map of the gut microbiota in Atlantic salmon (Salmo salar): A basis for comparative gut microbial research. Sci Rep 2016; 6: 30893.

Ghanbari M, Kneifel W, Domig KJ. A new view of the fish gut microbiome: advances from next-generation sequencing. Aquaculture 2015; 448: 464-475.

Ginestet C. ggplot2: elegant graphics for data analysis. JOURNAL-ROYAL STATISTICAL SOCIETY SERIES A 2011; 174: 245-245.

Guivier E, Pech N, Chappaz R, Gilles A. Microbiota associated with the skin, gills, and gut of the fish Parachondrostoma toxostoma from the Rhône basin. Freshwater Biology 2020; 65: 446-459.

Heuvel CE, Haffner GD, Zhao Y, Colborne SF, Despenic A, Fisk AT. The influence of body size and season on the feeding ecology of three freshwater fishes with different diets in Lake Erie. Journal of Great Lakes Research 2019; 45: 795-804.

Heys C, Cheaib B, Busetti A, Kazlauskaite R, Maier L, Sloan WT, et al. Neutral Processes Dominate Microbial Community Assembly in Atlantic Salmon, Salmo salar. Appl Environ Microbiol 2020; 86.

Hu C, Huang Z, Liu M, Sun B, Tang L, Chen L. Shift in skin microbiota and immune functions of zebrafish after combined exposure to perfluorobutanesulfonate and probiotic Lactobacillus rhamnosus. Ecotoxicol Environ Saf 2021; 218: 112310.

Huang Q, Sham RC, Deng Y, Mao Y, Wang C, Zhang T, et al. Diversity of gut microbiomes in marine fishes is shaped by host-related factors. Mol Ecol 2020; 29: 5019-5034.

Ives JT, McMeans BC, McCann KS, Fisk AT, Johnson TB, Bunnell DB, et al. Food-web structure and ecosystem function in the Laurentian Great Lakes—Toward a conceptual model. Freshwater Biology 2019; 64: 1-23.

Jami M, Ghanbari M, Kneifel W, Domig KJ. Phylogenetic diversity and biological activity of culturable Actinobacteria isolated from freshwater fish gut microbiota. Microbiol Res 2015; 175: 6-15.

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002; 30: 3059-66.

Khurana H, Singh DN, Singh A, Singh Y, Lal R, Negi RK. Gut microbiome of endangered Tor putitora (Ham.) as a reservoir of antibiotic resistance genes and pathogens associated with fish health. BMC Microbiol 2020; 20: 249.

Kirchman DL. The ecology of Cytophaga-Flavobacteria in aquatic environments. FEMS Microbiol Ecol 2002; 39: 91-100.

Klinard NV, Fisk AT, Kessel ST, Halfyard EA, Colborne SF. Habitat use and small-scale residence patterns of sympatric sunfish species in a large temperate river. Canadian Journal of Fisheries and Aquatic Sciences 2018; 75: 1059-1069.

Kohl KD, Dearing MD, Bordenstein SR. Microbial communities exhibit host species distinguishability and phylosymbiosis along the length of the gastrointestinal tract. Mol Ecol 2018; 27: 1874-1883.

Krotman Y, Yergaliyev TM, Alexander Shani R, Avrahami Y, Szitenberg A. Dissecting the factors shaping fish skin microbiomes in a heterogeneous inland water system. Microbiome 2020; 8: 9.

Lapointe NW. Effects of shoreline type, riparian zone and instream microhabitat on fish species richness and abundance in the Detroit River. Journal of Great Lakes Research 2014; 40: 62-68.

Lee JJ, Lee YH, Park SJ, Lee SY, Park S, Lee DS, et al. Deinococcus knuensis sp. nov., a bacterium isolated from river water. Antonie Van Leeuwenhoek 2017; 110: 407-414.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science 2008a; 320: 1647-51.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 2008b; 6: 776-88.

Li W, Liu J, Tan H, Yang C, Ren L, Liu Q, et al. Genetic Effects on the Gut Microbiota Assemblages of Hybrid Fish From Parents With Different Feeding Habits. Front Microbiol 2018; 9: 2972.

Li X, Zhou L, Yu Y, Ni J, Xu W, Yan Q. Composition of Gut Microbiota in the Gibel Carp (Carassius auratus gibelio) Varies with Host Development. Microb Ecol 2017; 74: 239-249.

Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. Proc Biol Sci 2020; 287: 20192900.

Liu H, Guo X, Gooneratne R, Lai R, Zeng C, Zhan F, et al. The gut microbiome and degradation enzyme activity of wild freshwater fishes influenced by their trophic levels. Sci Rep 2016; 6: 24340.

Liu Y, de Bruijn I, Jack AL, Drynan K, van den Berg AH, Thoen E, et al. Deciphering microbial landscapes of fish eggs to mitigate emerging diseases. ISME J 2014; 8: 2002-14.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.

Lutz HL, Jackson EW, Webala PW, Babyesiza WS, Kerbis Peterhans JC, Demos TC, et al. Ecology and Host Identity Outweigh Evolutionary History in Shaping the Bat Microbiome. mSystems 2019; 4.

Mallott EK, Amato KR. Host specificity of the gut microbiome. Nat Rev Microbiol 2021.

Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome 2015; 3: 31.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 2011; 17: 10-12.

McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013; 8: e61217.

MEGA X. molecular evolutionary genetics analysis across computing platforms; S Kumar, G Stecher, M Li, C Knyaz, K Tamura. Molecular Biology and Evolution: 1547-1549.

Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway: a community resource for phylogenetic analyses. Proceedings of the 2011 TeraGrid Conference: extreme digital discovery, 2011, pp. 1-8.

Minich JJ, Petrus S, Michael JD, Michael TP, Knight R, Allen EE. Temporal, Environmental, and Biological Drivers of the Mucosal Microbiome in a Wild Marine Fish, Scomber japonicus. mSphere 2020a; 5.

Minich JJ, Poore GD, Jantawongsri K, Johnston C, Bowie K, Bowman J, et al. Microbial Ecology of Atlantic Salmon (Salmo salar) Hatcheries: Impacts of the Built Environment on Fish Mucosal Microbiota. Appl Environ Microbiol 2020b; 86.

Miyake S, Ngugi DK, Stingl U. Diet strongly influences the gut microbiota of surgeonfishes. Mol Ecol 2015; 24: 656-72.

Miyake S, Ngugi DK, Stingl U. Phylogenetic Diversity, Distribution, and Cophylogeny of Giant Bacteria (Epulopiscium) with their Surgeonfish Hosts in the Red Sea. Front Microbiol 2016; 7: 285.

Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 2011; 332: 970-4.

Nelson JS, Grande T, Wilson MVH. Fishes of the world.

Nguyen CDH, Amoroso G, Ventura T, Minich JJ, Elizur A. Atlantic Salmon (Salmo salar L., 1758) Gut Microbiota Profile Correlates with Flesh Pigmentation: Cause or Effect? Mar Biotechnol (NY) 2020; 22: 786-804.

Nyholm SV. In the beginning: egg-microbe interactions and consequences for animal hosts. Philos Trans R Soc Lond B Biol Sci 2020; 375: 20190593.

O'Brien PA, Tan S, Yang C, Frade PR, Andreakis N, Smith HA, et al. Diverse coral reef invertebrates exhibit patterns of phylosymbiosis. ISME J 2020; 14: 2211-2222.

Paver SF, Newton RJ, Coleman ML. Microbial communities of the Laurentian Great Lakes reflect connectivity and local biogeochemistry. Environ Microbiol 2020; 22: 433-446.

Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, et al. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. Nat Commun 2018; 9: 4921.

Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010; 5: e9490.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013; 41: D590-6.

Ricaud K, Rey M, Plagnes-Juan E, Larroquet L, Even M, Quillet E, et al. Composition of Intestinal Microbiota in Two Lines of Rainbow Trout (Oncorhynchus Mykiss) Divergently Selected for Muscle Fat Content. Open Microbiol J 2018; 12: 308-320.

Riiser ES, Haverkamp THA, Varadharajan S, Borgan O, Jakobsen KS, Jentoft S, et al. Metagenomic Shotgun Analyses Reveal Complex Patterns of Intra- and Interspecific Variation in the Intestinal Microbiomes of Codfishes. Appl Environ Microbiol 2020; 86.

Ross AA, Muller KM, Weese JS, Neufeld JD. Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. Proc Natl Acad Sci U S A 2018; 115: E5786-E5795.

Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature 2018; 555: 210-215.

Sadeghi J, Chaganti SR, Shahraki AH, Heath DD. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. Sci Total Environ 2021; 781: 146771.

Shahraki AH, Chaganti SR, Heath DD. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microb Ecol 2021; 81: 1-13

Shahraki AH, Chaganti SR, Heath D. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. J Water Health 2019; 17: 37-49.

Slade D, Radman M. Oxidative stress resistance in Deinococcus radiodurans. Microbiol Mol Biol Rev 2011; 75: 133-91.

Steiner K, Heasman K, Laroche O, Pochon X, Preece M, Bowman JP, et al. The microbiome of Chinook salmon (Oncorhynchus tshawytscha) in a recirculation aquaculture system. Aquaculture 2021; 534: 736227.

Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rosen GL, Knight R, et al. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Mol Ecol 2012; 21: 3363-78.

Sylvain FE, Cheaib B, Llewellyn M, Gabriel Correia T, Barros Fagundes D, Luis Val A, et al. pH drop impacts differentially skin and gut microbiota of the Amazonian fish tambaqui (Colossoma macropomum). Sci Rep 2016; 6: 32032.

Sylvain FE, Holland A, Bouslama S, Audet-Gilbert E, Lavoie C, Val AL, et al. Fish Skin and Gut Microbiomes Show Contrasting Signatures of Host Species and Habitat. Appl Environ Microbiol 2020; 86.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution 2007; 24: 1596-1599

Team RC. R: A language and environment for statistical computing; 2018, 2018.

Tena D, Martinez NM, Casanova J, Garcia JL, Roman E, Medina MJ, et al. Possible Exiguobacterium sibiricum skin infection in human. Emerg Infect Dis 2014; 20: 2178-9.

Uren Webster TM, Consuegra S, Hitchings M, Garcia de Leaniz C. Interpopulation Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity. Appl Environ Microbiol 2018; 84.

Uren Webster TM, Rodriguez-Barreto D, Castaldo G, Gough P, Consuegra S, Garcia de Leaniz C. Environmental plasticity and colonisation history in the Atlantic salmon microbiome: A translocation experiment. Mol Ecol 2020; 29: 886-898.

Vishnivetskaya TA, Kathariou S, Tiedje JM. The Exiguobacterium genus: biodiversity and biogeography. Extremophiles 2009; 13: 541-55.

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 2017; 5: 27.

White RA, 3rd, Soles SA, Gavelis G, Gosselin E, Slater GF, Lim DSS, et al. The Complete Genome and Physiological Analysis of the Eurythermal Firmicute Exiguobacterium chiriqhucha Strain RW2 Isolated From a Freshwater Microbialite, Widely Adaptable to Broad Thermal, pH, and Salinity Ranges. Front Microbiol 2018; 9: 3189.

Woodhams DC, Bletz MC, Becker CG, Bender HA, Buitrago-Rosas D, Diebboll H, et al. Host-associated microbiomes are predicted by immune system complexity and climate. Genome Biol 2020; 21: 23.

Wu S, Wang G, Angert ER, Wang W, Li W, Zou H. Composition, diversity, and origin of the bacterial community in grass carp intestine. PLoS One 2012; 7: e30440.

Xiao F, Zhu W, Yu Y, He Z, Wu B, Wang C, et al. Host development overwhelms environmental dispersal in governing the ecological succession of zebrafish gut microbiota. NPJ Biofilms Microbiomes 2021; 7: 5.

Xing M, Hou Z, Yuan J, Liu Y, Qu Y, Liu B. Taxonomic and functional metagenomic profiling of gastrointestinal tract microbiome of the farmed adult turbot (Scophthalmus maximus). FEMS Microbiol Ecol 2013; 86: 432-43.

Yuille MJ, Fisk AT, Stewart T, Johnson TB. Evaluation of Lake Ontario salmonid niche space overlap using stable isotopes. Journal of Great Lakes Research 2015; 41: 934-940.

Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev 2017; 81.

CHAPTER 4: REGULATION OF HOST GENE EXPRESSION BY GASTROINTESTINAL TRACT MICROBIOTA

4.1 Introduction

The digestive tracts of nearly all animals examined to date are inhabited by microbes. It is evident from various studies in humans (Davison et al., 2017; Dayama et al., 2020; Meisel et al., 2018) and animals (Fuess et al., 2021; Muehlbauer et al., 2021; Naya-Catala et al., 2021) that there are bidirectional interactions between the gut microbiome and the host. These interactions affect a wide range of host phenotypes such as metabolism, immunity, and physiology (McFall-Ngai et al., 2013). Recent studies have also shown that host genetics can also change and shape the gut microbiome (Lopera-Maya et al., 2022; Piazzon et al., 2020). For example, genome-wide association studies in humans have revealed an association of lactose (*LCT*) gene variants (that hydrolyzes lactose) with multiple microbial taxa in the gut microbiome (Goodrich et al., 2016; Lopera-Maya et al., 2022). The evidence for benefits provided by the gut microbiota is increasing rapidly, for example gut microbiota can improve nutrition absorption by epithelia (Krajmalnik-Brown et al., 2012), facilitate colonization resistance against pathogens (Ducarmon et al., 2019), train the immune system and perhaps even modify behaviour and mental health (Surana and Kasper, 2017). Moreover, the gut microbiota gains substantial benefits from their gut habitat (e.g., available nutrients and suitable habitat) resulting in a mutualistic relationship with the host. This provides the context for a unique coevolved process in which host and their gut microbial communities coexist and coevolve in a mutualistic adaptive scenario (Escalas et al., 2021). In fact, the growing evidence for such close interactions is often used to conclude that the host and their gut symbiont microbiomes have co-evolved (Groussin et al., 2020). Although coevolution is defined as the reciprocal adaptation process experienced by two organisms as the result of reciprocal selection pressures that they apply on each other, it is possible for the microbiome to

evolve as both a collection of individual species as well as a community response to host-mediated selection (Koskella and Bergelson, 2020).

Many studies have shown the importance of the gut microbiome in healthy and diseased host states, which ultimately affects host fitness (Bozzi et al., 2021; Manor et al., 2020; Yao et al., 2018). The gut microbiome has been shown to alter host gene expression (Davison et al., 2017; Nichols and Davenport, 2021), perhaps a mechanism for the effect of the microbiome on the host. One way that the gut microbiome can affect host gene expression patterns is by changing the expression level of the host's epigenetic modifying enzyme (e.g., histone deacetylase 3 (*HDAC3*) gene) (Nichols and Davenport, 2021). However, the mechanisms and direction of these effects is still not clear since the evidence is largely correlational. Does a change in microbiome composition cause changes in host gene expression, or vice versa? And if the microbiome can change host gene expression, which genes will be more affected? Therefore, it is important to characterize the biological mechanisms through which the host microbiome composition can cause changes in host gene expression and how this relationship is affected during pathological states.

Fish live in diverse aquatic environments, but they all harbour complex and diverse microbiomes, and those microbial communities start developing when the eggs are laid (Llewellyn et al., 2014). Among the microbial communities populating fish, the one occupying the gastrointestinal (GI) tract is the most stable, dense, and individually specific (Egerton et al., 2018; Perez et al., 2010). The bidirectional interaction between the host gut and its associated microbes may arguably be better established in fish, relative to terrestrial animals, as fish are in constant direct contact with the aquatic environmental microbiome through their gut, gills, and skin. Moreover, given the long evolutionary history of fish as a group, studying host–microbe co-evolution in fish may provide unique insights into the host–microbe relationships in general

(Montalban-Arques et al., 2015). Characterizing the mechanisms of how the gut microbiota and gene expression processes of the host interact in a symbiotic manner, will help explain the physiological processes that maintain the balance among these intricated cross-kingdom interactions and ultimately, prevent dysbiosis (Nichols and Davenport, 2021). Knowing the molecular mechanisms by which microbiota associated with host can drive host responses in the gut will lead to new strategies for preventing or treating microbiota-associated diseases.

Most studies on host- microbiome interactions are correlative or associative analyses without clearly defining cause and effect (Surana and Kasper, 2017). To move beyond correlational studies of host-microbiome interaction, microbial studies must address causation through perturbation experimental analyses (Xia and Sun, 2017). Using probiotics and antibiotics to alter gut microbiome in healthy hosts can provide valuable experimental insight into the mechanisms of host-microbiome interactions. Antibiotics can be used for antibiotic-induced microbiome depletion (AIMD), because not only can they change the structure of host gut microbial communities but also their function (Ferrer et al., 2017). Moreover, probiotics can be used to change the microbial community of the gut and stimulate the host intestinal immune system (Lee and Bak, 2011). The accepted definition of probiotics is live microorganisms that, when is given is sufficient amount, deliver a health benefit to the host (Sanders, 2008). Experimental perturbations of the gut microbial community with probiotic strains in human and animal disease treatment is well known (Azad et al., 2018). However, the effect of probiotics in healthy individuals is not as well characterized.

Chinook salmon (*Oncorhynchus tshawytscha*) are ecologically and economically an important migratory species, arguably a keystone species in some ecosystems (Meek et al., 2016). However, their populations are declining (Kareiva et al., 2000). As the gut microbiome has a

significant impact on host heath and fitness, using Chinook salmon as an animal model to describe host- microbiome interactions will not only allow us to characterize the salmon gut microbiome effect on the host, but will also have implications for the conservation and management of this species. Moreover, external fertilization of Chinook salmon eggs provides the opportunity to use controlled breeding designs with large numbers of offspring which will allow correction for host genome effects, something that is impossible for other organisms.

The direction and nature of host-gut microbiome interactions is still an open question in the study of the microbiome, although it is likely bidirectional. Experimental analyses of the mechanisms of the host-microbiome interactions are needed to shed light on the nature of host microbiome interactions. Here, our goal was to explore a broad range of host gut tissue responses induced by the experimental manipulation of the gut microbiome. Specifically, we used antibiotic and probiotic treatments, plus control fish, to manipulate the gut microbiome and then compare host hindgut tissue gene expression patterns to untreated fish (control fish). We used 16S rRNA metabarcoding of the gut bacterial community coupled with host gut tissue transcriptomics to; (i) quantify the host gut bacterial community composition changes (and holding water) resulting from the antibiotic and probiotic treatments, (ii) determine the response of the host gut tissue transcriptome to the treatments, (iii) use gene transcriptional profiling Taqman^{TM} qRT-PCR to characterize the host response to the treatment-altered gut microbiome. Given the long evolutionary history of the relationship between fish and their microbiomes, we expect strong bidirectional effects, but perhaps the effects of the microbiome on the host are more critical. We hypothesized that the host transcriptional responses to each treatment could be attributed to the abundance of specific bacterial taxa. The results obtained provide insight into the co-evolved symbiotic relationship between host and its associated microbiome that may inform future studies

exploring host-microbiome interactions and evolution. Additionally, it will help in better using microbiome manipulation (probiotics, antibiotics) to improve host health in fishes as well as humans and other animals.

4.2 Materials and methods

4.2.1 Study design

We used a domesticated line of Chinook salmon (*Oncorhynchus tshawytscha*) from Yellow Island Aquaculture Ltd, an organic salmon farm on Quadra Island, BC, Canada to create the nested breeding design. The breeding design was two sires crossed with one dam (2×1) replicated six times. Eggs were fertilized in October 2019 and the eggs for each family were separated into two groups and incubated in replicated cells of adjusted incubation trays in a flow-through system fed by well water. When the eggs hatched and the fish reached the stage of first feeding, in March 2020, individuals from replicated incubation tray cells were placed in separate 200 L tanks with water flow of 2 L per minute with continuous aeration in a standard hatchery-rearing environment (16:8 h light-dark cycle). Fish were fed ~3% of their body weight three times per day with 1.0 mm EWOS Harmony fry feed until October 24th, 2020. At that time, 5 fish per family were moved to new 200 L tanks for a total of 72 tanks (12 (families)*2 (replicates)* 3 (treatments – see below)).

4.2.2 Microbiome manipulation

We manipulated the gut microbiome of the fish in the tanks using control (untreated) feed, antibiotic treated feed and probiotic treated feed – details are described below.

Antibiotic treatment: Oxytetracycline (OTC), and Chloramphenicol (CAP), two broad spectrum antibiotics that frequently are used in aquaculture (Lai et al., 1995; Leal et al., 2019), were selected for the trial. Twenty-four tanks (for 12 (families)*2 (replicates)) were labelled as antibiotic and on the first day were treated with OTC (83 mg/kg/day concentration) (Kokou et al., 2020; Rosado et al., 2019). After six days of OTC-treated food, the fish were switched to a combination of CAP (42 mg/kg/day) (Bilandzic et al., 2012) plus the OTC for four more days for a total of 10 days of antibiotic treatment. Fish were fed three times a day at approximately 3% of their body weight.

Probiotic treatment: Twenty-four tanks (for 12 (families)*2 (replicates)) were labelled as probiotic treatment and the probiotic trial was for 10 days. The commercially available Jamieson Probiotic Complex with 60 billion colony forming units (CFU) (Jamieson Laboratories, Canada) was used for this experiment (Supplementary Table S4.1). Three probiotics capsules per 100 grams of fish feed were used. Probiotic-treated feed (3 capsules per 100 gram of feed) was coated with 10 mL of sodium alginate (1%) and 10 mL of 0.5 % calcium chloride prior to mixing with the probiotic powder. Fish were fed three times a day at approximately 3% of their body weight.

Control: Twenty-four tanks (for 12 (families)*2 (replicates)) were labelled as control group and fish were fed with regular feed without probiotic or antibiotic for ten days. Fish were fed three times a day at approximately 3% of their body weight.

4.2.3 Sampling

All fish were terminally sampled after the ten-day trial over one day (November 3, 2020). The fish were not fed during sampling. The final mean mass of the fish was 23.3 g (±7.2 SE) across all families and treatments (no treatment effect on fish body weight was detected). Three fish were dip netted from each tank and humanely euthanized immediately in an overdose solution of clove oil (Toews et al., 2019). Of the 72 tanks, four tanks (control) had 100 % mortality and those replicates were excluded from the study, bringing the total number of samples to 204 fish (72 probiotic treated fish, 72 antibiotic treated fish, 60 control fish). The sampled fish were immediately weighed and dissected, with the entire GI tract placed in a 50 mL tube with 35 mL of a highly concentrated salt buffer (ammonium sulfate, 1 M sodium citrate, 0.5 M EDTA, H_2SO_4 to bring the pH to 5.2) for preservation for later RNA extraction. Moreover, 500 mL water samples were collected from each of the tanks (N=68) before sampling the fish and filtered immediately using 0.22-micron pore size, 47 mm diameter polycarbonate filters (Isopore™, Millipore, MA).

All samples (tissue in preservative and the filters) were stored at –20 °C, until used for DNA or RNA extraction.

4.2.4 Bacterial DNA extraction and 16S rRNA gene library preparation

DNA was extracted from whole fish hindgut samples, which included content as well as the surrounding tissue (~3-5 gram) using a sucrose lysis buffer solution method previously described (Shahraki et al., 2019) and extracted DNA was subsequently stored at –20 °C, until further analysis. Additionally, the PCR conditions and 16S rRNA primer sets (1st and second PCR) were the same as those used in previously described methods (Sadeghi et al., 2021). Briefly, 16S rRNA variable regions of V5-V6 were targeted for PCR amplification and sequencing. The PCR cycle program was set at 95 °C for 3 min followed by 28 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 1 m, and a final step at 72 °C for 7 m. A second short-cycle round (7 cycles) of PCR was performed on purified first PCR products to ligate the adaptor and barcode (10 -12 bp) sequences to the amplicons as required for sample identification and sequencing. During the first and second PCR, nine samples failed amplification and 263 samples (195 gut samples, and 68 water samples) remained for the gel extraction. The second-round PCR products were gel extracted using the QIAquick Gel Extraction Kit (QIAGEN, Toronto, ON, Canada). For each 96 well PCR plate, one negative control consisting of PCR mix (of first and second PCR) with ultra-pure water instead of DNA template was included. The concentration of the pooled purified PCR product mix (i.e., the pooled library) was measured on an Agilent 2100 Bioanalyzer with a High Sensitivity DNA chip (Agilent Technologies, Mississauga, ON, Canada). The library concentration was then diluted to 60 pmol/μL and sequenced on an Ion S5™ sequencing system using the Ion S5™ sequencing reagents and an Ion 530™ Chip (Thermo Fisher Scientific, ON, Canada).

