Analysis of microbial communities associated with Tilapia Aquaculture in Malawi

Submitted by Shayma Abdulmuttaleb Kareem Alathari to the University of Exeter

as a thesis for the degree of

Masters by Research in Biological Sciences

In July 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature …….

For Maryam

Acknowledgements

I would like to thank Ben Temperton and Charles Tyler for giving me this opportunity and being brilliant supervisors. I would also like to thank Dominique Chaput, for all her support and encouragement throughout this study.

I would like to thank Lauren Adams and Beth Dean for brightening my day with words of encouragement. A special thanks to Samantha Mitchell for guiding and welcoming me in the lab, and all of the members of the Temperton lab.

A big thank you must go to my family, for their understanding and support all year round.

Abstract

Aquaculture is of major and increasing importance to global food security, particularly in Low Income, Food Deficit Countries (LIFDCs), where it also serves as a significant contribution to poverty alleviation. Disease is widely acknowledged as the prominent bottleneck to achieving global food security and poverty alleviation targets relating to aquaculture, with annual losses exceeding >$6bn (Food and Agriculture Organization 2014). Outbreaks of disease caused by endemic and emerging pathogens impact directly on farmer income and their nutritional security. Avoidance of yield-limiting disease outbreaks is a fundamental requirement for future growth and resilience of aquaculture in LIFDCs. Advances in molecular techniques coupled with next-generation sequencing have provided a step-change in understanding the role of host-associated bacteria, archaea, protists and viruses (the microbiome) in host homeostasis. Shifts in microbiome communities under stressful conditions can contribute to disease states. However, the role of microbiomes in the emergence of diseases in aquaculture, where stressors include feeding, antibiotic and disinfectant use and over-stocking, is poorly studied.

Here our study presents an evaluation of the microbiomes (bacteria and viruses) associated with tilapia and their pond environments in aquaculture, using 16S rRNA community profiling techniques and viral amplicon sequencing. Samples investigated in this project were collected from Malawi tilapia fish farms; their skin community composition and diversity were examined across geographical scales. The high variability observed of the microbial communities in small geographic regions, showed that future sampling to detect shifts due to dysbiosis will require time-resolved sampling of ponds under study. Nanopore sequencing of full length 16S rRNA genes, using MinION, allowed us to examine the microbial communities at higher taxonomic resolution than short read sequencing techniques. Its success lays the foundation for *in-situ* microbial profiling of aquaculture ponds for disease, and offers independence to farmers to monitor their own ponds. Successful amplification of the T4-like Myoviridae phylogenetic markers from one rearing water sample was achieved, although the required degeneracy of the primers inhibited multiplexing. Therefore, our findings suggest that inclusion of bacteriophages in microbiome studies is better served using shotgun metagenomic methods, rather than amplicon sequencing.

Finally, we investigated the use of skin swabbing as an alternative to bucket incubations to minimise animal stress when categorising the fish skin microbiome. Skin swabbing successfully captured similar microbial communities in comparison to bucket incubations, with greater diversity and variance between fish.

Table of Contents

List of Figures

List of Tables

Abbreviations used in this study

AMP	Anti-Microbial Peptide
ANOSIM	Analysis of similarities
ASV	Amplicon Sequence Variant
BIV	Bohle Iridovirus
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CTAB	Cetyl trimethylammonium bromide
dNTP	Deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agriculture Organization
Hz	Hertz
IPNV	Infectious Pancreatic Necrosis Virus
ISKNV	Infectious Spleen and Kidney Necrosis Virus
iTOL	International Tree of Life
LIFDC	Low Income, Food Deficit Countries
MCP	Major Capsid Protein
NFW	Nuclear free water
NMDS	Nonmetric Multidimensional Scaling
OTU	Operational Taxonomic Units
PBS	Phosphate Buffered Saline
PHB	poly-β-hydroxybutyrate
PoU	Prevalence of Undernourishment
RBF	Running Buffer with Fuel Mix
RT	Room Temperature
SCFA	Short-Chain Fatty Acid
SDS	Sodium dodecyl sulfate
SIMPER	Similarity Percentage
SINA	SILVA Incremental Aligner
SPRI	Solid-phase Reversible Immobilization
TAE	Tris acetic acid + EDTA
TBE	Tris/Borate/EDTA

TiLV	Tilapia lake virus
TLEV	Tilapia Larvae Encephalitis Virus
USD	United States Dollar
VLP	Virus-like Particles
VNN	Viral Nervous Necrosis

1. Introduction

1.1. Importance of aquaculture and its challenges

1.1.1. Aquaculture feeding the world

The increasing importance of aquaculture to global food security has been recognised in the past few decades. Aquatic food represents an essential component of the global food basket that can help alleviate poverty, and improve the nutrition, health, and well-being of all people (Tacon and Metian 2013).

Sustainable fisheries and aquaculture are required to fill the demand for food, not met by terrestrial food systems (Lynch and MacMillan 2017). Yet, limited attention has been given so far to fish as an important element in food security and nutrition strategies at national levels and in wider development discussions and interventions. Discussions have often concentrated predominantly on questions of biological sustainability and on the economic efficiency of fisheries, neglecting issues linked to their contribution to reducing hunger and malnutrition and to supporting livelihoods. Increased consumption of fish and other aquatic derived products, and their addition to the diets of low-income populations offers an important means for improving food security and nutrition (Béné et al. 2015) .

Fish is an animal protein which contains several essential amino acids, lipids and is also an important source of essential micronutrients, which makes it particularly attractive in the fight against malnutrition in LIFDCs (Low Income, Food Deficit Countries). Some countries (e.g., Zambia, Brazil, or Chile) have recognized this potential and have included fish in their national school-feeding programmes (Béné et al. 2015). With a growing global human population, rising from 7.3 billion to an estimated 9.8 billion between 2015 to 2050, fish will be a critical source of food (Walker 2016). Recent (2016) official statistics indicate that 59.6 million people are engaged in the primary sector of capture fisheries and aquaculture; 19.3 million people in aquaculture and 40.3 million people in fisheries (Food and Agriculture Organization of ...) As well as those directly involved in aquaculture production, there are many others whose livelihoods are at least partially connected to the aquaculture sector via the

supply of goods and services such as transportation, ice making, feed production and marketing. The increase in aquaculture over the period 1950-2014 is shown in Figure 1. Overall, it is estimated that more than 100 million people depend on aquaculture for a living, either as employees in the production and support sectors, or as their dependants (Food and Agriculture Organization 2014).

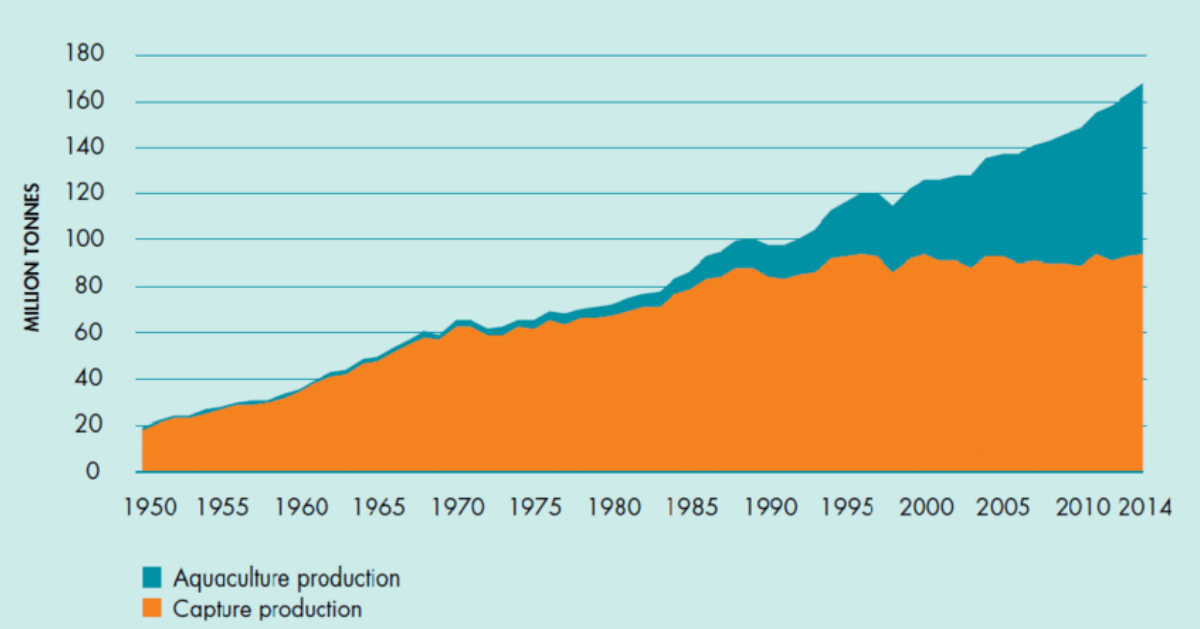


Figure 1. World capture fisheries and aquaculture production. The aquaculture production as shown in blue has increased from 20 million tonnes in 1950 to over 160 million tonnes in 2014 (Food and Agriculture Organization of ...).

Applied research and adaptive management techniques can assist with the necessary evolution of sustainable food systems to include a stronger emphasis on fish and other aquatic organisms (Lynch and MacMillan 2017). Additionally, necessary actions in fisheries and aquaculture must include effective governance, improved management and conservation, efforts to maximize societal and environmental benefits from trade, increased equitability of distribution and innovation in food production, and the continued development of low-input and low-impact aquaculture (Food and Agriculture Organization of ...).

The Food and Agriculture Organisation (FAO) estimates that the number of undernourished people in the world has been growing and in 2017 this reached 10.9% (Food and Agriculture Organization of ...) illustrating the need for enhanced effort needs to address the problem of world hunger.

1.1.2. Limitations: Threats from disease and overstocking

One of the main hurdles to sustainable finfish aquaculture, and indeed aquaculture more generally, is the management and control of infectious disease (Rodger 2016). This is exacerbated by overstocking and water quality pressures. Habitats for aquaculture development are further impacted by activities that affect water quality including mineral extraction, coastal development, waste disposal and water flow management (Béné et al. 2015). Disease outbreaks constraint aquaculture production by affecting both economic development of the country and socio-economic status of the local people in many countries (Sahu et al. 2008). These outbreaks are caused by endemic and emerging pathogens and impact directly on farmers income and their nutritional security. Thus, their avoidance is a fundamental requirement for future growth and resilience of aquaculture in LIFDCs (Food and Agriculture Organization of ...).

Viral infections are one of the most limiting factors of intensive aquaculture, leading to major losses in freshwater aquaculture and marine farming (Machimbirike et al. 2019). Currently, there is no cure for viral diseases in aquaculture and while vaccines and selective breeding have proved successful in reducing the severity of some viral diseases, there are currently severe knowledge gaps relating to Tilapia lake virus (TiLV), one of the most significant emerging pathogens in tilapia aquaculture with no effective, affordable vaccines yet available (Jansen et al. 2018). Furthermore, outbreaks of viral diseases are typically temperature dependent and water temperature affects the emergence of a wide range of parasitic, bacterial, and viral diseases of fish. The association of disease outbreaks with seasonality (when the water has relatively high temperatures) highlights the importance of understanding interactions between pathogens and environmental parameters that can exacerbate disease outbreaks. Studies on zebra fish have illustrated that mortality from viral nervous necrosis (VNN) is exacerbated under elevated temperature and crowding (Eyngor et al. 2014), (Binesh 2014), (Tsofack et al. 2017).

Environmental variability determines the distribution, migration, and abundance of fish (Jansen et al. 2018), while climate variability affects fisheries directly, as fish populations and fisheries activities are closely linked to weather and climate dynamics

(Lehodey et al. 2006). This is due to the fact that most aquatic animal species are poikilothermic (cold-blooded) and changes in water temperatures can rapidly and significantly influence distribution, prey availability, metabolism, growth and reproduction, with strong resulting impacts on fishing and aquaculture productivity (Cheung et al. 2009). At the same time however, the interconnectedness of aquatic systems allows many species to change spatial distribution more easily as ecosystems shift, to remain in their zones of preference. Clearly, therefore, the impact of global climate change on ocean capture fisheries will be important (Béné et al. 2015).

1.1.3. The importance of tilapia to current and future aquaculture

Tilapia are important for the sustainability of ecological systems and serve as the second most important group of farmed fish worldwide. The genus *Oreochromis*, is hugely important for providing food and employment, as well as domestic and export earnings to large populations worldwide (Machimbirike et al. 2019) (Eyngor et al. 2014). Their production has quadrupled over the past decade due to their relative ease of farming, marketability and stable market prices (Wang and Lu 2016). They are fast growing, efficient food convertors, and relatively disease resistant. These assets render them most suitable for farming; indeed, tilapines are one of the most significant groups of farmed fish worldwide, and their species have been introduced for farming purposes to more than ninety countries in the tropics and subtropics worldwide (Tsofack et al. 2017), (Machimbirike et al. 2019). World tilapia production in 2015, from both aquaculture and wild capture, amounted to 6.4 million tonnes, with an estimated value of United States Dollar (USD) 9.8 billion, and worldwide trade was valued at USD 1.8 billion (FAO, 2017). The common commercial Tilapia species in aquaculture are the Nile tilapia (*Oreochromis niloticus*), the Mozambique tilapia (*Oreochromis masssambicus*), the blue tilapia (*Oreochromis aureus*), *Oreochromis urolepsis hornorum,* also known as the Wami river tilapia, and various hybrid tilapia species (Machimbirike et al. 2019).

In the late 2000s, large losses of farmed tilapia were recorded throughout Israel. A novel RNA virus was subsequently identified and termed tilapia lake virus (TiLV).(Eyngor et al. 2014). Tilapia mortality, has also been described in Ecuador (Tsofack et al. 2017), and most recently, Egypt (Fathi et al. 2017), Thailand (Dong et

al. 2017) India, Malaysia (Amal et al. 2018) and the Philippines (OIE, 2017). Mortality levels between 20% and 90% have been reported (Jansen et al. 2018). In Egypt, unexplained mortality of tilapia during the summer months in recent years indicated a potential economic impact of around USD 100 million in 2015, with 37% of fish farms being affected (Fathi et al. 2017).

However, a lack of thorough investigation of all mortality incidents means that the geographic distribution of TiLV may be wider than currently described (Fathi et al. 2017). International trade of tilapia has taken place for more than 50 years with a resultant global distribution only exceeded by common carp. As a result, TiLV may have been circulating worldwide through movement of live fish for aquaculture in the absence of knowledge of the existence of an associated risk. It is speculated that over 40 countries may have a theoretical risk of inadvertent TiLV introduction through trade in aquaculture and this illustrates the importance of the need for surveillance activities in these countries (Dong et al. 2017), (Jansen et al. 2018). Further research including epidemiological and experimental infection studies will be needed to limit the negative impact of these infections, as well as improving the management of emerging diseases in the future. Furthermore, rapid action should be taken to control the spread of the TiLV, including immediate improvements to aquaculture biosecurity practices and in the longer term potentially, vaccines and breeding TiLV resilient strains of tilapia (Fathi et al. 2017). TiLV is only one of several viral threats to Tilapia production. Others include infectious pancreatic necrosis virus (IPNV, Aquabirnavirus), nervous necrosis virus (VNN, Betanodavirus), tilapia larvae encephalitis virus (TLEV, Herpesvirus), and the Iridovirus infections namely Bohle iridovirus (BIV, Ranavirus), infectious spleen and kidney necrosis virus (ISKNV, Megalocytivirus), Lymphocystivirus and another Iridovirus-like infection. These other viruses have been associated with relatively high mortalities (20–100%) in several cases of natural disease outbreaks or as studied through laboratory challenges (Machimbirike et al. 2019).

1.2. The Fish Microbiome (Focusing On the skin and mucus)

Microbiomes are the assemblage of microorganisms present in an environment. They refer to microbes, their genomes, and their environment, which are essential for growth and survival of the host. (Xia and Sun 2017), (Marchesi and Ravel 2015). The microbiome is typically characterised using molecular methods relying on marker genes and genomic regions, amplified and sequenced from given biological samples. Taxonomic assignments are performed using a variety of tools to assign each sequence to a microbial taxon at different taxonomic levels from phylum to species (Marchesi and Ravel 2015). Microbial communities are very dynamic and respond to local selective pressures via phenotypic plasticity, rapid mutation rates, short generation times and high intra -community gene flow. Microbial plasticity has been proposed to also enhance phenotypic plasticity in the host, such as improving thermoregulation capacity; enabling digestion of novel food sources or conferring increased resistance to local pathogens (Webster et al. 2018).

Different microbial communities are associated with different regions on the fish. Communities present on the skin, gills, and in the gastrointestinal tract of fish are key components of the host mucosal barrier defences (Merrifield and Rodiles 2015).
Some argue that the host and its microbiome (holobiont), along with the totality of genomes in the holobiont (hologenome) are a unit of selection, and therefore this unit has properties similar to an individual organism, i.e., it is a superorganism (Douglas and Werren 2016). Yet, evolutionary biologists debate whether a holobiont should be considered as a unit of selection. For opponents to this theory, holobionts are unstable entities and viewed as ecosystems in which antagonism, mutualism and chance determine the interactions between members of the holobiont (Simon et al. 2019).

The skin of fish is continuously exposed to a water environment, and unlike that of terrestrial vertebrates, it is a mucosal surface with a thin epidermis of live cells covered by a mucus layer. The mucosa plays an important role in maintaining the homeostasis of the fish and preventing the entry of invading pathogens (Brinchmann 2016). Fish skin mucosal surfaces (SMS) could constitute an adequate niche for the evolution of mucosal aquatic pathogens in natural environments (Carda-Diéguez et al. 2017). Mucosal epithelia act as a main organ of defence with the highest densities of

microbes and the most intimate interactions with the host (Carlson et al. 2017).These mucus layers are an optimal environment for microbial growth, providing structure as well as nutrients in the form of diverse, mucin-associated glycans (Barr et al. 2013). Through competition for adhesion sites and nutrients, they may limit or reduce the abundance of pathogens and facilitate the development and homeostasis of host immunity (Legrand et al. 2017). They may also antagonize pathogens via the production of various compounds such as organic acids, siderophores, bacteriocins, H_2O_2, and various antimicrobial peptides (Merrifield and Rodiles 2015). In contrast to the well-studied human skin microbiome, only a limited number of studies have focused on the complexity of the bacterial community associated with the fish skin-mucus (Minniti et al. 2017).

A study undertaken by Legrand *et al* (2017), revealed the importance of understanding fish skin and gills microbial assemblages, and the factors which drive their composition, as possible biomarkers to reveal changes in host health status. This study showed that disease onset was marked by an overall loss in diversity with the emergence of specific community members, likely reflecting changes in the immune states and the connectivity of these barrier systems. It also suggested that common mucosal responses exist with a direct link between the gut mucus and the skin (Legrand et al. 2017). Indeed in humans, the extraordinary success of faecal microbial transplantation to effectively cure patients from *Clostridium difficile* infection in up to 90% of cases by resetting the gut microbial community, confirms the importance of the structure of microbial communities for the health of the holobiont (Tacon and Metian 2013).

1.2.1. Factors Influencing the Microbiome Structure

The study of the beneficial effects of bacteria and their influence on host health is a growing field and analysing the microbial community composition alone may not give us full insight into their dynamic interactions among microbiome, host, environment, and disease intervention. These communities are diverse and dynamic, and knowledge of the factors that control the success of an individual population is important to understanding their influence on ecosystems and biodiversity**.** Research that has explored host microbiota variability in space and time suggests the presence

of a host genetic component in the development of mutualist communities (Needham et al. 2017), (Eichmiller et al. 2016), (Boutin et al. 2012). Distinct body sites are jointly influenced by host-specific organizing factors operating at the level of the host individual (Pratte et al. 2018) and greater difference was noticed in microbiome structure between more genetically distinct fish populations, adding that within a population genetic diversity is positively correlated with microbiome diversity (Webster et al. 2018). A decisive study with zebrafish, reported that a core microbial community is supported through host system selective pressures regardless of environmental parameters (Roeselers et al. 2011).

Diet shapes the relative abundance of specific taxa that colonise the host (Goodrich et al. 2014), while more drastic transitions, such as habitat changes from wild to captive can lead to dramatic changes in the gut microbiome of fish (Eichmiller et al. 2016). Many other factors may heavily alter the microbial composition of the environment and the fish host such as intensive fish farming, (Gomez et al. 2013), temperature, physiology, and taxonomy (Eichmiller et al. 2016). In addition, seasonality may significantly influence the skin microbiome structure (Larsen et al. 2015).

While the microbiome has been shown to aid in host nutrient acquisition and immune response, they may also influence host behaviour, development, reproduction and overall health (Colston and Jackson 2016). The immune system is considered to be one of the most important forces by which the host shapes the configuration of the microbiome. The past decade of research on the interactions between the microbiome and host immunity has clarified a large number of mechanisms by which the microbiota affects both innate and adaptive immune responses. Mechanistically, the communication between the microbiome and the innate immune system seems to rely on metabolites, such as short-chain fatty acids (SCFAs). In the case of adaptive immune cells, specific bacterial species have been shown to directly influence immune system development and differentiation (Levy et al. 2017). A study held by Larsen et al on gulf killifish, revealed the high similarity between the microbiomes of individual fish's skin and mucus, and it is different from that of the surrounding water, suggesting the presence of a resilient core microbiome (Larsen et al. 2015).

During the course of an infection, the relationships between the microbiota and the host immune system are vulnerable to changes, as some of the host genes are involved in the recruitment of specific bacterial strains (Boutin et al. 2012).

Additionally, the microbiota-host immune system interaction can significantly impact the outcome of an infection (Gomez et al. 2013). Microbiome homeostasis seems to be an important influence on resistance against some diseases previously considered exclusively influenced by genetic factors. As an example, *Methylobacterium rhodesianum* produces poly-β-hydroxybutyrate (PHB), that contain antimicrobial properties known to inhibit the growth of pathogens like enterobacteria and *Vibrio* species. This suggests that the reduction of abundance allowed for colonization or over-growth by other bacteria, some of which are pathogenic (Boutin et al. 2014). To determine precisely which genes are involved in the recruitment of specific bacterial strains, some studies have looked at gene expression in the presence of symbiotic bacteria. Thus, three-way interactions between pathogens, microbiota, and the animal host are complex and require a deep understanding of all three components (Reid et al. 2017).

Creating a list of infectious agents, their genetics, their symptoms and treatment can no longer considered as isolated processes. Contemporary models of disease require integrated studies between disciplines; to understand the role of the microbiome in the prevention of disease requires analysis of how they interact with each other and with the host that harbours them (Chow and Fuhrman 2012). Thus, diverse fields include microbiology, genomics, immunology, molecular biology, epidemiology, and clinical medicine must be combined to develop a more complete understanding of the microbes (Kong and Segre 2017), and their interaction by exploring their molecular components through genetic analysis and sequencing (Chow and Fuhrman 2012).

1.2.2. Dysbiosis

The stable relationship between host and its microbiome can be disrupted by several environmental factors as well as host-intrinsic factors leading to dysbiosis. Dysbiosis is defined as a change in relative composition of the different microbes compared to normal, altering the symbiotic relationship between the host and associated microbes,

which may result in disease (Cho and Blaser 2012). As an example, while the skin microbiome of unstressed fish is dominated by taxa known for their probiotic and antimicrobial activity, the microbiome of stressed fish is dominated by potential pathogens, which are an important factor that can tip the balance of host-microbiota relationships (Reid et al. 2017), (Williams and Gallo 2015), (Levy et al. 2017), (Rosado et al. 2019).

In recent years, there has been an increase in the incidence of many multifactorial diseases associated with dysbiosis, where the taxonomic composition and function of the microbial community were shifted to a less favourable health state (Levy et al. 2017). A study by Reid *et al*, showed that a disruption of the microbiome by viral infection rendered fish more susceptible to secondary infections by opportunistic bacterial pathogens present in the environment or within the host indigenous microbial reservoir (Reid et al. 2017), to the detriment of host health. These communities usually contain low levels of primary pathogens, which are mainly suppressed under normal conditions, but can infect immunocompromised hosts (Merrifield and Rodiles 2015).

The use of antibiotics in aquaculture for disease prevention and treatment has been widely explored, as these drugs may disrupt the microbial communities and increase disease susceptibility (Butt and Volkoff 2019). They have shown to have drastic effects on the microbiome structure, by lowering community diversity, change in community composition, loss of some rare members, alteration of enzymatic functions, selection of resistant members, and, in some but not all cases, lowering of total numbers (Carlson et al. 2017). These aquaculture practices can increase the risk of dysbiosis, which is known to facilitate pathogen infections, and the proliferation of opportunistic pathogens. These traditional practices are typical in fish farms and include, along with antibiotic misuse and abuse, overcrowding and low oxygen (Rosado et al. 2019). It has been noticed that crowding and net handling are commonly believed to affect the swimming activity levels of salmon, which may be the expression of lowered activity due to lack of movement space to increased activity due to agitation and stress (Føre et al. 2018), inducing dysbiosis in the skin microbiome, facilitating the proliferation of opportunistic pathogens and may affect the associated microbial communities (Rosado et al. 2019).