4.2.5 16S Matabarcode Sequence Data Processing

The resulting FASTQ file was analyzed using the Quantitative Insights Into Microbial Ecology (QIIME2-2020.11) platform (Bolyen et al., 2019). The FASTQ sequence file was demultiplexed and the DADA2 pipeline was used to denoise single-end sequences, dereplicate and filter chimeras. This was followed by Amplicon Sequence Variant (ASV) picking using the removeBimeraDenovo function with the "consensus" method, while default values were used for the other parameters (Callahan et al., 2016). Taxonomic classification was done through the feature-classifier plugin (Bokulich et al., 2018) using the SILVA 138-99 reference database (Quast et al., 2013). This plugin supports taxonomic classification of features using the Naive Bayes method. All ASVs were aligned with mafft (Katoh et al., 2002) and used to construct a phylogeny with fasttree (Price et al., 2010). A total of 8,820,568 sequences with 19,776 ASVs were obtained for the 267 samples (195 gut samples, 68 water samples, and 4 negative controls). The four negative controls had 1 to 7 reads and were excluded from the rest of the study. Using a taxon filter-table, ASVs related to eukaryotes, mitochondria, chloroplasts (combined ~ 1%), and unassigned (1%), were removed resulting in a total of 8,655,659 (98%) sequences remaining. Furthermore, to have a better estimate of diversity, samples with low sequence depth (less than 3000 reads), low abundance taxa (less than 10 ASVs) and ASVs that showed up in only one sample were removed. This decreased the total number of samples to 255 samples (189 gut samples, 66 water samples) with 8,217,478 sequences and 2888 ASVs. The 8 deleted samples were not related to specific treatment type or family (antibiotic treatment (one water sample), probiotic treatment (4 gut samples, and one water sample), control (two gut samples)). Alpha diversity indices of bacterial communities were calculated using the QIIME2 alpha diversity plugin. The ASV table was rarefied to 3000 reads per sample for the alpha diversity estimation because most of the rarefaction curves plateaued at 3000 reads. The calculated alpha diversity indices were Chao1 (a

metric for species richness), and Faith's phylogenetic diversity (PD) (a metric that incorporates both species richness and species evenness). A Bray-Curtis dissimilarity matrix was calculated to estimate β-diversity.

4.2.6 RNA extraction

RNA was extracted from host hindgut tissue using TRIzol® reagent (Life Technologies, Mississauga, ON, CAT=15596018) following the manufacturer's protocol. RNA was dissolved in sterile water and treated with TURBO™ DNase (Life Technologies, Mississauga, ON) to remove genomic DNA contamination and preserved at –80°C until RNA sequencing or cDNA synthesis and quantitative real-time PCR were performed (see below).

4.2.7 RNA sequencing and transcriptome assembly

A total of 18 samples from one family but different treatments (6 probiotic treated fish, 6 antibiotic treated fish and 6 control fish) were used for transcriptome analyses by RNA sequencing. Fish from one family were used to minimize differences due to genetic variability among individuals. RNA quality was assessed using the Eukaryotic RNA 6000 Nano assay on a 2100 Bioanalyzer (Agilent, Mississauga, ON). Only samples with an RIN > 7 and a 28S:18S rRNA ratio >1.0 were used for RNA sequencing. RNA sequencing libraries were prepared and sequenced at the McGill University and Genome Quebec Innovation Centre (McGill University, Montreal, QC) using the Illumina NovaSeq 6000 S4 PE100 protocol and 100-bp paired-end sequencing. To remove potentially contaminating rRNA sequences, raw sequences were filtered against eight default rRNA databases using SortMeRNA v2.1 (Kopylova et al., 2012). The non-rRNA sequences were then quality-filtered using the default parameters with Trimmomatic v0.38 (Bolger et al., 2014). This filtering step removed poor-quality sequences as well as adapter sequence that was required for RNA sequencing. The non-rRNA sequences were aligned to the Chinook salmon

(GCF_002872995.1_Otsh_v1.0; https://www.ncbi.nlm.nih.gov/assembly/GCF_002872995.1/) reference genome using the splicing aligner HISAT2 (Kim et al., 2015). FeatureCounts (Liao et al., 2014), was used to calculate the number of transcript sequence fragments assigned to each gene.

4.2.8 Differential expression gene analysis

The output from FeatureCounts was imported into DESeq2 (version '1.32.0') (Love et al., 2014) in R (R version 4.1.1) (Team, 2013) for normalization and differentially expressed genes analysis. After importing the count data into DESeq2, calling *DESeq*2 command will run several steps; briefly these steps are: (i) the estimation of size factors, controlling for differences in the counts due to varying sequencing depth of the samples (ii) the estimation of dispersion values which captures how much the counts for the gene will vary around an expected value (iii) fitting negative binomial generalized linear models for each gene and using the Wald test for significance testing of deviation from expected read count values.

4.2.9 qRT-PCR Primer/probe optimization and cDNA synthesis

Primer and probe optimization: Fifty transcripts (genes) that were significantly differentially expressed between antibiotic and probiotic groups versus control group in the DESeq2 analysis were selected for OpenArray Taqman qRT-PCR chips. The selected transcripts are listed in supplementary Table S4.2. Four endogenous control genes (β-2-microglobulin, β-Actin, ribosomal protein L13, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH)) were selected to normalise the transcription profiles of the 50 candidate transcripts for each sample. These four genes were selected based on previous studies (Geffroy et al., 2021; Limbu et al., 2018; Toews et al., 2019). Initial primers for the candidate transcripts were designed using Geneious Software v7.1.5 (http://www.geneious.com) and optimized on DNA from Chinook salmon fry. After PCR

optimization, the primers were tested on a subset of our cDNA samples with SyBr® Green Dye I (Thermo Fisher Scientific) following the manufacturer's protocol on the QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific). After testing positive for amplification of the expected sized fragment using SyBr® Green assays, new quantitative real-time PCR (qRT-PCR) primers and Taqman® probes were developed using Primer Express® Software v3.0.1 (Thermo Fisher Scientific) for all 54 genes (Supplementary Table 1). The qRT-PCR primers were developed across intron-exon boundaries, to reduce the chance of amplifying genomic DNA or pseudogene RNA and for a short amplicon length (50–100 bp). The Taqman® probe was designed for a melting temperature between 57 and 60 °C.

cDNA synthesis: RNA was quality tested on a random subset of the samples both on a 2100 Bioanalyzer to determine the RNA Integrity Number (RIN) and on 2% agarose gels. RIN values were consistent among samples, ranging between 7 and 9.8. Gel images also showed the expected rRNA bands, indicative of good RNA integrity. The RNA concentration for each sample was estimated by Spark® multimode microplate reader and NanoQuant Plate™ (Tecan, Morrisville, NC, USA). All total RNA preparations had purity values of 1.8 – 2.1 (A260/A280) with concentrations ranging from 2,000 to 5,000 ng/µL. TURBO DNA-free™ Kits (Thermo Fisher Scientific, cat. no. AM1907) were used to remove genomic DNA contamination. Total RNA treated with the TURBO DNA-free™ kit was converted to cDNA using High Capacity cDNA Kits (Applied Biosystems, Burlington, Ontario, Canada), following the manufacturer's protocol. Reverse transcriptase (RT) reactions contained 10 µL of total RNA at a concentration of 200 ng/µL, 2 µL of 10X RT random primers (Applied Biosystems), 0.8 µL of dNTP (100mM), 50 U of MultiScribe RT (Applied Biosystems) and 40 U of RNase Inhibitor (Applied Biosystems) in a 2 µL of 10X RT buffer at a final volume of 20 µL. RT reactions were incubated at 25°C for 10

min followed by 37°C for 120 min and were stopped by incubating at 85°C for 5 min. cDNA samples were stored at –20°C until further analysis.

4.2.10 OpenArray high-throughput qRT-PCR

TaqMan® OpenArray® chips from Applied Biosystems (Burlington, ON, Canada) were used to quantify transcription at the 54 genes (50 candidate and 4 endogenous control genes) on a QuantStudio 12K Flex Real-Time PCR System following the manufacturer's protocol. Each chip included 48 subarrays, in a total of 2,688 through-holes per chip. Thus 48 cDNA samples were run (two chips for 48 samples) for each of the 54 genes on each chip. A 5 µL reaction volume which includes 1.2 µL of cDNA (100ng/µL/per sample), 1.3 µL of ddH_2O and 2.5 µL of TaqMan® OpenArray® Real-Time PCR Master Mix (Applied Biosystems, Burlington, ON, Canada) was used, aliquoted across a 384-well plate and then loaded onto the TaqMan® OpenArray® chips using the OpenArray® AccuFill System. The through-holes on the chips were preloaded with the primer and probe sequences for each of the 54 genes by the manufacturer. A total of 10 chips were used for 213 cDNA samples. The samples were randomly distributed among the chips. ExpressionSuite Software (Applied Biosystems, Thermo Fisher Scientific, Carlsbad, CA, USA) was used to analyse the endogenous control genes. Of four endogenous control genes, β-Actin was selected for normalization due to high among-sample variation for the three other endogenous control genes. Subsequently, all 10 chips were normalized with the selected endogenous control gene (β-Actin) together in ExpressionSuite Software v1.0.3 (Applied Biosystems, Burlington, Ontario, Canada). Moreover, ExpressionSuite Software was used to calculate raw critical threshold (C_T) values and the relative critical threshold values (ΔC_T). Values produced by this platform are already corrected for the efficiency of the amplification (Molina-Lopez et al., 2020). We tested for replicate effect using Paired sample T test in SPSS (IBM SPSS Statistics for Windows, Version

27.0. Armonk, NY: IBM Corp). As we found no evidence for a replicate effect (P value > 0.05), C_T and ΔC_T values were averaged between the replicate and only one CT or ΔC_T value was used for each gene.

4.2.11 Statistical analysis

Treatment effects on bacterial community composition

Aquatic bacterial community composition- To test for the effect of treatment on the bacterial community composition in the hold tank water, taxonomical compositions of the bacterial community were visualized using stacked barplots and Pie charts of the relative abundance of the bacteria at the phylum and family level using the online tool MicrobiomeAnalyst (Chong et al., 2020). Moreover, differences in alpha diversity indices (Chao1 and PD) among the treatments (antibiotic, probiotic, control) for the tank water BCs were statistically tested using a Kruskal-Wallis (KW) rank test. In the case of a significant association, a *post hoc* Dunn tests with Bonferroni corrected P values were done. To visualize among-treatment divergence in the tank water BCs, a Principal-coordinate analysis (PCoA) using the Bray Curtis distance matrix was used. Subsequently, the significance of the observed clusters was assessed using permutational multivariate analysis of variance (PERMANOVA) analysis permutations in Primer 6 (v6.1.15). Pairwise comparisons were performed in cases of significant PERMANOVA among treatment groups.

Fish gut bacterial community composition- The effect of treatment on taxonomic composition of the gut sample BCs was visualized using Pie charts and stacked barplots of the relative abundance of the bacterial taxa at the family and phylum level (Chong et al., 2020). To identify the treatment, and parental effects on gut microbial community, alpha (Chao1 and PD) diversity indices for gut samples were compared using the KW rank test. Moreover, to visualize treatment effects on BC

structure, a PCoA using the Bray-Curtis distance matrix was used to generate scatterplot of the first two PCoA axes. Moreover, PERMANOVA analysis were performed in Primer 6 (v6.1.15) to test treatment and parental (dams, sires) effect on bacterial community composition. Pairwise comparisons were performed when significant differences among the treatment groups were detected to identify specific treatment effects.

Comparison between fish gut and aquatic bacterial community composition- Fish gut microbiome BC composition was compared against the environmental tank microbiome BC at both the alpha and beta diversity level. Alpha diversity measures (Chao1 and PD) of gut and water samples were compared using Mann-Whitney U test in SPSS (IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp). PCoA first and second axes were used to visualize clustering of the samples based on sample type (gut or water). Subsequently, PERMANOVA analysis were performed in Primer 6 (v6.1.15) to test sample type effect on BC composition.

Gut transcriptome response to treatment: The DESeq2 (version '1.32.0') package in R (version 4.1.1) was used to identify differentially expressed transcripts in the host gut transcriptome between any of the treatment groups in three pairwise comparisons (antibiotic vs control, probiotic vs control, antibiotic vs probiotic). The package uses a Wald test to test the significance of gene transcription differences. To identify differentially expressed transcripts, Benjamini–Hochberg corrections for multiple testing was used (false discovery rate (FDA) < 0.05). We identified differentially expressed transcripts as those genes with thresholds of FDR < 0.05 and $|\log_2 FC| > 1$. Volcano plots of differentially expressed genes between the treatments were generated by using the FC and the log-scaled adjusted *p* value using the EnhancedVolcano package (K et al., 2021) in R.

Transcriptional profile (qRT-PCR) response to treatment: The 50 selected candidate transcripts (hereafter "genes") were tested to determine which genes showed a transcription response to either of the treatments. Two genes (*cfap58*, *ubr4*) were dropped from the analysis due to failure of PCR amplification for most of the samples, thus 48 candidate genes were included for the rest of the study. To reduce the number of independent variables and to avoid over fitting the models, we used Principal Component Analyses (PCA) on the qRT-PCR data for the 48 selected genes using "prcomp" (which is a part of the R statistical analysis package) and factoextra package (1.0.7) (Kassambara and Mundt, 2017) in R (version 4.1.1). Based on a threshold of Eigenvalue > 1, and % variance explained > 2%, the first nine PC axes were selected. We used Linear mixed models (LMM) (lmerTest package (v3.1.3)) (Kuznetsova et al., 2017) in R with the selected PC axes to test for the effect of treatment (fixed effect), and the random effects of dam, sire, fish body weight, tank ID and chip effect, with all interaction terms for fixed and random factors on gene transcription patterns. Chip ID, body weight, dam, treatment×dam, treatment×sire effects were nonsignificant before FDR correction and were removed from the model. When any of the nine PCs were found to exhibit significant effects with any of the independent variables (treatments, dam, sire, body weight, tank ID, or chip effect), we examined the individual gene transcription loading values. We used *fviz_contrib* within the factoextra package (1.0.7) to identify genes with contributions to the PC greater than expected (Kassambara and Mundt, 2017). The identified genes were included in a second analysis that used LMM with the ΔC_T values for the selected genes and the same independent variables (treatment (fixed effect), and random effect of dam, sire, body weight, tank ID and chip effect), including all interaction terms for fixed and random factors. Nonsignificant factors (Chip ID, body weight, dam, and all interactions) were removed from the

model and the analysis was re-run. Lastly, a sequential Bonferroni P value correction was applied for multiple testing correction (Rice, 1989).

Correlation between gut bacterial community and host transcriptional profile: To investigate the direct effect of variation in the gut microbiome BC composition on host gene expression patterns, regression analyses were performed using the function *lm* in R (R version 4.1.1). We selected common bacterial taxa (bacteria families with more than 5% contribution to total sequence reads counts within each treatment; (7 taxonomic families) and individual genes with evidence for treatment effects (*P* value <0.1 (9 genes)) from the gene-level analysis described above. We included treatment and family ID as covariates to account for specific treatment and family effects on the relationship between host gut microbiome BC composition and gene transcription. Moreover, a sequential Bonferroni P value correction was applied for multiple testing correction (Rice, 1989). We visualized the pattern of correlation across all genes and bacterial taxa using a heatmap generated in the pheatmap function in R (in the package pheatmap).

4.3 Results

4.3.1 Impact of antibiotics and probiotics on aquatic and fish microbiome.

Microbial community associated with water: We characterized the tank water bacterial community at two taxonomic levels; the phylum and family levels. Aquatic (tank) BC diversity was divergent among the treatments, with the top 10 most abundant families making up the majority of reads. Proteobacteria were the most common phylum among all the treatment waters (control (70%), antibiotic (68%), probiotic (51%)). Bacteroidota (13 %), and Actinobacteriota (17%) were also common phylum in the control water. Moreover, in the antibiotic treated water, Firmicutes (24%) and Bacteroidota (12%) were common phyla after Proteobacteria. On the other hand, in the probiotic treated water, Bacteroidota (12%) and Firmicutes (8%) were the common phyla after Proteobacteria (Supplementary Figure S4.1). At the family level, the most common aquatic associated bacterial taxa were members of *Comamonadaceae*), a family of the Betaproteobacteria (accounting for 30%, 28%, and 35% bacterial taxa in control, antibiotic, and probiotic waters, respectively. *Mycoplasmataceae* were found in all samples, but at relatively higher abundance in antibiotic challenge water compared to probiotic and control waters. Members of *Oxalobacteraceae* were also found in all sampled tanks but at higher abundance in the probiotic and control tanks relative to the antibiotic tanks. Other notable freshwater-associated bacterial taxa at the family level were *Flavobacteriaceae*, *Pseudomonadaceae*, *Sporichthyaceae* and *Aeromonadaceae* (Figure 4.1A).

To quantify treatment effects on the aquatic BCs, alpha and beta diversity indices for water samples were compared for the three treatment groups (antibiotic, probiotic, control). Alpha diversity analysis (Chao1, PD) showed no significant differences among the groups (Chao1: KW 5, $P > 0.05$; PD: KW 3, $P > 0.05$). However, our PCoA plot showed clear separation between the

water samples based on treatments (Figure 4.1B). PERMANOVA results confirmed that the overall community structures were significantly different for the three groups (F-value 8.9; R-squared: 0.22; p-value < 0.001). Pairwise comparison also showed that the three groups are different from each other, but with the probiotic treatment group compared to antibiotics treatment group showing the highest dissimilarity (probiotic- control F: 2.17, P<0.001; probiotic- antibiotic F: 2.86, P< 0.001; control-antibiotic F: 2.77, P < 0.001). Moreover, the average dissimilarity within treatments was higher for the control tanks (73.2%) compared to our probiotic (65.4%) and antibiotic treatment tanks (61.8%).

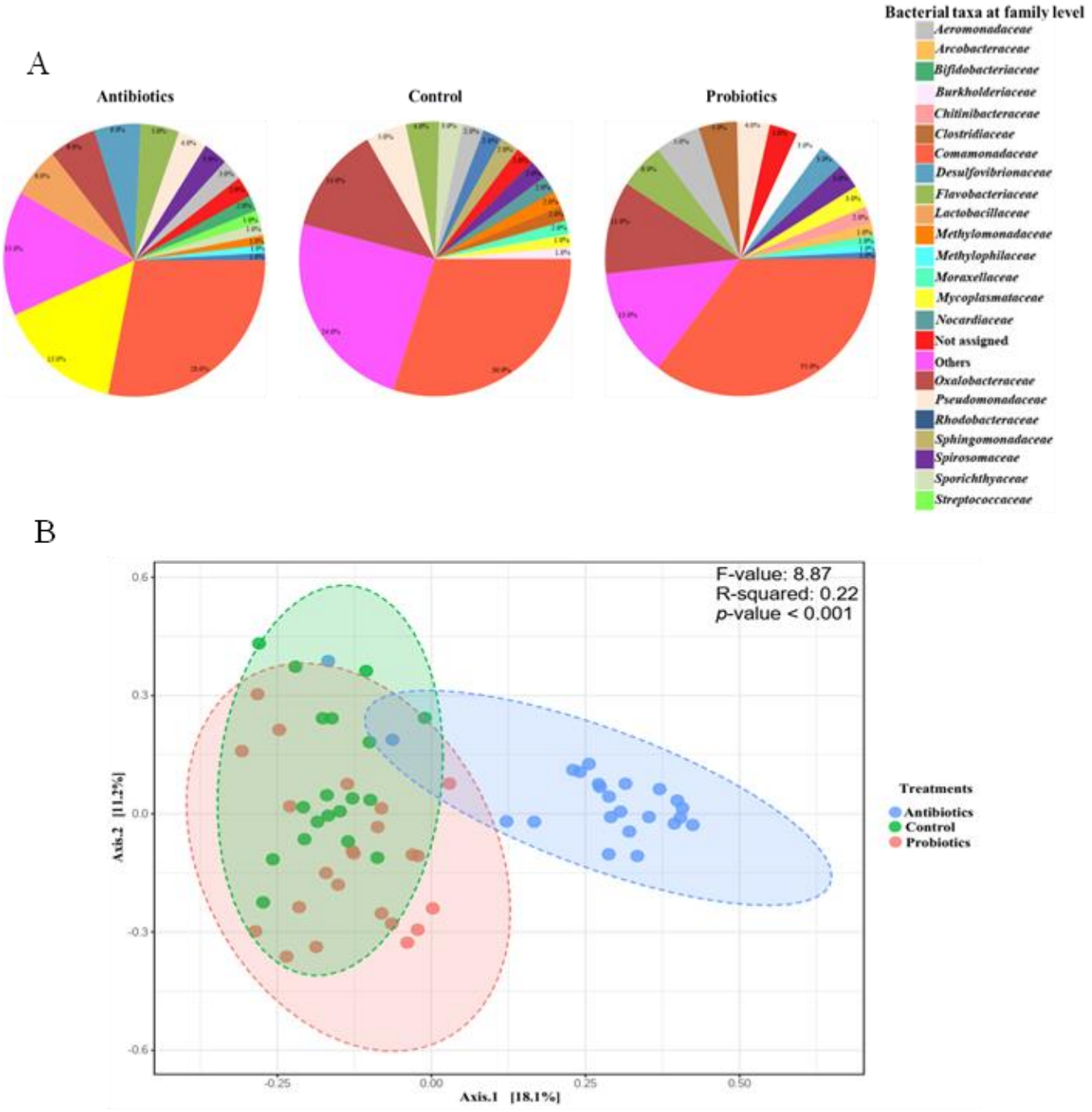


Figure 4.1. Panel A: Pie charts showing BC diversity from tank water samples for the three treatments with the relative abundances of the most abundant bacteria families (15 families). Other less abundant taxa (here less than 1%) were merged and renamed as "Others" in the pie chart. Panel B: Scatterplot of the first two axes from the PCoA of the tank water microbiome BC where the treated fish were held. Treatment is shown by colour with the 95% ellipses.

Microbial community associated with gut: Firmicutes were the most common phylum for the control and probiotic group fish (46%, and 49%, respectively). On the other hand, members of Desulfobacterota were the most common bacteria in the antibiotic treated fish gut microbiomes (Supplementary Figure S4.2A). We also compared members of Firmicutes phylum among the treatments at the family level. Within the Firmicutes phylum, *Mycoplasmataceae* was the most common gut associated bacterial taxa across all treatments, in addition to other important taxa (Supplementary Figure S4.2B). For example, control and probiotic treated fish had *Mycoplasmataceae* (control (65%), probiotic (50%)), *Streptococcaceae* (control (30%), probiotic (28%)), and *Lactobacillaceae* (control (2%), probiotic (17%)) present. However, in the antibiotic group, different families were present within Firmicutes phylum (*Mycoplasmataceae* (68%), *Streptococcaceae* (14%), and *Leuconostocaceae* (5%) (Supplementary Figure S4.2B). At the family level, the most common gut associated bacterial taxa across all treatment groups were members of *Desulfovibrionaceae* (related to Desulfobacterota phylum) and *Mycoplasmataceae* (Figure 4.2A). While *Streptococcaceae had* high relative abundances in control group, samples *in* probiotic groups had high relative abundances *Lactobacillaceae.* Moreover, members of *Pseudomonadaceae* had high relative abundances in antibiotic group (Figure 4.2A). Unlike in the tank water microbiome, *Mycoplasmataceae* was higher in the control and probiotic groups compared to the antibiotic group. We also found two important fish associated pathogens, *Enterovibrio* and *Photobacterium,* in the fish gut microbiome, but at low abundance.

To identify the treatment and parental (dams and sires) effects on the gut microbial BC, alpha diversity indices for gut samples were compared. Alpha diversity analysis (Chao1, PD) for the gut microbiome BC showed no significant differences among the treatments (Chao1: KW 2.8, $P > 0.05$; PD: KW 3.2, $P > 0.05$), sires (Chao1: KW 6.9, $P > 0.05$; PD: KW 6.8, $P > 0.05$), and

dams (Chao1: KW 5.3, $P > 0.05$; PD: KW 8.9, $P > 0.05$) effects. Beta diversity variation was also explored using Bray Curtis distance matrices and a PCoA plot. The PCoA plot showed clear separation among the samples based on treatments (Figure 4.2B). PERMANOVA results confirmed that the overall BC structures were significantly different among the treatments (Table 1). Treatment alone had the highest influence on the gut microbial community (Pseudo-F:6.1, *P* value < 0.05). Pairwise comparisons also showed that the three treatment groups exhibit significant difference in beta-diversity, with the probiotic versus control treatment samples showing the highest dissimilarity (probiotic- control F: 3.01, $P<0.001$; probiotic- antibiotic F: 2.85, $P< 0.001$; control-antibiotic F: 1.52, $P < 0.05$). Moreover, the average within treatment group dissimilarity was higher for the control gut BC (82.2%) than the probiotic (77%) and antibiotic treatments (80.5%), indicating that the control group had higher diversity than the other two groups. Dams alone did not have significant effects. However, sires had marginal significant effects on BC structures (Table 4.1).

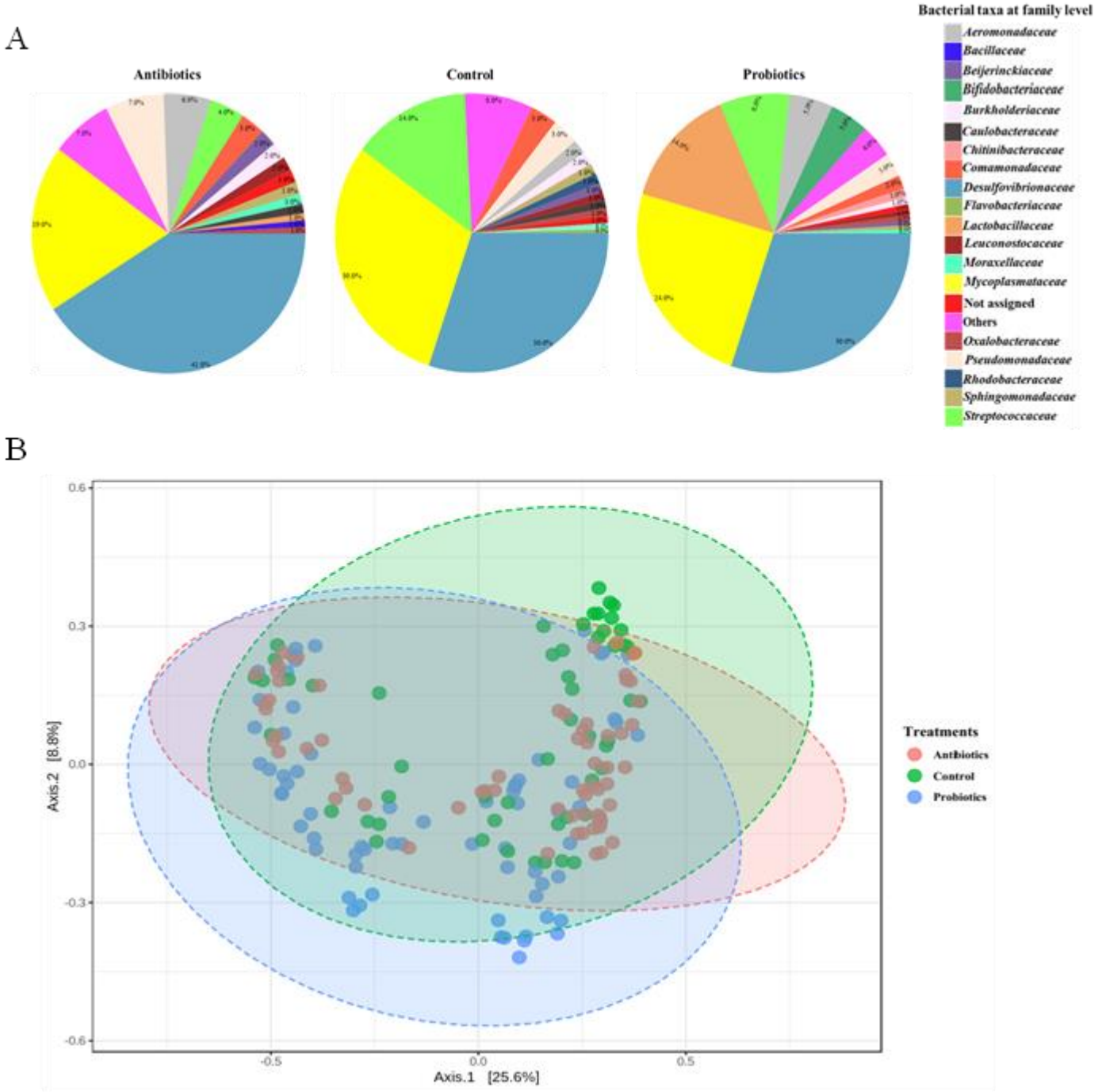


Figure 4.2. Panel (A) Pie charts showing BC diversity from Chinook salmon hindgut samples for the three treatments with the relative abundances of the most abundant bacteria families (15 families). Other less abundant taxa (here less than 1%) were merged and renamed as "Others" in the pie chart. Panel (B) Scatterplot of the first two axes from the PCoA of the Chinook salmon gut BC. Treatment is shown by colour with the 95% ellipses.

Table 4.1. Multivariate statistical testing (PERMANOVA) of effects of treatment, dams, and sires (nested within dams) on microbial community beta diversity (Bray-Curtis dissimilarity matrix).

Source	df	SS	MS	Pseudo-F	P(perm)
Treatment	2	41246	20623	6.1	**0.001**
Dams	5	23505	4700	1.1	0.22
Sires (Dams)	6	24259	4043	1.3	0.06
Res	151	4.6	3107.6	-	-
Total	186	6.5	-	-	-

Association between gut and aquatic microbial community: We evaluated the relationship between the tank water microbiome BC and the fish gut microbiome BC. Chao1 and PD (diversity measures) showed significant differences in the species richness of the two sample types; overall, diversity was significantly higher in the water samples than gut samples ($P < 0.001$, Mann-Whitney U test: 2191.5). The PCoA plot (Figure 4.3) showed clear separation between the gut and water samples. Moreover, PERMANOVA test also revealed that the clusters showed in PCoA plot were significantly different (Pseudo-F: 39.6, *P* value< 0.05).

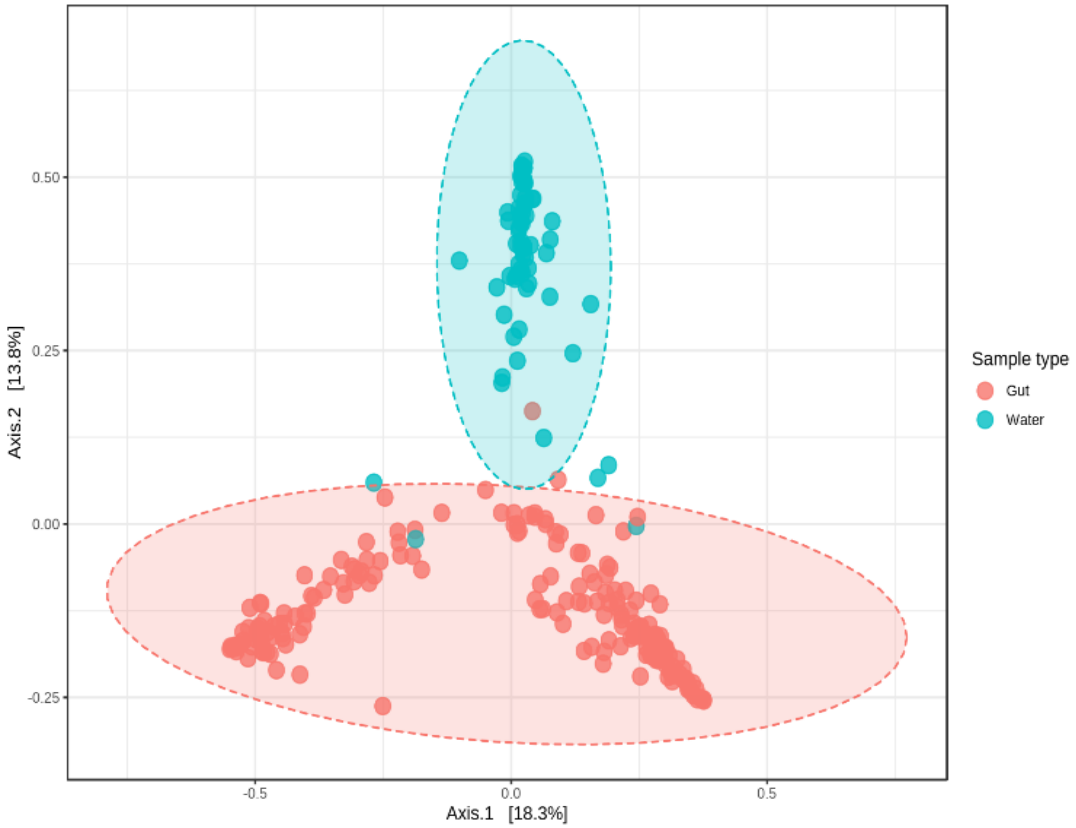


Figure 4.3. Scatterplot of the first two axes from the PCoA of the Chinook salmon gut as well as water BC. Sample type is shown by colour with the 95% ellipses.

4.3.2 Treatment effects on the host gut transcriptome

To determine if antibiotic and probiotic-induced changes in the microbiome led changes in the host gut transcriptome, RNA-Seq was used to determine host transcript levels in the hindgut. Pairwise treatment comparisons resulted in 96 (control vs antibiotic; 35 control upregulated and 61 control downregulated), 105 (control vs probiotic; 61 control upregulated, and 44 control downregulated), 120 (antibiotic vs probiotic; 84 antibiotic upregulated, and 36 antibiotic downregulated) transcripts that were differentially expressed among treatments (Benjamini-Hochberg false-discovery rate (BH FDR) 0.1, $|\log_2 FC| > 0.25$). However, for selecting candidate genes for the OpenArray high-throughput qRT-PCR analyses, we took a conservative approach and we only selected genes with transcripts that were significantly

expressed at $|\log_2 FC| > 1$ and FDR P value < 0.05 (Figure 4.4). This decreased the differentially expressed transcripts to 29 (control vs antibiotic), 29 (control vs probiotic), and 27 transcripts (antibiotic vs probiotic) (Supplementary Table S4.3). For the control versus antibiotic group comparisons, the selected genes related to cellular process (e.g., cell activation, cell communication, cell cycle, and cell death) were upregulated and genes related to metabolism and response to stimuli and stress were downregulated in antibiotic group (Supplementary Table S4.3). While in control versus probiotic group genes related to regulation of different functions (intracellular protein transport, angiogenesis, transmembrane transporter, cell adhesion, negative regulation of apoptotic process) were downregulated and genes related to post-translation modifications were highly expressed (Supplementary Table S4.3). Moreover, when we compared antibiotic against probiotic group genes related to cellular process (mostly apoptotic process) were highly expressed in antibiotic group while genes related to cell adhesion, regulation of transcription (tcf12) were highly expressed in probiotic group (Supplementary Table S4.3).

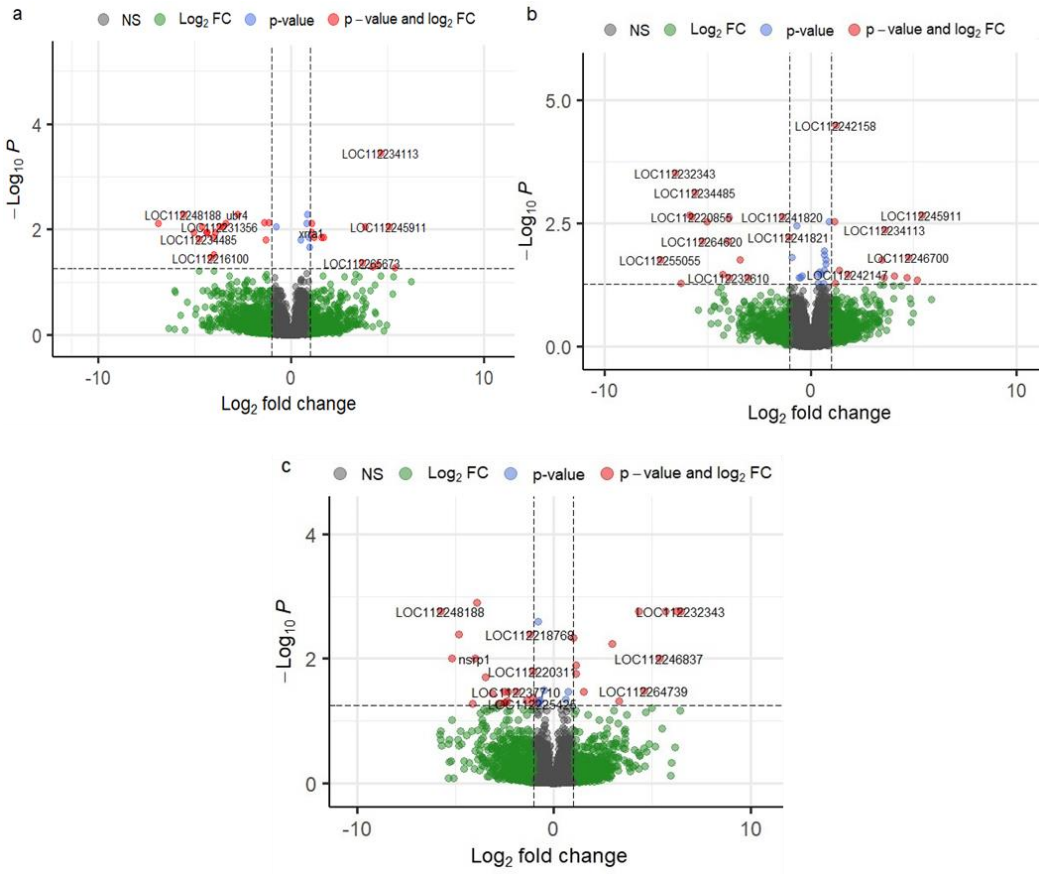


Figure 4.4. Volcano plots of differentially expressed transcripts (genes) between (a) control vs antibiotic (positive FC = downregulation of antibiotic and negative FC = upregulation of antibiotic), (b) control vs probiotic (positive FC = down regulation of probiotic, negative FC = upregulation of antibiotic) and (c) probiotic vs antibiotic (positive FC= upregulation of probiotic, negative FC = upregulation of antibiotic). X-axis indicates the FC (log scaled), whereas the Y-axis shows the p values (log scaled). Each symbol represents a different gene, and symbols colour means different criteria based on p value and FC threshold; gray (NS, not significant), green (log fold >1), blue (p value < 0.05), red ($|\log_2 FC| > 1$ and FDR P value < 0.05). FDR p value< 0.05 is considered as statistically significant, whereas ($|\log_2 FC| > 1$ as the threshold.

4.3.3 OpenArray high-throughput qRT-PCR

The LMM analysis showed PCs 4, 5, 6, 7 and 9 were significantly affected by treatment (Table 4.2). We identified only those genes whose contributions to the significantly affected principal component axes were important (Supplementary Figure S4.3) and selected them for analyses. In our analysis we also included tank, body weight, and OpenArray chip ID as random effects to correct for possible technical, environmental, and body size effects. Chip and body

weight were not significant for any of the genes and were dropped from our analyses. Sire effects (nested within dam) was not significant effect after FDR correction. Moreover, tank effect was observed for only one gene (*anxa1*, $p < 0.05$) before FDR correction. Moreover, we found no significant effect for dam-by-treatment or sire-by-treatment interactions. When considered FDR correction into our model *aifm3*, *manf*, and *prmt3* still showed significant effect for treatment (Table 4.3).

Table 4.2: LMM model of PC1-9 (Eigenvalue > 1, and % variance explained > 2%) on the qRT-PCR data for the 48 selected genes test for the effect of treatment.

PCA axes	Type III Sum of Squares	df	Mean Square	F	Sig.
PC1	64.65	2	32.325	1.7583	0.1752
PC2	5.1956	2	2.5978	0.445	0.6431
PC3	7.9187	2	3.9594	1.4297	0.285
PC4	27.45	2	13.725	9.592	0.0002274 ***
PC5	31.933	2	15.967	14.448	1.038e-05 ***
PC6	12.648	2	6.3241	5.9915	0.009771 **
PC7	12.193	2	6.0965	4.4468	0.01295 *
PC8	3.1608	2	1.5804	1.6472	0.2324
PC9	10.621	2	5.3108	5.9366	0.00316 **

a. Significant codes: $0.01 < P \leq 0.05$*, $0.001 < P \leq 0.01$**, $P \leq 0.001$***

Table 4.3. Results of the LMM analysis for significance levels for treatment, dam, sire (nested in dam), tank (nested in sire nested in dam) effects for each. Body weight, dam, treatment×dam, treatment×sire effects were nonsignificant before FDR correction and were removed from the model. Treatment was considered as fixed effects, with body weight, dam, and sire effects as random effects. The dependent variable was log transformed ΔC_T.

Genes	Probiotics vs Control	Antibiotics VS Control	Treatment	Sire (nested within dam)	Tank (sire(dam))
uqcrh	0.9	0.14	0.11	0.036 *	0.32
sidt2	0.07	0.17	0.08	0.09	0.43
rabep2	0.05*[a]	0.60	0.015*	0.78	0.45
piezo1	0.06	0.25	0.18	1.00	1.00
ffar2	0.71	0.14	0.09	0.83	0.89
trpv5	0.52	0.88	0.62	0.33	0.98
aifm3	0.04*	0.89	**0.002** **[b]	0.009 **	0.99
ub	0.78	0.14	0.05*	0.62	1.00
dspa2b	0.4	0.20	0.06	0.07	1.00
pml	0.58	0.39	0.60	0.01 *	0.99
nkpd1	0.27	0.33	0.47	0.04 *	1.00
tmem38b	0.41	0.02*	0.07	0.13	0.98
pknox1	0.63	0.44	0.43	1.00	1.00
manf	**0.0001****	0.87	**5.6e-06** ***	1.00	0.14
ifitm3	0.44	0.35	0.25	0.17	0.49
ifnar2	0.4	0.80	0.77	0.02 *	0.99
anxa1	0.57	0.31	0.19	0.13	0.02*
prmt3	**0.0027****	0.16	**0.001** ***	0.37	1.00

a. Significant codes: $0.01 < P \leq 0.05$*, $0.001 < P \leq 0.01$**, $P \leq 0.001$***

b. Significant bold *P* value indicates significant after *P* value correction.

4.3.4 Correlation between gut bacterial community and host transcriptional profile

A multiple regression analysis was carried out to evaluate the potential link between bacterial taxon abundance (at the family level) for taxa common to the gut and differentially transcribed genes, while controlling for the treatment, family effect. The abundance of *Lactobacillaceae, Aeromonadaceae, Streptococcaceae* were positively correlated with several gene transcription levels (Figure 4.5).

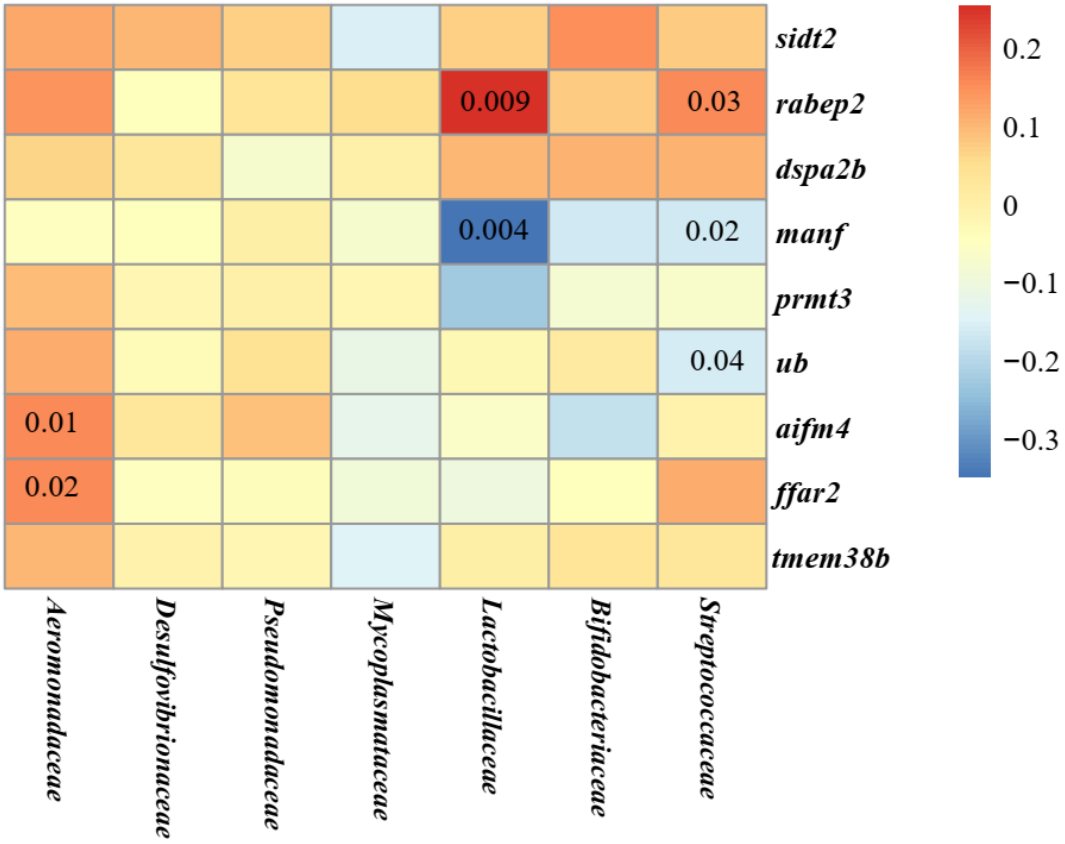


Figure 4.5. Hierarchical clustering of the 7 core bacterial taxa and association with gene expression. Columns correspond to the 7 core bacterial taxa; rows correspond to 9 selected differentially expressed genes. Red and blue denote positive and negative associations, respectively. The intensity of the colors represents the degree of association between the genus abundance and bacterial taxa without considering treatment effect. Numbers in each square represent significant *P*-values (unadjusted) with treatment and Family ID effects included in our model.

4.4 Discussion

Interactions between fish hosts and their microbiomes have been an under-studied area of research, perhaps due to the complexity of the host-microbiome relationship making the detection of specific microbial features that impact the host phenotype challenging. We approached this problem by manipulating gut microbiomes and measured the impact on key candidate gene regulation – such effects are likely mechanism for microbes to affect host phenotype and health. To the best of our knowledge this study is the first to characterize the response of the fish host intestinal gene expression to variation in gut microbiome composition driven by antibiotic and probiotic microbiome manipulations. We found that our treatments resulted in changes in host gene expression patterns, and those changes were mostly related to immune function and cell motility/integrity. By correcting for the direct effects of the treatment, as well as the quantitative genetic effects of inheritance, we showed that changes in microbial communities do lead to changes in host gene expression. Given the putative function of the responding genes, our work indicates a likely effect on host fitness as well. Indeed, many recent studies have shown that microbial symbionts are critical biological components for host traits closely associated with fitness, such as immune system development and function (Fuess et al., 2021; Langlois et al., 2021; Rosshart et al., 2017).

To the best of our knowledge, this is the first study to consider and compare the impact of probiotics and antibiotics fed to fish on the aquatic microbial communities. Our analysis showed that aquatic microbial communities in the holding tanks had unique compositions that were significantly influenced by our treatments. This was not expected as the fish food was only 3% of the fish's body weight with relatively low levels of antibiotic and probiotic additives. One possible factor is that up 90% of the administered antibiotic dose is excreted in the urine and faeces of the

fish, still in the active form (Polianciuc et al., 2020). Our common bacterial phyla in the tank water were reported in other studies that showed Proteobacteria, Bacteroidota, Firmicutes are the dominant taxa in water where fish are held, including both hatchery and wild populations (Chiarello et al., 2015; He et al., 2018; Stevick et al., 2019; Uren Webster et al., 2018; Zhang et al., 2019). However, there were significant differences among our treatment group tank waters. One reason for this could be that antibiotics can cause antibiotic-associated diarrhea thus more microbial associated with the fish gut could be exported to the water. Another reason could be because antibiotic-susceptible taxa are replaced by those taxa that were resistant to antimicrobial agents (e.g., *Mycoplasmataceae* (Firmicutes) (antibiotic (15%), control (1%), probiotic (3%)). Moreover, the aquatic microbiome itself is important for maintaining fish health (Blancheton et al., 2013), and thus, quantifying the unexpected effects of feed-based treatment of the fish on the tank water is important as the changes we documented may contribute to dysbiosis in the fish, and ultimately affect fish health negatively. Although the negative effects of antibiotics on healthy fish have been reported before, few studies have considered the effect of antibiotic treatment targeting fish on the water microbiome. Moreover, our study showed that probiotic treatment of the fish did change the water microbiome. Previous studies revealed that feeding with probiotics can effectively treat the water with probiotics and can improve water quality (e.g., dissolved oxygen, free ammonia, and pH) (Elsabagh et al., 2018; Tabassum et al., 2021). Further studies are needed to explore the range of potential effects of fish feed treated with antibiotics and probiotics on the rearing water microbiome.

The microbial communities present in fish tank water are thought to determine the initial colonization of the fish microbiota via direct seeding and by promoting the colonization of other species (Llewellyn et al., 2014; Talwar et al., 2018). However, similar to other studies (Uren

Webster et al., 2018; Wu et al., 2018), our fish gut microbiomes were distinct from the water sample microbiomes. This indicates that the fish host gut microbiome is likely independent of the water microbial community and while the water microbiota are likely contributing to the fish microbiomes, other factors such as diet and host genome also are contributing to the microbial communities (Talwar et al., 2018).