Dysbiosis cannot be defined as an altered state of the microbial community. As enormous interindividual variability occurs in the taxonomic microbiota composition between healthy individuals across geography, age and dietary habits, this raises the question of what can be considered a reference population and allows for almost any given microbial configuration to be considered 'dysbiotic' when compared with a particular control. Therefore, it is crucial that studies in both humans and animal models are very carefully controlled to avoid conclusions of a false dysbiotic state, caused by interindividual variability, vertical transmission, housing effects, variations in pathogen screening and other factors accounting for incidental deviations of microbiome composition from a given reference population (Levy et al. 2017). Beyond the mechanisms of pathogen adhesion, and the subsequent mucosal inflammatory responses and possible mucosal translocation, very little is known of the process of cross-talk between the fish host and its microbiome. Much of our understanding of this relationship is inferred from mammals, and to a lesser extent derived from a small number of germ-free and gnotobiotic studies in fish. Many questions need to be addressed in the search for evidence of dysbiosis, as it is still unknown whether the presence or absence of certain microbes or both may cause disease. As we move away from the one-microbe-one-disease model and into the multifaceted world of dysbiosis, the lines of cause and effect become much harder to untangle (Merrifield and Rodiles 2015) (Yong 2016).

1.2.3. Probiotics and other treatments

Due to the complex structure of microbial communities, disentangling interactions and identifying important taxa for specific functions is enormously challenging, especially when environmental influences on population dynamics and activities are taken into account. Microbiome research of humans and other eukaryotes typically describe correlations between microbiome composition and diseases, but it remains a challenge to establish which taxonomic or functional changes in the microbiome are actually causal for disease development or disease protection (de Bruijn et al. 2018).

Disease control in aquaculture industry historically has been achieved by the use of synthetic chemicals and antibiotics. The use of expensive chemotherapeutants for controlling diseases however has been widely criticized. Residues of these

compounds, accumulate in the water environment and fish lipids, and can be harmful to human health as well as to other aquatic organisms and farmed species. The side effects include reproductive and developmental toxicity, neurotoxicity, and bioaccumulation, depending on the specific chemical/agents used (Carro et al. 2012).

Due to the growing awareness of the problems of antibiotics, strict regulations have been established in some sectors of the aquaculture industry and alternative methods are being developed and tested to maintain a healthy microbial environment in the cultured organisms. The use of probiotics is one of such methods that is gaining importance in controlling potential pathogens (Sahu et al. 2008). 'Probiotics are defined as 'live microorganisms which, when administered in adequate amounts, provide a health benefit to the host' (WHO and FAO, 2006). Their use in aquaculture, as an alternative to antibiotics, rose significantly as legislations were introduced to restrict the widespread use of chemicals in animal rearing. They have been found to reduce the cost of fish farming through improvements in fish welfare and nutrition (Egerton et al. 2018). However, some commercially available probiotic products are often ineffective in reducing pathogenicity in aquatic species. Therefore, it may be dangerous to call them "probiotic" for commercial purposes, as they bring no benefits, and may create harmful or unexpected effects on the hosts, and may actually enhance economic losses (Van Hai 2015).

Advances in the field of probiotic development indicate that endogenous bacteria are able to outcompete pathogens (Boutin et al. 2014). Such implications have been validated in studies showing that these treatments resulted in a decrease of mortality (54-86%) across fish families (Boutin et al. 2012). Despite the large number of promising probiotic candidates, only a few have been commercialized or licensed for aquaculture in Europe (Ricci et al. 2017), (de Bruijn et al. 2018). The lack of consistency in performance and persistence under different environmental conditions, their relatively narrow activity spectrum as well as current constraints for commercial application of probiotics, such as strict regulations for registration, shelf life, high production costs, and highly specific storage and transport conditions of the microbial inoculants, are major concerns to the industry (de Bruijn et al. 2018), as is the lack of understanding of underlying mechanisms as to how probiotics confer a health benefit in fish.

Other treatments that have been identified as new disease control agents, include the bacterial storage compound PHB, that has been based on the knowledge that this biopolymer can be degraded into short-chain fatty acids (SCFAs) that contain antimicrobial properties (Laranja and Bossier 2019). The beneficial effects of PHB in farmed aquatic animals has been shown in a study held by Sugana *et al* 2014, where it was shown to be an overall stimulant of the immune system in experimental tilapia species, and to protect against disease challenges (Suguna et al. 2014). Examples of these PHB producing bacteria are species belonging to the phylum Proteobacteria (García-Pérez et al. 2018).

1.2.4. Pitfalls in microbiome research

Relatively few epidemiological studies have evaluated the role of microbiota in health with the degree of standardization necessary for translational applications. Inter-study technical variability in microbiome measurements can sometimes outweigh biological effects (Sinha et al. 2017) and advances in microbiome research depend on carefully executed, controlled, and reproducible studies (Goodrich et al. 2014).

Factors that may influence a microbiome study could be numerous and achieving sufficient reproducibility and accurate determination of the abundance of microbes on the skin of the fish is challenging. First, it is difficult to avoid contamination of the epidermal tissues during the capture and handling of the fish when conducting sampling. Second, a major issue is the standardization of the precise sampling site on the skin. (Sinha et al. 2017; Merrifield and Rodiles 2015). Variability depends greatly on biospecimen type and origin, followed by DNA extraction, sample handling environment, and bioinformatics (Sinha et al. 2017).

The presence of contaminating DNA is a particular challenge for researchers dealing with samples containing a low microbial biomass, as the low amount of starting material may be effectively swamped by the contaminating DNA and generate misleading results (Salter et al. 2014). A number of methods have been devised to treat reagents in order to reduce potential contamination, including: gamma or

Ultraviolet radiation and DNase treatment, with varying levels of success (Sheik et al. 2018). Microbiologists, statisticians, and bioinformaticians need to collaborate to develop standardised practices and recognise potential measurement variability in each step of microbial profiling (Sinha et al. 2017). Issues of contamination have plagued studies, with high-profile examples in the fields of novel virus discovery. For instance, a research study trying to identify a population of bacterial microbiota in the amniotic fluid from healthy term pregnancies, found the concentration and content of these communities were similar to the sequences amplified from negative controls. (Lim et al. 2018).

Previous studies have focused on the “kit-ome” or contamination associated with DNA extraction kits (Sheik et al. 2018). Kit-ome contamination is a concern for both 16S rRNA gene sequencing projects, which require targeted PCR amplification and enrichment. Awareness of this issue is important to ensure that studies are adequately controlled and erroneous conclusions are not drawn from culture independent investigations (Salter et al. 2014).

An important component in limiting the impact of contamination on investigated samples, is running negative controls, consisting of `blank' DNA extractions. This is considered the first line of contamination identification. Despite adding no sample template at the DNA extraction step, these negative control samples often yield a range of contaminating bacterial species. Therefore, incorporating controls with extraction kits, PCRs, and sequencing are all necessary to trace the source when contamination may have been introduced (Sheik et al. 2018).

As the field trends towards larger data sets, understanding subtle confounding factors long known to epidemiologists and taking more care with longitudinal study designs will become increasingly essential. The value of interventional studies over observational studies is considerable, especially when human, animal models and in vitro data can be correlated across scales and systems. Increased standardization of techniques and dissemination of methods with low noise and bias will increase the ability of the microbiome field to deliver on the promise of translatability from lab-scale studies to the clinic, field or natural environment (Knight et al. 2018).

1.3. Missing a key ingredient: Interactions between Host, Bacteria and Bacteriophages

Shared bacterial species have been proposed to form a "core bacterial community" that provides beneficial functions. This concept, however, is based almost entirely on bacterial sequence data and has not considered the role of bacteriophages (Manrique et al. 2016). With a burst of studies of microbiomes and their influence on their host, changes in the viral community structure may strongly influence shifts in the bacterial and eukaryotic communities through the interactions of viruses with their potential hosts (Robinson and Pfeiffer 2014). Therefore, the influence of phages on the microbiome is not restricted to their hosts, but also extends to their host's host (Cordingley 2017).

Bacteriophages, also known as phages, are viruses that are specific to bacteria and they are the most common biological entities on earth. They are regulators of bacterial abundance, diversity and metabolism, holding a central role in the maintenance and function of many ecosystems (Chow and Fuhrman 2012). Twort and d'Hérelle were the first to describe bacteriophages, but it was d'Herelle who applied the term to a bacteriolytic substance that he isolated from faeces. This finding has led to several studies and the creation of the "phage group" (Steier et al. 2019).

The majority of viruses in aquatic environments are thought to be bacteriophages. Several studies have shown their ecological importance as a key factor influencing microbial diversity, shaping community structure and enhancing nutrient cycling in the ocean (Liu et al. 2017). Among the most widespread and abundant phage groups, are T4-like phages, a member of the *Myoviridae* family. They have been isolated from a variety of hosts, including typical heterotrophic bacteria such as *Escherichia coli*, *Vibrio*, *Pseudomonas* and *Pelagibacter ubique*, and autotrophic cyanobacteria such as *Prochlorococcus* and *Synechococcus*. Most of the isolated cyanophages have been characterized to be T4-like viruses (Liu et al. 2017; Ignacio-Espinoza and Sullivan 2012).

It is essential to understand the changes in these viral communities, as they may strongly influence shifts in the bacterial and eukaryotic communities through their interaction with their potential hosts (Fuhrman 1999). Yet, a unified molecular taxonomy is impossible because viruses contain no universally shared sequences. Consequently, studies on phage ecology and diversity require group-specific molecular markers (Filée et al. 2005). Therefore, the viruses present in the microbiome require a different approach because, creating challenges especially for community-wide studies (Sullivan 2015). The current approach for studying these communities is to isolate virus-like particles (VLPs) using size fractionation and to sequence those using metagenomics, or alternatively microarrays (Goodrich et al. 2014), (Callahan et al. 2016).

Previous surveys in both marine and non-marine environmental studies support the significance of T4-like myoviruses in the environment and the use of g23 major capsid protein (MCP) as a marker for them (Filée et al. 2005). Although genomic analysis has discovered other conserved genes within this phage family, g23 remains one of the key core single-copy genes (Sullivan et al. 2010). The MCP is the foundation of both the T4-like phage's capsid structure and also of the operant phylogeny of the T4 superfamily. In a study conducted by Comeau et al, they indicate that within the T4 superfamily there are some clear phylogenetic groups with regard to the more conserved and more variable domains of the MCP. These groups could be a result of differences in capsid morphology, the arrangement of the MCP lattice, and the presence of different capsid accessory proteins between the subgroups of the T4 superfamily (Comeau and Krisch 2008). Currently, deep sequencing using high-throughput shortread-based technologies (Roche 454, Illumina MiSeq/NextSeq/ HiSeq/NovaSeq platforms, Ion Torrent platforms) remains the primary approach to characterizing unculturable viral communities in the gut (Shkoporov and Hill 2019).

It is yet far from clear what specific role phages play maintaining the healthy microbiome or initiating dysbiosis, but it is self-evident by analogy with their profound influence on marine ecosystems, that they have the potential to dramatically affect the abundance and composition of the gut microbiome (Cordingley 2017).

1.3.1. ***Microbial ecosystems***

Over the past decade, new and exciting findings in the field of environmental virology have spearheaded a recent revival into research of bacterial viruses. Despite the recent increased interest in environmental viruses, our knowledge remains sparse. A series of recent studies has shown that viruses have the ability to manipulate the life histories and evolution of their hosts in remarkable ways, challenging our understanding of this almost invisible world (Rohwer and Thurber 2009). Viruses, have proven to be powerful modulators of locally and globally important microbes through mortality, horizontal gene transfer, and metabolic reprogramming (Sullivan 2015).

Many studies have shown that bacteria reshape their symbiotic hosts, by activating a wider range of genes to create a healthier environment, and the easiest way of checking if an animal needs microbes to develop properly is to deprive it of them. The result of this is that some of them cannot survive, an example is the dengue-carrying mosquito *Aedes aegypti*, which makes it to larval stages but fails to progress beyond that (Coon et al. 2014). The existence of defensive microbes isn't surprising, as protecting their hosts is an obvious way of guaranteeing their own success, moreover antibiotics are created from bacteria. Yet, in many cases, part of the DNA of these bacteria belong to phages, by opting for a more passive lifestyle, by integrating their DNA into a bacterium's genome and stay there for many generations. Without these phages, bacterial symbionts become useless (Yong 2016).

1.3.2. ***Phage and bacteria dynamics***

Animals rely on their associated microbial communities to aid with digestion, immunity, and other aspects of physiology and disease. To understand the structure and dynamics of the microbiota, researchers use molecular techniques, including metagenomics, metatranscriptomics and metabolomics, but these methods only provide a narrow, correlative view of bacteria and phage dynamics (Myhrvold et al. 2015). However, in combination with model system studies, they have enabled the development of numerous ecological theories that characterize complex bacteria and

phage dynamics in the gut (Mirzaei and Maurice 2017). Indeed, mucosal surfaces of fish appear to be particularly rich in bacteriophage, with concentrations typically four-fold higher than those found in surrounding environments (Barr et al. 2013).

Widespread bacteriophage predation and lysogenic conversion in bacterial populations plays a major role in regulating bacterial biomass, maintaining biodiversity, horizontal gene transfer and driving biogeochemical cycles in the earth biosphere (Shkoporov and Hill 2019). Lysogenic conversion is an interaction in which expression of phage-encoded proteins contributes to bacterial fitness with no apparent value to the phage (Feiner et al. 2015). Additionally, they drive microbial evolution by natural selection for microbes resistant to infection and via lateral gene transfer. Many viruses are strain- specific predators. Therefore, as a particular microbial strain becomes dominant in a system, density-dependent predation posits that the viral predators will also increase in abundance, reducing the abundance of its prey. This 'kill- the- winner' hypothesis may explain much of the observed microbial diversity and changes in community structure, as niches are opened for other microbial strains to grow and dominant microbial species within a system are constantly turned over (Rohwer et al. 2009).

We are still only scratching the surface of discovery of global viral diversity, have little understanding of the functionality of the majority of genes in the global viral gene pool and the roles they play in the interaction with their hosts, and are still grappling with understanding viral impact on ecological and evolutionary processes (Rohwer and Thurber 2009) (Rohwer et al. 2009). Future research may provide and assist in answering questions regarding taxonomic composition, community structure, and population dynamics (Shkoporov and Hill 2019).

1.4. Improving sampling techniques for fish microbiomes to reduce animal stress

There is increased focus on the need for reducing invasive sampling of animals in research. For studies in fish, swabs can be considered as an alternative and less stressful method to the traditional fin clipping method for molecular and biochemical studies in to organism health and responses of fish to their environment. Responsible research has an obligation to address the principles of the "3Rs"—replacement, reduction, and refinement of animals in research (de Wolf et al. 2007). Although it is not yet confirmed that skin swabbing may yield the same DNA concentrations as the fin clipping method, it has been shown in an earlier study held by Carl Breaker *et al* 2017, that fin clipping is an invasive method, and could cause side effects that may affect the outcome of experimental studies, including the potential for secondary infections and an elevated nonspecific immune response (Breacker et al. 2017). Fin clipping could also affect behavioural interactions, reproductive success, or locomotion, as some fish use fins in communication or in reproductive behaviours (Breacker et al. 2017). Furthermore, nonterminal anaesthesia treatment required for fin clip collection could affect the responses of the fish (and the associated microbiomes) behavioural and physiological data recorded from fin-clipped fish (Le Vin et al. 2011), by increasing levels of cortisol, causing unintended consequences for studies that are influenced by stress responses, such as the immune function (Breacker et al. 2017).

Nevertheless, the presence of external stressors can cause changes in the bacterial community composition in the mucus of aquatic organisms, resulting in lower species diversity with an associated increased occurrence of opportunistic pathogens (Larsen et al. 2015). A previous study by Legrand *et al* 2017, has revealed the bacterial assemblages associated with the outer surface mucosal barriers of healthy Yellowtail Kingfish using swabs. It was possible through this type of sampling to identify the core bacterial community associated with both skin and gills, which was different to that of the surrounding environment (Legrand et al. 2017).

1.5. Next generation Sequencing

The realisation of the importance of microbial communities to human and environmental health has motivated researchers to develop methods for the efficient characterization of these communities (Callahan et al. 2016). The development of next-generation sequencing has led to a surge in an effort to characterize the microbiomes of various vertebrate hosts, a necessary first step to determine the functional role these communities play in host evolution or ecology (Cho and Blaser 2012), and the continual developments in sequencing and analyses provide increasingly sophisticated tools to interrogate microbiomes, ultimately to improve our understanding of health and disease (Kong and Segre 2017). High throughput sequencing technologies have enabled major advances in microbiome research by generating enormous new data sets that can be mined for information on the composition and functional properties of vast numbers of microbial communities (Kuczynski et al. 2011)(Vale and Little 2010), (Cho and Blaser 2012).

The most cost-effective of these methods is the amplification and sequencing of targeted genetic elements, and although microbiome data are obtained in several ways, amplicon sequencing of taxonomic marker genes such as the 16S rRNA gene provides an economically viable census of a community scalable to large projects. (Callahan et al. 2016). It is the most common strategy to assess bacterial communities, developed initially by Woese and colleagues to establish a comprehensive outline of prokaryotic phylogeny, by using RNase T1 oligonucleotide catalogues of 16S ribosomal RNA gene (Fox et al. 1980). This bacterial marker gene is found in all bacteria, and has nine hypervariable regions (V1-V9) that can be used to infer taxonomy (Cusco et al. 2017). It has been proven the most useful for establishing distant relationships because of its high information content, conserved nature, and universal distribution (Lane et al. 1985) . Another advantage that the gene encoding 16S rRNA is supported by the availability of several large databases of reference sequences and taxonomies (Callahan et al. 2016) (Kuczynski et al. 2011).

The ability to classify sequences to the genus or species level from amplicon data is a function of read length, sample type, the reference database, and the quality of the

sequence (Cusco et al. 2017). Ribosomal genes contain both slowly evolving regions that can be used to design broad-spectrum PCR primers and fast-evolving regions that can be used to classify organisms at finer taxonomic levels, although species-level resolution might be unfeasible using this information alone (Kuczynski et al. 2011). This pitfall can be overcome sequencing the full length 16S rRNA gene by third generation sequencers (Cusco et al. 2017), and one way to assess the usefulness of the various hypervariable regions of the 16S rRNA gene is to compare the taxonomic assignments of short fragments (for example, 100–250 base pairs) in these regions against those of the full 16S rRNA gene (Kuczynski et al. 2011). With the arrival of 3rd generation single-molecule technology sequencers, these short-length associated problems can be overcome by sequencing the full or almost full-length of 16S rRNA gene with different sets of universal primers. Their higher information content has the potential to overcome some of the short-read limitations by allowing for longer alignments that potentially can contribute to higher taxonomic specificity, functional characterization, and resolution (Brown et al. 2017).

An example of this is the MinION (Oxford Nanopore Technologies), which is a portable, single-molecule genome sequencing instrument which directly senses native, individual DNA fragments by translating disruptions in the current across a membrane as single-stranded DNA passes through a nanopore (Cusco et al. 2017) (Warwick-Dugdale et al. 2018). A major advantage of the MinION is low cost, small size, and ability to analyse the nucleic acid data in real time in the field by connecting to a laptop. (Kerkhof et al. 2017).

Long read technology has also recently been used to improve recovery of viral genomes from metagenomic samples, improving the scaffolding of large novel viral genomes, obtaining information on methylation patterns, and for studying population structure at a single-virion level, since long reads can, in some cases, represent complete or near complete viral genomes (Warwick-Dugdale et al. 2018) (Shkoporov and Hill 2019). With further improvement in sequence throughput and error rate reduction, the MinION platform shows great promise for precise real-time analysis of the composition and structure of more complex microbial communities (Brown et al. 2017).

The aims of this study were to:

- Characterize the skin microbiomes of tilapia fish in Malawi, to improve our understanding of fish skin microbiomes in health, through short read sequencing targeting the v4 16S rRNA hypervariable region.
- Assess the accuracy and efficiency of MinION sequencing for enabling real time analysis of the microbial community structure, targeting the full length 16S rRNA gene.
- To evaluate inclusion of bacteriophages in aquaculture microbiome analyses -- an often ignored, yet crucial part of the microbiome — using amplicon sequencing. This was achieved by investigating the presence of T4-like myovirus communities.
- Evaluate whether the use of skin swabs is a successful and reliable method in capturing communities, to reduce animal stress and handling during sampling.

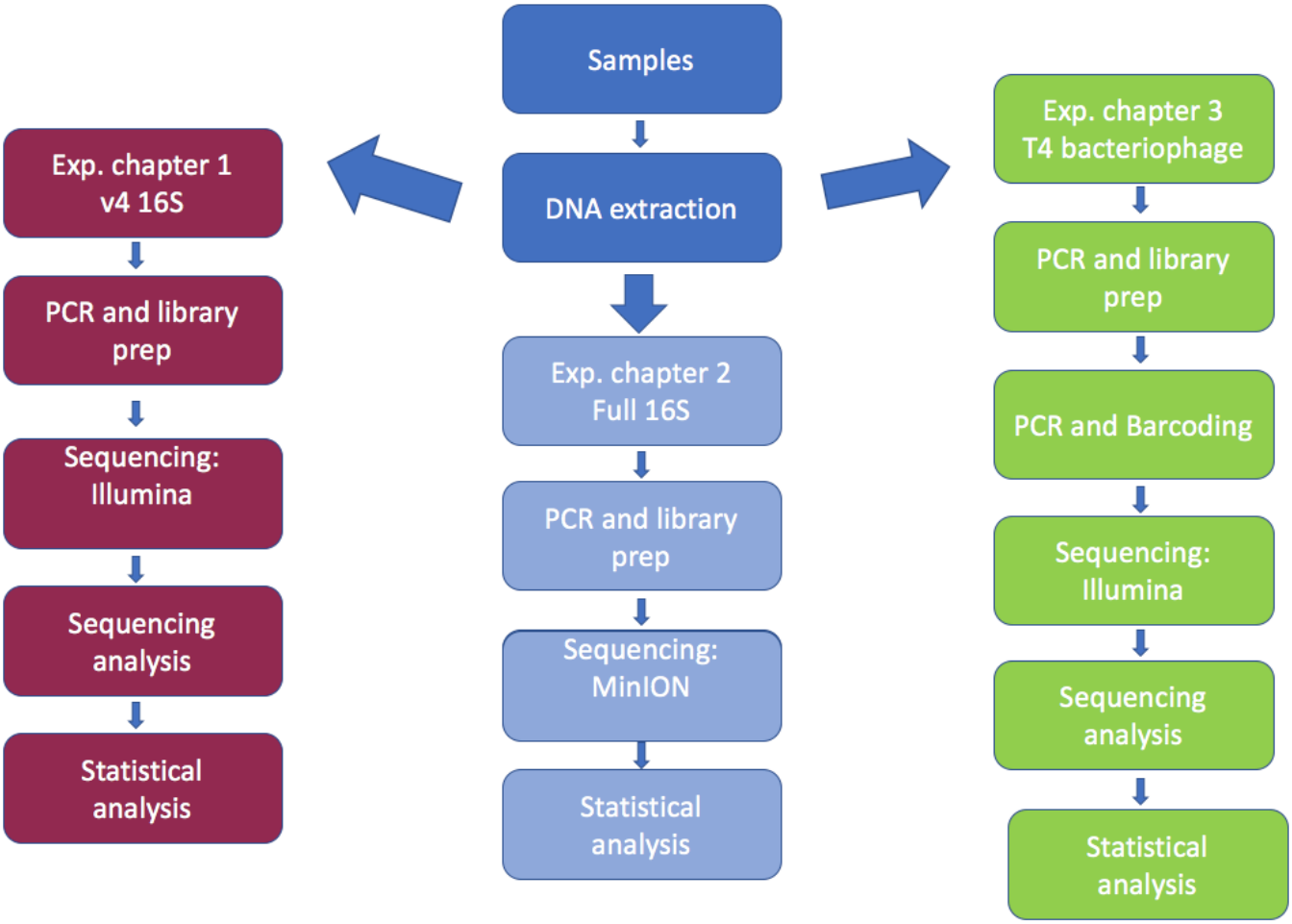


Figure 2. A schematic representation of the workflow of the Experimental Chapters described in this thesis.

2. Materials and Methods

2.1. Ethics

This study complied with the 3Rs (Reduction, Refinement and Replacement) https://nc3rs.org.uk/the-3rs. The following non-invasive procedures were conducted in accordance with the approved documentation and no animals were harmed throughout this study.

2.2. Sampling and preservation

Microbiome samples were collected in October 2017 from Malawi fish farms, from two geographically distinct sites, Maldeco and Blantyre. Five ponds (ponds 1-5 inclusive) were located in Blantyre and two (pond 6 and 7) were located at Maldeco, see Figure 3.