Our principal goal was to use probiotic and antibiotic treatments to alter the gut microbiome to determine the potential role of gut microbiota composition variation in host-microbiome interactions. However, we also assessed how the gut microbial community reacted to the treatments. We found that, while fish gut BC alpha diversity was not affected by the treatments, beta diversity was significantly different among all three groups. Similar results were reported in other studies, indicating community richness (alpha diversity) did not respond to treatment with probiotics and antibiotics, but beta diversity did (Hernandez-Perez et al., 2022; Kokou et al., 2020; Laursen et al., 2017). One possible reason for this is that using antibiotics does not necessarily mean a reduced diversity of bacterial taxa, indeed a review showed that individuals with dysbiosis (potentially caused by treatment) can have even more diverse microbial community compared to healthy individuals (Berg et al., 2020). For example, Rosado *et al*. (2019) showed that treatment of farmed seabass (*Dicentrarchus labrax*) with OTC caused a decrease in core diversity in the gill and an increase in the skin microbiome BCs. One reason that our probiotic treatment did not change the community richness (alpha diversity) could be that we treated healthy fish with probiotics. Previous studies have shown that probiotic supplements in healthy humans does not substantially impact the resident microbial populations (Eloe-Fadrosh et al., 2015; Lahti et al., 2013). In general, external stimuli that affect the intestinal environment can drive a hierarchical series of microbiome responses; resistance, resilience, redundancy or finally dysbiosis−depending

on if the disturbance overcomes the intestinal microbial ecosystem (Lozupone et al., 2012; Moya and Ferrer, 2016; Sommer et al., 2017). BC compositions are generally sensitive to disturbance (Allison and Martiny, 2008). It appears that the microbial response to probiotic treatment in our healthy fish study was resilience, as previous studies have shown that the BCs tended to be more resilience to external stimuli if the stimuli are not substantial. On the other hand, treatment with antibiotics, depending on how strong the dose is, will drive the BC to either of resilience, redundancy or dysbiosis.

It is worth noting that several potentially pathogenic (*Vibrionaceae*, *Aeromonadaceae*) and beneficial (*Lactobacillaceae*, *Bifidobacteriaceae*) microbes were abundant in the antibiotic and probiotic treatment groups – an outcome that has critical implications for aquaculture. Previous studies have indicated that lactic acid bacteria (*Lactobacillus*, *Streptococcus* and *Lactococcus*) and *Bacillus* include important agents for biological control in aquaculture (Ringo et al., 2020; Vieco-Saiz et al., 2019). We found that the relative abundance of lactic acid bacteria was high in probiotic group but low in the other two groups. Another important family that was observed across all fish was *Aeromonadaceae*. The *Aeromonas* genus in the *Aeromonadaceae* family contains two important fish pathogens: *Aeromonas hydrophila* and *Aeromonas salmonicida* both of which have been shown to pathogenically infect fishes (Ringø et al., 2010). Lastly, another important fish pathogen, *Photobacterium damselae* was detected in the gut microbiome in this study, but at low relative abundance. Overall, our results generally agreed with the previously described microbiomes of healthy Salmonid species (Minich et al., 2020; Uren Webster et al., 2018).

We predicted that the gut microbial community would respond to the treatments through an increase in beneficial gut bacteria (probiotic treatment) or through a decrease in the beneficial microbes with a related increase in the number of potential pathogens (antibiotic treatment). This

was based on the expectation that antibiotics can cause dysbiosis in the gut, resulting in elevated levels of opportunistic pathogens (Dethlefsen and Relman, 2011; Francino, 2015), while prebiotics and probiotics are expected to increase the frequency of gut barrier-protecting bacteria such as *Lactobacillaceae* and *Bifidobacteriaceae* (Xiao et al., 2014). In this study bacteria with potential probiotic properties (*Lactobacillaceae*, *Bifidobacteriaceae*, *Streptococcaceae*) were higher in the probiotic group compared to other treatment groups. On the other hand, *Pseudomonadaceae* and *Aeromonadaceae* had high relative abundances in antibiotic group. Similar patterns of response to probiotics and antibiotics in BC structure and composition have been reported by others (Falcinelli et al., 2016; Kokou et al., 2020; Navarrete et al., 2008; Rutten et al., 2015). For example, Kokou *et al.* (2020) showed that after seven days of administration of antibiotics via the diet, the European seabass (*Dicentrarchus labrax*) microbiome increased in taxa belonging to *Staphylococcus*, *Pseudomonas* genera (Proteobacteria). OTC treatment was seen to reduce gut microbial diversity in Atlantic salmon, while enhancing possible opportunistic pathogens belonging to *Aeromonas* spp likely due to eliminating competing microorganisms (Navarrete et al., 2008). Moreover, Falcinelli *et al.* (2016) showed that metabarcoding sequence reads for Firmicutes, specifically *Lactobacillus* genus, were significantly higher in probiotic treated Zebrafish (*Danio rerio*) larvae than control fish.

Studies in humans (Qin et al., 2010) and fishes (Boutin et al., 2014) have reported that the gut microbiome varies substantially at the individual and population level, and the transcriptome of the fish gut appears to be correlated with this variation (Franzosa et al., 2014; Qin et al., 2010). Moreover, Thaiss *et al.* (2016) showed that treatment with antibiotics will change the mouse gut microbiome, and that the microbiome in turn regulates fluctuations in the host transcriptome and epigenome. In our study, we showed that our treatment altered the gut microbiota, then we tested

if these changes were associated with changes in host gene expression. Specifically, we showed that several genes related to cellular processes such as cell activation, cell communication, and cell death were upregulated after treatment with antibiotics in the feed. Although previous studies have shown a direct effect of antibiotic treatment on gene transcription in human (Ryu et al., 2017) and mice (Morgun et al., 2015). However, Ruiz *et al.* (2017) showed that antibiotic treatment had a limited effect on gene expression in germ-free mice, providing evidence that the microbiome mediates the effects of orally administered antibiotics on the host. Morgun *et al*. (2015) argued that most antibiotic-induced alterations in the gut can be explained in three scenarios: depletion of the microbiota, direct effects of antibiotics on host tissues, and the effects of antibiotic resistant microbes, or possibly a combination of all three. In this study we found that our antibiotic treatment resulted in the upregulation of genes related to cell death. Moreover, bacteria from the Firmicutes and Bacteroidetes phyla were reduced while members of the Proteobacteria phylum increased. Zarrinpar *et al.* (2018) showed a similar shift the BC in the mouse cecal; however, a cecal transcriptome analysis showed that the changes in the BC resulted in changes in the expression of genes related to cellular growth and proliferation, as well as cell death and survival pathways. This suggests that colonic remodeling after treatment with antibiotics is directly driving changes in the host transcriptome. Additionally, in our antibiotic treatment group, we showed increased transcription of the *mrp7* (multidrug resistance-associated protein 7-like) gene. Multidrug resistance-associated proteins are a group of closely related gene products that prevent intracellular accumulation of certain drugs, thus the observed change in gene transcription may be an example of direct effects of antibiotics on the host tissues. However, our experiment design did not allow us to capture the third scenario proposed by Mogun *et al.* (2015). Moreover, our qRT-PCR method showed upregulation of *aifm3* gene in antibiotic group. A study by Stoddard *et al.* (2019) in

zebrafish showed that after introducing antibiotics to fish, inflammatory gene transcription was downregulated and apoptotic genes such as *aifm3* were upregulated within 24 hours.

Antibiotics are designed to pass the gut barrier and become systemic; however, probiotics are live microorganisms that are not able to pass the lumen barrier. Probiotics can directly modulate the host physiology by interacting with host cells (mostly immune cells) at mucosal surface, and indirect changes of gut microbiome (Langlois et al., 2021). We showed (transcriptomic data) that genes related to post-translation modifications were over-expressed in the probiotic treatment group, relative to the control and antibiotic treatment groups. Previous studies showed that probiotic diet supplements elicit a proinflammatory response in fish (Nayak, 2010) and honeybees (Daisley et al., 2020) which promotes more effective pathogen clearance and improved disease resistance. Previous studies have shown that the administration of probiotics might have beneficial effects by (i) competing with pathogenic bacteria for habitat and nutrients, (ii) enhancing epithelial barrier function and prevention of apoptosis of epithelial cells, and (iii) modulation of host immune responses (Dawood et al., 2018; Lebeer et al., 2008). In this study we found that our treatment with probiotics indeed changes the BC composition by increasing the number of potential probiotic taxa (*Lactobacillaceae* and *Bifidobacteriaceae*). Moreover, our treatment with probiotics showed upregulation of transcription of genes related to membrane trafficking (e.g., *rabep2)* and post-translation modifications (e.g., *prmt3*, *cops6*, *psmb4*), angiogenesis (rspo3), compared to control groups. Finally, we noticed that our probiotic treatment changed the transcription of several genes (*hsbp1*, *commd10*, *blnk*) related to the immune system when compared to control group, similar to previous studies (Petrof et al., 2004; Tomosada et al., 2013). For example, Tomosada *et al.* (2013), showed that Bifidobacteria strains can have immunoregulatory effect on the host intestinal epithelial cells by modulation the ubiquitin-editing enzyme. Moreover, similar to this study,

Willms et al (2022) also showed that beneficial bacteria can promote intestinal angiogenesis in Zebrafish. The precise mechanism of action of probiotics remains to be elucidated, especially in heathy states.

The mechanisms of bidirectional interaction between the host and their microbiomes have been an underexplored area of research. The complexity of microbiomes encourages a shift from a reductionist approach to a more holistic one. The reductionist approach focuses on identifying specific microbial taxa or host factors that affect the two-way interactions which is ineffective for complex systems. However, a holistic approach focus on interactions among networks or groups of taxa in the microbiome and multiple host genes with diverse function. Complex ecosystems exhibit properties such as nonlinearity and uncertainty and are defined as a system that cannot be explained by the sum of its interacting components. One approach to characterize the bidirectional interactions between the host and the microbiome BC is to perturb the gut and measure the response of the host (such as in AIMD studies). In this study, we used antibiotics and probiotics to modify the microbial communities within the gut and measured host gene transcription responses to those modifications. We explored this effect using correlation between multiple common bacterial taxa and host gene transcription. The results of that analysis were consistent with a microbiome-mediated effect on the host. We found that specific microbial taxa are affecting the regulation of several host genes, for example, the abundance of *Lactobacillaceae* was positively and negatively associated with the transcription of the *rabep2* and *manf* host genes, respectively. Previous work has shown that a single-nucleotide polymorphism (SNP) in the *rabep2* gene in humans is associated with ulcerative colitis, consistent with a strong association between *rabep2* gene and gut bacterial taxa (Jostins et al., 2012). Moreover, upregulation of *manf* gene can activate innate immune cells and facilitate repair of damaged tissue (Neves et al., 2016; Sereno et al., 2017).

However, further studies will be required to determine the specific association of *Lactobacillaceae* with the expression of the *manf* host gene.

Finally, we cannot ignore the potential influence of the host's genetic background on the gut microbiome. Several bacteria taxa have shown to be associated (both abundance levels and presence/absence) with human genome content (e.g., *LCT*) (Kurilshikov et al., 2021; Lopera-Maya et al., 2022), which strongly suggests that the human gut microbiota can be influenced by host genetics (Lopera-Maya et al., 2022). Heritability studies for human gut microbiome composition have estimated that the host genetic background could explain between 2% to 8% of gut microbiome variation (Goodrich et al., 2016; Rothschild et al., 2018). Unfortunately, there is little published information on the heritability of the BC composition in fish species. However, our study breeding design was incorporated more to correct for host genetic signal than to estimate heritability or specific quantitative genetic variances. Moreover, Dvergedal *et al.* (2020) reported weak associations between host (Atlantic salmon) genetics and gut BC composition, similar to our findings. A more focussed analysis of the inheritance of the microbiome in fish is overdue, since, unlike mammals and other animals, most fishes do not have an obvious vertical transmission mechanism. Similar to other organisms (Gutierrez Lopez et al., 2021; Frazier and Chang, 2020), the fish microbiome can go through diurnal changes in community composition and hence metabolic activity that are impacted by both the host's circadian rhythms and diet (Willms et al., 2022). As a result, the time that we collected the samples during the day (one day sampling) might have had an effect on host gene transcriptome. Further studies on how host microbiome can change during the day and night can its influence on host transcriptome is needed. Moreover, fish subjected to environmental stressors such as pharmaceuticals and anesthetic products can change

the gut BC composition (Parrott et al., 2022). In this study we used clove oil to euthanize the fish, and this might have effect on both fish gut microbiome as well as host gene expression.

4.5 Conclusion

The direction of interaction between fish gut and microbiome is not clear. In this study we experimentally modified the fish gut microbiome and evaluated host gut tissue responses to those perturbations using transcriptome analysis and transcriptional profiling (qRT-PCR). Short term (10 days) perturbation of the juvenile Chinook salmon gut microbiome with antibiotics and probiotics affected the microbiome BC composition and host gene expression patterns. This study has achieved a number of important goals: (1) characterized the effects of antibiotics and probiotics on the aquatic BC (2) characterized juvenile Chinook salmon gut microbiome BC response to antibiotic and probiotic treatment (3) characterized the host gut tissue transcriptional response to antibiotic and probiotic treatments. We showed that our treatments with antibiotics and probiotics not only changed the Chinook salmon microbiome BC (composition), but we also observed significant changes at the gene expression level in the gut tissue of the fish. This study provides insight into a long-standing co-evolved symbiotic relationship between fish gut tissue and its associated microbiome. Moreover, understanding factors influencing the fish gut microbiome and its influence on host health and fitness will help in better sustainable growth for the aquaculture.

4.6 Reference

Allison SD, Martiny JB. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences 2008; 105: 11512-11519.

Azad MAK, Sarker M, Li T, Yin J. Probiotic Species in the Modulation of Gut Microbiota: An Overview. Biomed Res Int 2018; 2018: 9478630.

Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome 2020; 8: 103.

Bilandzic N, Tankovic S, Varenina I, Kolanovic BS, Smajlovic M. Chloramphenicol residues in muscle of rainbow trout following two different dose treatments. Bull Environ Contam Toxicol 2012; 89: 461-6.

Blancheton J, Attramadal K, Michaud L, d'Orbcastel ER, Vadstein O. Insight into bacterial population in aquaculture systems and its implication. Aquacultural engineering 2013; 53: 30-39.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 2018; 6: 90.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114-20.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019; 37: 852-857.

Boutin S, Sauvage C, Bernatchez L, Audet C, Derome N. Inter individual variations of the fish skin microbiota: host genetics basis of mutualism? PLoS One 2014; 9: e102649.

Bozzi D, Rasmussen JA, Caroe C, Sveier H, Nordoy K, Gilbert MTP, et al. Salmon gut microbiota correlates with disease infection status: potential for monitoring health in farmed animals. Anim Microbiome 2021; 3: 30.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016; 13: 581-3.

Chiarello M, Villeger S, Bouvier C, Bettarel Y, Bouvier T. High diversity of skin-associated bacterial communities of marine fishes is promoted by their high variability among body parts, individuals and species. FEMS Microbiol Ecol 2015; 91.

Chong J, Liu P, Zhou G, Xia J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. Nat Protoc 2020; 15: 799-821.

Daisley BA, Pitek AP, Chmiel JA, Gibbons S, Chernyshova AM, Al KF, et al. Lactobacillus spp. attenuate antibiotic-induced immune and microbiota dysregulation in honey bees. Commun Biol 2020; 3: 534.

Davison JM, Lickwar CR, Song L, Breton G, Crawford GE, Rawls JF. Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor Hepatocyte nuclear factor 4 alpha. Genome Res 2017; 27: 1195-1206.

Dawood MA, Koshio S, Esteban MÁ. Beneficial roles of feed additives as immunostimulants in aquaculture: a review. Reviews in Aquaculture 2018; 10: 950-974.

Dayama G, Priya S, Niccum DE, Khoruts A, Blekhman R. Interactions between the gut microbiome and host gene regulation in cystic fibrosis. Genome Med 2020; 12: 12.

Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 2011; 108 Suppl 1: 4554-61.

Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev 2019; 83.

Dvergedal H, Sandve SR, Angell IL, Klemetsdal G, Rudi K. Association of gut microbiota with metabolism in juvenile Atlantic salmon. Microbiome 2020; 8: 160.

Egerton S, Culloty S, Whooley J, Stanton C, Ross RP. The Gut Microbiota of Marine Fish. Front Microbiol 2018; 9: 873.

Eloe-Fadrosh EA, Brady A, Crabtree J, Drabek EF, Ma B, Mahurkar A, et al. Functional dynamics of the gut microbiome in elderly people during probiotic consumption. mBio 2015; 6.

Elsabagh M, Mohamed R, Moustafa EM, Hamza A, Farrag F, Decamp O, et al. Assessing the impact of Bacillus strains mixture probiotic on water quality, growth performance, blood profile and intestinal morphology of Nile tilapia, Oreochromis niloticus. Aquaculture nutrition 2018; 24: 1613-1622.

Escalas A, Auguet J-C, Avouac A, Seguin R, Gradel A, Borrossi L, et al. Ecological Specialization Within a Carnivorous Fish Family Is Supported by a Herbivorous Microbiome Shaped by a Combination of Gut Traits and Specific Diet. Frontiers in Marine Science 2021; 8: 91.

Falcinelli S, Rodiles A, Unniappan S, Picchietti S, Gioacchini G, Merrifield DL, et al. Probiotic treatment reduces appetite and glucose level in the zebrafish model. Sci Rep 2016; 6: 18061.

Ferrer M, Mendez-Garcia C, Rojo D, Barbas C, Moya A. Antibiotic use and microbiome function. Biochem Pharmacol 2017; 134: 114-126.

Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol 2015; 6: 1543.

Franzosa EA, Morgan XC, Segata N, Waldron L, Reyes J, Earl AM, et al. Relating the metatranscriptome and metagenome of the human gut. Proc Natl Acad Sci U S A 2014; 111: E2329-38.

Frazier K, Chang EB. Intersection of the Gut Microbiome and Circadian Rhythms in Metabolism. Trends Endocrinol Metab 2020; 31: 25-36.

Fuess LE, den Haan S, Ling F, Weber JN, Steinel NC, Bolnick DI. Immune Gene Expression Covaries with Gut Microbiome Composition in Stickleback. mBio 2021; 12.

Geffroy B, Gesto M, Clota F, Aerts J, Darias MJ, Blanc MO, et al. Parental selection for growth and early-life low stocking density increase the female-to-male ratio in European sea bass. Sci Rep 2021; 11: 13620.

Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe 2016; 19: 731-43.

Groussin M, Mazel F, Alm EJ. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. Cell Host Microbe 2020; 28: 12-22.

Gutierrez Lopez DE, Lashinger LM, Weinstock GM, Bray MS. Circadian rhythms and the gut microbiome synchronize the host's metabolic response to diet. Cell Metab 2021; 33: 873-887

He X, Chaganti SR, Heath DD. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microb Ecol 2018; 75: 140-151.

Hernandez-Perez A, Zamora-Briseno JA, Soderhall K, Soderhall I. Gut microbiome alterations in the crustacean Pacifastacus leniusculus exposed to environmental concentrations of antibiotics and effects on susceptibility to bacteria challenges. Dev Comp Immunol 2022; 126: 104181.

Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012; 491: 119-24.

K B, S R, M L. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. . In: 1.12.0 Rpv, editor, 2021.

Kareiva P, Marvier M, McClure M. Recovery and management options for spring/summer chinook salmon in the Columbia River basin. Science 2000; 290: 977-9.

Kassambara A, Mundt F. Factoextra: extract and visualize the results of multivariate data analyses. R package version 2017; 1: 337-354.

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002; 30: 3059-66.

Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015; 12: 357-60.

Kokou F, Sasson G, Mizrahi I, Cnaani A. Antibiotic effect and microbiome persistence vary along the European seabass gut. Sci Rep 2020; 10: 10003.

Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 2012; 28: 3211-7.

Koskella B, Bergelson J. The study of host-microbiome (co)evolution across levels of selection. Philos Trans R Soc Lond B Biol Sci 2020; 375: 20190604.

Krajmalnik-Brown R, Ilhan ZE, Kang DW, DiBaise JK. Effects of gut microbes on nutrient absorption and energy regulation. Nutr Clin Pract 2012; 27: 201-14.

Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet 2021; 53: 156-165.

Kuznetsova A, Brockhoff PB, Christensen RH. lmerTest package: tests in linear mixed effects models. Journal of statistical software 2017; 82: 1-26.

Lahti L, Salonen A, Kekkonen RA, Salojarvi J, Jalanka-Tuovinen J, Palva A, et al. Associations between the human intestinal microbiota, Lactobacillus rhamnosus GG and serum lipids indicated by integrated analysis of high-throughput profiling data. PeerJ 2013; 1: e32.

Lai HT, Liu SM, Chien YH. Transformation of chloramphenicol and oxytetracycline in aquaculture pond sediments. Journal of Environmental Science & Health Part A 1995; 30: 1897-1923.

Langlois L, Akhtar N, Tam KC, Dixon B, Reid G. Fishing for the right probiotic: host-microbe interactions at the interface of effective aquaculture strategies. FEMS Microbiol Rev 2021; 45.

Laursen MF, Laursen RP, Larnkjaer A, Michaelsen KF, Bahl MI, Licht TR. Administration of two probiotic strains during early childhood does not affect the endogenous gut microbiota composition despite probiotic proliferation. BMC Microbiol 2017; 17: 175.

Leal JF, Santos EB, Esteves VI. Oxytetracycline in intensive aquaculture: water quality during and after its administration, environmental fate, toxicity and bacterial resistance. Reviews in Aquaculture 2019; 11: 1176-1194.

Lebeer S, Vanderleyden J, De Keersmaecker SC. Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev 2008; 72: 728-64, Table of Contents.

Lee BJ, Bak YT. Irritable bowel syndrome, gut microbiota and probiotics. J Neurogastroenterol Motil 2011; 17: 252-66.

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923-30.

Limbu SM, Zhou L, Sun SX, Zhang ML, Du ZY. Chronic exposure to low environmental concentrations and legal aquaculture doses of antibiotics cause systemic adverse effects in Nile tilapia and provoke differential human health risk. Environ Int 2018; 115: 205-219.

Llewellyn MS, Boutin S, Hoseinifar SH, Derome N. Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. Front Microbiol 2014; 5: 207.

Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sanchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet 2022; 54: 143-151.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15: 550.

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature 2012; 489: 220-30.

Manor O, Dai CL, Kornilov SA, Smith B, Price ND, Lovejoy JC, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 2020; 11: 5206.

McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 2013; 110: 3229-36.

Meek MH, Baerwald MR, Stephens MR, Goodbla A, Miller MR, Tomalty KMH, et al. Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. Ecol Evol 2016; 6: 7706-7716.

Meisel JS, Sfyroera G, Bartow-McKenney C, Gimblet C, Bugayev J, Horwinski J, et al. Commensal microbiota modulate gene expression in the skin. Microbiome 2018; 6: 20.

Minich JJ, Poore GD, Jantawongsri K, Johnston C, Bowie K, Bowman J, et al. Microbial Ecology of Atlantic Salmon (Salmo salar) Hatcheries: Impacts of the Built Environment on Fish Mucosal Microbiota. Appl Environ Microbiol 2020; 86.

Molina-Lopez J, Ricalde MAQ, Hernandez BV, Planells A, Otero R, Planells E. Effect of 8-week of dietary micronutrient supplementation on gene expression in elite handball athletes. PLoS One 2020; 15: e0232237.

Montalban-Arques A, De Schryver P, Bossier P, Gorkiewicz G, Mulero V, Gatlin DM, 3rd, et al. Selective Manipulation of the Gut Microbiota Improves Immune Status in Vertebrates. Front Immunol 2015; 6: 512.

Morgun A, Dzutsev A, Dong X, Greer RL, Sexton DJ, Ravel J, et al. Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks. Gut 2015; 64: 1732-43.

Moya A, Ferrer M. Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. Trends Microbiol 2016; 24: 402-413.

Muehlbauer AL, Richards AL, Alazizi A, Burns MB, Gomez A, Clayton JB, et al. Interspecies variation in hominid gut microbiota controls host gene regulation. Cell Rep 2021; 37: 110057.

Navarrete P, Mardones P, Opazo R, Espejo R, Romero J. Oxytetracycline treatment reduces bacterial diversity of intestinal microbiota of Atlantic salmon. J Aquat Anim Health 2008; 20: 177-83.

Naya-Catala F, do Vale Pereira G, Piazzon MC, Fernandes AM, Calduch-Giner JA, Sitja-Bobadilla A, et al. Cross-Talk Between Intestinal Microbiota and Host Gene Expression in Gilthead Sea Bream (Sparus aurata) Juveniles: Insights in Fish Feeds for Increased Circularity and Resource Utilization. Front Physiol 2021; 12: 748265.

Nayak SK. Probiotics and immunity: a fish perspective. Fish Shellfish Immunol 2010; 29: 2-14.

Neves J, Zhu J, Sousa-Victor P, Konjikusic M, Riley R, Chew S, et al. Immune modulation by MANF promotes tissue repair and regenerative success in the retina. Science 2016; 353: aaf3646.

Nichols RG, Davenport ER. The relationship between the gut microbiome and host gene expression: a review. Hum Genet 2021; 140: 747-760.

Parrott JL, Restivo VE, Kidd KA, Zhu J, Shires K, Clarence S, et al. Chronic Embryo-Larval Exposure of Fathead Minnows to the Pharmaceutical Drug Metformin: Survival, Growth, and Microbiome Responses. Environ Toxicol Chem 2022; 41: 635-647.

Perez T, Balcazar JL, Ruiz-Zarzuela I, Halaihel N, Vendrell D, de Blas I, et al. Host-microbiota interactions within the fish intestinal ecosystem. Mucosal Immunol 2010; 3: 355-60.