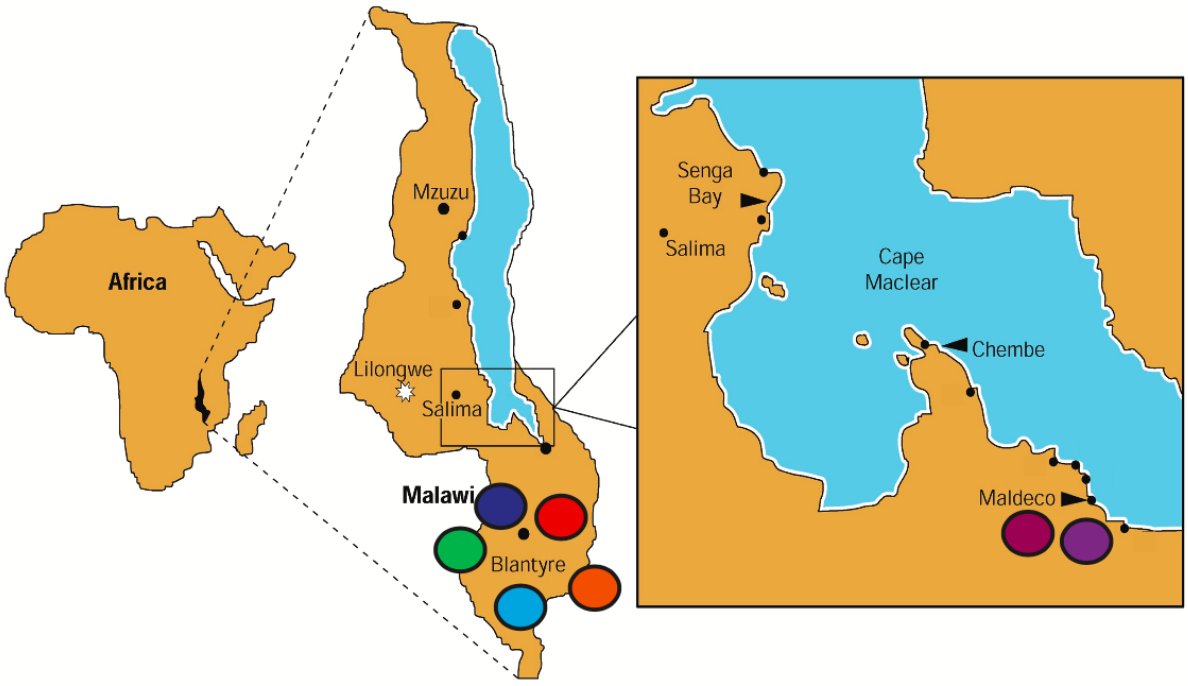


Figure 3. Locations of the seven aquaculture ponds in Malawi, where fish samples were collected for this study. Five ponds were located in Blantyre (1-5) as indicated by the green, dark blue, red, orange and light blue circles and two ponds were located in in Maldeco (6 & 7) as indicated by the pink and purple circles. Figure adapted from (Cetron et al. 1996).

32 tilapia fish were sampled belonging to two different species, *Tilapia sparmanii* and *Tilapia rendalli*. The length of the sampled fish varied between 103-250cm (Table 1)

Fish sample	Length	Fish sample	Length	Fish sample	Length	Fish sample	Length
T1c	135mm	T8	140mm	T19(s)	138mm	Ts 26 g	235mm
T1f	135mm	T9	120mm	T20(s)	130mm	Ts 27 g	230mm
T1g	135mm	T10	103mm	T21(s)	140mm	Ts 28	225mm
T2c	103mm	T11	130mm	T22(s)	140mm	Ts 29	220mm
T2f	103mm	T11 c	130mm	T21	154mm	Ts 30	192mm
T2g	103mm	T12	145mm	T22	190mm	Ts 31	210mm
T3c	112mm	T12 c	145mm	T23	162mm	Ts 32	214mm
T3f	112mm	T13(s)	151mm	T24	164mm	Ts 28 g	225mm
T3g	112mm	T14(s)	156mm	T25	141mm	Ts 29gx2	220mm
T4c	112mm	T15(s)	160mm	Ts 23	225mm	Ts 30 g	192mm
T4f	112mm	T16(s)	151mm	Ts 24	170mm	Ts 31 g	210mm
T4g	112mm	T17(s)	155mm	Ts 25	250mm	Ts 32 g	214mm
T5c	115mm	T17	115mm	Ts 26	235mm		

T5f	115mm	T18	123mm	Ts 27	230mm		
T5g	115mm	T19	124mm	Ts 23 g	225mm		
T6	149mm	T20	110mm	Ts 24 g	170mm		
T7	113mm	T18(s)	135mm	Ts 25 g	250mm		

Table 1. Sampled tilapia fish and their associated sample name and length.

Three different sample types were collected as shown in Table 2: For pond (rearing water) samples, 5 replicates of each of the seven ponds were obtained. 35-100 ml of pond water was pushed through polycarbonate filters (47 mm diameter, 0.4 μm pore size) until the filter clogged preventing further water extrusion. For incubation samples, 23 out of the 32 fish (Table 1) were incubated individually in buckets, containing five litres of an inlet water source, for a period of approximately 20 minutes and the fish then subsequently released (after skin swab sampling – see below). Water from buckets was then filtered as described above (see Figure 4). Samples from inlet water source only were also taken as background controls (these were accidentally omitted from ponds 5 and 7). Filters from pond and incubation samples were preserved in 100% molecular-grade ethanol, in 2 mL screw-cap tubes. Sterile swabs were used to capture the skin and mucus microbial communities, from the gills and caudal, and these samples were preserved in RNALater (initially this was done to enable the recovery of the ssRNA tilapia lake virus, but no evidence of disease was subsequently identified). Swab samples were stored at ambient temperature during transit back to the UK and stored at -20°C upon arrival. A schematic of sample processing for this project can be found in Figure 2.

Type of Sample	Number of Samples	Preservation
Skin Swabs (*T)*	52	RNAlater
Pond (rearing water) (*P)*	35	Ethanol
Incubation (*I)*	24	Ethanol
Incubation controls	5	Ethanol

Table 2. Different sampling methods and preservatives used to collect tilapia microbiome samples. Rearing water samples were derived from the ponds where the fish were farmed. Incubation control samples represent the water, before fish incubation.

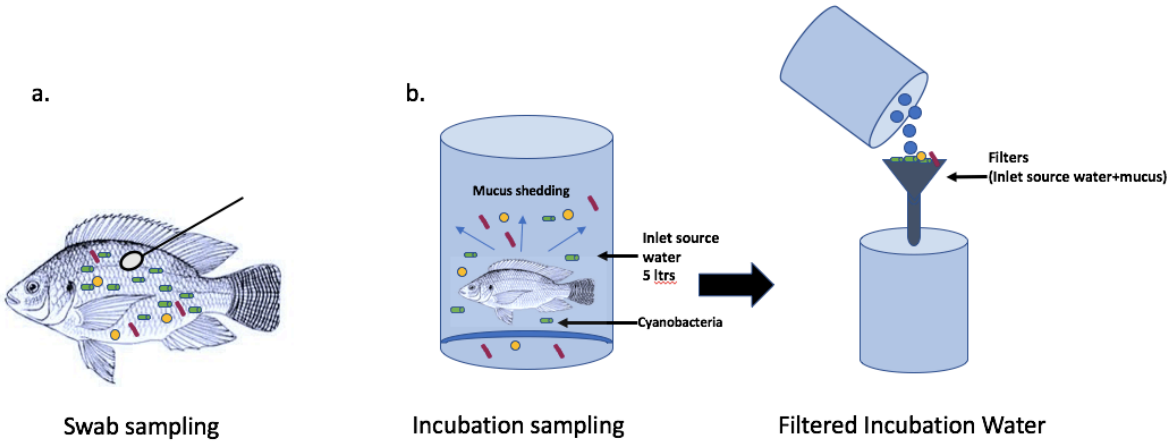


Figure 4. Sampling methods used to measure microbial communities and the presence of different taxa on the skin and in the mucus secreted by the fish. (a) Swab sample: using a sterilised swab to capture the skin microbial community. **(b)** Incubation sample: allowed the fish to swim and shed mucus in a 5 litre bucket filled with water from an inlet water source.

2.3. DNA extraction protocols for the analysis of microbiome

The DNA extraction protocol was used on all the samples and was used for the following 3 experimental chapters.

2.3.1. CTAB / EDTA DNA extraction protocol for filters

Preparation

Prior to DNA extraction, ethanol was removed from the incubation samples and rearing pond water samples, using a freeze dryer (ScanVac CoolSafe Pro with a 4L, -110°C condenser). Tubes were kept cold at all times to prevent DNA degradation. After freeze drying, tubes were stored at -80°C until processing. To protect samples from degradation, lysis buffer (30 mM Tris with 30 mM EDTA, pH 8) was added immediately after samples were removed from the freeze dryer.

DNA Extraction

Sample tubes were removed from -80°C freezer, and 570 µL lysis buffer was added immediately. Tubes were vortexed briefly, and returned to -80°C freezer (freeze/thaw cycle) for a minimum of one hour.

Samples were removed from the freezer and thawed briefly in pre-heated 37°C water bath. Pre-weighed beads of 0.6g were added and processed on Qiagen TissueLyser II bead-beater for 40s at 30 Hertz (Hz). 1µL of Ready-Lyse Lysozyme Solution (epicentre)/stock concentration 1000 U/µL, was added and mixed well by inversion. Incubation at 37°C took place for 30 mins with frequent mixing, to allow lysozyme to work. 30µL of SDS stock (10 %) and 3 µL proteinase K stock were added to each tube. An additional incubation took place at 55°C for 1 hr with frequent mixing using a rotating incubator. Tubes were transferred to 65°C water bath. 120µL pre-warmed, (at 65°C) 5 M NaCl was added and mixed well by inversion. 96µL of pre-warmed CTAB was placed in each tube, and mixing by inversion and incubating at 65°C for 10 minutes.

In a fume hood, 800 µL of chloroform was added and vortexed for 5 seconds. Tubes were then centrifuged for 5 minutes at 14 000 x g, 4°C, and kept chilled on ice. The

top aqueous layer was removed (approx. 700 uL) into a clean 1.5 mL tube, without disrupting the interface. An additional 700µL of chloroform was vortexed for 5 seconds. Tubes were centrifuged for 5 minutes at 14 000 x g, 4°C. With a 100 µL pipette, the top aqueous layer was transferred to a clean 1.5 mL tube. Total volume obtained was recorded from each tube without disturbing the interface. 1 µL of Linear polyacrylamide solution (GenElute LPA Sigma) was added to each 1.5 mL tube, vortex briefly. For each tube, the volume of aqueous layer was recorded and 0.7x v/v isopropanol was added to precipitate the DNA. Tubes were incubated in the dark at room temperature for at least 1 hour or overnight.

To pellet the DNA, tubes were centrifuged at maximum speed (~21 000 x g) for 30 minutes. All tubes were oriented carefully and consistently in the centrifuge, e.g. with hinges facing outward, so that pellets are in the same place in each tube.

Supernatant was removed by decanting and pipetting, with caution not to disturb the pellet. Pellets were washed with 400 µL ice-cold 70% ethanol and tubes were centrifuged at maximum speed for 10 minutes at 4°C. Ethanol was removed and the pellets were air-dried for 5 minutes. DNA was resuspended in 50 µL of TE buffer, mixing gently by flicking the tube.

Each sample was quantified using a Qubit fluorometer, using the Qubit 2.0 broad-range kit. Samples were stored at -20°C.

Removal of inhibitory RNALater from tissue swabs to enable 16S rRNA PCR

DNA extraction from swabs required extensive protocol development to remove inhibitory RNAlater from the samples into which they had been collected. Fourteen swab samples randomly selected from all ponds were used in these trials to optimise extraction protocols.

The whole ~2 ml contents of the sample including the swab was added to a 50 mL Falcon tube containing 23 mL of 1x PBS and vortexed at maximum speed for 30 sec, to allow DNA material to detach from the swab. The swab and liquid were transferred into a syringe, where it was pushed through a Sterivex filter unit (0.22 µm). Filters were then removed by cutting the tip with a pipe cutter, and in a sterile petri dish filters were

cut from their inner casing. Filters were then cut into smaller pieces using sterilized forceps and a scalpel (Le Vin et al. 2011).

Finally, the small filter pieces were transferred into a 1.5 ml Eppendorf containing lysis buffer, and DNA was extracted following the same procedure described previously.

2.4. Investigating bacterial skin microbiomes in tilapia aquaculture using Illumina

Primers used for V4 16S rRNA, were based on those adopted by the Earth Microbiome project:
(EMP; http://www.earthmicrobiome.org/emp-standard-protocols/16s/). Previously, both primers had been modified to reduce bias against Thaumarchaeota and SAR11 by adding additional degeneracy to the forward primer (Walters and et al. 2016). For this work, we used the updated versions here:
515fB: GTGYCAGCMGCCGCGGTAA
806rB: GGACTACNVGGGTWTCTAAT.

These primers produce an amplicon of ~291 bp, enabling near-complete overlap of forward and reverse reads on a paired-end run using v2 chemistry (2x250 bp). During merging of paired-end reads, the overlapping region between them can be deployed for correcting sequencing errors and potentially yield sequences of higher quality, resulting in a very low sequencing error rate. This is preferred over a longer amplicon with potentially more taxonomic resolution but a higher error rate (Rohwer et al. 2018) (Zhang et al. 2014).

2.4.1. PCR and library preparation

The DNA extracted was diluted to 2ng/μL and arrayed in 96 well plate format, using a cold rack to keep the samples cool. A temporary seal was used as a cover to avoid contamination. The multi-welled plate was placed in a centrifuge briefly to remove air bubbles.

The PCR reaction in each well contained a total volume of 50µL: 2.5µL DNA, 25µL NEBnext mix, and 2.5µL of each 10µM primer. The thermocycler was preheated to 98ºc, and cycling parameters were: 98°C for 30s, 30 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 30 s with a final extension at 72°C for 2 min. Each plate contained a negative control, nuclear-free water, and a positive control (mock community), from ZymoBIOMICS® Microbial Community DNA standard, lot number ZRC190811.

Amplified DNA was barcoded using a dual-barcode approach (Kozich et al. 2013) prior to sequencing. 4 µL from each sample underwent electrophoresis on a 1% agarose gel in TAE buffer (Tris-acetate EDTA) to verify successful amplification.

2.4.2. Library clean up and input normalization

Library clean-up was performed using pre-prepared Solid-phase Reversible Immobilization (SPRI) magnetic beads, following the (DeAngelis et al. 1995) protocol. The bead mix was kept at room temperature for 30 min before use. 0.8x volume of the bead mix was added into each tube (3 in total). The mix was kept for 2 minutes on the magnetic beads to allow the DNA to bind to the magnets. The supernatant was then removed and discarded. An ethanol 80% wash took place, by adding 200 µL of ethanol, and incubated on the magnets for 2 minutes. Ethanol wash was repeated once. Ethanol residue was removed with a fine tip, and left to air dry for 10 min on magnetic stand. Magnets were removed, and 22.5 µL TE + Tween were added. These were mixed by pipetting up and down, and incubated for 2 minutes at room temperature, and another 2 min on the magnet.

The samples were individually quantified using the Promega QuantiFluorONE kit and each was diluted to 4 nM with 10 mM Tris (pH 8.5). Each sample was pooled into single wells (i.e. 1 well/plate) and the plates were pooled to equal concentration into a single well (i.e. 1 well per run). The pooled library of samples included control extractions with sterile swabs, and negative controls for each DNA extraction plate and PCR plate.

2.4.3. Sequencing

The DNA was sequenced, in one sequence run using 250 bp paired-end sequence reads, on MiSeq (v.2.0), (Illumina Inc.), run by the Exeter Sequencing Service (University of Exeter, UK).

2.4.4. Sequencing analysis

In order to determine the bacterial microbiome of tilapia fish skin, amplicon sequencing of the V4 16S rRNA hypervariable region gene was employed, and then processing the samples for taxonomic composition. Sequences that are much longer or shorter than the expected product size for V4 may be the result of non-specific priming. After quality filtering, the median sequence length of each read was 253 bp, therefore sequences shorter than 252 bp and longer than 254 bp were removed in accordance with protocols established in the DADA2 pipeline (Callahan et al. 2016), This was to assist in improving taxonomic assignment.

Microbial community analysis was performed using DADA2 pipeline (Callahan et al. 2016) (1.8), followed by PhyloSeq to determine bacterial community composition. The analysis of microbiological communities brings many challenges: the integration of many different types of data with methods from ecology, genetics, phylogenetics, network analysis, visualization and multivariate statistics. Phyloseq is a package used as a tool to import, store, analyse and graphically display complex phylogenetic sequencing data that has already been clustered into Operational Taxonomic Units (OTUs) or Amplicon Sequence Variants (ASVs), which serve as taxonomic classifiers. Phyloseq seeks to facilitate the use of R for efficient interactive and reproducible analysis of OTU-clustered high-throughput amplicon sequencing data. This is particularly useful when there is associated sample data for evaluating ecological drivers of community composition and taxonomic assignment of the OTUs from many communities for comparison (Koblitz et al. 2017) (McMurdie and Holmes 2013).

In order to assess sampling depth (i.e. whether the community had been sufficiently sampled to capture the majority of taxa), rarefaction analysis was performed in PhyloSeq. Rarefaction Curves are a model for estimating the number of species in any assembly, and are used to predict the relationship between sampling effort and

species to be found, allowing a meaningful standardization and comparison of datasets (Gotelli and Colwell 2001), (Koblitz et al. 2017). Seven samples were discarded due to low sampling depth, including negative controls comprising filtered nuclear free water processed exactly as other samples to remove any bacteria associated with kit contamination from the samples.

2.4.5. Statistical analysis

To summarize and compare the bacterial community structure with respect to its richness (number of taxonomic groups) and evenness (distribution of abundances of the groups), alpha diversity metrics were measured. In microbial ecology, analysing the alpha diversity of amplicon sequencing data is a common first approach to assessing differences between environments (Willis 2019). To compare the alpha-diversity and the differential abundance of microbiome data, Shannon diversity of all samples was calculated within PhyloSeq, and mean Shannon diversity for each sample type were compared using pairwise t-tests to evaluate bacterial population abundance and alpha diversity between sample types.

Beta-diversity and statistical differences between sample types were assessed using Non-metric multidimensional scaling (NMDS) and ANOSIM (Primer-e software (v.7) https://www.primer-e.com/), respectively. All *P*-values were corrected for multiple hypothesis testing with the Benjamini–Hochberg false discovery rate method (Benjamini and Hochberg. 1995). Ordination analyses for all sample types were generated with the Bray-Curtis distance metric. One-way crossed SIMPER analysis was calculated, between different sample types, using the Primer-e software (v.7) https://www.primer-e.com/, to estimate the per-taxon contribution to Bray-Curtis distances between sample types. A cut off for low contributions was set at a ranked cumulative value of 70% dissimilarity. All figures were created in RStudio (169) with the PhyloSeq (v. 1.14.0) and ggplot2 (v. 2.1.0) packages.

Cyanobacterial amplicon sequence variants were selected for further study to determine if enrichment in swab samples was evidence of host-specific, cyanobacterial taxa exclusively associated with the mucosal layer. Classification of selected ASVs were aligned using SILVA Incremental Aligner (SINA) (Pruesse et al. 2012).

2.5. Sequencing full length 16S rRNA using the Oxford Nanopore Technologies MinION™ sequencer

For DNA and the library preparation, the Oxford Nanopore protocol 1D PCR barcoding amplicons (SQK-LSK109) was followed. Specifically, amplifying ~1,500bp fragments of the full 16S rRNA gene. Bacterial DNA was amplified by PCR for 11 samples, chosen based on available DNA quantities across incubation, swab and rearing water (pond) sample types. The primer set 27F-1492R was used to amplify V1-V9 hypervariable regions of the 16S rRNA gene. This universal primer set is the most commonly used when assessing full length 16S rRNA gene, because it has been shown to successfully amplify all known bacterial taxa (Mao et al. 2012). The primers used in this study are listed in Table 3, and contain some ambiguous bases (2, 5) for forward and reverse primers respectively, to make the primers more universal (Cusco et al. 2017).

Primer name	Sequence	Target	Reference
27F	**GTGYCAGCMGCCGCGGTA*A (19)**	**Full 16S rRNA**	(Frank et al. 2008)
1492R	**CCGYCAATTYMTTTRAGTT*T (20)**	**Full 16S rRNA**	

Table 3. Primers used to amplify V1-V9 hypervariable regions of the 16S rRNA gene (full-length). Asterisks refer to the location of the modified phosphorothioate bonds, to prevent nuclease degradation (Skerra 1992). The number in brackets after the sequence refers to the length of the primer.

2.5.1. PCR amplification and clean-up

2.5.1.a PCR

The bacterial DNA was amplified by creating a final PCR mixture of (50 µL), which contained 3 µL (2ng/µL) of DNA for each sample, 25 µL of NEBnext, and 1 µL of each primer (10µM). The PCR thermal profile consisted of an initial denaturation step for 30s at 98°C, followed by 35 cycles for 30s at 98°C, 45s at primer-adjusted annealing

temperature at 59°C, followed by 90s at 72°C, 30s at 72°C for extension. To assess possible reagent contamination, each set of PCR reactions included a no template control (NTC) sample.

The PCR thermal profile consisted of an initial denaturation step for 30s at 98°C, followed by 35 cycles for 30s at 98°C, 45s at 59ºC and 90s at 72°C for annealing, and a final extension for 30s at 72°C. Primer concentrations for 27F-1492R primer was 0.4μM of 27F and 0.8μM of 1492R. extension, and a final extension step for 7 min at 72°C. Following the PCR round, a clean-up step using AMPure XP beads at 0.5X concentration was used to discard short fragments and the DNA quantity was assessed using Qubit fluorometer.

2.5.1.b PCR clean-up (Ampure Beads)

The library clean-up and normalization was performed using AMPure magnetic beads, using a 40 % bead wash (17.2 μL of beads in 43 μL of PCR product). Beads were brought to room temperature for 5 minutes before use. The mixture was placed on a magnetic holder for 10 minutes. The supernatant was discarded. Two ethanol washes took place, by adding 200 μL of freshly prepared 70 % ethanol to the beads. The supernatant was removed with a fine tip to allow the beads to air dry. Finally, 26 μL of NFW (heated at 55°C) was added to elute the DNA, and 1 μL was used to quantify the DNA in each sample using the Qubit fluorometer.

2.5.1.c Library preparation and barcoding

Following the protocol, we performed an end-repair/ dA tailing of fragmented DNA, by gently mixing 45 μL < 1 μg end-repaired DNA, 7 μL of Ultra II End-prep reaction buffer and 3 μL Ultra II End-prep enzyme mix. Samples were transferred to a 0.2 μL PCR tube, and incubated in the thermocycler for 5 min, at 20°C, followed by 5 min at 65°C. 60μL of resuspended AMPure XP beads were added to the end-prep reaction, and mixed by pipetting, followed by a rotator mixer at room temperature for 5 min. Samples were spun down and pelleted on a magnet. Supernatant was pipetted and discarded while the tubes were on the magnet. An ethanol wash was performed twice by washing the beads with 200 μL of freshly prepared 70 % ethanol. The ethanol was removed and discarded, without disturbing the pellet. The tubes were spun down and placed back on the magnet. Any residual ethanol was removed by using a fine pipette. It was

allowed to dry for ~10 min. Tubes were removed from the magnetic rack and pellet was resuspended in 25 µL nuclease-free water at 55 °C. The tubes were kept at room temperature for 2 min. Beads were pelleted on the magnet until the eluate was clear and colourless. 25 µL of eluate was removed and transferred into a clean 1.5 ml Eppendorf DNA LoBind tube. 1 µL of DNA was quantified using a Qubit fluorometer.

To multiplex sequences on the MinION run, native barcode oligomers (Oxford Nanopore) were ligated onto amplicons prior to final library preparation. Native barcodes were thawed at room temperature, using for one barcode per sample. 500 ng of each end-prepped sample was diluted to 22.5 µL in nuclease-free water and 2.5 µL of native barcode and 25 µL of blunt/TA ligase master mix were added to the 22.5 µL 500 ng blunted DNA, mixed by flicking the tube after each addition and spun down. The reaction was incubated for 10 minutes at room temperature. Resuspended AMPure XP beads (50 µL) were added to the reaction and mixed by pipetting. Tubes were incubated for 5 minutes, on a rotator mixer at room temperature, followed by 2 ethanol washes as described previously. The tubes were removed from the magnetic rack and the pellets were resuspended in 26 µL nuclease-free water. Tubes were incubated for 2 minutes at room temperature and beads were pelleted on the magnet until the eluate was clear and colourless.

26 µL of the eluate was removed and transferred into a clean 1.5 ml Eppendorf DNA LoBind tube. Finally, the beads were disposed of and 1 µL of each barcoded DNA sample was quantified, using a Qubit fluorometer. Equimolar amounts of each barcoded sample were pooled into a DNA LoBind 1.5 ml Eppendorf tube to produce a pooled sample of >700 ng total. 1 µL of pooled and barcoded DNA was used for quantification on a Qubit fluorometer. 700 ng of the pooled sample was diluted into 50 µL in Nuclease-free water for adapter ligation.

The adapter ligation step was performed by adding 50 µL 700 ng pooled barcoded sample to 20 µL Barcode Adapter Mix (BAM), 20 µL NEBNext Quick Ligation Reaction Buffer (5X), and 10 µL Quick T4 DNA Ligase. The tubes were gently flicked and spun down. All the reagents used in this step were initially thawed and mixed with a quick spin down and pipetting, to ensure the mixing of the contents of the reaction. The reaction was incubated for 10 minutes at room temperature. 40 µL of resuspended

AMPure XP beads were added to the adapter ligation reaction from the previous step and mixed by pipetting.

Using a rotator mixer, the reaction was incubated for 5 minutes at room temperature and placed on a magnetic rack, to allow beads to pellet. Supernatant was removed by using a fine pipette. This was followed by adding 140 µl of ABB Buffer (ABB) to the beads. The beads were resuspended by flicking the tube, and then pelleted on the magnetic rack, and the supernatant was removed. The previous step was repeated, and the beads were suspended in 15 µL Elution Buffer, followed by an incubation of 10 min at room temperature. The beads were placed on the magnet until the eluate was clear and colourless. As a final step, 15 µL of eluate, which contains the DNA library was retained and transferred into a clean 1.5 ml Eppendorf DNA LoBind tube. The DNA was quantified using a Qubit fluorometer.

2.5.1.d Priming the SpotON Flow Cell

The flow cell was inspected to confirm no air bubbles were present, and to ensure the buffer was continuous throughout the flow cell. A priming mix was made by mixing 480 µL RBF (Running Buffer with Fuel Mix), and 520 µL nuclease-free water. The flow cell priming port was then loaded with 800 µL priming mix. After 5 min, the remaining primer mix was then loaded and left at room temperature for a further 5 minutes.

2.5.1.e Library loading

A library loading mix was made consisting of RBF (25.5 µL), DNA library (11.0 µL), LLB (26.6 µL) and nuclease-free water (12 µL) resulting in a total volume of 75 µL. The final mix was mixed by inversion and spun down. The mix was finally loaded dropwise onto the SpotON sample port. The sequencing run was then initiated through the MinKNOW software and ran uninterrupted for 48 hours. After 24 hours, an additional 70 µL of RBF was added via the SpotON port to refuel the flowcell and improve yield.

2.5.1.f Sequencing analysis

Base calling was performed using Guppy v.2.3.5. The new Flip-flop algorithm was used to maximise the accuracy of reads (92-95% accuracy). Adapters and barcodes

were trimmed using Porechop (https://github.com/rrwick/Porechop), and reads were assigned into groups, according to their barcodes. Reads were only included if a barcode was detected at both ends to increase stringency and minimise cross-talk between samples. Reads with barcodes and/or adapters in the middle of the read were also discarded (these can occur if two reads go through the same pore in quick succession).

As a final step, reads were filtered to remove sequences shorter than 1,200bp and longer than 1,800bp (Cusco et al. 2017), using seqtk (https://github.com/lh3/**seqtk**). Trimmed reads were classified using SINA (Pruesse et al. 2012) against the SILVA database (https://www.arb-silva.de/). The workflow of the of this study is shown in Figure 5.

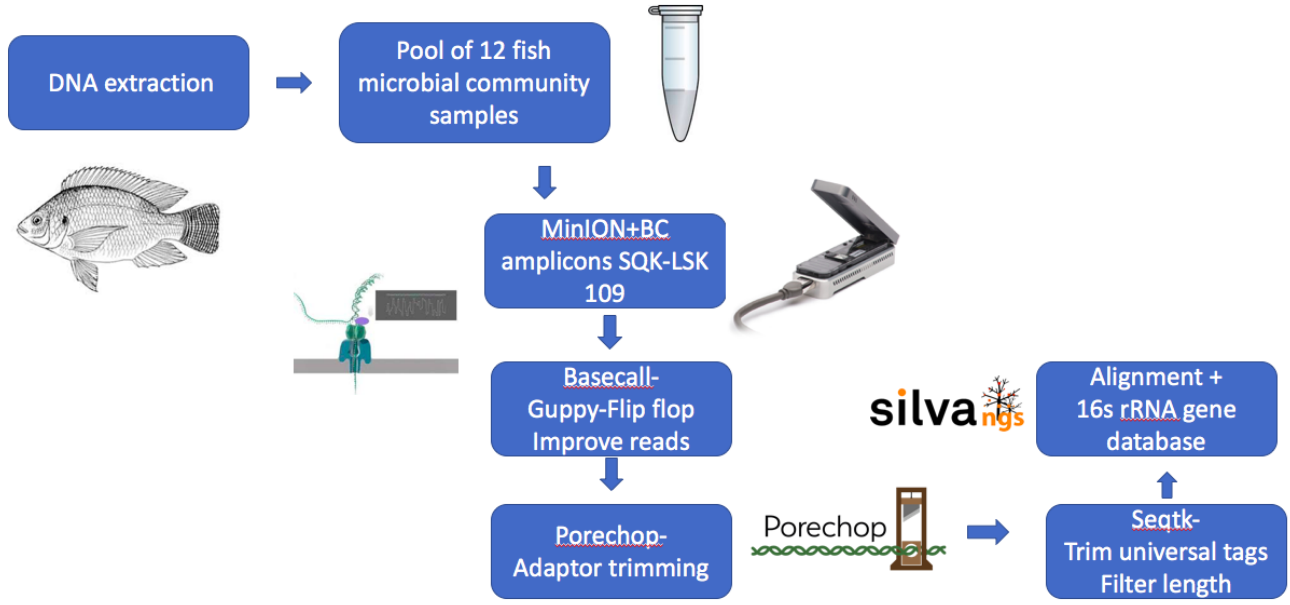


Figure 5. Work flow for MinION sequencing of full length 16S rRNA primer sets. Swab, incubation, and rearing water samples were selected from ponds six and seven. The sequencing kit SQK-LSK109 was used, and different types of software were applied at different steps to filter reads based on quality. This figure was adapted from (Cusco et al. 2017).

2.5.1.g Statistical analysis

All quantitative and qualitative statistical analyses were conducted in “R” version 3.5.1. All samples were converted to relative abundance by dividing the number of reads assigned to each taxon by the total number of reads assigned to all taxa within a sample. Alpha-diversity of all samples was measured using Shannon diversity, followed by the mean Shannon diversity for all samples types excluding the negative control. Sample MLW_I6b was discarded due to a low number (<1000) of reads.

2.6. Comparing microbial populations in samples using Illumina versus MinION sequencing

A comparison was made to compare the use of full length, but error prone 16S rRNA amplicons to shorter, but high-accuracy Illumina reads for capturing microbial communities**.** Each sample was sequenced using the full length 16S rRNA on the MinION and using the standard V4 2x300 bp Illumina run. The abundance of the top five phyla for each sequencing method was evaluated to assess whether the two methods recovered similar microbial community profiles.

.

2.7. Evaluating T4 Bacteriophages in Tilapia Skin Microbiomes

The conserved gene encoding the major capsid protein (*g23)*, was utilised to investigate T4- like myovirus communities present in the previously extracted and quantified bacterial microbiomes, following the protocol carried out by Chow et al. (Chow and Fuhrman 2012). To evaluate the use of these primers for assessing bacteriophage diversity, a trial was first conducted on samples selected from each of the three sampling types (rearing pond water, incubation, and skin swabs) based on available DNA quantities. T4 (ATCC 11303-B4) DNA extracted from a pure culture was used as a positive control.

2.7.1. Isolation of DNA from T4 Phage lysate

T4 (ATCC 11303-B4) was propagated in a 50 mL overnight culture of *E. coli* (ATCC11303) in LB medium with 0.5 % NaCl. Cells were pelleted by centrifugation and the supernatant was filtered using Sterivex filter (0.22 µL filter units). 6.5 g of NaCl was added to 100 µL filtered phage suspension (a final concentration of 1 M), and incubated for 1 hour on ice, followed by a 10 min centrifugation at 11,000xg. The supernatant was then transferred to a new bottle. 10g of Polyethylene Glycol 8000 at 100g/l, was added and mixed by shaking.

The mix was incubated for 1 hour on ice, followed by another centrifugation step for 10 min. at 10,000xg. The supernatant was discarded carefully. Bottles were set upside down on paper towels to drain the remaining liquid. The inside of the bottles was rinsed twice with 1-2 ml SM buffer. 1 ml was collected in two (2 ml) microcentrifuge tubes. 1 ml of resin was added to each 2 ml tube after shaking to resuspend and TE was heated to 80°C. A column was attached to a 3 ml syringe and samples were pushed through, 1 ml at a time. Subsequently, 2 ml of 80 % isopropanol was pushed through each column, 1 ml at a time. The column was placed back in the original 2 ml tube, and then transferred in a new 1.5 ml microcentrifuge tube, where 100 µL of 65°C TE was added to elute. This was followed by a brief vortex and centrifuge for 30 sec. at 10,000xg. Finally, the concentration of isolated DNA was quantified using a Qubit fluorometer, using the Qubit 2.0, broad-range kit.

2.7.2. PCR of T4 (g23 major capsid protein gene) (No barcodes)

To amplify and sequence the g23 major capsid protein gene on the Illumina MiSeq platform, the initial PCR was performed with g23 primers as follows:

Primer name	Sequence	Target	Reference
T4-SuperF1	**GAYHTIKSIGGIGTICARCCIATG**	**g23 major capsid protein**	

T4-SuperR1	**GCIYKIARRTCYTGIGCIARYTC**	**g23 major capsid protein**	(Chow and Fuhrman 2012)

Table 4. T4 primers used to amplify g23 major capsid protein gene in T4 bacteriophages, with degeneracies of 48 and 128, respectively. The expected fragment size ranges from 400–500 bp (Chow and Fuhrman 2012).

The PCR reaction run for each sample contained 2 ng of DNA (3µL), 1x ThermoPol Buffer (includes 20 mM MgSO4, NEB), additional 1.75 mM $MgCl_2$ (final concentration is 3.75 mM), 0.25 mM each dNTP, 400 nM each primer, 10 mg BSA (Sigma #7030) and 5 units Taq DNA polymerase (NEB) with a final volume of 50 µL.

Cycling parameters were as follows: 95°C for 5 min, 35 cycles of 95°C for 30 s, 54°C for 30 s and 72°C for 30 s with a final extension at 72°C for 9.5 min. Five microlitres of each PCR reaction was electrophoresed in 1% agarose in 1x TBE, pH 8.0 at 100 V for 60 min to verify amplification.

A concentration gradient PCR was performed by repeating the same protocol above for other samples using the same T4 DNA extract. DNA concentrations were as follows: 1, 2, 5, 10 ng/µL.

2.7.3. T4 Barcoded Primers

To multiplex samples on a single run, barcodes were added to the T4-like *Myoviridae* amplicons. Phased spacers were used to increase library yields by increasing the heterogeneity of the base calls during the cycle. The same protocol as above was followed**.** Primers were customized and ordered from *Eurofins (*Figure 6).

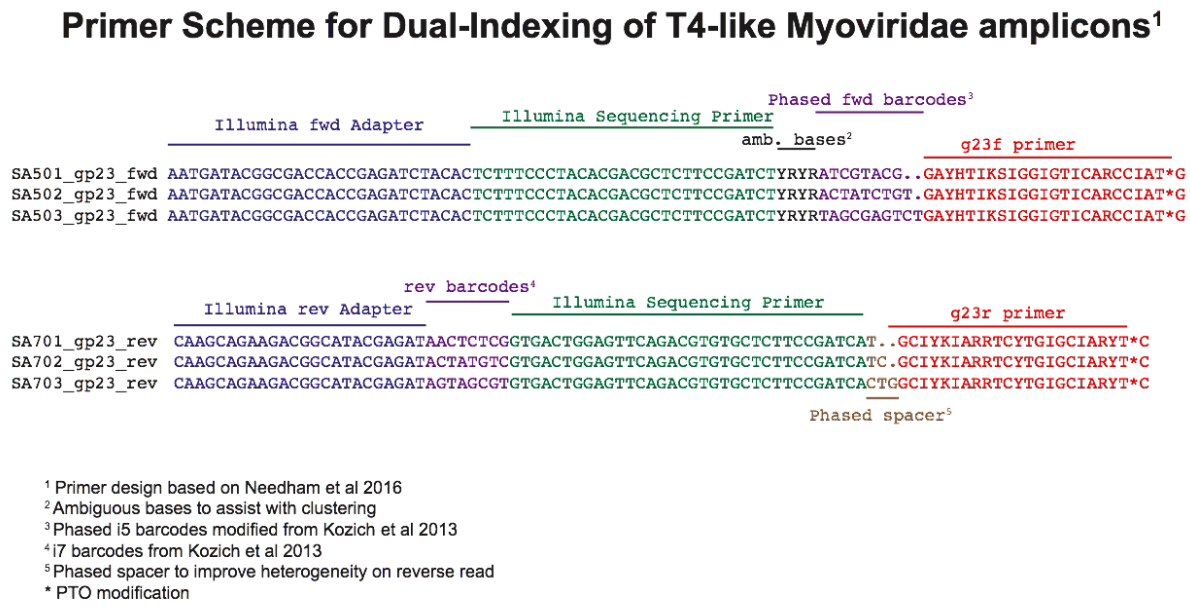


Figure 6. Primer scheme for dual indexing of T4-like *Myoviridae* amplicons using Illumina.

PCR conditions were the same as the protocol followed for the non-barcoded primers, except the concentration of primers were 2.5 µM, and the volume of the primers used for each sample was 2 µL.

2.7.4. Temperature gradient PCR

To optimise the PCR product and reduce the amount of primer dimers present as by-products, a temperature gradient PCR was performed, concentration and the amount of primer added to the PCR reaction was modified; the concentration of primers were 10µM, and the volume of the primers used for each sample was 1 µL. Cycling parameters were as follows: 95°C for 5 min, 35 cycles of 95°C for 30 s, 57°C for 30 s and 72°C for 30 s with a final extension at 72°C for 9.5 min. Five microlitres of each PCR reaction was electrophoresed in 1% agarose in 1x TBE, pH 8.0 at 100 volts, for 60 min to verify amplification.

Following the temperature gradient, a PCR was performed with the chosen annealing temperature (57.5ºC) to reduce primer dimers, for selected samples and T4 DNA extract also. This was performed several times. The concentration of the primers ranged from 1- 2.5 µM.

Finally, our last attempt to amplify the g23 gene using the T4 barcoded primers, was by creating a concentration gradient for the barcoded primers, using the same successful sample “P4a”. An annealing temperature 54 ºC was chosen for this cycle. Concentration of the primers was as follows: 1µM, 1.4 µM, 2.5µM.

Attempt	*primer*	*Type*	*Conc P. (10μM)*	*Vol. p.*	*Annealing temp.*	*DNA con.*	*Product size*
1	**Super T4**	**Pond Swab**	**1μM**	**2μL**	**54 ºC**	**2ng/μL**	**400bp**
2	**Super T4**	**Pond Swab Incub.**	**1μM**	**2μL**	**54 ºC**	**2ng/μL**	**400bp**
3	**Super T4**	**Swab Incub.**	**1μM**	**2μL**	**54 ºC**	**2ng/μL**	**400bp Pos 600**
4 *conc. Gradient	**Super T4**	**T4 DNA**	**1μM**	**2μL**	**54 ºC**	**2ng/μL**	**600bp**
5	**BC**	**T4, Pond**	**2.5μM**	**2μL**	**54 ºC**	**2ng/μL**	**400 Pos 500**
6 *temp Gradient	**BC**	**T4**	**1μM**	**1μL**	**(44-60)ºC**	**2ng/μL**	**600bp**
7	**BC**	**T4**	**1μM**	**1μL**	**57.5ºC**	**2ng/μL**	**Non, Pos 500**
8*conc. Gradient	**BC**	**Pond**	**2.5μM, 1.4μM, 1μM**	**1μL**	**54 ºC**	**2ng/μL**	**Non Pos 500**

Table 5. Optimizing methods for PCR with T4 Myoviridae barcoded primers, showing the volumes and concentration of the primers used in each step.

2.7.5. PCR Purification and Sequencing

Due to the failure to get successful amplification using barcoded primers, we selected one sample (P4a) to proceed with evaluating the T4-primers without barcoding for ground-truthing future studies. Amplicons for sample P4a (a rearing water sample from pond 4) was selected, which was amplified using the barcoded primers. A gel extraction was performed using the QIAquick PCR purification kit (Qiagen, Courtaboeuf, France), by following the manual instructions. The g23 products were prepared for Illumina sequencing by adding 1 µL of each purified PCR product to the reaction mix, except g23-only primers were replaced with an Illumina compatible g23 primer construct (Illumina Inc.), as shown in Figure 6, by the Exeter Sequencing Service (University of Exeter, UK).

2.7.6. Sequencing Analysis

Paired-end sequence reads were filtered, trimmed, concatenated and processed with DADA2 as described previously. Primers were removed from each sequence. Sequences (*g23* gene) were clustered into ASVs as described previously. Nucleotide sequence similarity searches were performed for 10 randomly selected ASVs using BLASTN, on the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/BLAST/) website against the NT database, to validate that ASVs were indeed representative of T4-like myoviridae capsid genes.

A phylogenetic tree of ASV sequences was created using the Interactive tree of life (iTOL) v.4 online tool with default parameters, and plotted with ASV abundance in the single sample to evaluate phylogenetic diversity and community evenness. Alpha-diversity was also calculated for this single sample.

3. Results and Discussion

v4 16S rRNA gene amplification

Three different types of samples were obtained from the field for sequencing: skin swab samples to capture the mucosal microbial communities, fish incubation samples and rearing water samples (ponds), obtained from seven different ponds in Malawi (Figure 3 and Table 6). These samples were sequenced using Illumina MiSeq (2 × 250 bp) paired-end reads, targeting the V4 region of the 16S rRNA gene. 2,139,615 raw reads were obtained from 116 samples. Relative abundances of taxa were estimated in each sample using the DADA2 v1.8 (Callahan et al. 2018) and PhyloSeq software packages to resolve taxonomic units at a resolution of as little as one nucleotide per amplicon. The end product is an amplicon sequence variant (ASV), and the number of times each ASV was observed in each sample was recorded. Despite the popularity of clustering-based methods that identify taxon- independent operational taxonomic units (OTUs) using a sequence similarity threshold (Eren et al. 2013), researchers are forced to employ relatively low similarity thresholds (usually 97%) to minimize inflation of the number of OTUs because of random sequencing errors. ASVs are reusable across studies, reproducible in future data sets and are not limited by incomplete reference databases. These improvements are sufficient enough for ASVs to replace OTUs as the standard unit of marker-gene analysis and reporting (Callahan et al. 2017). Following the pre-processing of the reads, which included quality-based trimming, paired end merging, de-replication and chimera removal, 1,689,361 reads were retained for further analysis.

To remove non-specific priming, ASVs with a length less than 252bp, and greater than 254bp (based on ASV length distribution across all remaining samples) were removed. 514 chimeric sequences (<1% of the total) were identified, leaving a total of 14,424 unique ASVs.

Sample type	*No. of samples*	*No. of reads*
Ponds	35	676,740
Incubations	24	289,513
Incubation control	5	78,206
Swabs	52	644,902
Total Reads	**116**	1,689,361

Table 6. The number and percentage of final reads for each sample type. Results obtained after pre-processing the reads for the V4 16S rRNA genes of the microbial community.

3.1. Community compositional analysis in the fish and water samples: (skin swabs, and fish incubation and rearing pond waters)

Although the number of reads that were assigned to a taxonomy was 75.5%, only 2.3% of the reads reached species level classification. Overall, the top 20 taxa in all sample types, were represented by the following seven phyla: Cyanobacteria, Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia, Planctomycetes (Table 7). In total, they represented 88% of all reads. Proteobacteria dominated the top 20 ASVs, (up to 37.9%), followed by Cyanobacteria (11.57%), Actinomycetes (8.6%), Bacteroidetes (7.88%), Planctomycetes (6.98%), and Verrucomicrobia (5.01%). Three of the Cyanobacteria ASVs could only be assigned to Family level, as "*Family I*" due to limitations in the SILVA database.

ASV	*Phylum*	*Further Taxonomic Classification*
ASV 1	Cyanobacteria	Subsection III, Family I
ASV 2	Cyanobacteria	Subsection III, Family I, *Microcystis*
ASV 3	Cyanobacteria	Subsection III, Family I
ASV 4	Cyanobacteria	Subsection III, Family I
ASV 5	Proteobacteria	Burkholderiaceae*, Ralstonia*
ASV 6	Cyanobacteria	Family I, *Synechococcus*
ASV 7	Actinobacteria	Frankiales, Sporichthaceae, *HgcI clade*
ASV 8	Bacteroidetes	Cytophagales, Cytophagaceae
ASV 9	Proteobacteria	Rhizobiales, Methylocystaceae, *Methylocystis*
ASV 10	Proteobacteria	Burkholderiaceae, *Polynucleobacter*
ASV 11	Proteobacteria	Sphingomonadaceae, *Sphingomonas*
ASV 12	Verrucomicrobia	Chthinobacteriales, *LD29*
ASV 13	Planctomycetes	Plactomycetaceae, *Pirellula*
ASV 14	Proteobacteria	Comamonadaceae, *Pelomonas*
ASV 15	Proteobacteria	Comamonadaceae
ASV 16	Cyanobacteria	Subsection III, Family I, *Cyanobium*
ASV 17	Actinobacteria	Frankiales, Sporichthaceae, *HgcI clade*

ASV 18	Proteobacteria	Comamonadaceae
ASV 19	Verrucomicrobia	Chthinobacteriales, *LD29*
ASV 20	Actinobacteria	Frankiales, Sporichthaceae, *HgcI clade*

Table 7. Top 20 ASVs of tilapia microbiome communities (fish skin, rearing water (pond), incubation and incubation control samples). Showing the highest abundant taxa represented across all sample types, and the last taxonomic classification reached for each ASV.

3.1.1. Species Richness

Species richness in each sample was evaluated using rarefaction. Across all samples, rarefaction curves were asymptotic, shown in Figure 7, suggesting sufficient sampling depth had been achieved to saturate sample diversity. Less variation (spread of the asymptotes) was shown in pond (rearing water) samples, and the highest variation was seen in swab samples.

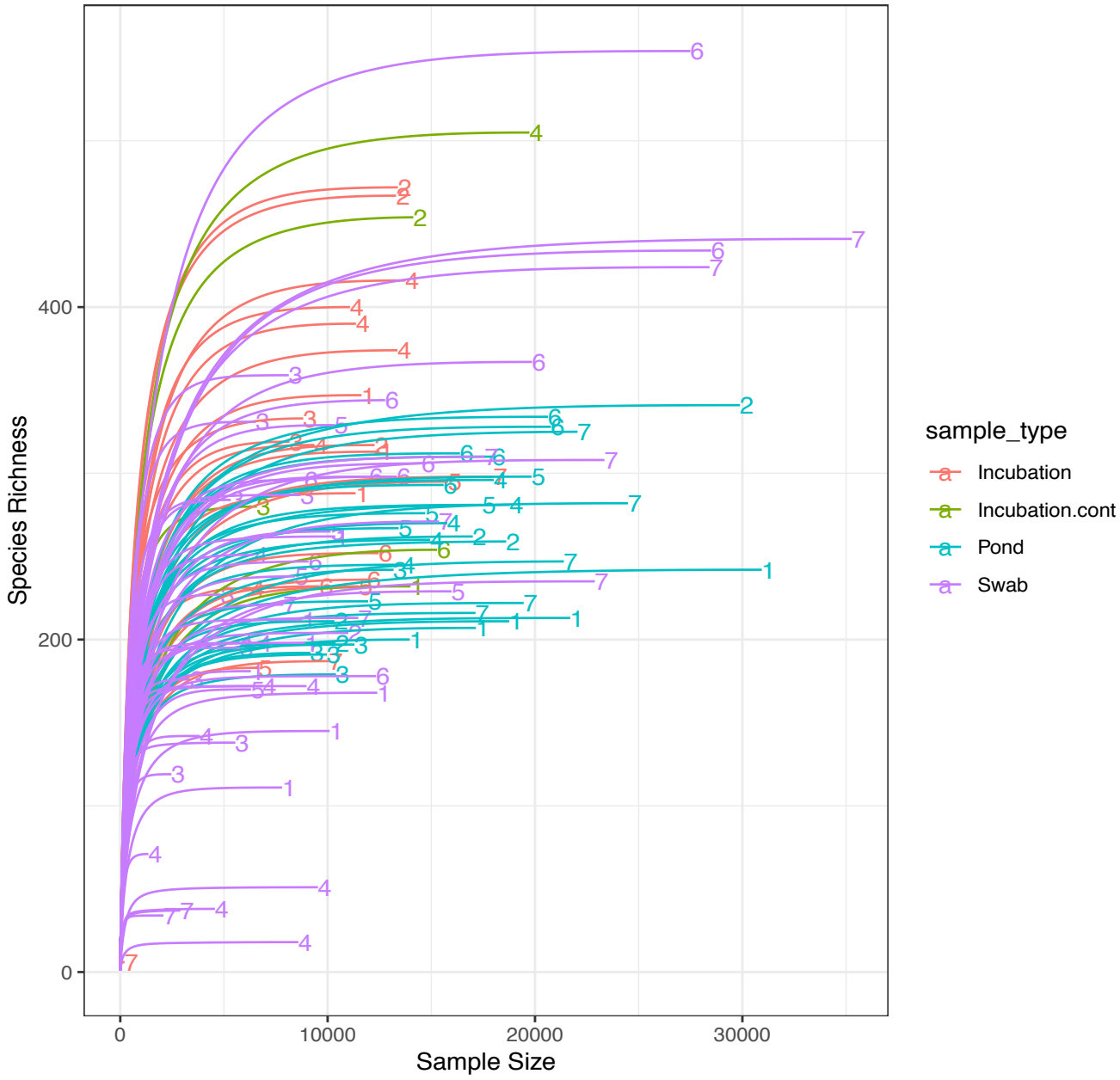


Figure 7. Sequence coverage represented by rarefaction curves. Sampling effort captured the majority of microbial diversity in each sample. Each colour represents the sample type, and the numbers represent the different ponds they were sampled from.