Petrof EO, Kojima K, Ropeleski MJ, Musch MW, Tao Y, De Simone C, et al. Probiotics inhibit nuclear factor-kappaB and induce heat shock proteins in colonic epithelial cells through proteasome inhibition. Gastroenterology 2004; 127: 1474-87.

Piazzon MC, Naya-Catala F, Perera E, Palenzuela O, Sitja-Bobadilla A, Perez-Sanchez J. Genetic selection for growth drives differences in intestinal microbiota composition and parasite disease resistance in gilthead sea bream. Microbiome 2020; 8: 168.

Polianciuc SI, Gurzau AE, Kiss B, Stefan MG, Loghin F. Antibiotics in the environment: causes and consequences. Med Pharm Rep 2020; 93: 231-240.

Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 2010; 5: e9490.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010; 464: 59-65.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013; 41: D590-6.

Rice WR. Analyzing tables of statistical tests. Evolution 1989; 43: 223-225.

Ringø E, Løvmo L, Kristiansen M, Bakken Y, Salinas I, Myklebust R, et al. Lactic acid bacteria vs. pathogens in the gastrointestinal tract of fish: a review. Aquaculture Research 2010; 41: 451-467.

Ringo E, Van Doan H, Lee SH, Soltani M, Hoseinifar SH, Harikrishnan R, et al. Probiotics, lactic acid bacteria and bacilli: interesting supplementation for aquaculture. J Appl Microbiol 2020; 129: 116-136.

Rosado D, Xavier R, Severino R, Tavares F, Cable J, Perez-Losada M. Effects of disease, antibiotic treatment and recovery trajectory on the microbiome of farmed seabass (Dicentrarchus labrax). Sci Rep 2019; 9: 18946.

Rosshart SP, Vassallo BG, Angeletti D, Hutchinson DS, Morgan AP, Takeda K, et al. Wild Mouse Gut Microbiota Promotes Host Fitness and Improves Disease Resistance. Cell 2017; 171: 1015-1028 e13.

Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature 2018; 555: 210-215.

Ruiz VE, Battaglia T, Kurtz ZD, Bijnens L, Ou A, Engstrand I, et al. A single early-in-life macrolide course has lasting effects on murine microbial network topology and immunity. Nat Commun 2017; 8: 518.

Rutten NB, Gorissen DM, Eck A, Niers LE, Vlieger AM, Besseling-van der Vaart I, et al. Long Term Development of Gut Microbiota Composition in Atopic Children: Impact of Probiotics. PLoS One 2015; 10: e0137681.

Ryu AH, Eckalbar WL, Kreimer A, Yosef N, Ahituv N. Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. Sci Rep 2017; 7: 7533.

Sadeghi J, Chaganti SR, Shahraki AH, Heath DD. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. Sci Total Environ 2021; 781: 146771.

Sanders ME. Probiotics: definition, sources, selection, and uses. Clin Infect Dis 2008; 46 Suppl 2: S58-61; discussion S144-51.

Sereno D, Muller WEG, Bausen M, Elkhooly TA, Markl JS, Wiens M. An evolutionary perspective on the role of mesencephalic astrocyte-derived neurotrophic factor (MANF): At the crossroads of poriferan innate immune and apoptotic pathways. Biochem Biophys Rep 2017; 11: 161-173.

Shahraki AH, Chaganti SR, Heath D. Assessing high-throughput environmental DNA extraction methods for meta-barcode characterization of aquatic microbial communities. J Water Health 2019; 17: 37-49.

Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol 2017; 15: 630-638.

Stevick RJ, Sohn S, Modak TH, Nelson DR, Rowley DC, Tammi K, et al. Bacterial Community Dynamics in an Oyster Hatchery in Response to Probiotic Treatment. Front Microbiol 2019; 10: 1060.

Stoddard M, Huang C, Enyedi B, Niethammer P. Live imaging of leukocyte recruitment in a zebrafish model of chemical liver injury. Sci Rep 2019; 9: 28.

Surana NK, Kasper DL. Moving beyond microbiome-wide associations to causal microbe identification. Nature 2017; 552: 244-247.

Tabassum T, Mahamud ASU, Acharjee TK, Hassan R, Snigdha TA, Islam T, et al. Probiotic supplementations improve growth, water quality, hematology, gut microbiota and intestinal morphology of Nile tilapia. Aquaculture Reports 2021; 21: 100972.

Talwar C, Nagar S, Lal R, Negi RK. Fish Gut Microbiome: Current Approaches and Future Perspectives. Indian J Microbiol 2018; 58: 397-414.

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

Thaiss CA, Levy M, Korem T, Dohnalova L, Shapiro H, Jaitin DA, et al. Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. Cell 2016; 167: 1495-1510 e12.

Toews SD, Wellband KW, Dixon B, Heath DD. Variation in juvenile Chinook salmon (Oncorhynchus tshawytscha) transcription profiles among and within eight population crosses from British Columbia, Canada. Mol Ecol 2019; 28: 1890-1903.

Tomosada Y, Villena J, Murata K, Chiba E, Shimazu T, Aso H, et al. Immunoregulatory effect of bifidobacteria strains in porcine intestinal epithelial cells through modulation of ubiquitin-editing enzyme A20 expression. PLoS One 2013; 8: e59259.

Uren Webster TM, Consuegra S, Hitchings M, Garcia de Leaniz C. Interpopulation Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity. Appl Environ Microbiol 2018; 84.

Vieco-Saiz N, Belguesmia Y, Raspoet R, Auclair E, Gancel F, Kempf I, et al. Benefits and Inputs From Lactic Acid Bacteria and Their Bacteriocins as Alternatives to Antibiotic Growth Promoters During Food-Animal Production. Front Microbiol 2019; 10: 57.

Willms RJ, Jones LO, Hocking JC, Foley E. A cell atlas of microbe-responsive processes in the zebrafish intestine. Cell Rep 2022; 38: 110311

Wu ZB, Gatesoupe FJ, Li TT, Wang XH, Zhang QQ, Feng DY, et al. Significant improvement of intestinal microbiota of gibel carp (Carassius auratus gibelio) after traditional Chinese medicine feeding. J Appl Microbiol 2018; 124: 829-841.

Xia Y, Sun J. Hypothesis Testing and Statistical Analysis of Microbiome. Genes Dis 2017; 4: 138-148.

Xiao S, Fei N, Pang X, Shen J, Wang L, Zhang B, et al. A gut microbiota-targeted dietary intervention for amelioration of chronic inflammation underlying metabolic syndrome. FEMS Microbiol Ecol 2014; 87: 357-67.

Yao Z, Yang K, Huang L, Huang X, Qiuqian L, Wang K, et al. Disease outbreak accompanies the dispersive structure of shrimp gut bacterial community with a simple core microbiota. AMB Express 2018; 8: 120.

Zarrinpar A, Chaix A, Xu ZZ, Chang MW, Marotz CA, Saghatelian A, et al. Antibiotic-induced microbiome depletion alters metabolic homeostasis by affecting gut signaling and colonic metabolism. Nat Commun 2018; 9: 2872.

Zhang Z, Li D, Xu W, Tang R, Li L. Microbiome of Co-cultured Fish Exhibits Host Selection and Niche Differentiation at the Organ Scale. Front Microbiol 2019; 10: 2576.

CHAPTER 5: GENERAL CONCLUSION

5.1 Summary

An ecosystem is usually defined as a particular spatial region that houses elements of biodiversity (such as animals, plants, and microorganisms) interacting with each other and the abiotic environment (Levin, 2013). Part of what sustains and stabilizes an ecosystem is the complex networks of interactions among and within those two elements (biotic and abiotic). Aquatic microbes (microeukaryotes and prokaryotes) are vital for the function and stability of aquatic ecosystems (Trombetta et al., 2020). The definition of a microbiome includes the genome of all the microorganisms plus their surrounding and interacting environment (Berg et al., 2020). In this dissertation, I focused on bacteria (chapters 2, 3, 4), and to a lesser extent, microeukaryotes (chapter 2) due to their role influencing bacterial diversity and community composition (Audebert et al., 2016; Beghini et al., 2017).

Due to the complex interactions within the microbial community, a shift from a reductionist view that focuses on individual taxa, to a more holistic approach that highlights interactions among members of the community and their environments seem necessary. The diverse and complex microbial communities that we see today were probably necessary for the evolution of life (Ley et al., 2008). Indeed, microorganisms created spatially structured communities as early as more than 3 billion years ago (Ley et al., 2008) and have thus been an integral part of the evolution of all biodiversity. Comparative studies of microbial communities are starting to define how biotic (e.g., species interactions) and abiotic features (e.g., salinity or pH) contribute to the structure, organization, and function of microbial diversity. Ontario, Canada has more than 250,000 lakes that include the largest freshwater ecosystems on the earth - those aquatic ecosystems are characterized by diverse and complex trophic interactions, with additional terrestrial input from

forests, wetlands, and metropolitan areas. Thus, the aquatic ecosystems of Ontario alone provide a powerful natural ecological laboratory to study important questions about microbial ecology, host-microbial interactions and co-evolution (Gandhi et al., 2015).

Microbial community composition also varies substantially within a diverse range of living organisms, including human and fish microbiomes (He et al., 2018; Lozupone et al., 2012). More than 32 000 fish species live in an environment that is dominated by aquatic microbiomes (Huang et al., 2020). Thus, interactions between fish and environmental microbes are unavoidable and ongoing, and due to this close association, fish, more than any other group of animals, may exhibit close associations with symbiotic microbiomes. Moreover, since fish evolved earlier than other vertebrates, fish and their associated microbes have interacted longer than terrestrial vertebrates (Gibson, 2018). Thus, studying host-microbe co-evolution in fish is a logical starting point to understanding host-microbe co-evolution in general. Exogenous abiotic and biotic factors (e.g., habitat, spatial variation, microbial biodiversity, diet, etc.) and endogenous host-related factors (e.g., genetics, physiology, immunity) are driving the composition of the aquatic microbiome in both lake ecosystems and host fish species (Minich et al., 2020a; Minich et al., 2020b; Sadeghi et al., 2021). Despite the well-documented effects of endogenous and exogenous factors on both the host fish microbiomes and the surrounding water microbiome, the disagreement over the relative contribution of the factors that drive microbial community composition reflects a critical knowledge gap in the field of microbial ecology (Zhou and Ning, 2017). Additionally, to determine how host-associated BCs may influence host phenotype, and hence contribute to evolutionary processes, quantifying the degree and nature of among-taxa microbiome variation, and the systematic drivers behind it, seems crucial.

My thesis provides separate, but connected, lines of evidence for the roles of exogenous (abiotic and biotic factors) and endogenous (host-related) factors in determining the composition of aquatic microbiomes in both lakes and host fish species. I incorporated multidisciplinary field, lab, bioinformatic and statistical approaches that allowed me to address significant knowledge gaps in the microbiome field, with the most impactful being; 1) characterizing the biotic and abiotic effects on aquatic bacterial communities in north temperate lakes (chapter 2), 2) determining the environmental and host species (phylogenetic) effects on the fish microbiome (chapter 3), and 3) demonstrating the regulation of host gene expression by the gut microbiome using perturbation experiments (chapter 4). The overall goal of my dissertation was to address how environmental, as well as host-related factors, can affect microbial communities in an aquatic ecosystem. To address that goal, my thesis made several technical, analytical, and conceptual contributions to our understanding of the factors that can change the microbial community associated with lakes as well as fish.

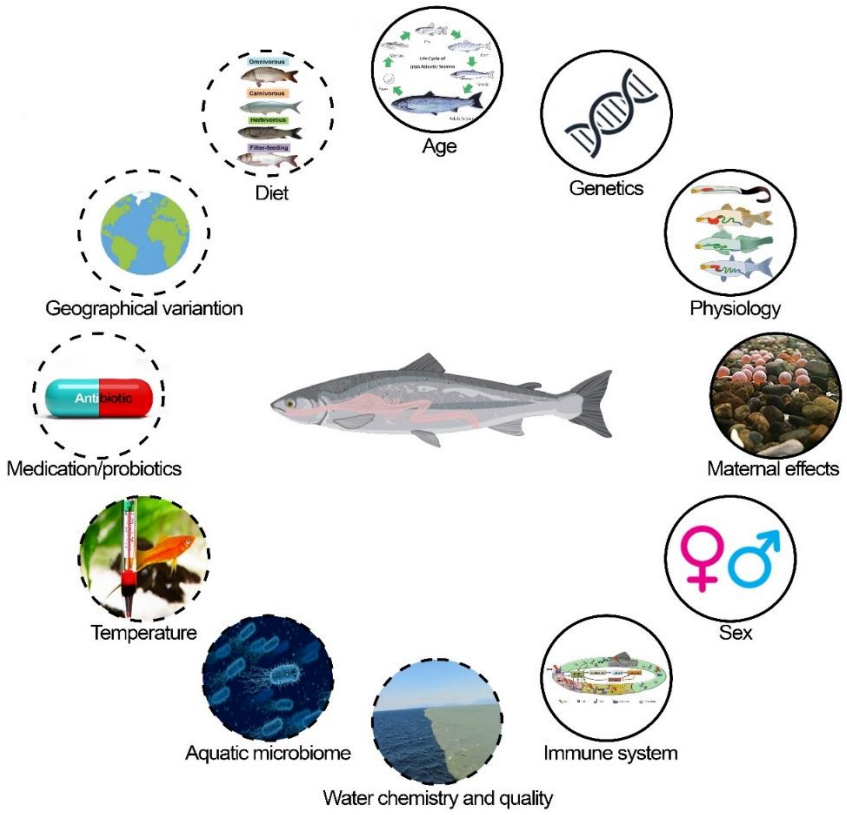


Figure 5-1. Schematic diagram (modified from Kers et al., 2018) showing the endogenous and exogenous factors that have been shown to affect fish microbiomes. Solid circles indicate endogenic factors while dashed circle indicate exogenic factors affecting the fish microbiome.

Based on Baas Becking's hypothesis "Everything is everywhere, but the environment selects" (Baas-Becking, 1934), environmental factors should dominate in determining the composition of microbial communities. The regulation of community structure by environmental factors could be through the abiotic (e.g., pH, nutrients, etc.) and/or biotic compartments (e.g., competition, cooperation, and predation) (chapter 2 (Sadeghi et al., 2021)) which generally is referred to as the niche theory or deterministic outcomes. However, recent studies have revealed that stochastic processes alone or coupled with the deterministic process can also affect microbial community composition (Sadeghi et al., 2021; Zhou and Ning, 2017). Based on stochastic theory, all microbial taxa are functionally equivalent, and their community composition is driven by

ecological drift, coupled with dispersal limitation with little or no environmental input (Oliphant et al., 2019). While there is no doubt that both stochastic and deterministic processes can affect microbial community composition, their relative contribution and quantitative strength in structuring the microbial composition and temporal/spatial variation in natural communities are poorly known (Aguilar and Sommaruga, 2020; Grilli, 2020; Stegen et al., 2012).

Chapters 2 and 3 were designed to test for how biotic and abiotic variables shape the bacterial communities associated with freshwater lakes and fish, respectively. As shown in chapters 2 (Sadeghi et al., 2021) and 3, significant effects of biotic and abiotic factors indicating deterministic processes affecting BC composition were identified, with greater effects from abiotic than biotic factors. Likely the freshwater BC composition is determined by a combination of deterministic and stochastic processes (Yuan et al., 2019); however, my analyses indicate that deterministic processes are dominant in the lakes I sampled. Similar to my results, deterministic processes have been reported to be the major driver shaping BC composition in freshwater lakes (Aguilar and Sommaruga, 2020; Llames et al., 2017). The trophic states of those lakes ranged from oligotrophic to hypereutrophic lakes. Moreover, exogenous deterministic factors not only have been reported to drive microbial community composition of the freshwater lakes but also the microbial community within the fish living in those lakes (Kim et al., 2021; Schmidt et al., 2015). However, the role of the factors in determining the composition of the teleost fish microbiome remains poorly understood. For example, Heys *et al.* (2020) showed that neutral processes dominate microbial community assembly in the Atlantic Salmon gut (*Salmo salar*). Further research in this area is needed to determine how endogenous versus exogenous factors are affecting the fish microbiome. Additionally, in chapter 3, I showed that the fish gut and skin microbiomes were substantially divergent from the surrounding water. Although fish are exposed to their aquatic

microbiome all their life, microbiome studies in fish generally report that fish skin and gut environments harbor different microbiome than the water (Li et al., 2015; Reinhart et al., 2019; Schmidt et al., 2016).

Similar to humans (Lopera-Maya et al., 2022) and other organisms (Gogarten et al., 2018; Rudman et al., 2019; Wen and Duffy, 2017), the composition of the gut and skin microbiome in fish is dependent on endogenous factors, as exogenous factors alone only explain a small part of the variation associated with fish microbiomes. Host endogenous factors such as genomic variation (Riiser et al., 2020), gut and skin morphological features (Egerton et al., 2018), and fish age (Minich et al., 2020a) are known to drive variation at the individual level (Boutin et al., 2014), the population level (Uren Webster et al., 2018) and the species-level (chapter 3; (Kim et al., 2021).

Based on "the Red Queen hypothesis" (Van Valen, 1973), biotic interactions should dominate abiotic forces in driving evolution as biotic effects can adaptively evolve in both directions. This hypothesis has improved our understanding of the evolutionary process as biotic conflict can cause microevolutionary responses which subsequently can accumulate into macroevolutionary changes (Brockhurst et al., 2014). This is the basis of coevolution, which is the process of reciprocal evolutionary changes that appears in pairs of species (O'Brien et al., 2019). The Red Queen hypothesis is centered on the concept of antagonistic coevolution, or –host-parasite/pathogen coevolution, and presumes that adaptations that increase the fitness of one species will decrease the fitness of another and thus drive selection (O'Brien et al., 2019). Moreover, mutualistic/symbiotic relationships can also cause coevolutionary patterns, resulting in benefits on both sides of the relationship (Herre et al., 1999).

"Phylosymbiosis", a pattern of coordinated microbiome-host divergence across host species, has been documented among diverse organisms such as within the class Mammalia (Ross

et al., 2018), and fish (Doane et al., 2020; Ross et al., 2018). According to phylosymbiosis theory, two closely related species (phylogenetically) will have more similar microbiomes than distantly related hosts (Lim and Bordenstein, 2020). Evidence for phylosymbiosis in non-mammalian vertebrate animals, including amphibians and fish, are not consistent. For example, one study showed the presence of phylosymbiosis in coral reef fishes (e.g. (Pollock et al., 2018), whereas others report no relationships in 12 co-occurring species of teleost (Escalas et al., 2021) or weak relationships (Chiarello et al., 2018) in coral reef fishes. In chapter 3, I found skin and gut samples were showed phylosymbiosis patterns. However, the phylosymbiosis signal for skin BCs was stronger than for gut BCs. This was surprising as my expectation was that the fish skin microbiome would be less impacted by evolutionary changes in the host than the fish gut microbiome, similar to what has been reported in terrestrial animals (Ley et al., 2006). One reason for this could be related to the aquatic environment (with a diverse group of microorganisms and environmental variation) that fish interact with. As a result, skin microbiomes are exposed to harsher environments compared to the gut. For example, bacterial communities on the skin are exposed to a diverse micro-predator (including protists, predatory bacteria, and bacteriophages) that are not able to live in the fish gastrointestinal environment. The success of the fish skin microbiome BC (highly differentiated from the water microbiome) probably would not be possible without close association with control factors from the host. Moreover, only a small subset of aquatic microbial taxa can colonize the skin. The skin microbial community likely has strict conditions for membership such as various enzymes to utilize available nutrients on fish skin mucus, being able to attach to the ‘‘right’’ location on the skin, or the ability to grow rapidly to avoid washout. This implies that the successful taxa are highly adapted to their host and there is a strong selection from

the fish for a specific skin community. Such a scenario would explain the strong association between fish phylogeny and the skin microbiome BC.

In chapter 4, I found that modification of the gut microbiome in fish caused changes in host gene expression patterns (Dethlefsen and Relman, 2011; Douglas, 2019; Zheng et al., 2020; Figure 5-2). Co-evolution requires symmetrical bidirectional selection pressures from each of the two species. I speculate that the peculiar structure and composition of bacterial communities in the fish gut and skin tissue developed from natural selection acting at two levels. At the host level, selection on the community provides stable environments in gut and skin with a high degree of functional redundancy. At microbial level, strong selection pressures would be applied for functional specialization to succeed in the skin and gut environments. Interactions between bacteria with bacteria or other organisms are often described as commensal (one side benefits and the other side appears unchanged) rather than mutualistic (benefits for both sides), or parasitic (benefit for one side at the expense of the other). It seems that the bidirectional interaction between the fish and its associated microbes is more mutualistic rather than commensal or parasitic. Microbial communities in fish benefit their host by protecting the fish from pathogens, aid in the development of the host immune system and facilitate nutrition absorption. In compensation fish also provide a habitat and nutrition for their associated bacterial communities.

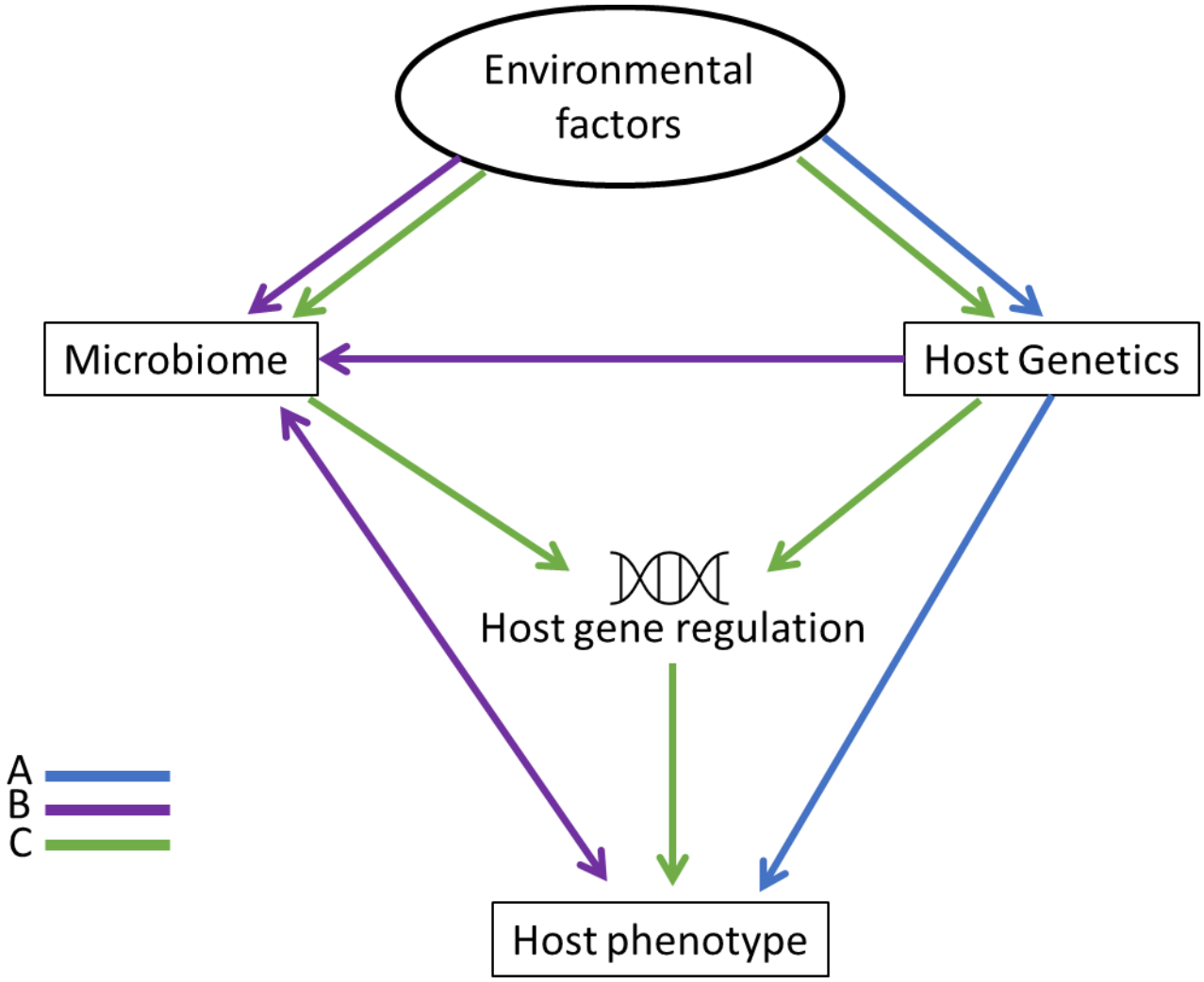


Figure 5-2. Schematic diagram (modified from Awany et al., 2018) of the host-microbiome interaction; a possible scenario that the three-way interactions between environment-microbiome-host can determine the microbiome composition, host gene expression pattern, and finally host phenotype. A (blue lines): host genetic background (genome composition) directly influences the host phenotype, mediated by environmental signals (as shown in chapters 2, 4). B: (purple lines) endogenous and exogenous factors can modulate the microbiome and the microbiome will affect the host's phenotype but note the indirect effect of the host's genome (chapters 3 & 4). C (green lines): another mechanism for three-way environment-host-microbiome interactions affecting host gene expression which results in changes in host phenotype (as shown in chapters 2,3,4). Probably C is the main way that the environment-host-microbiome interactions can change host gene expression patterns.