3.1.2. Microbial alpha diversity across sample types and location

Shannon diversity was employed to measure the richness and evenness across all samples types taken from the seven different ponds. Four out of seven ponds had

swab samples with a higher diversity than incubation samples as shown in Figure 8. Rearing water samples were generally less diverse than incubation and swab samples. Pond 4 showed the lowest diversity for swab samples. Alpha-diversity in the incubation control samples was variable, with low diversity observed in ponds 1 and 6 (3.94 and 4.0), compared to other sample sites (5.26, 5.33 and 5.67 for ponds 2, 3 and 5, respectively). The mean Shannon diversity for pond samples (4.63 ± 0.05 s.e), incubation samples (4.71 ± 0.21 s.e.), incubation control samples (4.84 ± 0.36 s.e.) and swab samples (4.41 ± 0.12 s.e.) were not significantly different (Q> 0.05, pairwise t-test, adjusted for multiple testing with Benjamini-Hochberg).

Guided by the similarity in alpha diversity between incubation control samples and incubation samples; and the fact that two of the ponds (5 and 7) were missing incubation controls, incubations and incubation controls were pooled together for each pond as 'incubation' samples for evaluating statistical differences in community composition. One-way analysis of variance (ANOSIM) was performed in Primer-e software v.7, https://www.primer-e.com, using the type of sample as factors to determine if communities were significantly different between sample types. A pairwise analysis of differences in community composition was performed (ANOSIM, 999 permutations). There was no significant difference between communities associated with incubation and swab samples (R=0.023, P=0.204). Incubation and rearing water communities were significantly different (R=0.114, P=0.006), as were rearing water and swab communities (R=0.082, P=0.009).

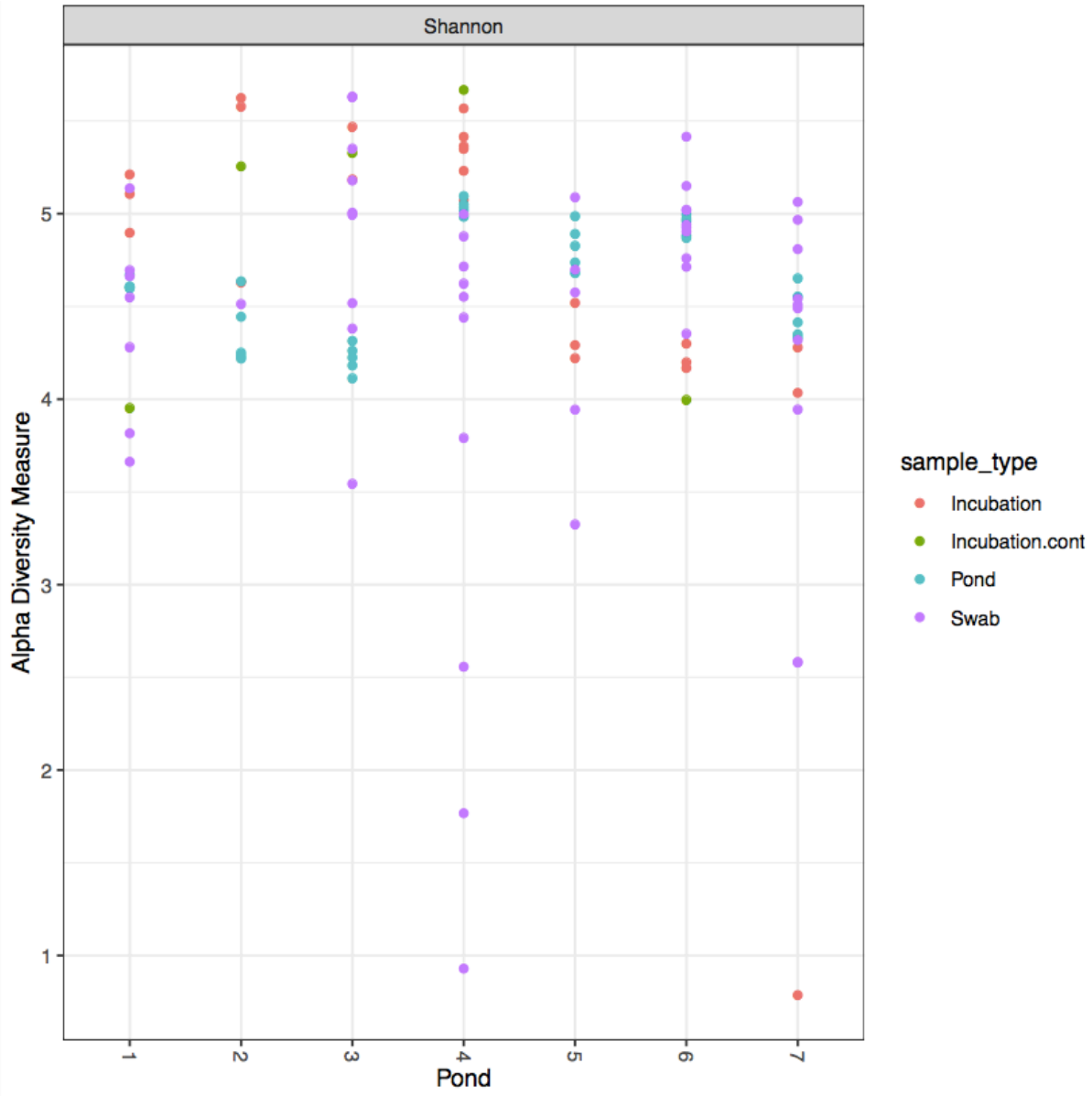


Figure 8. Alpha-diversity, measured by Shannon Diversity Index. Alpha diversity was plotted for all 7 ponds, with the colour representing different sample types (incubation, incubation control, pond (rearing water), and swab). Incubations and swabs generally had a higher diversity than rearing water samples.

3.1.3. Structural diversity of microbial communities

Beta-diversity analysis was measured using Nonmetric multidimensional scaling (NMDS) ordination analysis for all sample types, showing notable differentiation

among bacterial communities, as they grouped mainly by origin, as seen in Figure 9. Rearing water samples had the lowest variance in community structure. Swab samples showed the greatest variance in community structure and diversity, potentially as a result of greater community variance across individual fish. Additionally, rearing water samples and incubation samples from the same geographical location showed low within-sample-type variance, suggesting community structure was correlated with sampling site, with an obvious lower variance (tighter clustering) in ponds six and seven, located in Maldeco. Similar to alpha-diversity, Bray-Curtis distances between incubation controls and post-incubation samples varied across sample sites. Incubation controls for Pond 4 clustered with their respective post-incubation environment, suggesting either high microbial load in the incubation control sample or little influence of fish microbiome in post-incubation samples. In contrast, distances between controls and post-incubation samples in ponds 1 and 6 were much greater (Figure 9).

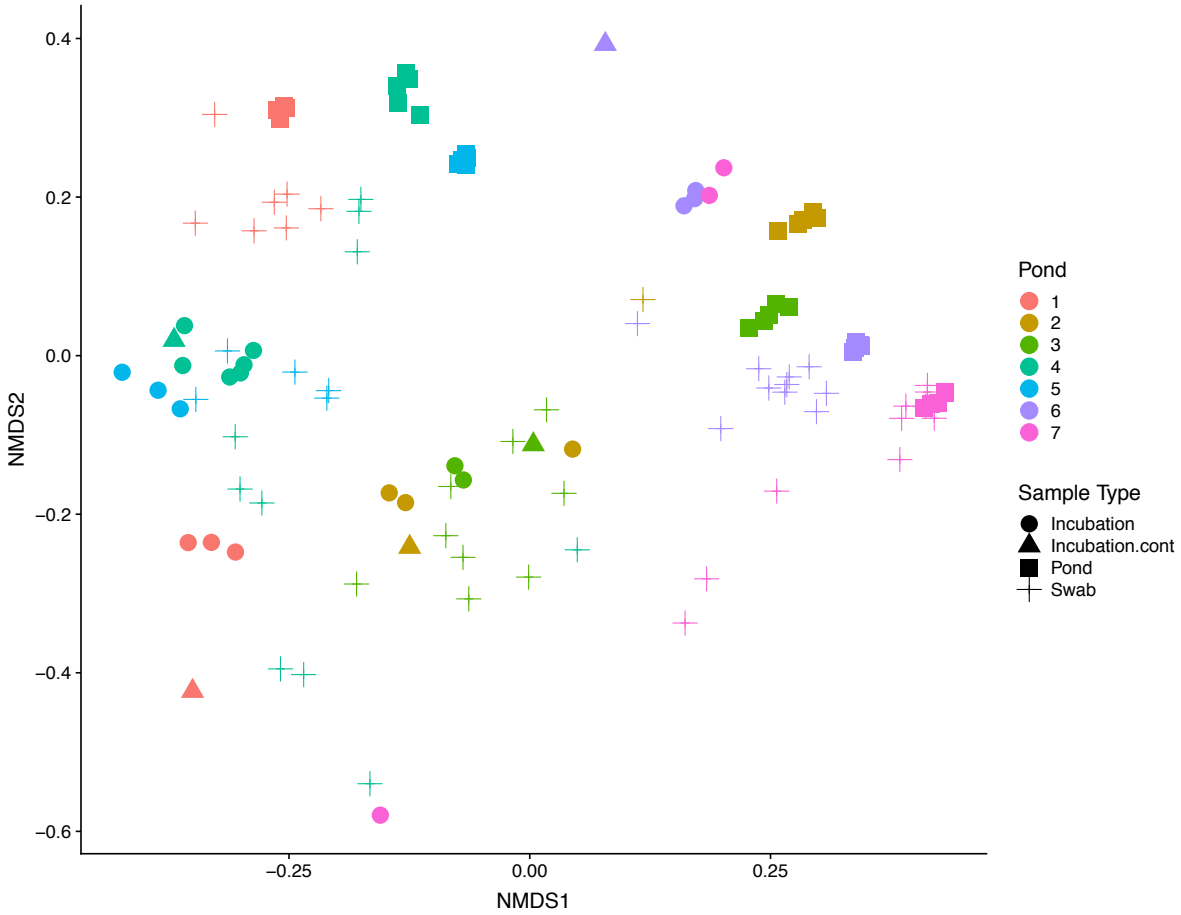


Figure 9. Beta diversity of all samples as measured using Bray-Curtis distances. Greater variance between swab samples compared to incubation samples is clearly visible. Distances between incubation controls and incubation samples varied according to sampling site.

3.1.4. Dissimilarity among sampling methods (Rearing pond water, incubations, and swabs)

One-way crossed SIMPER analysis (Primer-e v.7) was employed to measure the relative contribution of ASV abundance to the Bray-Curtis distance between different sample types. The largest contributors to the dissimilarity between all sample types were members of Proteobacteria, Cyanobacteria, and Bacteroidetes. Dissimilarity was measured pair-wise between each sample type; between the incubation samples and pond samples the average dissimilarity was 91.32%, whereas the average dissimilarity between the incubation and swab samples was 90.94%. The ASVs contributing most greatly to dissimilarity tended to be those assigned to the Cyanobacteria and

Proteobacteria. Finally, the dissimilarity of the pond (rearing water samples), and swab samples, was accounted mainly for Cyanobacteria and Bacteroidetes ASVs. The average dissimilarity between the swab and pond samples was 83.34%. A detailed table below, Table 8, shows the top phyla and percentage of their contributions to dissimilarities between sampling types.

Incub. & Pond taxa	***Contrib %***	***Incub. & Swab Taxa***	***Contrib %***	***Rearing water & Swab***	***Contrib %***
Proteobacteria Burkholderaceae	15.15	Cyanobacteria Family I	11.38	Cyanobacteria Family I	14.66
Cyanobacteria Family I	15.03	Proteobacteria Burkholderaceae	10.99	Bacteroidetes Cytophagaceae	11.59
Bacteroidetes Cytophagaceae	12.64	Proteo comm	9.84	Cyanobacteria Family I	11.28
Cyanobacteria Family I	9.89	Cyanobacteria Family I	9.51	Proteo Comamonadaceae	11.23
Proteobacteria Sphingomonas	7.37	Cyanobacteria Microcystis	8.64	Proteobacteria Burkholderaceae	10.93

Table 8. One-Way SIMPER analysis showing the top phyla for dissimilarities between sample types, and the percentage of contribution to these dissimilarities.

*(Contrib% = % of contribution for each variable)

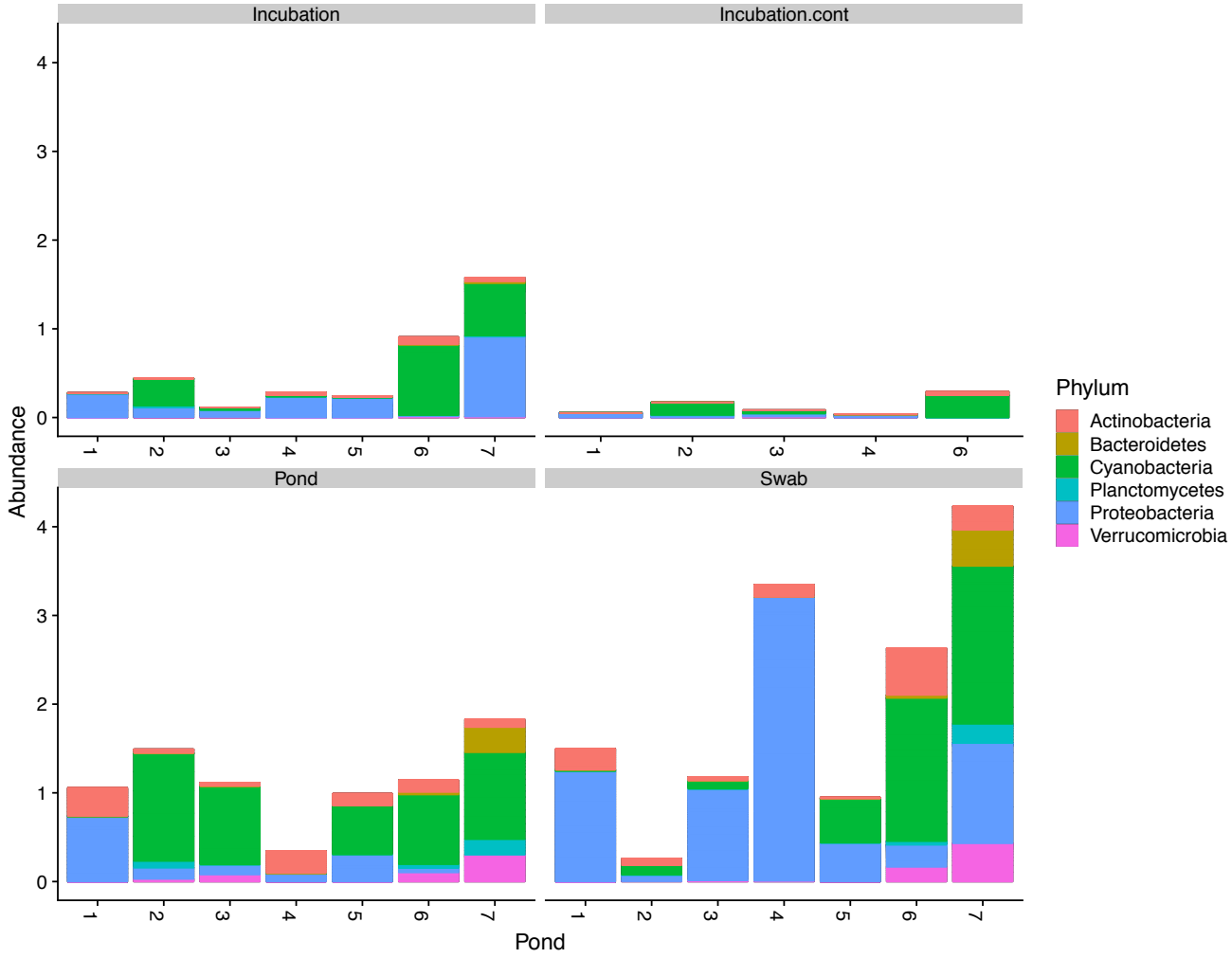
Community composition was performed at three selected taxonomic ranks: phylum, family, and genus. In general, all sample types from different geographical locations in Malawi seemed to compose of similar communities at phylum level as observed in Figure 10. Differences were more apparent at lower taxonomic classification. Proteobacteria showed a high percentage in the incubation samples, representing (52.57%), with similar numbers obtained from swab sample at (50.41%), compared to (28.28%) in rearing water samples, while incubation control samples were (43.57%). Cyanobacteria was highest in pond samples (22.18%), incubations (13.86%), incubation control samples (13.73%) and swab samples (14.44%). Verrucomicrobia was represented by (2%) from incubation samples, (2.23%) incubation control, (6.42%) rearing water, and (3.39%) for swab samples.

At family level, *Burkholderiaceae* was captured best by swab samples, shown in Figure 11, at (4.95%), and (1.97%) in rearing water samples and rarely appeared in the incubation samples and incubation control samples (0.6%). At higher taxonomic resolution - genus level, it was noted that *Ralstonia* was highly represented in incubation samples (3.81%) in comparison to swab samples (1.01%) and much less in rearing water samples (0.08%) and absent in incubation control samples. *Methylocystis*; another member of Proteobacteria, was found to be higher in rearing water samples (4.68%), incubations (0.98%), incubation control (0.9%) and swab samples (1.35%). While incubation control samples were enriched with *Synechococcus* (9.15%), when compared to incubation samples (1.06), swabs (1.16%) and rearing water samples (2.5%). *Shingomonas* was higher in swabs (4.29%), incubations (1.57%), incubation control (0.15%), and very low in the rearing water (0.008%). While *Polynucleobacter* was present in all sample types; (2.5%) for swabs, (0.54%) for incubations, (0.37%) for incubation controls, but highest captured in the rearing water samples (2.94%). Interestingly, *Cetobacterium* was also found in a higher percentage in the swab and incubation samples, (1.62%), (1.75%) respectively, with only (0.04%) in the incubation control samples and (0.23%) in the rearing water.

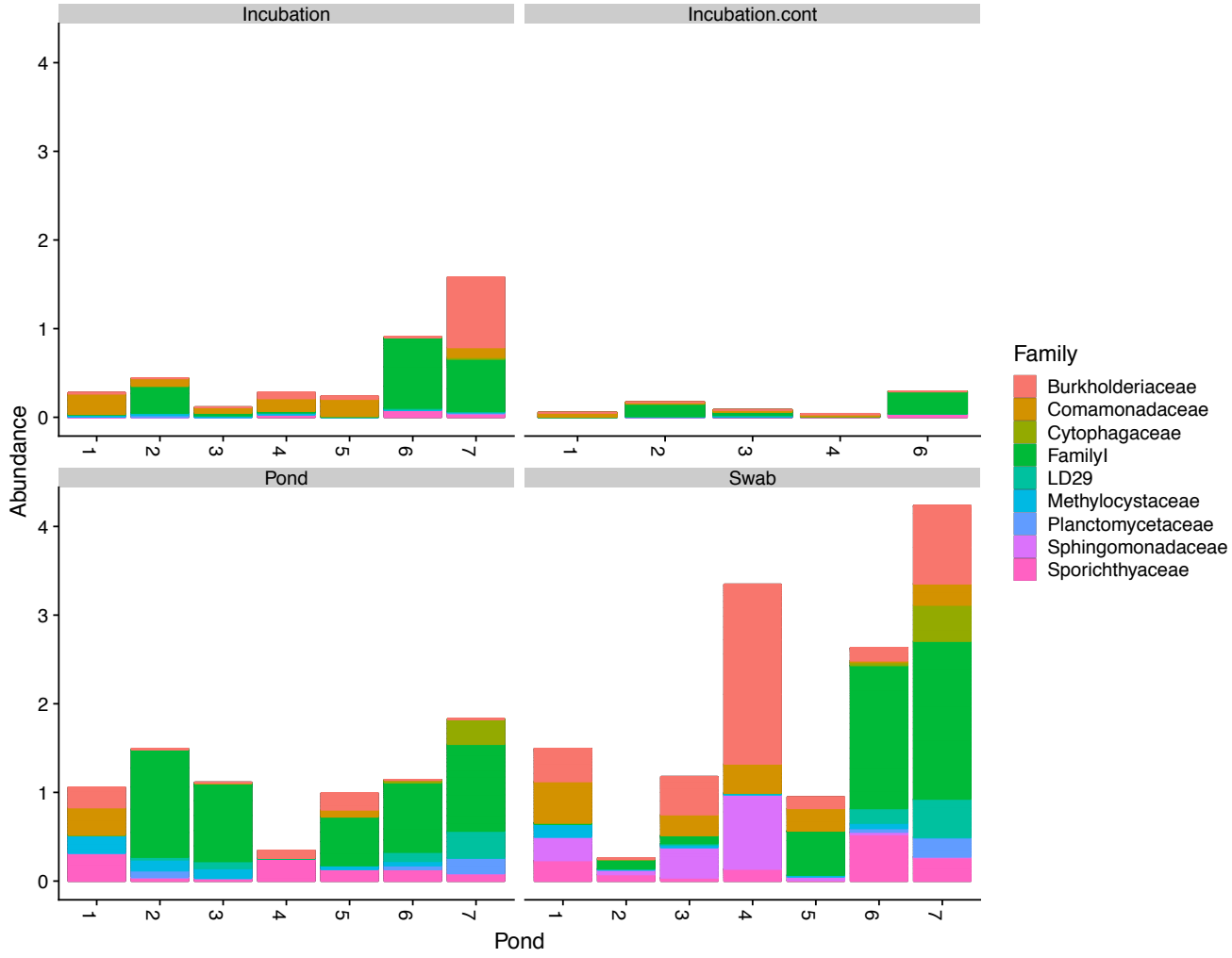
In the samples from Maldeco, *Synechococcus,* an abundant member of the Cyanobacterial genus, represented (5.06%), but was extremely low in samples obtained from Blantyre (0.55%). Cytophagaceae was also noticeably higher in

Maldeco samples (3.1%), compared to (1.47%) from Blantyre. On the other hand, *Shingomonas* was better represented in the samples collected from Blantyre (1.21%) when compared to the ones collected from Maldeco (0.07%).

(a)


Incubation
Incubation.cont
Pond
Swab
Abundance
Pond
Phylum
Actinobacteria
Bacteroidetes
Cyanobacteria
Planctomycetes
Proteobacteria
Verrucomicrobia

(b)


Incubation
Incubation.cont
Pond
Swab
Abundance
Pond
Family
Burkholderiaceae
Comamonadaceae
Cytophagaceae
FamilyI
LD29
Methylocystaceae
Planctomycetaceae
Sphingomonadaceae
Sporichthyaceae

(c)

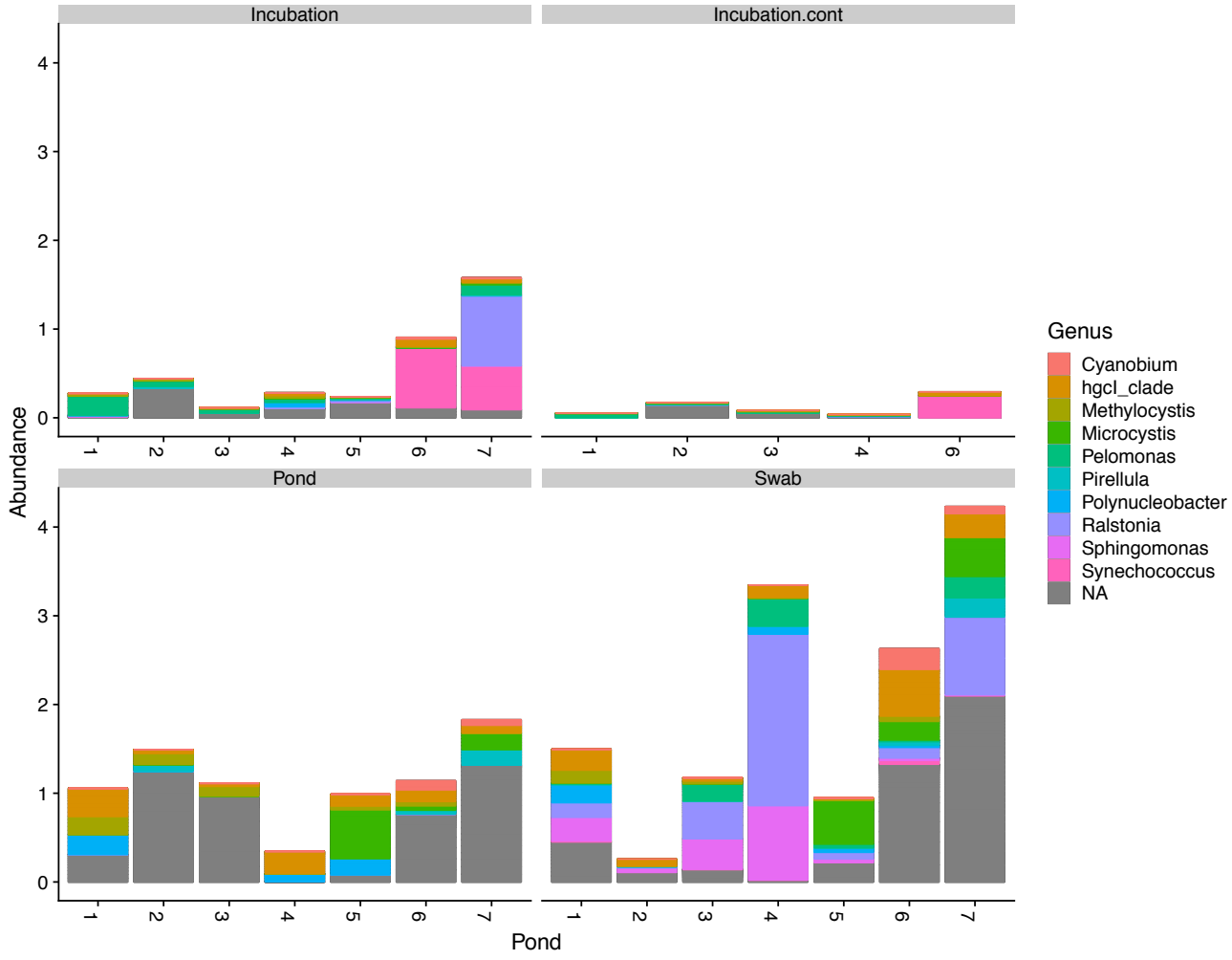
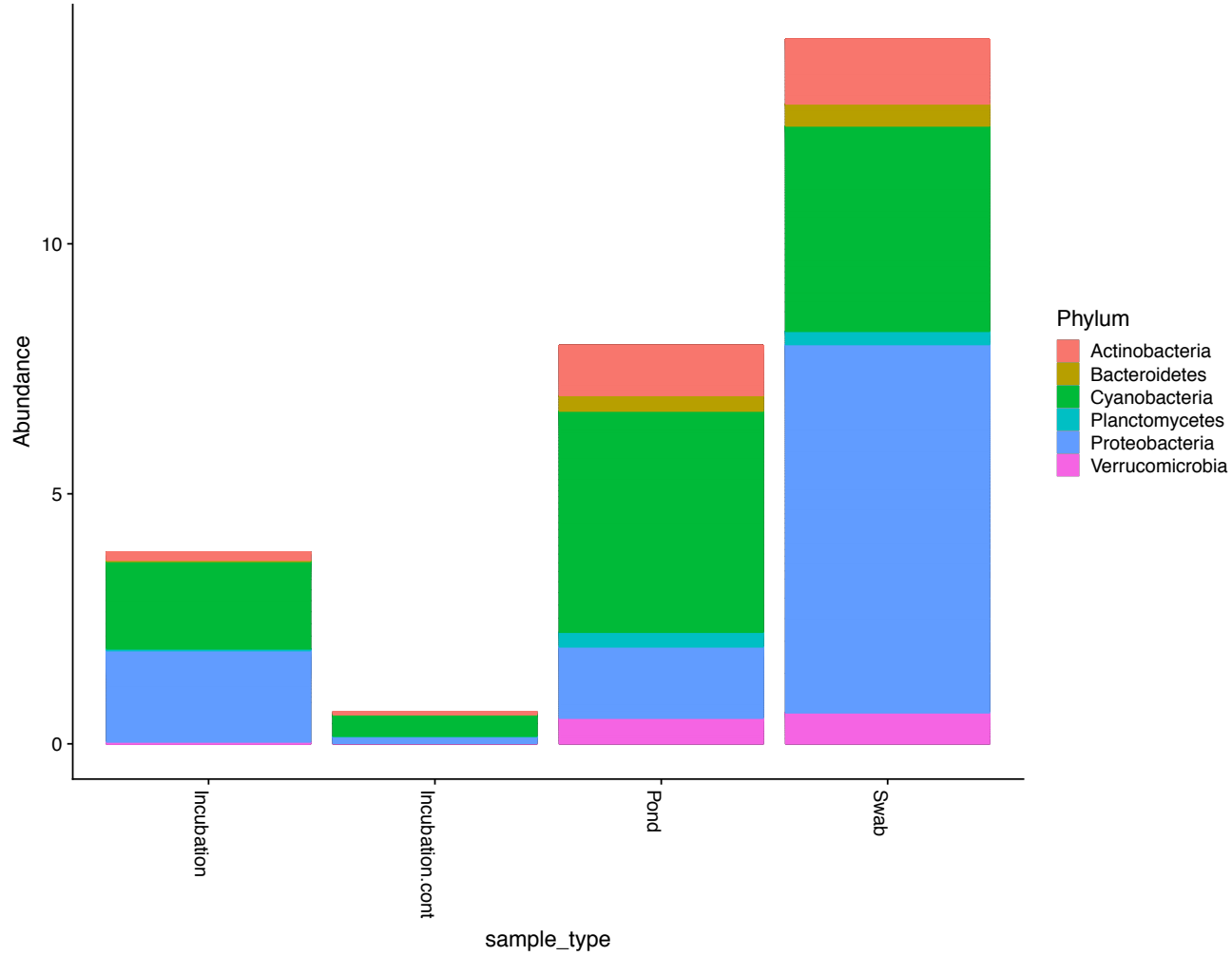


Figure 10. Skin microbiota taxonomic profile using short reads. Only top 20 most abundant taxa included in all sample types in seven different ponds. Bar plots of the most abundant taxa on all samples sequenced with V4 primers in short-reads **(a)** at Phylum level; **(b)** at Family level; and **(c)** at Genus level.

(a)


Abundance
10
5
0
Incubation
Incubation.cont
Pond
Swab
sample_type
Phylum
Actinobacteria
Bacteroidetes
Cyanobacteria
Planctomycetes
Proteobacteria
Verrucomicrobia

(b)

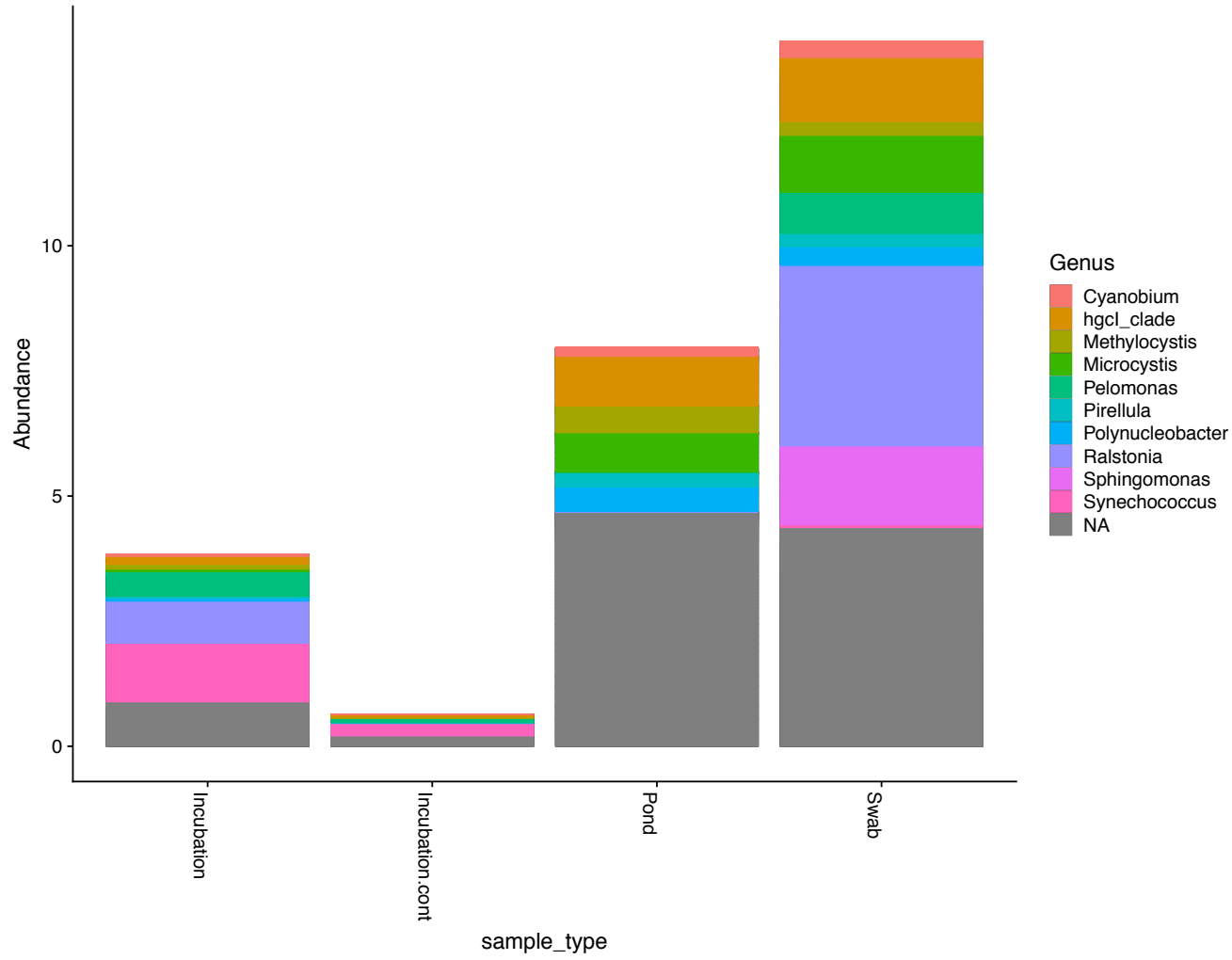


Figure 11. Taxonomic profile of each sample at different classification levels: a. Phylum: some similarity across sampling types. **B.** Genus: some taxa missing in rearing water samples, and incubation control samples.

3.1.5. Assessing sampling types through capturing microbial diversity

Early analysis suggested that there were differences in the Cyanobacterial communities between swab and other sample types, raising the possibility of autotrophy specifically associated with the epidermal mucosal layer. Cyanobacteria were selected to identify if their abundance differed according to sampling methods, as they were extremely low in incubation samples. The percentage of the Cyanobacteria for the top 10 ASVs for each sample type was 49.5%, 44.5% and 5.5% in rearing water, swab and incubation samples, respectively.

ASV	*Swabs*	*Ponds*	*Incubations+ Incubation cont.*	*Total*
ASV1	26,703	30,171	2,451	59,325
ASV2	12,467	6,158	675	19,300
ASV3	3,213	10,296	1224	14,733
ASV4	295	1,020	1,132	2,447
ASV6	3,904	4,412	459	8,775
Total	46,582 (44.5%)	52,057 (49.7%)	5,841 (5.5%)	104,580

Table 9. Abundance of Cyanobacteria in each sample type. This table shows the number of reads for each sample type, and their percentage. ASV numbering was according to the top 10 phyla in Table 7.

A presence/absence matrix of cyanobacterial ASVs was created as an initial screen to see if there were any ASVs that were found exclusively in swab samples, but absent in pond (rearing water) samples. The initial observations revealed that all ASVs were present in pond and swab samples, and none were exclusive to skin microbial communities. It is noted that many strains belonging to the phylum Cyanobacteria were frequently observed in the swab samples, and in some cases twice as many in comparison to the rearing water (pond) samples. In contrast, the incubation samples which should represent the communities associated with the mucus of the fish, had few or no Cyanobacteria, as shown in Table 10. Further Specific ASVs were selected according to their low diversity or absence in incubation samples, in comparison to other sampling types, and classification of these selected ASVs was obtained using the **SILVA** Incremental Aligner, (**SINA v1.2.11**) (Pruesse et al. 2012).

ASV	Classification	Incub	Pond	Swab
Cyano ASV 4	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	4	10	19
Cyano ASV 6	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	5	10	19
Cyano ASV 7	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	3	10	19
Cyano ASV 10	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	2	12	12
Cyano ASV 11	Cyanobacteria; Oxyphotobacteria; Synechococcales, Prochlorotrichaceae; prochlorothrix; PCC-9006	0	2	9
Cyano ASV 12	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	3	12	21
Cyano ASV 13	Cyanobacteria, Oxyphotobacteria; Nostocales; Microcystaceae; Synechocystis; CCALA 700;	5	17	17
Cyano ASV 14	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	1	12	10
Cyano ASV 16	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	3	9	10
Cyano ASV 19	Cyanobacteria, Oxyphotobacteria; Synechococcales;	0	10	9

	Cyanobiaceae; Cyanobium; PCC-6307			
Cyano ASV 20	Cyanobacteria; Oxyphotobacteria; Pseudanabaenales; Psuedanabaenaceae; Pseudanabaena; PCC-7429	2	10	10
Cyano ASV 23	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	2	5	13
Cyano ASV 24	Unclassified	1	5	8
Cyano ASV 28	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	0	7	10
Cyano ASV 30	Cyanobacteria, Oxyphotobacteria; Nostocales; Gloeocapsaceae; Gleocapsa;	0	7	11
Cyano ASV 37	Cyanobacteria, Oxyphotobacteria; Synechococcales; Cyanobiaceae; Cyanobium; PCC-6307	0	4	8
Cyano ASV 42	Cyanobacteria, Oxyphotobacteria; Leptolyngbyaceae; Leptolynbyaceae; Limnolyngbya; CHAB4449	2	6	13

Table 10. The number of samples in which Cyanobacteria ASVs were observed. The samples chosen from short read v4 16S rRNA-Illumina aligned using SINA against the SILVA 16S rRNA database. Cyanobacteria were highly represented in swab samples. Incub: incubation samples; Pond (rearing water samples).

Fish skin mucus acts as the first line of physical defence against pathogens (Wang et al. 2011). Its major function is entrapment and sloughing of microbes and it contains innate immune components, acting as a barrier between fish and its immediate niche (Dash et al. 2018).

Exploring the microbial community through incubation sampling has been assessed by examining Cyanobacteria as one of the top six most abundant phyla found in tilapia microbiomes. Earlier analysis of the percentage of cyanobacteria reads showed very few when sampling through incubation, in comparison to the other sampling methods. It was initially suspected that the presence/absence data revealed that most of the ASV assigned to Cyanobacteria, were repetitively seen in swabs and less in rearing water samples, and in some cases, completely absent in the incubation samples. Although the amount of ASVs present in the incubation samples, should reflect the communities present in the fish microbiome community obtained by the swabs, but this was not observed and swabs captured host specific and more diverse communities. A theory behind this reduction of microbial communities in the incubation samples, might be due to the high production of innate immune components within the mucus, due to stressing the fish while sampling. Such components include antimicrobial peptides (AMPs), lectines, lysozymes and proteases, which will be secreted to the environment; the surrounding water in the bucket, as epidermal mucus in fish provides an innate and fast acting protection which is non-specific (Dash et al. 2018). The time required to perform an incubation sample is ~20 minutes. One possibility is that within that time, the microbial community structure is adversely altered by components of the fish innate immune response. Lysosymes and proteases would result in cell lysis and thus loss of DNA from the filtered samples. Another possibility is as the source water used for this method (labelled as 'incubation controls' above) has lower microbial counts, there would be significant dilution of the microbiome. Thus, the filtering of 2-3 ml of incubation water per sample to maintain consistency with pond samples, was insufficient to accurately capture the fish skin microbiome. Since there was inconsistency in the similarity of the incubation samples to their incubation controls, it would be ideal to use a defined sterile medium as incubation background water for sampling, to reduce variance. These results would also suggest filtering greater volumes of water during incubation sampling may increase the Bray-Curtis distances between incubations and their controls.

Cyanobacteria are rarely reported as inhabitants of fish skin microbiome (Larsen et al. 2015), and there was no evidence from our results of skin-specific Cyanobacterial species, as all ASVs present in skin swab samples, were also seen in the rearing water

samples of these fish. Therefore, the cyanobacterial communities identified in this study are likely transients from the surrounding water and their increased abundance in our samples may be due to cyanobacterial blooms. These blooms may have far-reaching ecological effects on aquatic ecosystems, and their formation typically could lead to massive fish kills, as it causes reduction of light penetrating through the water column, and an increase in pH due to carbon dioxide depletion(Demeke 2016). Nevertheless, Cyanobacteria can produce toxic metabolites known as cyanotoxins. Common and frequently investigated cyanotoxins include microcystins, nodularin and saxitoxins. Histopathological analyses of fish grown in fishponds with cyanotoxin production showed damage to the liver, kidney, gills, intestines and muscle tissues (Drobac et al. 2016).

Some tilapia species such as Nile tilapia (*Oreochromis niloticus*) have been selected for aquaculture due to their ability to consume cyanobacteria, and its effectiveness in cyanobacterial bloom control. The stomach of the Nile tilapia is known to have an extremely low pH, therefore enhanced damage to ingested cyanobacteria is increased, and may be causing the reduction of their presence in rearing water samples (Demeke 2016). It is also likely that the increase of cyanobacteria in swab samples, may be the result of the reduction of light penetrating through the water, causing them to utilise mixotrophy with the presence of essential organic and inorganic compounds in the mucus(Wan et al. 2015), and become part of these rich mucosal niches.

However, in some cases these cyanobacteria blooms are dominated by colonial and filamentous forms, which may cause mechanical interference and clogging of gills in fish (Bwanika et al. 2006)(Demeke 2016). Our community composition results showed the presence of *Microcystis*, in both geographical sampling sites; Maldeco and Blantyre. *Microcystis* is the most common bloom forming unicellular cyanobacterium in ponds, and it has been reported worldwide in animal poisoning (Kumar and Sinha 2014). Other cyanobacteria were represented by "*Family I*" and this family has usually been referred to *Synechococcus*, *Microcystis* (which have already been seen in the top 20 taxa in this work) and *Anabeana* (Gołębiewski et al. 2017). The absence of Cyanobacteria in incubation samples in this study may suggest that, assessing mucus communities through incubation sampling may not reflect the actual skin microbiome

of fish, as this practice may have its impact on microbial communities in the epidermal mucosa of fish resulting from stress. In comparison, swabbing body mucus of fish may provide a less invasive method of DNA collection, making it less stressful for fish and less likely to adversely alter the microbiome (Le Vin et al. 2011). Within the scientific community there is an increased awareness that fish may experience pain, stress, or lasting harm as a consequence of invasive procedures, driving the development of alternative methods (Breacker et al. 2017).

From our study, we were able to collect significant amounts of DNA and sequencing yielded high numbers of reads for swab samples, even greater than those obtained from incubation sampling. This work therefore supports the use of skin swabs as a viable alternative to incubation sampling for fish microbiome research.

It is worth mentioning that when aligning the selected ASVs to the SINA database, for the selected Cyanobacteria, it was possible to obtain a lower taxonomic level (species and strains were identified), whereas some of these taxa were only able to be assigned to genus level during taxonomic assignment by DADA2. This is most likely due to the use of k-mer abundances by DADA2 to classify sequences by supervised machine learning, compared to sequence alignment used by SINA. (Callahan et al. 2016)

To quantify the microbial diversity within each sample, alpha diversity was measured for the four sample types across the 7 ponds. This revealed that the rearing water samples showed lower diversity in comparison to the incubation and swab samples (figure 8). This is expected as the communities present in the skin and mucus are host specific and selective. Beta diversity analysis showed that the differences in community composition between different sampling locations in Malawi was typically greater than the distances between rearing water and swab sample types at the same location (Figure 9). In contrast, incubation controls and incubation samples were much more varied in their distances to other sample types from the same sample location. Importantly, incubation samples did not cluster together, suggesting that the microbiomes derived from these samples were highly variable. ANOSIM, used to examine statistical differences between sample types, confirmed our NMDS plot results in figure 9; picking up the similarity between incubation and swab samples; showing higher resemblance compared to the rearing water samples. This trend was expected as the incubation samples contain the same communities present in the

mucus captured by the swabs sampling the skin. ANOSIM tested the null hypothesis that the average rank similarity between microbial communities within a group are the same as the average rank similarity between communities between groups. These results reject the null hypothesis that there is no difference between rearing water and swab samples, as well as the differences between incubation samples and rearing water samples. On the contrary, we cannot reject the null hypothesis for the swab samples and incubation samples as they showed less difference, thus capturing similar communities. This is evidence that swabs are a good alternative to incubations, since no significant difference was found between swabs and incubations.

As samples were collected from two different locations in Malawi, Blantyre (ponds 1-5) and Maldeco (ponds 6 & 7), there is evidence that the microbial composition of the fish skin is highly influenced by the geographical locations, and these findings clearly demonstrate a geographical discrimination between clusters from ponds located in Blantyre to the ones in Maldeco.

With the presence of structural differences between the microbiomes from different sample types, SIMPER analysis was employed to detect the largest taxa contributors to the dissimilarity between sample types. Briefly, SIMPER analysis explains which taxa are most responsible for the Bray-Curtis distance between the two communities. The overall dissimilarity between sample types was attributed to three major phyla: Proteobacteria, Cyanobacteria, and Bacteroidetes. This finding assists us in understanding the differences between microbial communities, with dissimilarities as high as 91% between incubation samples and water samples. It was able to reveal that these microbial communities showed a dissimilarity according to sampling methods. This finding was expected, for the water used for the incubation sampling method (spring water) was different to the rearing water of the fish ponds. The contribution to each phyla is shown in table 11. Since some of the different ASVs in our taxa belonged to the same phyla, broader scale taxonomic resolution shows similarity at phylum level, but when examined at ASV level, a higher discrimination is shown between microbiomes according to sampling method. This fine scale shows similarity at phylum level, but when examined at ASV level, a higher discrimination is shown between microbiomes according to sampling method. This fine scale resolution may suggest having different ecotypes, which are distinct genetic lineages with

ecologically relevant physiological differences, but are typically defined as a single species on the basis of their DNA similarity (Rocap et al. 2003), and may allow for greater fitness within an ecological niche. This concept gives weight to common taxonomic properties between members of the same species, but differentiates based upon small changes in gene content or expression (Boon et al. 2014).

The following table (Table 11), is a summary of the top 20 taxa representing the fish bacterial microbiome communities across all sample types and their previous selected studies in fish microbiomes.

ASV	*Phylum*	*Family/Genus*	*Description*	*Ref.*
ASV 1 ***ASV 3*** ***ASV 4***	**Cyanobacteria**	**Family I**	**Cyanobacteria blooms- leads to massive fish kills.**	**(Demeke 2016)**
ASV 2	**Cyanobacteria**	***Microcystis***	**Toxic activity Inhibits gill activity**	**(Kumar and Sinha 2014) (Drobac et al. 2016)**
ASV 5	**Proteobacteria**	***Ralstonia***	**PHB* Immunity**	**(Van Hung et al. 2019)**
ASV 6	**Cyanobacteria**	***Synechococcus***	**Found in well–lit surface waters**	**(Reverter et al. 2017)**
ASV 7 ***ASV 17*** ***ASV 20***	**Actinobacteria**	**Hgcl clade**	**Improves skin microbiota in humans**	**(Lee et al. 2019)**
ASV 8	**Bacteroidetes**	**Cytophagaceae**	**Symbiotic bacteria- impose positive effects on aquatic animal**	**(Chen et al. 2017)**

ASV 9	**Proteobacteria**	***Methylocystis***	**PHB* production**	**(García-Pérez et al. 2018) (Franke et al. 2017)**
ASV 10	**Proteobacteria**	***Polynucleobacter***	**Associated with fish skin and gills microbiomes**	**(Rosado et al. 2019)**
ASV 11	**Proteobacteria**	***Sphingomonas***	**Found in gut microbiota of healthy and unhealthy fish**	**(Wang et al. 2018)**
ASV 12 ASV 19	**Verrucomicrobia**	**LD29**	**Associated with Cyanobacteria blooms in ponds; Found on salmon**	**(Lokesh and Kiron 2016)**
ASV 13	**Planctomycetes**	***Pirellula***	**Marine bacterium**	**(Givens et al. 2015)**
ASV 14	**Proteobacteria**	***Pelomons***	**Core gut microbiome in fish**	**(Nikouli et al. 2018)**
ASV 15 ASV 18	**Proteobacteria**	**Commomanoda-ceae,**	**Associated with fish skin microbiomes**	**(Carlson et al. 2017)**
ASV 16	**Actinobacteria**	**Hgcl clade**	**Improves skin microbiomes in humans.**	**(Lee et al. 2019)**

Table 11. A summary of the top 20 taxa representing the bacterial microbiome communities in all sample types and their previous selected studies in fish microbiomes. *PHB: poly-β-hydroxybutyrate

** **Family I** classification of Cyanobacteria, when identified in the long reads it has shown to possibly belong to the following lower taxa (strains): MTP1, Leptolyngbya VRUC 135, CENA359, Phormidium SAG 37.90. - equivalent to Nostocaceae in the Greengenes database. (Aschenbrenner et al. 2017)

Proteobacteria is generally the predominant phylum in the skin microbiome of teleosts. Importantly, several skin Proteobacteria isolates from salmonids have been shown to have inhibitory effects against bacterial and fungal pathogens (Reid et al. 2017). *Methylocystis,* has also been known for its high production of PHB and was seen much more abundant in rearing water samples. Some species in this phylum are known fish pathogens, while some of them could be part of a healthy skin microbiome of the fish (Lokesh and Kiron 2016).

The presence of *Cetobacterium* is an indication that faecal microbial communities were present in the surrounding environment. Finding *Cetobacterium* in relatively high abundance as part of the skin microbiome is suspected, since the vast majority of reports for this genus have typically been associated with the gastrointestinal tracts of other fish. *Cetobacterium* has been known to be an indigenous bacterium in the intestinal tract of freshwater fish including goldfish, common carp, tilapia and ayu (Tsuchiya et al. 2008). Yet interestingly, their percentages were high in swab samples and much lower in pond samples. *Cetobacterium,* a protease producing bacteria, is capable of producing high amounts of vitamin B12 and also inhibiting growth of some other bacterial taxa. *Cetobacterium* is microaerotolerant thus fish skin is unlikely to be a natural habitat for this genus. Although most reports of this genus have been associated with the gastrointestinal tracts of various organisms, *Cetobacterium* has also been reported in small percentages in earthen pond water samples from an aquaculture facility in China (Arias et al. 2019).