5.2 Conclusion

In conclusion, my thesis contributes to the characterization of the modulators of microbiome composition in aquatic systems, including both freshwater lakes and fish host habitats in those lakes. My study design provided a robust test of the relative effects of exogenous abiotic

(chapter 2), biotic (chapter 2, and 3), and endogenous host-specific factors (chapter 3, and 4) on aquatic microbiomes. Moreover, I also characterized the effect of the gut microbiome on host phenotype (chapter 4). In general, I found a higher impact from exogenous factors relative to endogenous factors. Investigations of the nature of fish-microbe associations, and whether they are sustained, functional relationships, or transient effects of fish and habitat associations are critical to further our understanding of the potential beneficial interactions between the host and their microbiomes.

5.3 Future directions

Deterministic (mainly) and stochastic processes drive the microbial communities associated with freshwater lakes (chapter 2). However, I am aware of confounding interaction effects (either stochastic or deterministic). Partitioning those suspected and unknown effects in natural systems would be difficult, if not impossible. By implementing controlled experiments in the lab or semi-natural mesocosms, then adding or changing individual environmental factors (such as nutrients) and/or biotic factors (i.e. change the microeukaryote community) in separate steps, we would have a better quantitative estimate of the effect of deterministic versus stochastic processes on microbial community composition. Moreover, such an approach would allow the quantification of the mutualistic and parasitic relationships among bacteria and microeukaryotes that contribute to BC composition (chapter 2) but are so complex in most natural systems as to make detailed assessments impossible. Specifically, an important next step building on my work would be to test the positive and negative correlations identified in my network analysis under controlled (lab) conditions.

Furthermore, I showed evidence for phylosymbiosis for gut and skin samples across 17 different fish species. An important extension of this work would be to include fish species that

diverged at specific time scales or selecting taxa that have known evolutionary relationships. Alternatively, we can select species within a single genus (to control for phylogeny) but with different diets (e.g., carnivore, omnivore, and herbivore), or species from different taxonomic orders with similar diet preferences to quantify the role of possible phylogeny-associated diets in phylosymbiosis. Moreover, it will be interesting to compare the phylosimbiotic patterns among ovoviviparous and viviparous fish species with the prediction that ovoviviparous fish species will have weaker phylosimbiosis due to the higher influence from the environment and less selection pressure from the host.

In chapters 3 and 4, I showed that the host-microbiome relationship in aquatic ecosystems is strong. The next step should be to characterize how the host genome variation affects the BC composition, and more importantly, the mechanisms of interaction between the host and microbiome. This can be through heritability assessment (controlled breeding), using genetically modified host fish, or using clonal or highly inbred fish. Previous studies have shown that between 2-8 percent of the microbiome in humans is heritable. However, those studies have been done in humans, with greater maternal effects (including vertical transmission of the vaginal microbiome (Mueller et al., 2015) compared to fish. Although some recent studies have shown the possible vertical transmission of the microbiome from the dam to offspring in fish (Ziab, 2020), still there is a lack of strong evidence of vertical transmission of the microbiome for aquatic species.

Finally, in chapter 4, I showed positive and negative correlation between some bacterial taxa and host gene expression. For example, members of *Lactobacillaceae* were positively and negatively associated with the transcription of the *rabep2* and *manf* host genes, respectively. An exciting next step would be to further test the mechanisms driving these associations between

different bacteria taxa and the host genome transcription using knock-down experiments (miRNA) or knock-out experiments (CRISPR-cas9) with control and or germ-free fish.

5.4 References

Aguilar P, Sommaruga R. The balance between deterministic and stochastic processes in structuring lake bacterioplankton community over time. Mol Ecol 2020; 29: 3117-3130.

Audebert C, Even G, Cian A, Blastocystis Investigation G, Loywick A, Merlin S, et al. Colonization with the enteric protozoa Blastocystis is associated with increased diversity of human gut bacterial microbiota. Sci Rep 2016; 6: 25255.

Awany D, Allali I, Dalvie S, Hemmings S, Mwaikono KS, Thomford NE, et al. Host and Microbiome Genome-Wide Association Studies: Current State and Challenges. Front Genet 2018; 9: 637.

Baas-Becking LGM. Geobiologie; of inleiding tot de milieukunde: WP Van Stockum & Zoon NV, 1934.

Beghini F, Pasolli E, Truong TD, Putignani L, Caccio SM, Segata N. Large-scale comparative metagenomics of Blastocystis, a common member of the human gut microbiome. ISME J 2017; 11: 2848-2863.

Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome 2020; 8: 103.

Boutin S, Sauvage C, Bernatchez L, Audet C, Derome N. Inter individual variations of the fish skin microbiota: host genetics basis of mutualism? PLoS One 2014; 9: e102649.

Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GD. Running with the Red Queen: the role of biotic conflicts in evolution. Proc Biol Sci 2014; 281.

Chiarello M, Auguet JC, Bettarel Y, Bouvier C, Claverie T, Graham NAJ, et al. Skin microbiome of coral reef fish is highly variable and driven by host phylogeny and diet. Microbiome 2018; 6: 147.

Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 2011; 108 Suppl 1: 4554-61.

Doane MP, Morris MM, Papudeshi B, Allen L, Pande D, Haggerty JM, et al. The skin microbiome of elasmobranchs follows phylosymbiosis, but in teleost fishes, the microbiomes converge. Microbiome 2020; 8: 93.

Douglas AE. Simple animal models for microbiome research. Nat Rev Microbiol 2019; 17: 764-775.

Egerton S, Culloty S, Whooley J, Stanton C, Ross RP. The Gut Microbiota of Marine Fish. Front Microbiol 2018; 9: 873.

Escalas A, Auguet J-C, Avouac A, Seguin R, Gradel A, Borrossi L, et al. Ecological Specialization Within a Carnivorous Fish Family Is Supported by a Herbivorous Microbiome Shaped by a Combination of Gut Traits and Specific Diet. Frontiers in Marine Science 2021; 8: 91.

Gandhi N, Bhavsar SP, Tang RW, Arhonditsis GB. Projecting Fish Mercury Levels in the Province of Ontario, Canada and the Implications for Fish and Human Health. Environ Sci Technol 2015; 49: 14494-502.

Gibson M. Food science and the culinary arts: Academic Press, 2018.

Gogarten JF, Davies TJ, Benjamino J, Gogarten JP, Graf J, Mielke A, et al. Factors influencing bacterial microbiome composition in a wild non-human primate community in Tai National Park, Cote d'Ivoire. ISME J 2018; 12: 2559-2574.

Grilli J. Macroecological laws describe variation and diversity in microbial communities. Nat Commun 2020; 11: 4743.

He X, Chaganti SR, Heath DD. Population-Specific Responses to Interspecific Competition in the Gut Microbiota of Two Atlantic Salmon (Salmo salar) Populations. Microb Ecol 2018; 75: 140-151.

Herre EA, Knowlton N, Mueller UG, Rehner SA. The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends in ecology & evolution 1999; 14: 49-53.

Heys C, Cheaib B, Busetti A, Kazlauskaite R, Maier L, Sloan WT, et al. Neutral Processes Dominate Microbial Community Assembly in Atlantic Salmon, Salmo salar. Appl Environ Microbiol 2020; 86.

Huang Q, Sham RC, Deng Y, Mao Y, Wang C, Zhang T, et al. Diversity of gut microbiomes in marine fishes is shaped by host-related factors. Mol Ecol 2020; 29: 5019-5034.

Kers JG, Velkers FC, Fischer EAJ, Hermes GDA, Stegeman JA, Smidt H. Host and Environmental Factors Affecting the Intestinal Microbiota in Chickens. Front Microbiol 2018; 9: 235.

Kim PS, Shin NR, Lee JB, Kim MS, Whon TW, Hyun DW, et al. Host habitat is the major determinant of the gut microbiome of fish. Microbiome 2021; 9: 166.

Levin SA. Encyclopedia of biodiversity: Elsevier Inc., 2013.

Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 2008; 6: 776-88.

Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 2006; 124: 837-48.

Li T, Long M, Gatesoupe FJ, Zhang Q, Li A, Gong X. Comparative analysis of the intestinal bacterial communities in different species of carp by pyrosequencing. Microb Ecol 2015; 69: 25-36.

Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. Proc Biol Sci 2020; 287: 20192900.

Llames ME, Huber P, Metz S, Unrein F. Interplay between stochastic and deterministic processes in the maintenance of alternative community states in Verrucomicrobia-dominated shallow lakes. FEMS Microbiol Ecol 2017; 93.

Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sanchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet 2022; 54: 143-151.

Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature 2012; 489: 220-30.

Minich JJ, Petrus S, Michael JD, Michael TP, Knight R, Allen EE. Temporal, Environmental, and Biological Drivers of the Mucosal Microbiome in a Wild Marine Fish, Scomber japonicus. mSphere 2020a; 5.

Minich JJ, Poore GD, Jantawongsri K, Johnston C, Bowie K, Bowman J, et al. Microbial Ecology of Atlantic Salmon (Salmo salar) Hatcheries: Impacts of the Built Environment on Fish Mucosal Microbiota. Appl Environ Microbiol 2020b; 86.

Mueller NT, Bakacs E, Combellick J, Grigoryan Z, Dominguez-Bello MG. The infant microbiome development: mom matters. Trends Mol Med 2015; 21: 109-17.

O'Brien PA, Webster NS, Miller DJ, Bourne DG. Host-Microbe Coevolution: Applying Evidence from Model Systems to Complex Marine Invertebrate Holobionts. mBio 2019; 10.

Oliphant K, Parreira VR, Cochrane K, Allen-Vercoe E. Drivers of human gut microbial community assembly: coadaptation, determinism and stochasticity. ISME J 2019; 13: 3080-3092.

Pollock FJ, McMinds R, Smith S, Bourne DG, Willis BL, Medina M, et al. Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. Nat Commun 2018; 9: 4921.

Reinhart EM, Korry BJ, Rowan-Nash AD, Belenky P. Defining the Distinct Skin and Gut Microbiomes of the Northern Pike (Esox lucius). Front Microbiol 2019; 10: 2118.

Riiser ES, Haverkamp THA, Varadharajan S, Borgan O, Jakobsen KS, Jentoft S, et al. Metagenomic Shotgun Analyses Reveal Complex Patterns of Intra- and Interspecific Variation in the Intestinal Microbiomes of Codfishes. Appl Environ Microbiol 2020; 86.

Ross AA, Muller KM, Weese JS, Neufeld JD. Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. Proc Natl Acad Sci U S A 2018; 115: E5786-E5795.

Rudman SM, Greenblum S, Hughes RC, Rajpurohit S, Kiratli O, Lowder DB, et al. Microbiome composition shapes rapid genomic adaptation of Drosophila melanogaster. Proc Natl Acad Sci U S A 2019; 116: 20025-20032.

Sadeghi J, Chaganti SR, Shahraki AH, Heath DD. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. Sci Total Environ 2021; 781: 146771.

Schmidt V, Amaral-Zettler L, Davidson J, Summerfelt S, Good C. Influence of Fishmeal-Free Diets on Microbial Communities in Atlantic Salmon (Salmo salar) Recirculation Aquaculture Systems. Appl Environ Microbiol 2016; 82: 4470-4481.

Schmidt VT, Smith KF, Melvin DW, Amaral-Zettler LA. Community assembly of a euryhaline fish microbiome during salinity acclimation. Mol Ecol 2015; 24: 2537-50.

Stegen JC, Lin X, Konopka AE, Fredrickson JK. Stochastic and deterministic assembly processes in subsurface microbial communities. ISME J 2012; 6: 1653-64.

Trombetta T, Vidussi F, Roques C, Scotti M, Mostajir B. Marine Microbial Food Web Networks During Phytoplankton Bloom and Non-bloom Periods: Warming Favors Smaller Organism Interactions and Intensifies Trophic Cascade. Front Microbiol 2020; 11: 502336.

Uren Webster TM, Consuegra S, Hitchings M, Garcia de Leaniz C. Interpopulation Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity. Appl Environ Microbiol 2018; 84.

Van Valen L. A new evolutionary law. Evol theory 1973; 1: 1-30.

Wen L, Duffy A. Factors Influencing the Gut Microbiota, Inflammation, and Type 2 Diabetes. J Nutr 2017; 147: 1468S-1475S.

Yuan H, Mei R, Liao J, Liu WT. Nexus of Stochastic and Deterministic Processes on Microbial Community Assembly in Biological Systems. Front Microbiol 2019; 10: 1536.

Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res 2020; 30: 492-506.

Zhou J, Ning D. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol Rev 2017; 81.

Ziab M. Transgenerational effects on the microbiome of Chinook salmon (Oncorhynchus tshawytscha). University of Windsor (Canada), 2020.

APPENDIX A; SUPPLEMENTARY INFORMATION OF CHAPTER 2

Supplementary Table S2.1. Sampled Ontario, Canada lakes with name, location, and sampling details

Lake	Lake ID	Latitude	Longitude	Number of sampling sites	Number of bottles per lake	Total[a]
Balsam	L27	44.6292	-78.846	1	2	4
Bass	L9	44.57445	-76.0843	1	2	4
Bernard	L50	45.75027	-79.4003	1	3	6
Big Blad	L32	44.58432	-78.3906	1	2	4
Big Rideau	L11	44.70099	-76.1859	1	2	4
Boshkong	L37	45.09121	-78.7399	1	2	4
Buck	L16	44.51214	-76.477	2	4	8
		44.53503	-76.44488			
Buckhom	L33	44.47568	-78.3689	1	2	4
Cameron	L26	44.54191	-78.7869	1	2	4
Charleston	L7	44.54705	-75.9842	1	2	4
Colonel	L1	44.29347	-76.4496	2	4	8
		44.30075	-76.43149			
Commanda	L54	46.01451	-79.6962	1	3	6
Couchiching	L23	44.73575	-79.3532	2	4	8
		44.72063	-79.34429			
Cranberry	L3	44.4509	-76.2856	1	2	4
Deer	L52	45.80905	-79.5654	1	3	6
Devil[b]	L15	44.586	-76.4684	3	5	10
		44.57821	-76.4251			
		44.57796	-76.51149			
Dog	L2	44.39791	-76.3552	1	2	4
Draper	L19	44.48043	-76.5234	1	2	4
Eagle	L51	45.83491	-79.5014	2	6	12
		45.80907	-79.5103			
Fairy	L49	45.33778	-79.1912	1	3	6
Four Mile	L34	44.69574	-78.7237	1	2	4
Gananoque	L5	44.45482	-76.1536	1	2	4
Grippen	L6	44.51514	-76.1516	1	2	4
Gull	L35	44.85951	-78.7665	1	2	4
Halls	38	45.11862	-78.7566	1	2	4
Kahshe	L44	44.85069	-79.3032	1	2	4
Kawartha	L28	44.52129	-78.6263	1	2	4
Kawigamog	L59	45.89188	-80.1814	1	3	6
Kushog	L40	45.11339	-78.8103	1	2	4

Lake Joseph	L60	45.1824	-79.7677	1	3	6
Lake of bays	L42	45.25506	-78.9787	2	4	8
		45.17768	-79.09778			
Little Blad	L30	44.57509	-78.4259	1	2	4
Little Franklin	L17	44.48964	-76.4821	1	2	4
Loon	L13	44.60412	-76.4015	2	4	8
		44.61845	-76.38437			
Loughborough	L20	44.37706	-76.529	1	2	4
Lower Beverley	L10	44.60152	-76.148	1	2	4
Lower Buckhom	L32	44.55433	-78.2818	1	2	4
Mary	L47	45.24395	-79.2646	1	3	6
Milton	L56	45.91049	-79.8664	1	3	6
Mosquito	L14	44.6055	-76.3658	1	2	4
Mountain	L36	44.98289	-78.7191	1	2	4
Mud	L18	44.48808	-76.4918	1	2	4
Muskoka	L45	45.02184	-79.454	2	6	12
		45.0643	-79.46823			
Nipissing	L53	46.29998	-79.4646	2	6	12
		46.12049	-79.5338			
Odessa	L22	44.29968	-76.7017	1	2	4
Penfold	L46	45.26739	-79.2815	1	3	6
Pigeon	L29	44.46662	-78.5045	1	2	4
Raven	L41	45.22878	-78.8625	1	2	4
Restoule	L55	46.05214	-79.7128	1	3	6
Rosseau	L58	45.93109	-80.0464	1	3	6
Saskachewan	L39	45.12937	-78.7792	1	2	4
Seagull	L57	45.90971	-79.9347	1	3	6
Simcoe	L25	44.58826	-79.4042	3	6	12
		44.55818	-79.25674			
		44.35104	-79.24893			
South	L4	44.44012	-76.2343	1	2	4
Sparrow	L43	44.8124	-79.375	1	2	4
St. John	L24	44.68702	-79.3115	1	2	4
Sydenham	L21	44.430485	-76.525623	1	2	4
Upper Beverley	L8	44.61509	-76.0742	1	2	4
Upper Rideau	L12	44.6872	-76.3449	1	2	4
Vernon	L48	45.31416	-79.2763	1	3	6

a. Number of samples after cutting each filter into half and extracting from both halves

b. For Devil Lake one bottle broke during transportation: 5 bottles remained.

Supplementary Table S2.2. Results from the nested ANOVA for various measure of microbial community composition testing for the effects of sampled lake, location within the lake, and the biological and technical replicates.

16S Chao1					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	57	2129490	2129490	28.704	4.03e-06 ***
Location (lake)	49	2383	2383	0.032	0.859
Biological replicate (Location)	51	88079	88079	1.187	0.283
Technical replicate (Biological replicate×Lake)	43	10318	10318	0.139	0.711
16S Observed OTUs					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	57	13581	13581	6.481	0.0150 *
Location (lake)	49	6016	6016	2.871	0.0982
Biological replicate (Location)	51	5850	5850	2.792	0.1028
Technical replicate (Biological replicate×Lake)	43	303	303	0.145	0.7057
16S PC1					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	57	0.04739	0.04739	14.215	0.000541 ***
Location (lake)	49	0.00491	0.00491	1.473	0.232146
Biological replicate (Location)	51	0.00049	0.00049	0.147	0.703671
Technical replicate (Biological replicate×Lake)	43	0.00066	0.00066	0.197	0.659400
16S PC2					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	57	0.01101	0.011011	4.379	0.0429 *
Location (lake)	49	0.00008	0.000082	0.033	0.8577
Biological replicate (Location)	51	0.00351	0.003506	1.394	0.2448
Technical replicate (Biological replicate×Lake)	43	0.00291	0.002909	1.157	0.2887
16S PC3					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	57	0.00278	0.0027802	1.305	0.260
Location (lake)	49	0.00094	0.0009354	0.439	0.511
Biological replicate (Location)	51	0.00129	0.0012919	0.607	0.441
Technical replicate (Biological replicate×Lake)	43	0.00068	0.0006821	0.320	0.575
18S Chao1					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	51	11482	11482	0.260	0.613
Location (lake)	43	66961	66961	1.516	0.225
Biological replicate (Location)	47	518	518	0.012	0.914
Technical replicate (Biological replicate×Lake)	42	1218	1218	0.028	0.869
18S Observed OTUs					

Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	51	14322	14322	1.739	0.194
Location (lake)	43	2182	2182	0.265	0.609
Biological replicate (Location)	47	572	572	0.069	0.793
Technical replicate (Biological replicate×Lake)	42	989	989	0.120	0.731
18S PC1					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	51	0.06746	0.06746	19.276	6.78e-05 ***
Location (lake)	43	0.00049	0.00049	0.140	0.710
Biological replicate (Location)	47	0.00374	0.00374	1.070	0.307
Technical replicate (Biological replicate×Lake)	42	0.00029	0.00029	0.084	0.773
18S PC2					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	51	0.02682	0.026819	5.045	0.0296 *
Location (lake)	43	0.00654	0.006538	1.230	0.2733
Biological replicate (Location)	47	0.00854	0.008543	1.607	0.2114
Technical replicate (Biological replicate×Lake)	42	0.00064	0.000642	0.121	0.7298
18S PC3					
Variable	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lake	51	0.09442	0.09442	71.017	8.31e-11 ***
Location (lake)	43	0.00064	0.00064	0.481	0.491
Biological replicate (Location)	47	0.00179	0.00179	1.345	0.252
Technical replicate (Biological replicate×Lake)	42	0.00055	0.00055	0.412	0.524

*** = $p<0.001$; ** = $p<0.01$; * = $p<0.05$

Supplementary Table S2.3: Results from the nested PERMANOVA for microbial (bacteria and micr-eukaryote) community beta diversity (Bray-Curtis dissimilarity matrix) testing for the effects of sampled lake, location within the lake, and the biological and technical replicates.

16SrRNA	Df	Sum Sq	Mean Sq	F value	R^2	Pr(>F)
Lake	57	1.7363	1.73627	9.2202	0.18271	0.036 *
Location (lake)	49	0.3014	0.30144	1.6007	0.03172	0.062
Biological replicate (Location)	51	0.3447	0.34471	1.8305	0.03627	0.083
Technical replicate (Biological replicate×Lake)	43	0.1529	0.15286	0.8117	0.01609	0.392
18SrRNA	Df	Sum Sq	Mean Sq	F value	R^2	Pr(>F)
Lake	51	1.7363	1.73627	9.2202	0.18271	0.029 *
Location (lake)	43	0.3014	0.30144	1.6007	0.03172	0.056
Biological replicate (Location)	47	0.3447	0.34471	1.8305	0.03627	0.084
Technical replicate (Biological replicate×Lake)	42	0.1529	0.15286	0.8117	0.01609	0.401

*** = p<0.001; ** = p<0.01; * = p<0.05

Supplementary Table S2.4: Results from the pairwise PERMANOVA for microbial (bacteria and micr-eukaryote) community beta diversity (Bray-Curtis dissimilarity matrix) testing for testing significant differences between NJ clusters

BC NJ clusters	T value	P value
C1-C2	2.44	0.001**
C1-C3	2.30	0.001**
C1-C4	1.00	0.432
C1-C5	1.27	0.02*
C1-C6	1.89	0.001**
C2-C3	1.72	0.001**
C2-C4	1.40	0.088
C2-C5	1.57	0.013*
C2-C6	2.17	0.001**
C3-C4	1.31	0.095
C3-C5	1.48	0.014*
C3-C6	2.19	0.001**
C4-C5	1.25	0.34
C4-C6	1.16	0.13
C5-C6	1.25	0.02*

MEC NJ clusters	T value	P value
C1-C2	2.06	0.001**
C1-C3	1.77	0.002**
C1-C4	1.83	0.002**
C1-C5	2.68	0.001**
C2-C3	1.30	0.014*
C2-C4	1.40	0.001**
C2-C5	1.65	0.001**
C3-C4	1.31	0.03*
C3-C5	1.20	0.012*
C4-C5	1.33	0.004**

*** = p<0.001; ** = p<0.01; * = p<0.05

Supplementary Table S2.5. Mean (and variance) of the physical and biogeochemical characteristics of the sampled lakes.

Factors	N	Min	Max	St. Dev.	Variance
Depth (m)	55	3	92	21.2	453
Depth (log)	55	1.4	4.53	0.7	0.5
Surface area (he)	58	18	88052	14855	220675893
Surface area (log)	58	2.9	11.3	1.61	2.7
Volume (m^3 x10^6)	47	0.4	11408	1744	3041246
Volume (log)	47	0.3	9	1.87	3.6
Shoreline length (km)	50	2.8	1013	180.61	32620
Shoreline length (log)	50	1.3	6.9	1	1.3
Altitude (m)	59	78	480	95.48	9118
Altitude (log)	59	4.3	6	0.4	0.2
Secchi depth (m)	50	0.4	8	1.8	3.4
Secchi depth (log)	50	0.3	2.1	0.4	0.1
TP (mg/L)	52	0.001	0.09	0.01	0.0002
TP (log)	52	0.0009	0.08	0.01	0.0002
Calcium (mg/L)	48	1.5	43.8	13	170
Calcium (log)	48	0.9	3.8	0.9	0.8
pH	51	6.4	8.5	0.6	0.4
pH(log)	51	2	2.2	0.07	0.005

Supplementary Table S2.6. Pearson correlation between bacteria (16SrRNA) alpha (Chao1, observed OTUs) and beta diversity indices with lake's physicochemical and biological variables.