Ralstonia was much higher in incubation and swab samples compared to other sampling methods, as they were rarely seen in the pond water, and not present in the inlet source water (incubation controls), which may suggest this genus is possibly host specific. A recent study has identified the importance of the presence of *Ralstonia* species, as they are rich in poly-β hydroxybutyrate (PHB) and may increase fish

immunity (Van Hung et al. 2019). Similarly, *Methylocystis*, from the phylum Proteobacteria, was identified as one of the top 20 phyla, is also known for its production of PHB (García-Pérez et al. 2018), with a higher percentage in rearing water samples. Other important members of Proteobacteria were *Pseudomonas*, *Burkholderia* and *Sphingomonas*. The latter two have been identified in healthy fish, while *Pseudomonas* is known to be a common residence of the skin microbiome, usually in high abundance (Larsen et al. 2015). Finally, Bacteroidetes was represented by *Cytophagaceae*, a bacterium frequently identified as a symbiotic microbe, and might impose positive effects on aquatic animals (Chen et al. 2017).

To sum up, the microbial communities sampled in this study contained previously identified members of fish microbiomes, as well as some putative pathogens in low abundance. This was supported by the absence of any signs of fish disease in the sample sites and supports the hypothesis of a 'healthy' microbiome containing low numbers of present pathogens.

4. Results and discussion from full length 16S MinION sequencing

4.1. Amplification of full length 16S rRNA and clean-up

To show that real-time genomic surveillance is possible in resource-limited settings and can be deployed rapidly to monitor outbreaks, a comparison was made to assess the use of MinION sequencing in capturing similar microbial communities to Illumina sequencing. To compare the number of reads obtained by different types of sequencing, 11 samples were selected for long read sequencing from the same samples previously run by short read sequencing, shown in Table 12. This selection was made from ponds located in Maldeco, and was based on available DNA quantities across incubation, swab and rearing water (pond) sample types.

Amplification of full-length 16S rRNA sequences was carried out, with primers targeting the V1-V9 hypervariable regions of the 16S rRNA bacterial gene. The PCR reaction included a no template control (NTC) sample, which did not amplify, as seen in Figure 12. Following every step of purification of the PCR product a quantification of the DNA was made using the Qubit fluorometer. The final DNA quantified from all the pooled samples was 423ng.

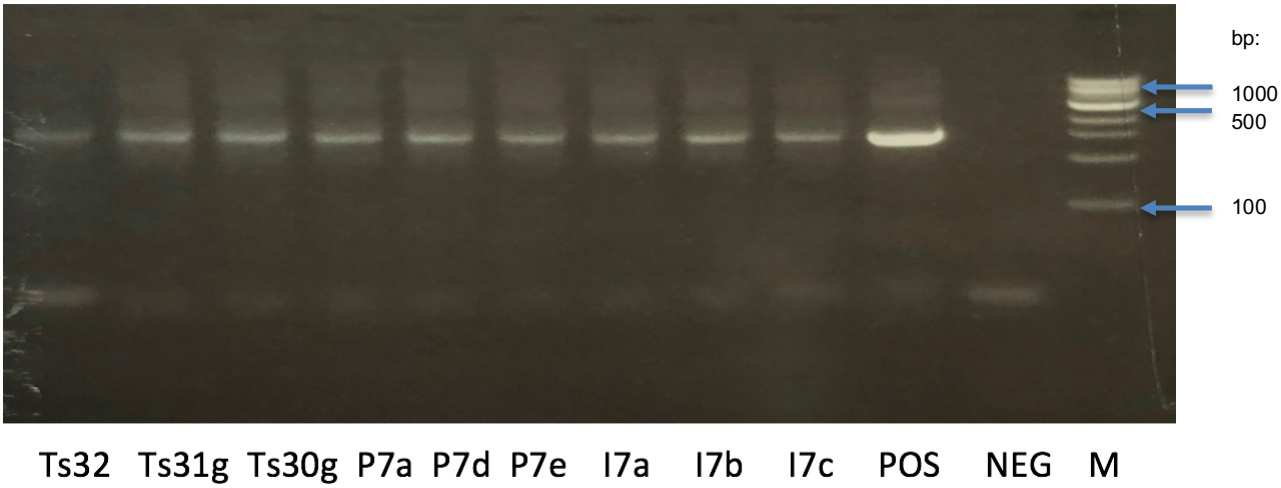


Figure 12. Gel electrophoresis analysis of PCR-amplification of the full length 16S rRNA gene fragments. Lines correspond to nine of the selected samples,

starting from the left; followed by two lanes representing an *E.coli* positive control and ; a negative control, and M denotes the 100bp ladder.

Sample	*Qubit Prior PCR Clean-up*	*Qubit ng/µL #1*	*Qubit ng/µL #2*	*Bar code #*
MLW_T32	**5.02**	**2.20**	**x**	**NB01**
MLW_T31g	**21.6**	**26.6**	**10.3**	**NB02**
MLW_T30g	**12.5**	**10.7**	**3.88**	**NB03**
MLW_P7a	**15.9**	**11.8**	**3.98**	**NB04**
MLW_ P7d	**50.6**	**41.8**	**21.4**	**NB05**
MLW_ P7e	**13.8**	**8.3**	**7.22**	**NB06**
MLW_I7a	**18.4**	**8.10**	**5.14**	**NB07**
MLW_I7b	**25**	**14.5**	**6.38**	**NB08**
MLW_I7c	**16.2**	**7.14**	**4.02**	**NB09**
NEG cont.	**x**	**x**	**x**	**NB10**
MLW_I6a	**7.5**	**13.4**	**3.80**	**NB11**
MLW_I6b	**21.4**	**25.4**	**11.9**	**NB12**

Table 12. DNA concentration following PCR clean-up for the 11 selected samples for MinION sequencing. Representing the amount of DNA quantified following each clean-up for amplified samples with their different assigned barcodes. Qubit #1: first DNA quantification; Qubit #2: second DNA quantification.

4.2. Sequencing analysis

A total of 3,438,234 reads were obtained after a single flowcell of MinION sequencing, of which 1,687,909 reads were assigned a taxonomy using SINA (Pruesse et al. 2012) against the SILVA v.135 16S rRNA database. The number of reads that were not assigned to a taxonomy (1,750,325 Unclassified reads) was ~ 50.9%, and the number of unique taxa obtained from all samples was 1349.

For taxonomic classification, 264 out of 1349 assigned taxa were at strain level. These samples were enriched with Proteobacteria. The most abundant phyla represented by long-read sequencing were similar to those observed in short read data (experimental chapter one): Proteobacteria represented 56.1%, followed by 14.6% for Bacteroidetes, 5.4% Planctomycetes, 11.9% Actinomycetes, 8.5% Cyanobacteria, 2.5% Verrucomicrobia, and 0.7% Fusobacteria, as seen in Figure 11.

Of note was the presence of known pathogenic bacteria in the long-read data: *Salmonella* in sample **MLW_I7b**, Shigella in sample **MLW_T30g**, and *Brucella* in sample **MLW_T31g**. The presence of *Shigella* was not detected in the short-read analysis conducted by Illumina. 20 different members of the family Clostridiaceae were represented in long read data.

4.3. Statistical analysis

The reads for each sample were summed and normalised, to calculate Shannon diversity for the long-read data. This was measured through R version 3.5.1, excluding the negative control. The mean Shannon diversity was calculated as seen in Table 13 for the remaining eleven samples and was 2.33.

Sample	*Total reads*	*Shannon Diversity*
MLW_T32	30715	2.213
MLW_T31g	211447	2.749
MLW_T30g	85766	2.654
MLW_P7a	111262	2.432
MLW_ P7d	6867	2.520
MLW_ P7e	15568	2.585
MLW_I7a	65403	2.083
MLW_I7b	53972	2.149
MLW_I7c	33115	2.117

NEG cont.	4	1.822
MLW_I6a	7626	2.134
MLW_I6b	116	2.027

Table 13. The number of reads produced and Shannon diversity for each sample. The negative control and MLW_I6b were excluded from calculating the mean shannon diversity analysis.

4.4. A comparison between long reads and short read, targeting the 16S rRNA gene.

The two most common specifications used to compare platforms are the number of reads produced in a given instrument run and the length of those reads (Levy and Myers 2016). In this study, we utilised Nanopore Sequencing technologies to assess the full-length 16S rRNA gene, and to retrieve increased richness estimates at higher taxonomic levels that were missed with short-reads obtained through Illumina sequencing technologies. To compare the number of reads obtained by different types of sequencing, 11 samples of long reads were selected with the same samples previously run by short read sequencing. Results shown in Figure 13, revealed that five out of the seven top phyla showed a higher percentage of reads obtained by long read sequencing.

When comparing the taxonomic information obtained by different types of sequencing, the amount of short reads that were not assigned to a taxonomy, was 24.5%, while approximately half of the reads were unclassified with MinION sequencing. On the contrary, it was possible to achieve a lower taxonomic classification using MinION sequencing than the one obtained by Illumina. In long read sequencing, 264 out of 1349 unique taxa reached strain level ~ 20% of total reads. In comparison, none of the short-read sequences reached species classification.

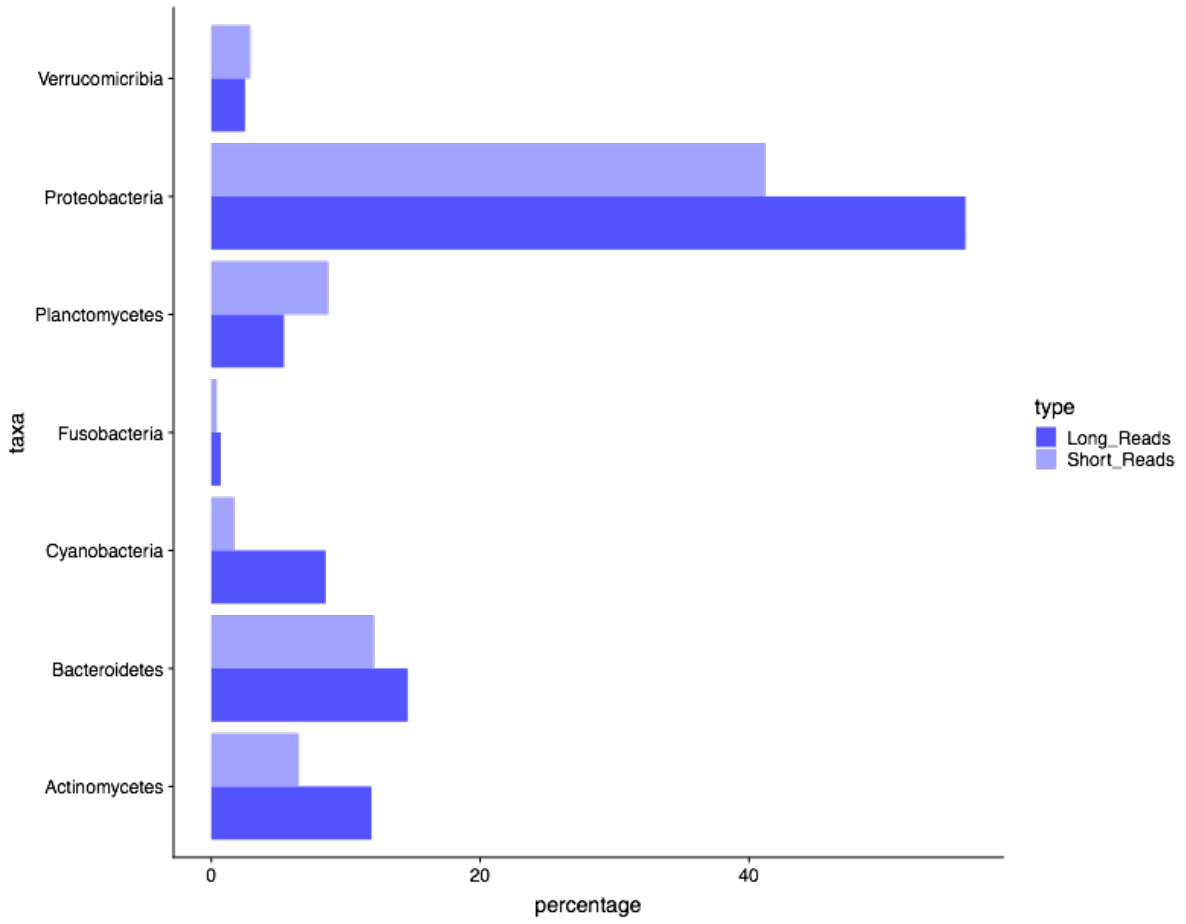


Figure 13. A comparison of the highest abundant phyla in short reads and long reads. The bar chart shows the percentage abundance between the communities captured by different sequencing methods; Illumina and the MinION.

Although short reads identified differences between tilapia skin microbiota in terms of bacterial abundances at each taxonomic level, it failed to reach species level for most the assessed samples. V4 16S rRNA gene is the most common region amplified with Illumina MiSeq for bacterial taxonomic classification, but this region fails to amplify some significant species for skin microbiota studies (Cusco et al. 2017).

The most abundant phyla represented by long-read sequencing were similar to those observed in short read data, and 5 out of 7 of these phyla showed a higher percentage of communities captured through long reads (Figure 13), with around 20% reaching strain level. Therefore, full-length 16S rRNA gene sequences could be used for accurate richness estimates especially at higher taxa, which are fundamental for microbiome studies (Cusco et al. 2017).

Some reads from our MinION sequencing analysis represented known pathogens such as *Salmonella, Shigella* and *Brucella*. We can speculate that since these taxa are low in their abundance and our samples were proved to be healthy, the stability of the microbial community could be due to the richness and evenness of the fish microbiome, which have kept these potential pathogens under control. Nevertheless, they are opportunistic and can be threatening. Therefore, sampling through incubations may increase the risk of dysbiosis, by disrupting the microbial community through the high production of innate immune components (mentioned earlier in this chapter), and facilitating the proliferation of opportunistic pathogens. Another important note will be that *Shigella* was not detected in short read analysis conducted by Illumina, indicating that short reads could be missing out the analysis of essential taxa, that are reforming these communities.

The MinION nanopore sequencer is a particularly promising technology research tool, making real-time genomic surveillance possible in resource-limited settings and can be established rapidly to monitor outbreaks and control disease, through its portability, low start-up costs, real-time data generation and straightforward application (Gallagher et al. 2018).

5. Results and discussion from T4 Bacteriophage amplification:

5.1. PCR Purification and Sequencing:

The DNA of the T4 phage lysate of was quantified using the qubit Fluorometer and 3.24 µg of DNA was used as a positive control in all subsequent attempts to derive T4-like Myoviridae amplicons from environmental samples.

Although the T4-Super primers, forward: **GAYHTIKSIGGIGTICARCCIATG** and reverse: **GCIYKIARRTCYTGIGCIARYTC** (Walters et al. 2016), generated amplicons of the appropriate size from environmental samples, their efficiency was greatly reduced once barcodes for multiplexing were added to the primers. Initially the non barcoded primers showed an obvious PCR product in all samples (selected from three different sample types), and the product size was at around 400-500bp (Figure 14).

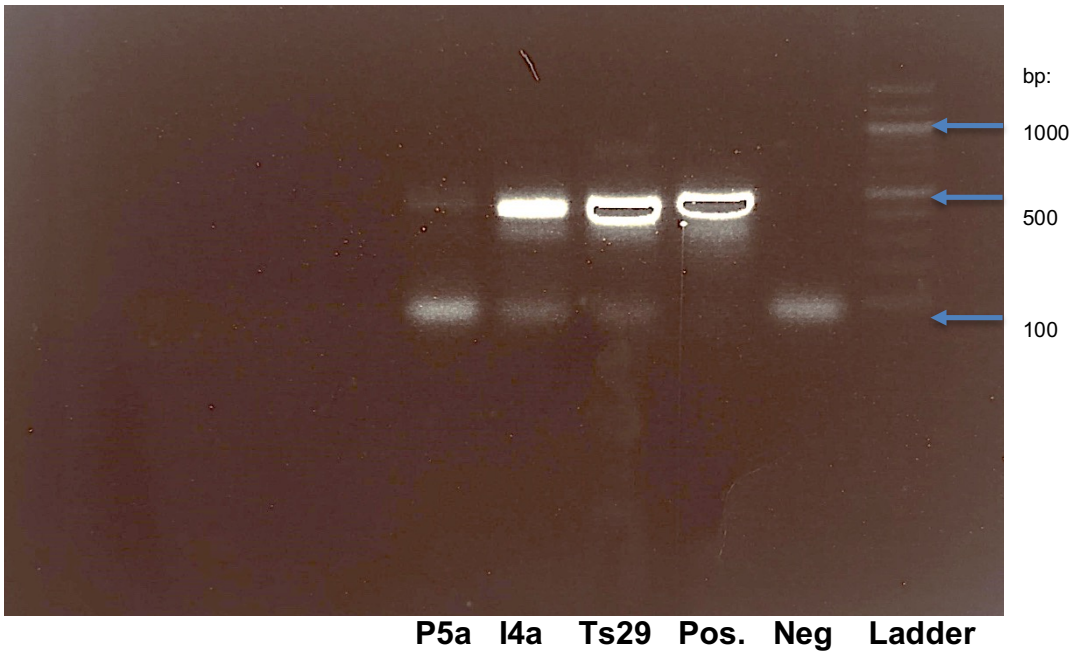


Figure 14. Gel electrophoresis analysis of PCR-amplified g23 gene fragments. The first three lanes correspond to PCR amplification of g23 gene from random, pond, incubation and swab samples respectively. The other lanes were a positive control, a negative control and a 100bp ladder.

Optimisation of PCR conditions was performed using a concentration gradient PCR from 1-10 ng/µL, and showed improved amplification at higher concentrations, as seen in Figure 15.

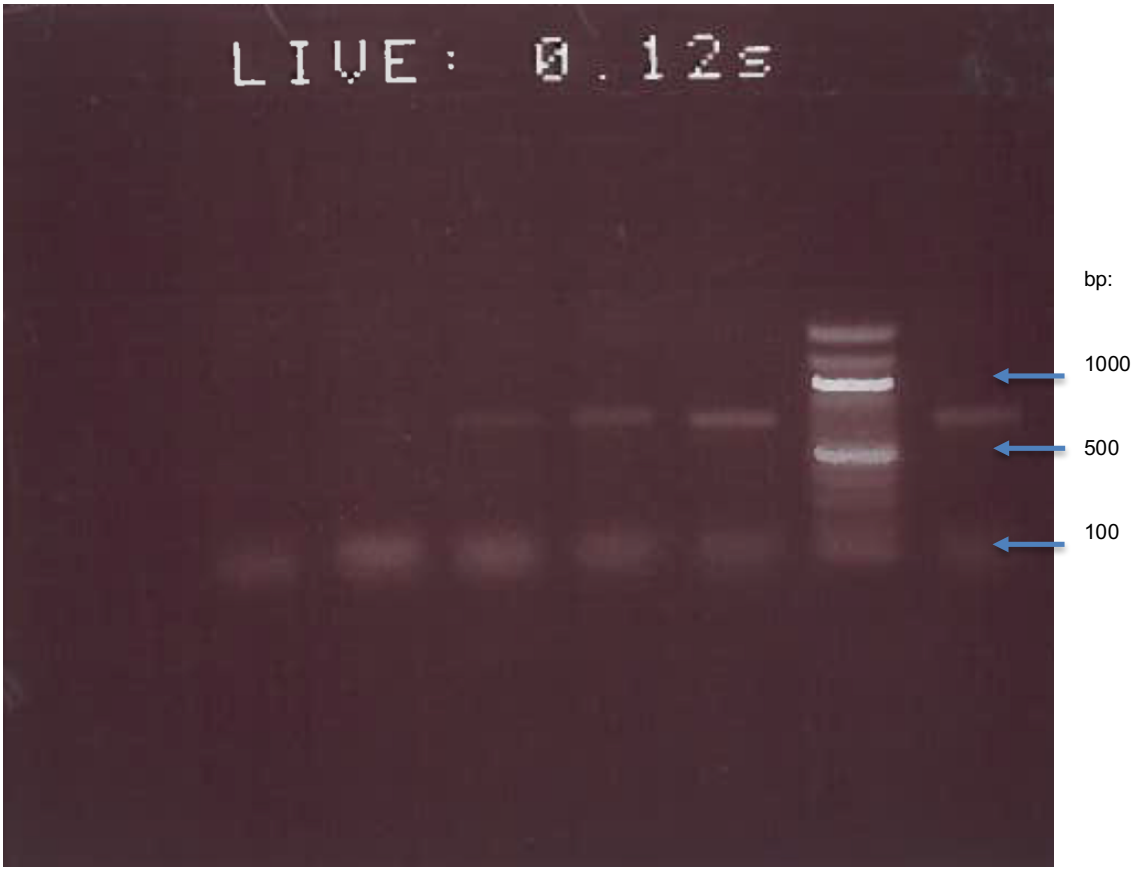


Figure 15. Gel electrophoresis analysis of PCR-amplified g23 gene fragments, showing a concentration gradient for T4, starting from the lowest concentration at the left-hand side. The 100bp ladder to the right, followed by a positive control**.**

Following addition of the barcodes, primers were evaluated again on the same environmental samples. However, only one rearing water sample (P4a) and the T4 positive control, showed amplification of a product. Visualization of the product by gel electrophoresis revealed this product to be ~100bp shorter than expected, and also showed the presence of primer dimers (Figure 16).

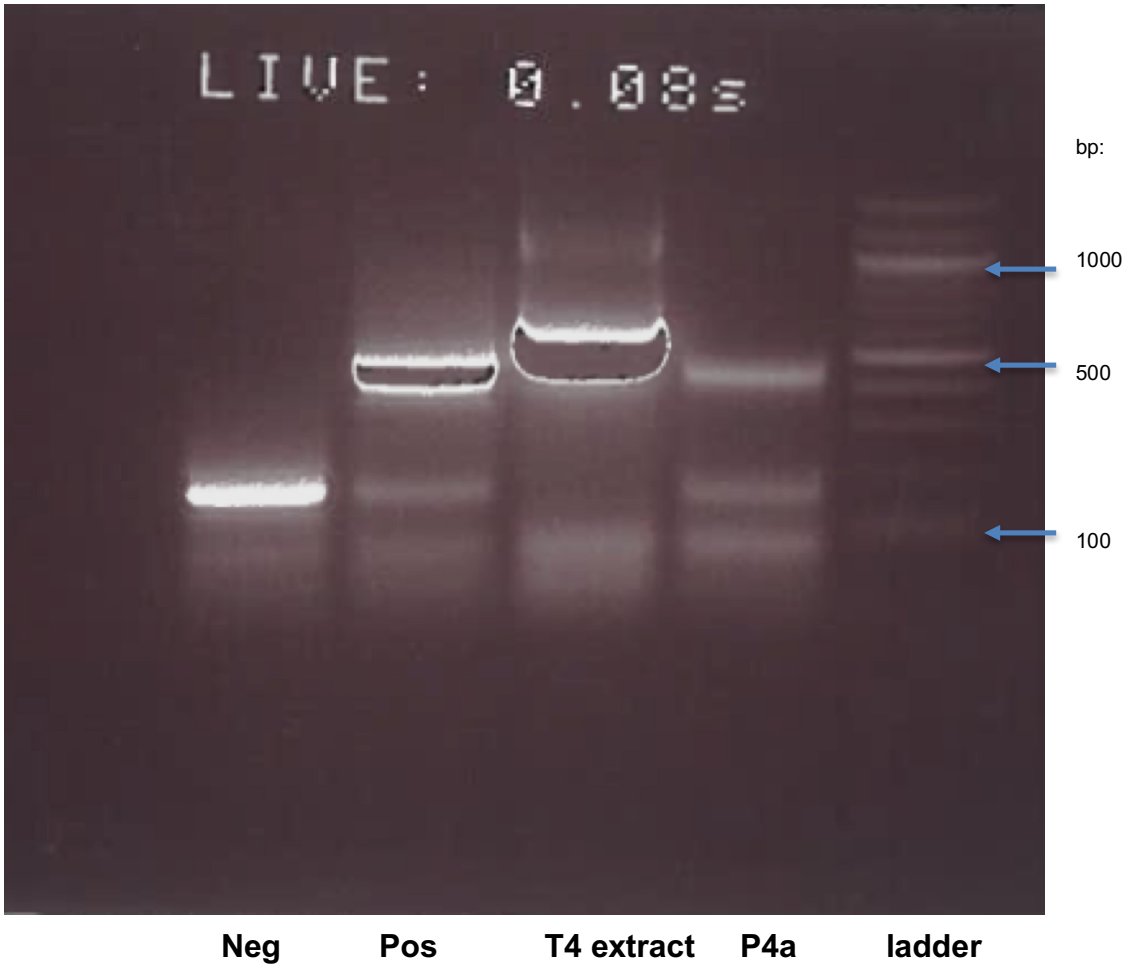


Figure 16. Gel electrophoresis analysis of PCR-amplified g23 gene fragments, using barcoded primers for T4. The first lane from the left, was a negative control, positive control and a rearing water sample (P4a), corresponding to PCR amplification of g23 gene and the 100bp ladder.

In an attempt to optimise the PCR product and reduce the amount of primer dimers present as by-products, a temperature gradient PCR was performed, from (44-60)ºC. This initially helped reduce the amount of primer dimers. The T4 DNA was amplified at the suspected location, indicating the correct size, although the product showed to be above 500bp, shown in Figure 17.