16s Observed OTUs			
Category	**Variables**	**Correlation**	**P value**
Geographical characteristic	Altitude	0.51	3.367e-05***[a]
Spatial variables	Latitude	0.40	0.001***
	Longitude	-0.52	1.716e-05***
Physical characteristics	PC1[b]	0.16	0.19
	PC2	-0.028	0.83
Chemical variables	TP	-0.26	0.059
	Secchi depth	0.167	0.24
	Calcium	-0.27	0.06
	pH	-0.21	0.13
Biological variables	18s Chao1	0.59	2.184e-06***
	18s observed OTUs	0.50	0.0001***
	18s PCoA1	-0.47	0.0002***
	18s PCoA2	0.15	0.26
	18s PCoA3	0.14	0.28
16s Chao1			
Category	**Variables**	**Correlation**	**P value**
Geographical characteristic	Altitude	0.48	0.0001***
Spatial variables	Latitude	0.35	0.006**
	Longitude	-0.52	1.716e-05***
Physical characteristics	PC1	0.09	0.45
	PC2	0.005	0.96
Chemical variables	TP	-0.20	0.1
	Secchi depth	0.09	0.51
	Calcium	-0.22	0.1
	pH	-0.17	0.22
Biological variables	18s Chao1	0.53	3.287e-05***
	18s observed OTUs	0.47	0.0002***
	18s PCoA1	-0.46	0.0003***
	18s PCoA2	0.15	0.26
	18s PCoA3	0.06	0.64
16S PCoA1			
Category	**Variables**	**Correlation**	**P value**
Geographical characteristic	Altitude	-0.65	1.827e-08***
Spatial variables	Latitude	-0.64	2.85e-08***
	Longitude	0.75	7.381e-12***
Physical characteristics	PC1	-0.08	0.54
	PC2	0.15	0.23
Chemical variables	TP	0.26	0.057
	Secchi depth	-0.02	0.85
	Calcium	0.57	2.119e-05***
	pH	0.50	0.00017***
Biological variables	18s Chao1	-0.49	0.00014***
	18s Observed OTUs	-0.49	0.00015***
	18s PCoA1	0.78	2.999e-12***
	18s PCoA2	-0.22	0.10
	18s PCoA3	-0.15	0.26
16S PCoA2			
Category	**Variables**	**Correlation**	**P value**
Geographical characteristic	Altitude	-0.19	0.13
Spatial variables	Latitude	-0.49	7.708e-05***
	Longitude	0.18	0.16
Physical characteristics	PC1	0.10	0.44
	PC2	0.23	0.07
Chemical variables	TP	-0.07	0.59
	Secchi depth	0.037	0.79
	Calcium	0.585	1.274e-05***
	pH	0.47	0.0004***
Biological variables	18s Chao1	-0.13	0.32
	18s Observed OTUs	-0.22	0.10
	18s PCOA1	0.46	0.0004***
	18s PCoA2	0.18	0.18

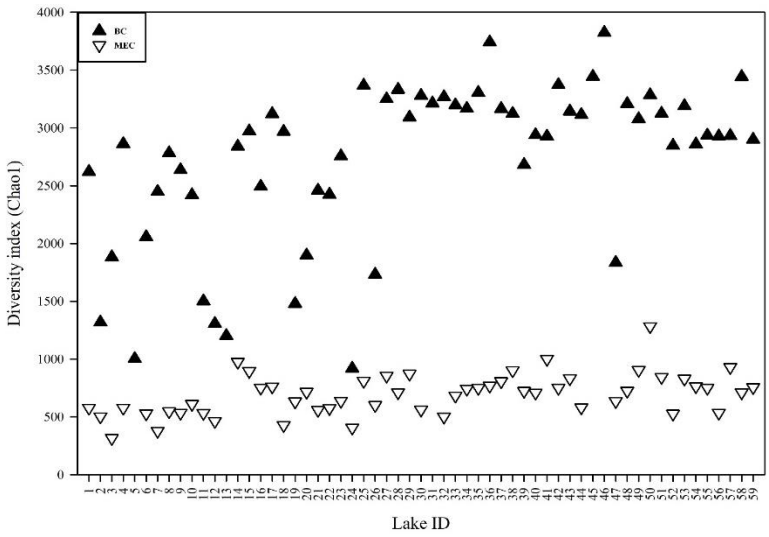
	18s PCoA3	-0.01	0.9
16S PCoA3			
Category	**Variables**	**Correlation**	**P value**
Geographical characteristic	Altitude	0.081	0.53
Spatial variables	Latitude	-0.05	0.65
	Longitude	-0.13	0.31
Physical characteristics	PC1	0.04	0.74
	PC2	0.25	0.051
Chemical variables	TP	-0.08	0.54
	Secchi depth	-0.12	0.39
	Calcium	0.19	0.19
	pH	0.16	0.24
Biological variables	18s Chao1	-0.13	0.32
	18s Observed OTUs	-0.10	0.46
	18s PCOA1	0.13	0.38
	18s PCoA2	0.38	0.004**
	18s PCoA3	0.4	0.002**

a. *** = p<0.001; ** = p<0.01; * = p<0.05

b. First and second principal component of physical characteristics (depth, volume, surface area, shoreline length).

Supplementary Table S2.7. Best solution for selected predictor variables based on DistLM "BEST" selection procedure for BC.

Number of variables	R^2	Adjusted R^2	Predictor variables
1	0.122	0.099	18s PCOA1
2	0.173	0.128	18s PCOA1, Longitude
3	0.208	0.142	18s PCOA1, Longitude, Latitude
4	0.237	0.150	18s PCOA1, Longitude, Latitude, Calcium
5	0.265	0.157	18s PCOA1, Longitude, Latitude, Calcium, Altitude
6	0.290	0.161	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP
7	0.314	0.164	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA3
8	0.337	0.166	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA2, 18S PCOA3
9	0.358	0.165	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA2, 18S PCOA3, Sechi depth
10	0.367	0.162	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA2, 18S PCOA3, Sechi depth, 18s Chao1
11	0.395	0.158	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA2, 18S PCOA3, Sechi depth, 18s Chao1, PC1-Physical variables
12	0.412	0.151	18s PCOA1, Longitude, Latitude, Calcium, Altitude, TP, 18S PCOA2, 18S PCOA3, Sechi depth, 18s Chao1, PC1-Physical variables, PC2-Physical variables
11	0.429	0.143	All



Supplementary Figure S2.1. Scatter plot of alpha diversity index (Chao1) for BC and MEC obtained from southern Ontario lakes.

APPENDIX B; SUPPLEMENTARY INFORMATION OF CHAPTER 3

Supplementary Table S3.1. Summary of Great Lakes fish species sampled for gut and skin microbiome. We provide a description of the normal diet of the host species along with the sample locations and total sample size.

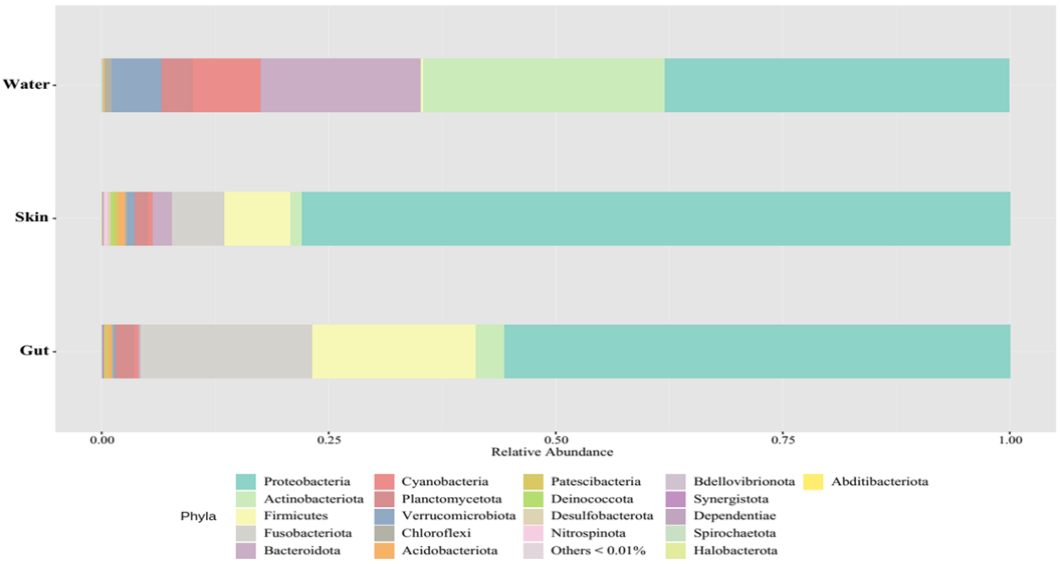
Species	Diet	Habitats			Number of samples
		Detroit River	Lake Erie	Lake Ontario	
Alewife (*Alosa pseudoharengus*)	Mostly zooplankton (such as copepods, Bythotrephes longimanus), Diporeia, or Mysis	0	0	22	22
Blacknose shiner (*Notropis heterolepis*)	Small aquatic invertebrates, Cladoceran (Chydoridae and Bosminidae) and ostracods	11	0	0	11
Brook silver (*Labidesthes sicculus*)	zooplankton, including copepods, cladocerans, and midge larvae	7	0	0	7
Brown trout (*Salmo trutta*)	Small fish (Alewives, rainbow smelt), aquatic and terrestrial insects, fish eggs, amphibians and crayfish	0	0	16	16
Emerald shiner (*Notropis atherinoides*)	Protozoans are important in the diet of the young-of-the-year shiners, and fish and insect larvae are eaten by adults	15	0	0	15
Freshwater Drum (*Aplodinotus grunniens*)	Initially planktonic cladocerans (zooplankton) and larval midges (Chironomidae) with increasing dependence on benthic invertebrates (primarily chironomids) and small fish	5	17	10	32
Gizzard Shad (*Dorosoma cepedianum*)	Fry feed primarily on copepods and cladocerans, whereas adults consume large amounts of phytoplankton and zooplankton	10	0	2	12
Lake trout (*Salvelinus namaycush*)	Alewife, round goby, rainbow smelt, slimy sculpin	0	0	24	24
Pumpkinseed (*Lepomis gibbosus*)	Benthic macroinvertebrates, zooplankton	11	0	5	16
Rock Bass (*Ambloplites rupestris*)	Aquatic insects, larval midges (Chironomidae), and small fish (round goby)	7	0	3	10
Round goby (*Neogobius melanostomus*)	Larval midges (Chironomidae), amphipods (Echinogammarus), dreissenid mussels	0	8	16	24
Spotfin shiner (*Cyprinella spiloptera*)	Aquatic and terrestrial insects; plants material and fishes also recorded in diet	12	0	0	12
Walleye (*Sander vitreus*)	Initially zooplankton, and benthic invertebrates and when they get older predominately fish (pumpkinseed and bluegill), some benthic invertebrates	0	19	7	26
White Bass (*Morone chrysops*)	Consume a variety of aquatic insects, amphipods, and zooplankton	0	10	11	21

White perch (*Morone americana*)	Zooplankton, benthic invertebrates, insect larvae, fish eggs, fish, dreissenid mussels	0	21	12	33
White sucker (*Catostomus commersonii*)	Larvae feed near surface on protozoans, diatoms, small crustaceans, and bloodworms. Adults feed opportunistically on bottom organisms, both plant and animal (e.g., chironomid larvae, zooplankton, small crayfishes)	10	0	2	12
Yellow perch (*Perca flavescens*)	Piscivorous, Shifts in diets from benthic invertebrates (mainly Diptera) in early spring to prey fish in the summer and fall	10	15	16	41
Total		98	90	146	334

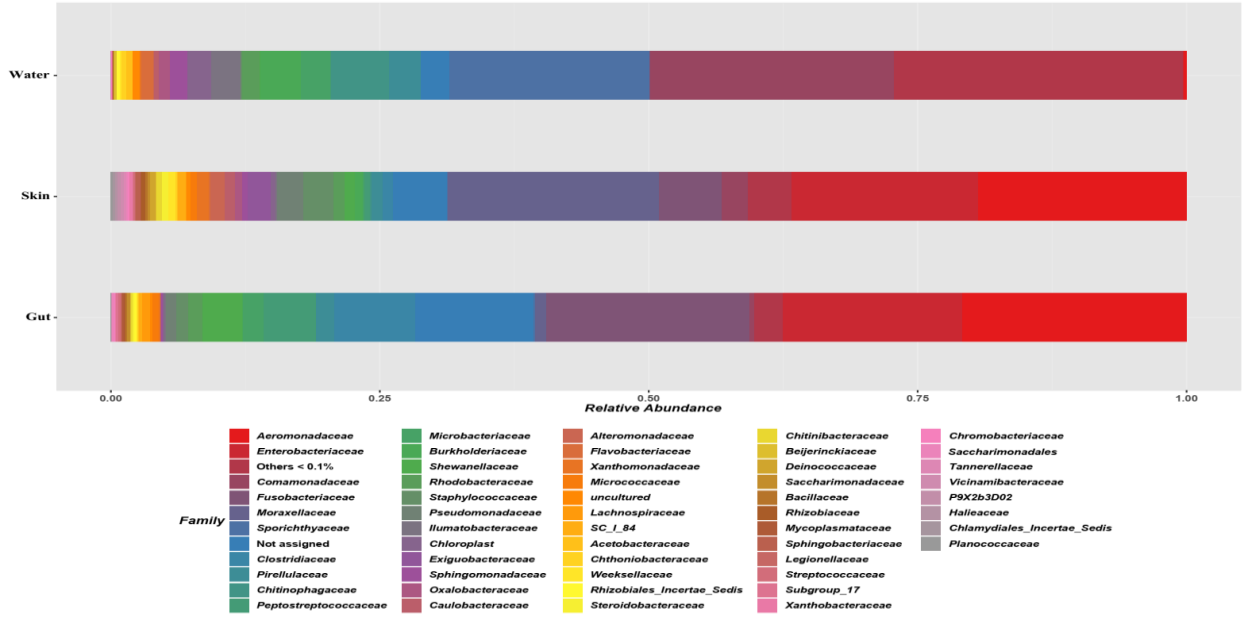
Supplementary Table S3.2. Comparison of differentially abundant bacterial taxa at the family level for gut and skin microbiomes across all fish species and sample locations using DESeq2 method (Benjamini-Hochberg false-discovery rate [BH FDR] 0.05, $|\log_2 \text{fold change}| > 2$). Positive $\log_2$ FC indicate higher abundance in skin samples and negative $\log_2$ FC specify higher abundance in gut samples.

Family	$\log_2$ FC	padj
Deinococcaceae	5.7	2.00E-169
Exiguobacteraceae	5.6	4.00E-144
Alteromonadaceae	5	2.00E-98
Moraxellaceae	4.9	2.00E-86
Oxalobacteraceae	4.6	1.00E-135
Caulobacteraceae	4	2.00E-84
Weeksellaceae	3.3	2.00E-50
Sphingobacteriaceae	3.2	3.00E-68
Xanthomonadaceae	2.7	4.00E-49
Rickettsiaceae	2.7	1.00E-44
Flavobacteriaceae	2.5	1.00E-30
Alcaligenaceae	2.4	1.00E-40
Devosiaceae	2.3	6.00E-41
Pseudomonadaceae	2.1	9.00E-28
Xanthobacteraceae	2.1	7.00E-25
Comamonadaceae	2	3.00E-36
P9X2b3D02 (Nitrospinota)	2	8.00E-35
Microbacteriaceae	-5.7	9.00E-139
Lachnospiraceae	-4.3	1.00E-72
PeM15 (Actinobacteria)	-4.3	7.00E-118
Rhizobiales	-4.1	1.00E-106
Cyanobiaceae	-3.4	2.00E-69
Clostridiaceae	-3.3	2.00E-40
Peptostreptococcaceae	-3.3	2.00E-36
Subgroup_17 (Vicinamibacteria)	-3.1	9.00E-65
Pirellulaceae	-3.1	6.00E-50
Caldilineaceae	-3	9.00E-68
Microcystaceae	-3	4.00E-53
Saccharimonadales	-2.9	6.00E-56
Mycoplasmataceae	-2.8	1.00E-38
Isosphaeraceae	-2.6	1.00E-46

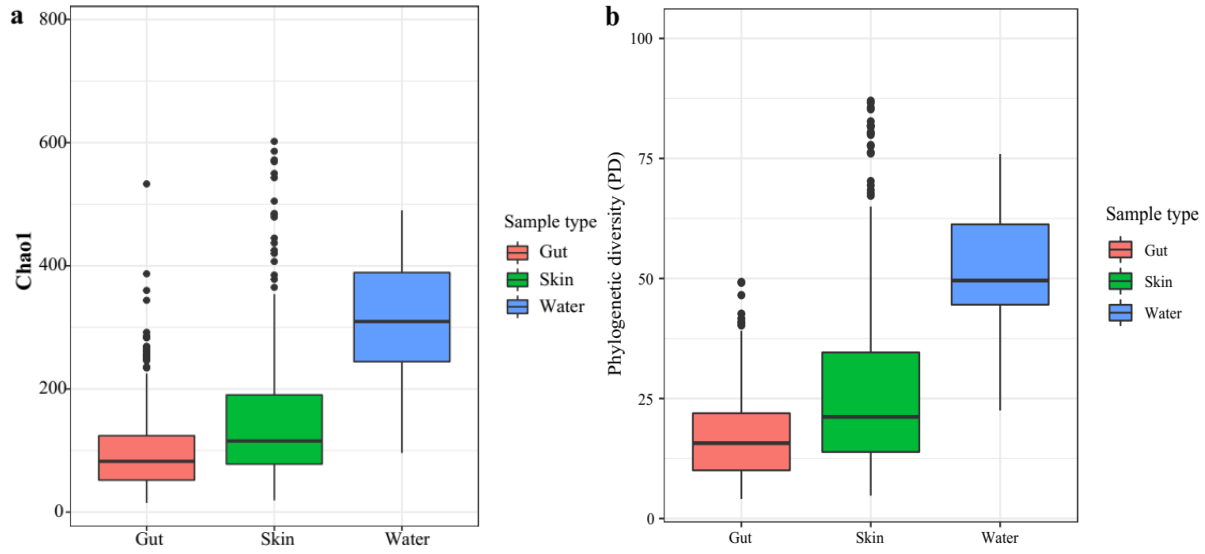
IMCC26256 (Acidimicrobiia)	-2.6	7.00E-58
Saccharimonadaceae	-2.6	5.00E-42
1_20 (Anaerolineae)	-2.4	1.00E-40
Gemmataceae	-2.3	3.00E-39
Holosporaceae	-2	5.00E-33
Fusobacteriaceae	-2	6.00E-15



Supplementary Figure S3.1. Relative abundance of bacterial community composition presented at the phylum level for gut, skin, and water microbiomes (samples are combined across sample types). Phyla with less than 0.01% of relative abundance are combined and presented as "others")



Supplementary Figure S3.2. Bacterial community composition (relative abundance at the family level) for gut, skin, and water microbiomes across all fish species collected at three sites in the Great Lakes (Lake Erie, Lake Ontario and Detroit River). Bacterial families with less than 0.1% relative abundance are combined and presented as "others".



Supplementary Figure S3.3. Box and whisker plots of alpha diversity indices (Cho1 (a) and Faith's phylogenetic diversity (b)) for gut, skin and water microbiomes. The colours reflect sample type and the black dots are outliers. The black line in each box plot is median.

APPENDIX C; SUPPLEMENTARY INFORMATION OF CHAPTER 4

Supplementary Table S4.1. Bacterial species with their proportion present in Jamieson Probiotic Complex with 60 billion colony forming units (CFU) (Jamieson Laboratories, Canada) used in this study.

Bacterial species	Total active cells CFU (billion)
Lactobacillus plantarum (R1012ND)	25.8
Lactobacillus casei (R0215ND)	19.8
Bifidobacterium breve (HA-129)	8.7
Bifidobacterium longum subsp. longum (HA-135)	3.0
Lactobacillus paracasei (HA-196)	2.1
Lactobacillus acidophilus (HA- 122)	0.15
Bifidobacterium animalis subsp. lactis (HA-194)	0.15
Bifidobacterium bifidum (HA-132)	0.15
Lactobacillus rhamnosus (HA- 111)	0.15

Supplementary Table S4.2: Primers and prob list for 56 genes selected based on RNA sequencing analysis

Function	Gene	Abbreviation	Probe (5'-3')	Primer F (5'-3')	Primer R (5'-3')
Cell integrity	vacuolar protein-sorting-associated protein 25-like	*vps25*	CCGCCATCACAAGC	CATGGTGCTCCCTTGCTCTC	GTGGTTGAACACAGGGCACTCT
	cilia and flagella associated protein 58	*cfap58*	GTCCATGCTGAACAAG	AATGATGAGCTGGCCCTACTCTAC	GTCCTCCACCCTCTGGTTGTAC
	FERM domain containing 4Bb	*frmd4bb*	AGCAGGTTGAAGACGAGAT	AGGACGCCATGAGGAAACTG	CAGAGGCTCTTTGGGTAGGC
	mesencephalic astrocyte-derived neurotrophic factor	*manf*	CCCTAGTCAAAACCTG	CACCAGCGCAGACATCGA	CCTTGCCCTTGGCGTCTT
	sodium-dependent multivitamin transporter-like	*slc5a6*	TGCCCTTCACTATAGCTGCA	CTTATGCTGGCCCTGACCAA	AAGGGGAGAAGACCACTGGA
	occludin a	*ocln*	CAAGTCCAACGTCCTGTGG	TGTTTGCCTTAAAGACGCGC	GGGCCTCCTCGTTGATGATC
	ano7	*ano7*	TATGTTCCGGGATTCCT	GCAGATGCTAAGCGACAGGA	AAGCGGTGTACTCAATGCCA
	piezo-type mechanosensitive ion channel component 1-like	*piezo1*	ACACCAAGGCGGATC	TAAAGAGGGCCGTCGGAAA	CACGCTTGTGGCTTCTCTTTT
Growth	WAS protein family homolog 1	*wash1*	TCCATCTTCAGTGGAGCCA	TCCAGAGCGTCTCCAGGATT	TTGGCCTGTATGCGGTATCG
	ubiquitin protein ligase E3 component n-recognin 4	*ubr4*	GACAGACCAGACTTCCA	TCAGCGGAGAGTGAGAGTGAGA	GTGTCCAGAACGTGTCCAGTCTC

	phosphatase and actin regulator 1-like	*phactr1*	TACGCCTTACATCACCGAGG	GGATGAGATAGACCGTCGGC	TCCAGGTGGAGAGAGAGACGC
	sorting nexin-10B-like	*snx10b*	TAACCTGAACAACGCCCAG	CTACTCATGGTCCAGCTGCCT	GGCCCTTCATCCTCTCTGTG
	rabaptin, RAB GTPase binding effector protein 2	*rabep2*	TCGACAGACTCTCCCACCA	GGTTGGAGGCAGGAGCTATG	TATCCCATCCGTCCGTTCCT
	trafficking kinesin-binding protein 1-like	*trak1*	GACGACTGGCTCCACA	GGACTGCATGTTTGGCTACG	ATCTGATCGTGCGTGAGTCC
Immune system	annexin A1-like	*anxa1*	AGTGCATGAGTGTCAAT	CCAGCAGACAATGTCCACCA	GGCAAGAGCATTTGAGGCAA
	macrophage-stimulating protein receptor-like	*mst1r*	GTCTGCCTGTCCAAAG	CCCCGTCAAGAGTGTCACTGA	GACAGGCACCTTCTTCACCAC
	transcription factor 12	*tcf12*	GAGTGCTGTCCTCAGTC	CCAGACCACCATCAAGCTGT	TCAACAACACGGTCTGGCTT
	vacuolar protein sorting-associated protein 33A-like	*vps33a*	CCCCACCATCAGGAAG	GGTCTTCTGAAGCCCCAGAC	CTGTTCGTTGGCATCCTCCA
	apoptosis inducing factor mitochondria associated 4	*aifm3*	TCTGAAAAGTGGTGCGGTC	GTGGTGGAGATCAGAGGGGA	CCCTGCTATCACCACATCCG
	HCLS1-binding protein 3-like	*hs1bp3*	TTGAAAGTGAGGAGACCGCT	GAACGTTGCAGATTGGCAGG	TTGGTTACTTTCGCAGCAGC
	COP9 signalosome complex subunit 6	*cops6*	ACTCCTTTGAGCTGCTTC	TAATCGGGAAGCAGGAGGGT	TGTCAATGTGTGCTCGGTCA
	protein arginine methyltransferase 3	*prmt3*	CGCACAGAAAGCTACAGG	GGCCACTACAGCATCCATGA	GAACACTTCCGGGTTGAGGT

	B-cell linker protein-like	*blnk*	ACGAGGAGCTTTTCAGTAGT	CGGAAAGGAGGGCAAGAAGA	TCGATCAGCAGCAGTTGGTT
	syndecan-1-like	*sdc1*	GCTGCCTCCCTGTTAA	TTATTGCAGGAGGAGTGGTGG	AGCCTCCCTCATCTTTCCTCT
	SID1 transmembrane family, member 2	*sidt2*	AATTTGGCCCCGTTTC	GAAATGGTTTTGAGGAGCTGTTG	CATAACCAAGCCGGAACAAGA
	proteasome subunit beta type-4-like	*psmb4*	TGGACCAAAACCCGGACA	TCCAAGCGGTTTGAAGCTCA	CATCCTGGGGCTAAACTGCT
	CD151 antigen-like	*cd151*	AACATGTGGGACCGTCT	CCAACAGCGTACGACGAGAA	TGCCAGCCAGAAAAGGAAGT
	dispanin subfamily A member 2b-like	*dspa2b*	GGAAGACTGCATGCTG	GGAGACTTGGAGGGAGCAAG	GAAGAGCAGCCCGAGGATTA
	protein PML-like	*pml*	AGCTTCGACGTCGTTATG	CATTCGACGGCAGACAGGAT	TGCTCCCATACCCAAAGAGG
	interferon alpha/beta receptor 2	*ifnar2*	AGCTACCACTGGACTGAC	CTCTAACCCCGAGACACACG	GTCCAAGGCCCCATCTTTCA
	ras-related protein Rab-34-like	*rab34b*	GTGTCTCTTGTGTGTGTGAT	TTGTTGGAGCGTGTGTGACTGT	GTGGAAGTTACGGGGTGGTAGG
Metabolism	NTPase KAP Family P-Loop Domain-Containing Protein	*kidins220*	GCTTTGTGGGAGGTGT	CAGAGATCGATGACGCCGTC	GAAGCTCCATCCTCGAGTCG
	neuronal acetylcholine receptor subunit alpha-3-like	*chrna3*	ACCAGTTTATCAGGCCGGTG	CAGGTTGTTCAGGAGGCTGTTC	ACCTCAAACTCCACGGTGACG
	golgin subfamily A member 7-like	*golga7*	TGAGGACGCTGAACAA	AGCAGCAGTTTGAGGAGACGG	CAGACAGCCCTCCAGATACGAC

	free fatty acid receptor 2-like	*ffar2*	GCTCTCTACTGCCCTGTG	GTGCAGGGGTTGTACTTCTGG	CATTGGAGCTTGGTGTGCTG
	multidrug resistance-associated protein 7-like	*mrp7*	CCCTGGAACACTGTCAC	GAGAACCTGGACCCATGTGG	GATGCAGTTGATGACAGCGC
	ER degradation-enhancing alpha-mannosidase-like protein 2	*edem2*	GCCAGGGCTACAGAG	AAGGGAACGGTCTCTATGCCT	TCAATGCACTGTCCAGCTCA
	NTPase KAP family P-loop domain-containing protein 1-like	*nkpd1*	CATCATGTCAGAACGC	CCTGCACAGTGGGCCTCTAC	CTCTCTGCCTCCTGCTCCAT
	homeobox protein PKNOX1-like	*pknox1*	AAACTGACAGCCCCTGTT	CTGCGCACCAAGATGAACAG	TGGGCTGGAGGTCTGAATCT
	WEE2 oocyte meiosis inhibiting kinase [	*wee2*	TCAGACGGGCACAGTTA	CCCCTCTCCCTCAGAATGGA	TTCTGAAAGCAGGGGAGAGC
	cytochrome b-c1 complex subunit 6, mitochondrial-like	*uqcrh*	GCTCCAAGTCCGAAAC	GCGAAACTGGAGGACTGTGA	CGCATGGAGGAAGTCGAAAAG
	low-density lipoprotein receptor-related protein 2-like	*lrp2*	GGCATCTCCTGTACCTGT	ACTGCAATGGGAATGGGGAG	GGTTTTGGCAGGAGTCTCCA
	lysophospholipid acyltransferase LPCAT4-like	*lpcat4*	CTCTCCTTCACATCGCCT	ATCTAAGTGTGGCGGCTGTCTC	GACTGCCACTACCCTCTCTGTCA
	6-phosphogluconate dehydrogenase	*pgd*	TCGCTATGGACCCTCACTGA	TTGTTTGTCGGCAGTGGAGT	TCACCAGGCCTCTTTGTGTC
	echinoidin	*echinoidin*	CCTTTTACCTGATATCTGC	AAACAGTGACTGGGCCTACAGAAG	AAGCCGTGTGATCCAGAGAGA
Stress-related genes	X-ray radiation resistance associated 1	*xrra1*	TGTTTCCTCAATCTGGCTGA	GTTTGCCACTACCAGAGCTTTGT	GCATAGGAAACAGGGCCACA

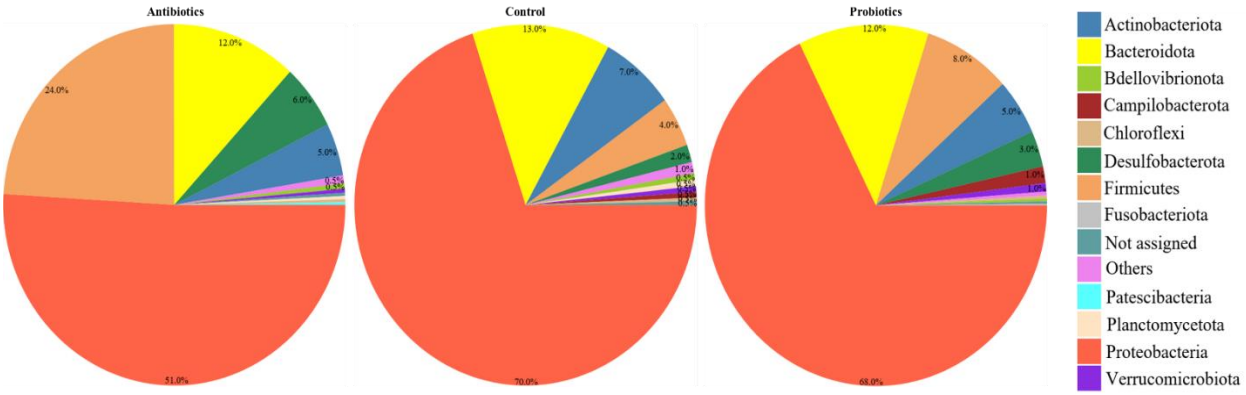
	heat shock factor-binding protein 1-like	*hsbp1*	ATGCAGGATAAATTC	CCAGACGCTGCTGCAACA	CGATGATCTGGTCTGACATGGT
	transient receptor potential cation channel subfamily V member 5-like	*trpv5*	ACAAGACCCCTGGAACAT	GATCCACCGCATGAAGGGAA	ATCCCGAACACAGCCATCAG
	polyubiquitin-like	*ub*	ACCCAGTTCAAAGCCAA	TGAGGTGGTGTCAGGTGAGACT	TGAATCAGCCTCTGTTGGTCG
	trimeric intracellular cation channel type B-lik	*tmem38b*	GATGGGCTAAAGGTGCT	CAAGGATGGACTGCTGGTGA	CCTCGTACCAGCTGCTCAAA
Endogenous control genes	β-2-microglobulin	*b2m*	ACACCCTGATCTGTCACGTG	AGGGGAACATGGGAAGGACA	ATGCCGTTCTTCAGGAGCTG
	β-Actin	*actb*	CCTGGTCGTTGATAAC	TTCCCGGTGCAGAAATGGA	ATCATCTCCTGCGAAACCGG
	ribosomal protein L13	*rpl13*	CAATGCACAGTTTTAG	TCTACCATTGGGTGCCATATCC	TGGCTGAAAGGAAAAAGGAAGT
	glyceraldehyde-3-phosphate dehydrogenase	*nadp*	CAGGGGCAGAACGACAGA	TTCAATGACGGCCAGGGA	ATTGAATCCCCCGAGCTGAC

Supplementary Table S4.3. Comparison of gene expression levels and differentially expressed gene distributions between the treatment groups (|log2 Fold Change| > 1 and FDR P value < 0.05)

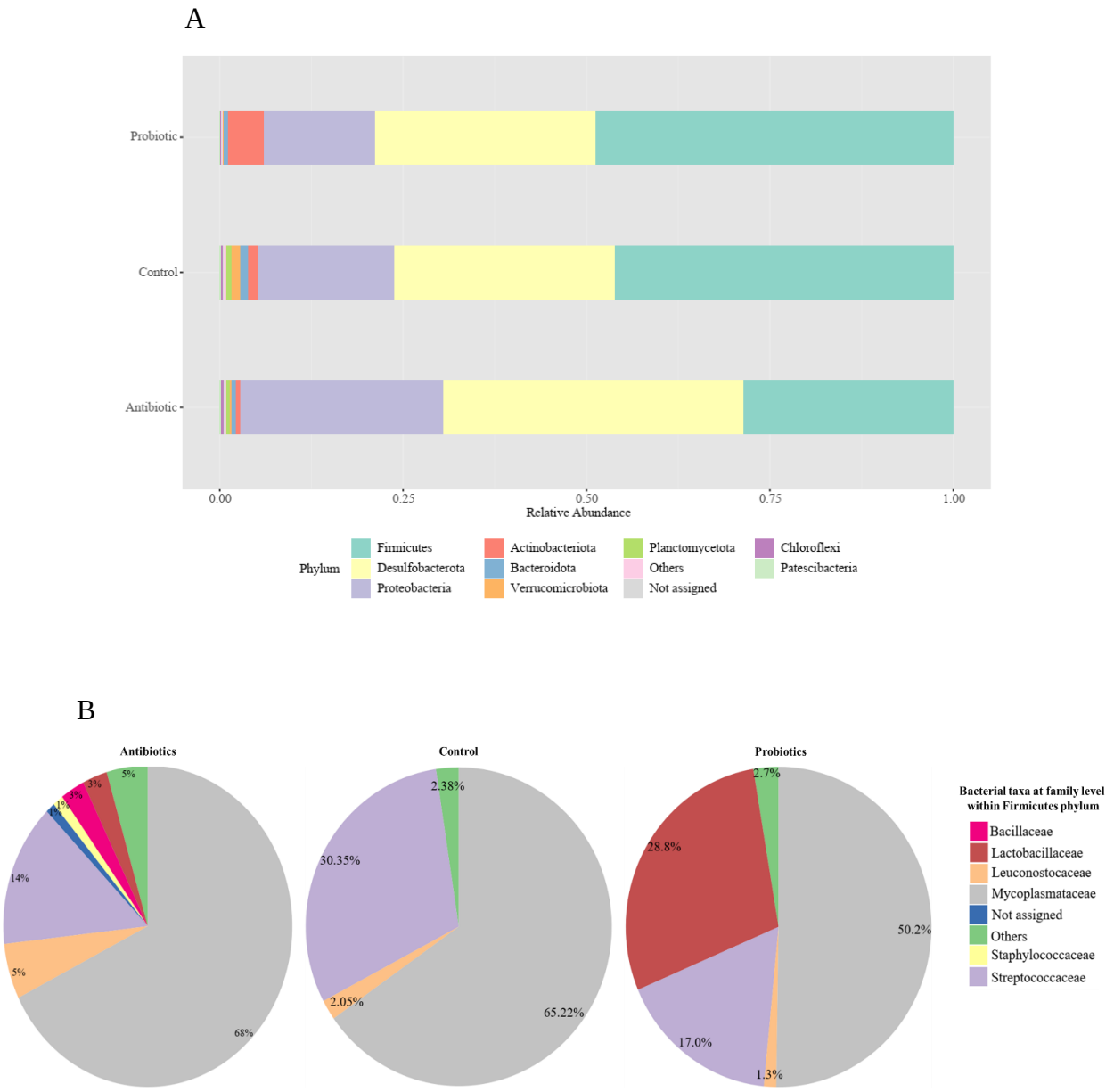
Genes	Gene name	Gene abbreviation	Base Mean	log2 FD	pvalue	padj
Control Vs Antibiotic						
LOC112248679	golgin subfamily A member 7-like	*golga7*	6.7	-6.86	1.97E-06	0.008
LOC112248188	annexin A1-like	*anxa1*	123	-5.56	6.34E-07	0.005
LOC112245911	vacuolar protein sorting-associated protein 33A	*vps33a*	57.2	5.09	3.59E-06	0.009
LOC112239977 *	apoptosis inducing factor mitochondria associated 3	*aifm3*	14.9	-5	6.54E-06	0.011
LOC112234485	uncharacterized	uncharacterized	192	-4.76	1.23E-05	0.015
LOC112234113	transmembrane protein 220-like	*tmem220*	40	4.68	1.09E-08	0
LOC112232636	ER degradation enhancer, mannosidase alpha-like 2	*edem2*	29.8	4.52	4.83E-05	0.047
LOC112246816	HCLS1 binding protein 3	*hs1bp3*	160.1	-4.37	6.66E-06	0.011
LOC112232613	peroxisomal biogenesis factor 14	*pex14*	17.2	-4.31	7.21E-06	0.011
LOC112216100	cyclin-dependent kinase 11B	*cdk11b*	32.7	-4.1	3.38E-05	0.03
LOC112245791	multidrug resistance-associated protein 7-like	*mrp7*	28.4	-4.01	9.57E-06	0.014
LOC112251319	phosphatase and actin regulator 1	*phactr1*	23.1	-3.97	2.82E-05	0.03
nsrp1	nuclear speckle splicing regulatory protein 1	*nsrp1*	21.1	-3.91	6.33E-06	0.011
LOC112265673	serine protease 16	*prss16*	41.2	3.69	4.34E-05	0.043
LOC112249106	piezo-type mechanosensitive ion channel component 1	*piezo1*	41.2	3.69	4.34E-05	0.043
LOC112258258	septin-8-A	*SEPTIN8*	42.1	3.84	4.24E-06	0.008
LOC112231356	free fatty acid receptor 2-like	*ffar2*	21.3	-3.66	4.05E-06	0.009
LOC112259970	vacuolar protein-sorting-associated protein 25-like	*vps25*	12.4	-3.52	3.82E-06	0.009
LOC112220969	macrophage-stimulating protein receptor-like	*mst1r*	47.2	-3.38	2.10E-06	0.008
ubr4	ubiquitin protein ligase E3 component n-recognin 4	*ubr4*	149.1	-2.75	5.55E-07	0.005
cfap58	cilia and flagella associated protein 58	*cfap58*	302.4	1.73	1.04E-05	0.014
LOC112215983	natriuretic peptide B	*nppb*	34.2	1.61	1.06E-05	0.014
LOC112241820	NTPase KAP family P-loop domain-containing protein 1-like	*nkpd1*	422.9	-1.36	1.20E-06	0.007
LOC112243336	SET domain containing 2, histone lysine methyltransferase	*setd2*	86.5	-1.29	1.37E-05	0.016
LOC112218626	WAS protein family homolog 1	*wash1*	58.3	1.22	1.11E-05	0.014
LOC112261371	echinoidin	*echinoidin*	58.3	1.2	1.18E-05	0.014
LOC112241821	kinase D-interacting substrate of 220 kDa B-like	*kidins220*	558.8	-1.12	1.35E-06	0.007
xrra1	X-ray radiation resistance associated 1	*xrra1*	2201.2	1.07	6.86E-06	0.011
LOC112255867	neuronal acetylcholine receptor subunit alpha-3-like	*chrna3*	88.5	1.1	1.73E-06	0.008
Control Vs Probiotics						
LOC112255055	LSM3 homolog, U6 small nuclear RNA and mRNA degradation associated	*lsm3*	114.8	-7.27	1.00E-05	0.017
LOC112219606	transmembrane protein 38B	*tmem38b*	114.8	-7.27	1.11E-05	0.017
LOC112232343	Heat shock factor-binding protein 1-like	*hsbp1*	131.3	-6.57	1.91E-08	0
LOC112246837	COMM domain containing 10	*commd10*	31.4	-5.88	3.12E-07	0.002
LOC112220855	COP9 signalosome complex subunit 6	*cops6*	138.4	-5.74	5.03E-07	0.002
LOC112234485	uncharacterized	uncharacterized	191.9	-5.59	7.11E-08	0.0007
LOC112245911	Vacuolar protein sorting-associated protein 33A	*vps33a*	57.2	5.4	3.37E-07	0.002
prmt3*	protein arginine methyltransferase 3	*prmt3*	138.8	-5.25	3.34E-06	0.007
LOC112264620	THO complex subunit 4	*alyref*	138.8	-5.25	3.00E-06	0.007
LOC112248204	low-density lipoprotein receptor-related protein 2	*lrp2*	12.9	5.16	5.71E-05	0.045
LOC112253929	homeobox protein PKNOX1	*pknox1*	18.1	-5.02	8.06E-07	0.003
LOC112246700	cytochrome b-c1 complex subunit 6, mitochondrial-like	*uqcrh*	97.9	4.75	9.35E-06	0.015
sidt2	SID1 transmembrane family, member 2	*sidt2*	19.7	4.66	4.79E-05	0.04
LOC112264739	proteasome subunit beta type-4-like	*psmb4*	97.9	-4.25	3.29E-05	0.034
LOC112263481	CD151 antigen	*cd151*	18.5	4.07	3.89E-05	0.037
LOC112248608	B-cell linker protein-like	*blnk*	37.6	-4.01	3.57E-06	0.007
LOC112232610	phosphogluconate dehydrogenase	*pgd*	155.3	-3.97	4.62E-05	0.04
LOC112234113	transmembrane protein 220-like	*tmem220*	40	3.6	1.73E-06	0.004
LOC112214559	lysophospholipid acyltransferase	*lpcat4*	42.1	3.46	1.27E-05	0.017
LOC112249934	syndecan-1	*sdc1*	71.6	-3.41	1.27E-05	0.017
rabep2*	rab GTPase-binding effector protein 2	*rabep2*	46.8	-3.05	4.56E-05	0.04
LOC112261503	sodium-dependent multivitamin transporter	*slc5a6*	291.4	1.8	3.44E-05	0.035
LOC112242147	uncharacterized	uncharacterized	291.3	1.8	3.44E-05	0.034
LOC112217687	R-spondin-3-like	*rspo3*	90.9	1.38	2.33E-05	0.028
LOC112241820	NTPase KAP family P-loop domain-containing protein 1-like	*nkpd1*	422.9	-1.35	5.06E-07	0.002
LOC112242158	Sorting nexin-10A	*snx10b*	7818.8	1.24	9.94E-10	0
LOC112218768	WEE2 oocyte meiosis inhibiting kinase	*wee2*	353.9	1.15	9.19E-07	0.003
LOC112254714	ras-related protein Rab-34-like	*rab34b*	558.7	-1.03	2.59E-06	0.005

LOC112241821	kinase D-interacting substrate of 220 kDa B-like	*kidins220*	558.7	-1.03	2.59E-06	0.005
Antibiotic Vs Probiotic						
LOC112232343	heat shock factor-binding protein 1-like	*hsbp1*	131.3	-6.47	1.22E-07	0.002
LOC112220855	COP9 signalosome complex subunit 6	*cops6*	138.4	-6.27	2.05E-07	0.002
LOC112248188	annexin A1-like	*anxa1*	123	5.74	2.91E-07	0.002
LOC112232636	ER degradation enhancer, mannosidase alpha-like 2	*edem2*	29.8	-5.67	3.23E-07	0.002
LOC112246837	COMM domain containing 10	*commd10*	31.4	-5.39	4.10E-06	0.01
LOC112239977	apoptosis inducing factor mitochondria associated 3	*aifm3*	14.9	5.17	4.30E-06	0.01
LOC112232613	peroxisomal biogenesis factor 14	*pex14*	17.2	4.82	1.14E-06	0.004
LOC112264739	proteasome subunit beta type-4-like	*psmb4*	97.9	-4.59	2.06E-05	0.033
LOC112265459	arsenite methyltransferase-like	*as3mt*	57.5	-4.33	2.39E-07	0.002
nsrp1	nuclear speckle splicing regulatory protein 1	*nsrp1*	21.1	4	4.35E-06	0.01
LOC112216146	FERM domain containing 4Bb	*frmd4bb*	24.1	3.93	3.97E-08	0.001
LOC112231356	free fatty acid receptor 2-like	*ffar2*	21.3	3.47	1.10E-05	0.02
ano7	anoctamin 7	*ano7*	62.4	-3.31	4.93E-05	0.049
LOC112218750	protein mono-ADP-ribosyltransferase	*parp12*	39.3	3.13	2.94E-05	0.036
LOC112222087 *	mesencephalic astrocyte-derived neurotrophic factor	*manf*	43.9	-2.96	1.98E-06	0.006
LOC112245658	trafficking kinesin-binding protein 1	*trak1*	36.8	2.52	2.38E-05	0.035
LOC112245441	uncharacterized	uncharacterized	20.3	2.37	4.77E-05	0.049
LOC112262831	polyubiquitin	*ub*	687.1	2.32	2.65E-05	0.035
LOC112237710	dispanin subfamily A member 2b	*dspa2b*	780.5	1.89	2.69E-05	0.035
LOC112215983	natriuretic peptide B	*nppb*	34.2	-1.54	2.60E-05	0.035
LOC112217407	protein PML	*pml*	442.7	1.36	4.40E-05	0.048
LOC112218768	WEE2 oocyte meiosis inhibiting kinase	*wee2*	353.9	1.19	1.16E-06	0.004
LOC112249580	transcription factor 12	*tcf12*	40.7	-1.1	5.97E-06	0.01
LOC112225266	interferon alpha/beta receptor 2	*ifnar2*	158.6	1.07	3.60E-05	0.043
LOC112220311	Occludin a	*ocln*	2096.9	1.05	7.94E-06	0.016
LOC112225425	T cell differentiation protein 2	*mal2*	155.6	1.05	6.50E-05	0.049
LOC112252883	transient receptor potential cation channel subfamily V member 5	*trpv5*	160.4	-1.01	1.43E-06	0.005

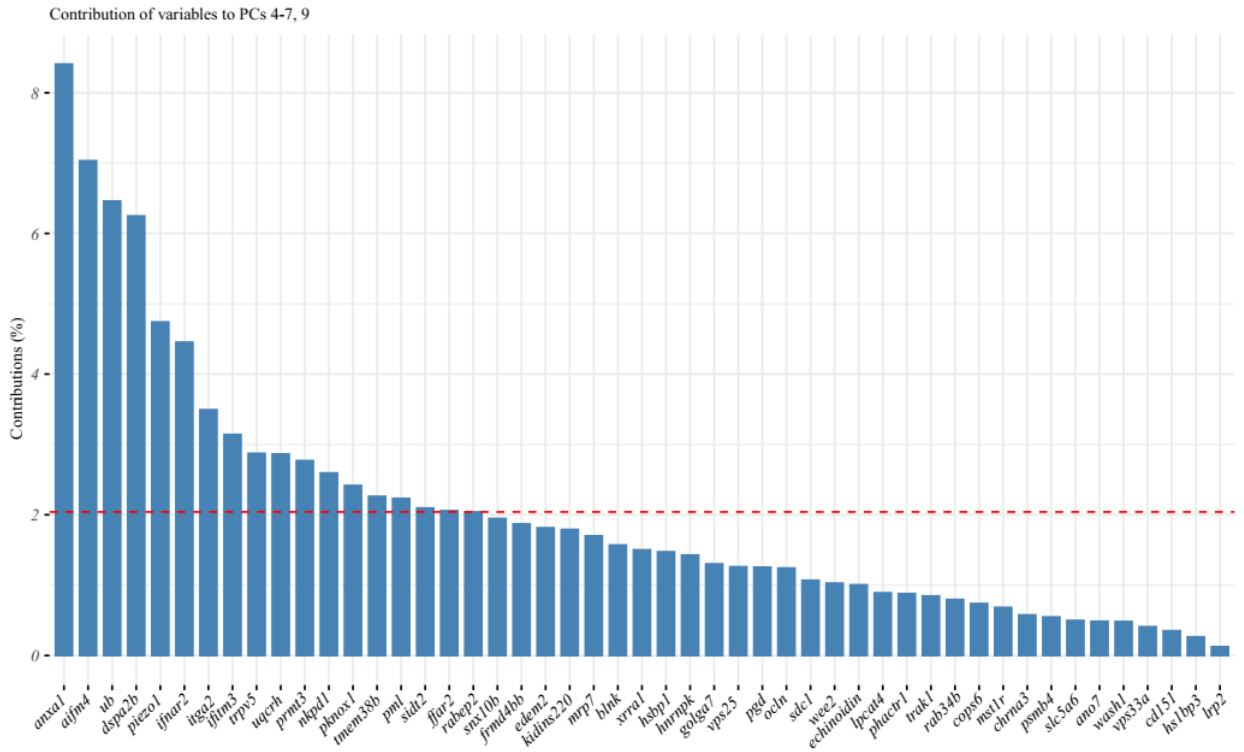
* Indicate that these genes were also significant in our OpenArray high-throughput RT-qPCR analysis.



Supplementary Figure S4.1. Pie charts showing microbial community composition of water samples with the relative abundances of the most abundant (top 10 families) bacteria phylum in control, antibiotic, and probiotic groups. Other less abundant taxa were merged and renamed as “Others”. Bacteria taxa with unidentified families were classified as “Not assigned”.



Supplementary Figure S4.2. (A) Bar plot comparing relative abundance of top 10 gut associated bacterial taxa at phyla level between the groups. All samples within each group are combined and average of abundance is presented for each group. Other less abundant phyla are classified as "Others". (B) Families presented within Firmicutes phyla are shown among the groups. Taxa at with low abundance (0.01%) were merged and renamed as "Others" in the pie chart.



Supplementary Figure S4.3. The red dashed line on the graph above indicates the expected average contribution. For a given component, a variable with a contribution larger than this cutoff could be considered as important in contributing to the component.

VITA AUCTORIS

NAME:	Javad Sadeghi
PLACE OF BIRTH:	Malayer, Hamedan Province, Iran
YEAR OF BIRTH:	1987
EDUCATION:	Tehran University of Medical Science, Tehran, Iran, M.Sc. in Medical microbiology, 2014