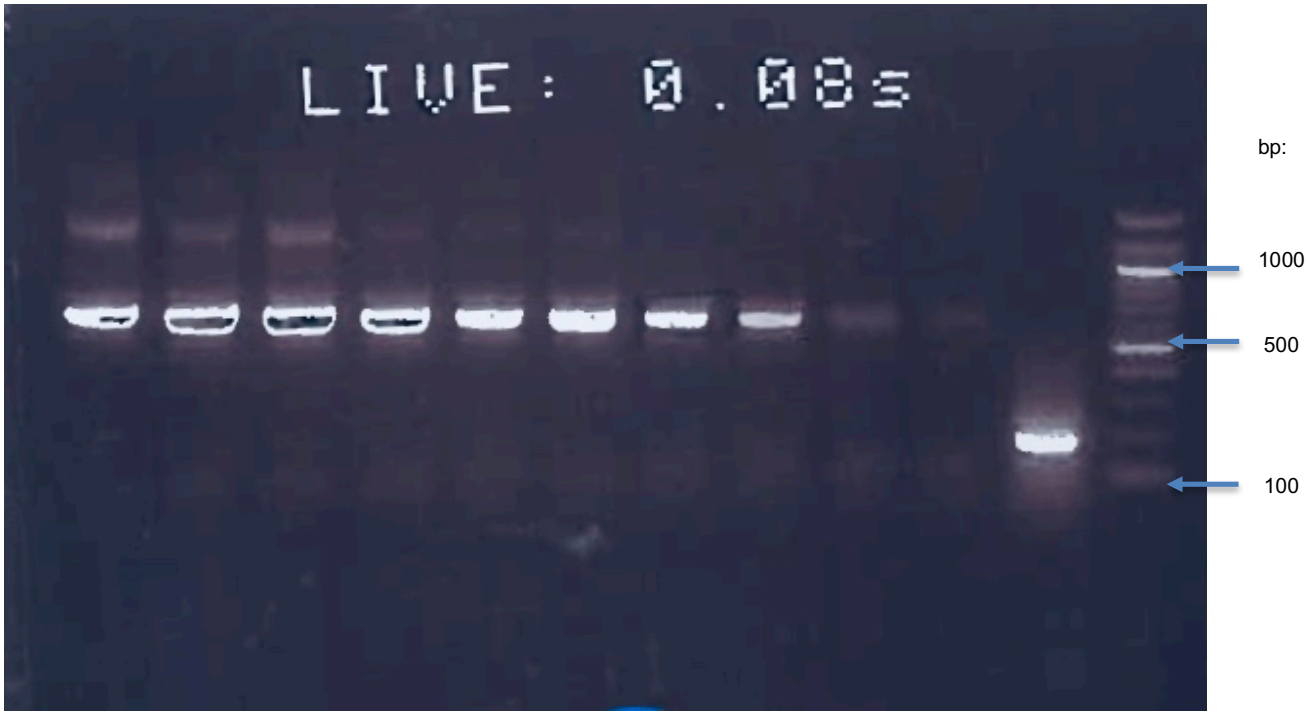
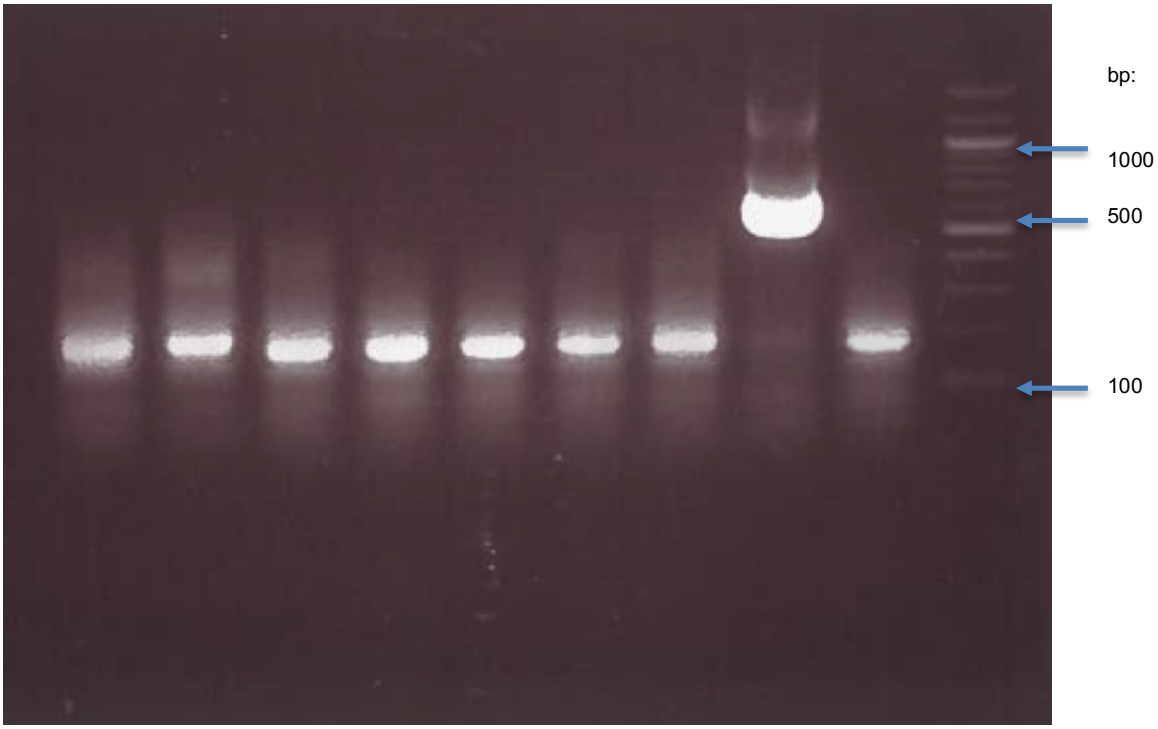


Figure 17. Gel electrophoresis analysis of PCR-amplified g23 gene fragments, for temperature gradient. Amplified bands are between 500-700 bp.

Following the temperature gradient, a PCR was performed with the chosen annealing temperature of 57.5ºC, with a primer concentration of 0.4µM, for randomly selected samples of different sample types, and the T4 positive control. This was performed several times, but no amplification of samples was detected. Light smears were shown for folds 1.4µM and 1µM, but these results, seen in figure 18, were not clear.



I5b 1µM **I5b 1.4µM** **I5b 2µM** **I5b 2.5µM** **T24 2.5µM** **Ts24g 2.5µM** **Ts23g 2.5µM** **Pos** **Neg** **ladder**

Figure 18. Electrophoresis analysis of PCR-amplified g23 gene fragments for different sample types. No bacteriophage amplification was seen, except for the clear band for the positive control (third lane from the right). Location of Primer dimers showing at around 100-200bp.

The final attempt for a concentration gradient for the barcoded primers, using the same successful sample “P4a”, and the previous annealing temperature 54ºC failed to successfully amplify a product.

The concentration and volume of the primers and DNA used in each step is explained in table 5.

5.2. Evaluating viral diversity using Amplicon Sequence Variants

Despite the difficulties with using barcoded primers, we wanted to evaluate whether a successfully amplified library would yield information about viral diversity in fish-associated microbiomes. By amplifying the viral fragment of one sample, targeting the g23 major capsid protein, we were able to process one rearing water sample P4a, with DADA2. 177,979, raw reads were initially obtained from one sample. Following the pre-processing of the reads, which included quality based trimming, paired end merging, de-replication and chimera removal, 20,315 reads were retained for further analysis; approximately 88% reads were lost after processing. Amplicon sequence variants (ASVs) were concluded, generating 1372 ASVs in total, and 296 as final unique ASVs.

The mean Shannon diversity within the one sample was calculated to be 4.97.

5.2.1. Abundance and diversity of T4-like Myoviridae phages

A phylogenetic tree was created to compare the strains within our selected rearing water sample. The phylogenetic placements were visualized on the iTOL v.4 webserver, (Letunic and Bork 2019), where the number of ASVs based on sequence identity were compared within one sample, as seen in Figure 20.

Figure 19. A phylogenetic analysis of g23 ASVs sequences. Results were obtained from the selected rearing-water sample P4a. Species names are indicated in the tips, and the purple bars represent the abundance of each ASV.

5.2.2. Sequence similarity analysis

Whilst the DADA2 software is primarily used for 16S rRNA analysis, in theory the generation of ASVs is applicable to any conserved region successfully amplified by PCR. However, to my knowledge, this is first use of DADA2 to investigate viral diversity using T4-like Myoviridae primers. The closest relatives of the *g23* sequences were examined using the Basic Local Alignment Search Tool (BLAST) search program on the NCBI website (http://www.ncbi.nlm.nih.gov/), 10 ASVs were selected randomly for the analysis against the NT database. All the tested ASVs returned a best-hit that was associated with T4-like Myoviridae; results shown in table 14.

Most of these sequences represented the major capsid protein gene, while they all belonged to the Myovirideae family. One was identified as an Enterobacteria T4 phage.

Name	*Accession no.*	*Description*
ASV1	KY685577.1	Uncultured Myoviridae clone N1-58 Major capsid protein genes
ASV 10	HM241951.1	Uncultured Myoviridae clone 1-4 Major capsid protein like (g23)
ASV 20	HM241951.1	Uncultured Myoviridae clone 1-4 Major capsid protein like (g23)
ASV 30	HM242208.1	Enterobacteria phage T4 clone s11-4 Major capsid protein like (g23)
ASV40	KY685577.1	Uncultured Myoviridae clone N1-58 Major capsid protein genes

ASV 50	MH161014.1	Myoviridae environmental samples clone XKH-S-48 capsid protein gene
ASV 60	JX185085.1	Uncultured Myoviridae clone K0809/9-5 Major capsid protein (gp23) gene, partial cds
ASV 70	HM241951.1	Uncultured Myoviridae clone 1-4 Major capsid protein like (g23)
ASV 80	MH161014.1	Myoviridae environmental samples clone XKH-S-48 capsid protein gene
ASV 90	MH161014.1	Myoviridae environmental samples clone XKH-S-48 capsid protein gene

Table 14. BLAST analysis for T4 bacteriophage. 10 random samples were chosen and assigned against BLAST database to identify the ASVs represented through DADA2.

To thoroughly understand the microbiome, it was essential to characterize all its components. Studies of microbiomes typically target only specific groups of microorganisms, with a main focus on bacteria through taxonomic classification of 16S rRNA gene sequences. In this study, we included bacteriophage analysis as a crucial part of the microbiome.

Following Nucleotide sequence similarity searches performed manually using BLAST(http://www.ncbi.nlm.nih.gov/), most sequences tested in our ten randomly selected samples were shown to be the major capsid protein genes, suggesting the likelihood of our extracted amplicons belonging to T4-like *Myoviridae.* One of the sequences was an enterobacteria T4 phage.

The majority of viruses in aquatic environments are thought to be bacterial virus (bacteriophages or phages). An important member of the *Myoviridae* family is the ‘T4-

like phages, and have been isolated from a variety of hosts such as enterobacteria and cyanobacteria (*Prochlorococcus* and *Synechococcus)*. It has been noted that most of the isolated cyanophages have been characterized to be T4-like viruses (Liu et al. 2017), and they commonly occur in marine and freshwater aquatic environment where they play an important role in determining the cyanobacteria during the season (Demeke 2016). Community composition analysis has shown the abundance of *Synechococcus*, which could be hosts for these T4-like phages.

Unfortunately, it was not possible to observe trends in bacteriophage abundance for different types of samples or create a comparison between their presence in different geographical locations in Malawi, due to the difficulties amplifying the targeted *g23* major capsid protein. Despite the high degeneracy of the primers (Chow and Fuhrman 2012), they successfully amplified products of appropriate size prior to the addition of the barcodes. Many reasons could be behind the unsuccessful amplification of these targeted genes; one potential solution could be increasing the sensitivity by using higher concentrations of barcode molecules (Severins et al. 2018). Another explanation for these failing primers could be attributed to their high degeneracy, and it is likely that they are interacting with the barcodes and phased spaces, as shown in (Figure 6), generating a thermostable secondary structure that inhibits binding, preventing the primers from attaching to the targeted sequences. Secondary structures in the primer can affect primer annealing, by spanning the region of primer annealing, and reduce the binding constant of that primer, leading to reduced PCR efficiency (Wang and Seed 2003; Fredman et al. 2004). Oligonucleotides that contain sequence elements that promote secondary structures and primer-primer interactions or hairpins can be avoided using numerous tools for primer design such as using the OligoAnalyser tool, *http://www.idtdna.com/analyzer/Applications/OligoAnalyzer*.

The resultant ASVs of the sample were analysed using a phylogenetic tree visualized by iTOL, allowed the resolution of the T4-like virus clades and the ASVs relationship and their abundance, seen in Figure 20. No other significant correlations were measured for microbial or environmental parameters.

The variability in viral abundance was shown in the phylogenetic tree, with most ASVs having low counts. Low abundance of these ASVs could indicate that they are prophages. While prophages can account for up to 20% of a bacterial genome, they are not simply passengers; they can be selectively advantageous or disadvantageous to the host cell, such as providing the bacterial host with protection from superinfection (Miller-Ensminger et al. 2018). On the contrary, there were fewer ASVs with higher abundance, and one has shown a distinctively high abundance on their own branch and considerably different to others, which may imply the presence of an ongoing lytic infection, increasing the abundance of this particular phage. Lytic phages are exclusively predatory, killing their host cell. Lysogenic phages have two lifestyle choices: they can be lytic immediately, or they can incorporate their genome into the host's bacterial genome or remain in the cytoplasm as a plasmid. The phage genome (now known as a prophage) generally replicates in synchrony with the host chromosome until some signal induces the phage to enter the lytic path (Miller-Ensminger et al. 2018).

As the most abundant entities in the biosphere, phages commonly outnumber bacteria by at least one order of magnitude. Considerable numbers of phages have been shown to be present in bacterial genomes, where these integrated phages (prophages) are at the heart of bacterial diversification processes (Pratama et al. 2018). A major mediator of phage evolution known as horizontal gene transfer, occurs frequently between diverse phages that share overlapping host ranges, as well as between phages and their host at a level where phages are considered as a potential major driving force of cellular evolution. This serves as the primary vector for horizontal genetic exchange in the environment. Phage genes have been frequently challenged to provide cellular functions, and the reverse is also true: host genes have been acquired by phages and subverted for their own purposes (Filée et al. 2006).

In summary, from a scientific and therapeutic perspectives, these complex dynamical relationships between the phage and commensal bacteria within the microbiome need to be further examined as they could assist us in creating a better understanding of their interactions and coevolution dynamics, having a major impact on the host microbiome, through health and disease.

6. General Discussion

An increase in global human population has focused our attention to utilise fisheries and aquaculture as a food staple. Fish farming could be an attainable source of income to many families, reaching remote and isolated areas, while overcoming struggles of malnutrition and hunger in these communities. Our attempt to use fish to feed a growing population is limited as we are faced with existing and emerging disease outbreaks in several parts of the world, causing massive mortalities in aquaculture.

The main aim of our study was to identify the microbiome of healthy tilapia; the second largest production finfish in the world (Machimbirike et al. 2019); and investigate alternative sampling techniques to minimise animal sampling stress, promoting the use of swabs for DNA sampling of the microbial communities of fish skin and mucus. A third major aim of this thesis work was to compare the efficiency of long read sequencing to short read sequencing as applied to microbiome research.

This study involved the characterization of skin microbiomes of tilapia on fish farms, sampled from different ponds in Malawi. Sampling was performed by measuring the microbial communities present in the fish skin and mucus. The composition of the bacterial community present in the rearing water was also investigated in order to evaluate its correlation with the community present on the fish skin. Fish tissues are home to diverse microbial communities and its composition is influenced by many factors.

Communities captured by swabs were not significantly different to those captured by incubation sampling. Thus, swabs can be considered as a suitable alternative to reduce fish stress during sampling. It is also essential to emphasise the increased stress caused by the incubation method, may lead to a disruption to the harmony of these communities and increase their susceptibility to infectious opportunistic pathogens by causing dysbiosis in the skin microbiome, facilitating the proliferation of opportunistic pathogens (Rosado et al. 2019). Studies comparing the gut and skin microbial niches has shown their similar core communities in healthy fish (Legrand et al. 2017), which could lead to understanding the hosts overall health with a stroke of

a swab, avoiding invasive methods. It would be interesting in future work to establish whether the sampling method itself induced a subsequent shift in the host microbiome as a result of stress from handling. If stress kills off bacteria in mucosal layer and pathogens are resistant to the innate immune response, they will be left with a pathogen and unoccupied niche which could trigger dysbiosis. Examining fish skin mucus could give us insight to the role these surfaces play as an important niche for mucosal pathogen evolution in nature, which include their ability to concentrate the bacteria present in water, and select bacteria with abilities to attach, resist innate immunity, and compete with other bacteria, by favouring for the exchange of genes that encode these functions (Carda-Diéguez et al. 2017).

Sampling locations were factored into these analyses, and showed the possibility for complex interactions between skin microbiomes and rearing water, along with the differences of these communities from one place to another. Rearing water samples collected from the same pond had similar microbial communities, while swab samples collected from fish located in the same pond were more distantly related. Therefore, to fully understand the microbiome associated with fish, it is necessary to capture the communities associated with both the skin itself and the surrounding water. Differences in community composition between the skin and water observed in this study, support the findings of Carlson and et al, when examining *Gambusia* skin microbiome. They suggested that these mucosal surfaces may be selective, not simply a passive recipient of bacteria from the water column (Carlson et al. 2017), confirming the idea of the host playing an essential role in recruiting specific microbes to be part of its mucosal niche.

With a continuous effort to find effective strategies to control diseases in farmed animals, new disease control agents are required to maintain the prosperity of aquaculture. This study confirmed *Ralstonia and Methylocystis* were part of the microbiome associated with tilapia fish skin and mucus. Since they are producers of PHB, which are capable of stimulating the immune system of fish early life stages, I speculate that they may be the cause of the increased resistance to diseases and robustness observed in previous studies (Franke et al. 2017). Outbreaks of viruses affecting wild and farmed fish, such as Tilapia lake virus, have been noticed in most surrounding countries of Malawi (Jansen et al. 2018), yet these lakes may be protected

by the production of this compound. It would be worth investigating in future studies, weather PHB play a role in stabilising microbiomes and preventing an increase in dysbiosis.

A comparison of the taxonomy assignment generated by long read sequencing and short read sequencing, has shown that long reads resulted in deriving a lower taxonomic level, reaching species and strain classification, while short reads were limited to exploration of an extremely low percentage of species level. This may lead to missing the identification of essential species that may be biomarkers of disease or causative agents of emerging outbreaks. Our study revealed the presence of potential pathogens, such as *Salmonella, Shigella* and *Brucella*, that were not identified through short read sequencing. The use of nanopore technologies in order to generate long read sequences through MinION, will enable real time, *in-situ* analysis of microbial community structure in aquaculture.

The presence of bacteriophages was confirmed through sequencing using Illumina, and the difficulties faced to generate amplicon products for these viruses, might be due to the degenerate primers used for targeting the g23 major capsid protein of these phages, followed by the addition of barcodes, which may be interfering with the amplification. Although new techniques in sequencing and analyses have driven bacteriophage studies into the era of big data, these data have raised new challenges in bacteriophage genomics, proteomics, and transcriptomics. Their huge diversity and extreme number in environmental samples, and capacity to recombine are major issues. Their taxonomy has become more and more difficult to define and it is now clear that classical dichotomous classification does not fit bacteriophage genomic data (Hulo et al. 2017). One option would be to improve databases for T4-like *myoviridae* for phylogenetic classification, similar to SILVA for 16S rRNA. However, this will still be limited to a small subset of total viral diversity. Therefore, improved sampling methods for accessing the diverse virome in fish skin mucosa will enrich our understanding of these systems.

7. Conclusion and future directions

In summary, this is the first study to explore skin microbiome from farmed tilapia collect from ponds in Malawi. These communities were characterized using high-throughput Illumina MiSeq technologies for v4 16S rRNA, along with Nanopore sequencing of the full-length 16S rRNA gene with MinION™. The presence of bacteriophages in these communities was identified through Illumina sequencing, targeting the g23 major capsid protein, of T4-like *Myoviridae.*

This thesis enhanced our knowledge of the composition of healthy Tilapia skin and mucus microbiomes, which has received little relative research given the huge importance of this fish species in aquaculture. Analysing fish skin microbiomes remains a continued challenge, since it is in direct and continual contact with the aquatic environment it lives in. Aquatic complex environments are constantly shifting with its surroundings of nutrients, temperature, salinity and other living organisms with their own distinctive microbial communities, bearing in mind the hosts' influence of its immunity on these inhabiting microbes. It is worth stating that our samples were collected in October, which is considered one of the warmest months of the year in Malawi. The factors influencing these communities are of major importance to investigate not only which microbes develop on the host's skin, but also their crucial role in regulating and affecting the host. This is critically important to assist us in understanding how perturbation might lead to dysbiosis in hosts.

With a variety of sampling methods used to investigate skin microbiomes, we suggest that incubation samples should be avoided in the future, as swabs were more effective in capturing the diverse microbial communities associated with fish skin and mucus. The lower diversity of communities in incubation samples could be an indication that the immune properties present in the fish might be degrading these communities when the fish is under stress.

Although I faced difficulties amplifying the T4- like Myoviridae, I was able to find the optimal temperature and concentration for the degenerate primers used for this experiment, and was able to successfully amplify these phylogenetic markers from

one rearing water sample. Future work would benefit from the collection of additional samples and precipitation of viruses from filtrate (denoted as the viral fraction) and primary concentration of virus particles, by iron chloride flocculation and collection on 1.0 µm polycarbonate filters (John et al. 2011). Importantly, phages shouldn't be ignored simply because they are challenging. Therefore, new approaches need to be considered for their exploration and incorporation into microbiome data. Further research on phages will assist us in vaccine development and phage therapy, to be one step ahead of outbreaks in aquaculture.

Raising awareness to unfavourable practices such as the use of antibiotics and stressful sampling methods and how they could affect these microbial communities through dysbiosis, would be of major importance to tackle this issue globally. Moreover, this study has revealed the high variability at the ASV level in a small geographic region, showing that future sampling to detect a shift to dysbiosis will require time-resolved sampling of ponds, as it is unlikely, given the variance, that there is a generic 'healthy' pond microbiome. Evolutionary dynamics of microorganisms allow for rapid change in these communities, making it difficult to directly conclude causal relationships from the correlated alterations in microbial community structures and host phenotypes, or directly pinpoint the core microbiome species, as we lack general purpose tools that could manipulate microbiomes at species level (Gao et al. 2018).

Adapting the idea of portable labs; (similar to the one created during the Zika virus epidemic in a roving bus-based mobile laboratory in Brazil) (Quick et al. 2016), using MinIONs, will allow real-time genomic surveillance in resource-limited settings and can be established rapidly to monitor outbreaks and track disease. MinIONs could be the answer for less fortunate, low equipped countries where conventional sequencing technologies are difficult to deploy.

References:

Amal, M. N. A., C. B. Koh, M. Nurliyana, M. Suhaiba, Z. Nor-Amalina, S. Santha, K. P. Diyana-Nadhirah, M. T. Yusof, M. Y. Ina-Salwany, and M. Zamri-Saad. 2018. “A Case of Natural Co-Infection of Tilapia Lake Virus and *Aeromonas Veronii* in a Malaysian Red Hybrid Tilapia (*Oreochromis niloticus*×*O. Mossambicus*) Farm Experiencing High Mortality.” *Aquaculture* 485 (February): 12–16.

Arias, Cova, Candis R. Ray, Wenlong Cai, Evelyn Willmon, And Cova R. Arias, and Sift Desk. 2019. “Fish Are Not Alone: Characterization of the Gut and Skin Microbiomes of Largemouth Bass (*Micropterus Salmoides*), Bluegill (*Lepomis Macrochirus*), and Spotted Gar (*Lepisosteus Oculatus*).” *SDRP Journal of Aquaculture, Fisheries & Fish Science* 2 (2).

Aschenbrenner, Ines Aline, Tomislav Cernava, Armin Erlacher, Gabriele Berg, and Martin Grube. 2017. “Differential Sharing and Distinct Co-Occurrence Networks among Spatially Close Bacterial Microbiota of Bark, Mosses and Lichens.” *Molecular Ecology* 26 (10): 2826–38.

Barr, Jeremy J., Rita Auro, Mike Furlan, Katrine L. Whiteson, Marcella L. Erb, Joe Pogliano, Aleksandr Stotland. 2013. “Bacteriophage Adhering to Mucus Provide a Non–host-Derived Immunity.” *Proceedings of the National Academy of Sciences of the United States of America*, May, 201305923.

Bauer, Robert J., Alexander Zhelkovsky, Katharina Bilotti, Laura E. Crowell, Thomas C. Evans Jr, Larry A. McReynolds, and Gregory J. S. Lohman. 2017. “Comparative Analysis of the End-Joining Activity of Several DNA Ligases.” *PloS One* 12 (12): e0190062.

Béné, Christophe, Manuel Barange, Rohana Subasinghe, Per Pinstrup-Andersen, Gorka Merino, Gro-Ingunn Hemre, and Meryl Williams. 2015a. “Feeding 9 Billion by 2050--Putting Fish Back on the Menu.” *Food Security* 7 (2): 261–74.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society* 57 (1): 289–300.

Binesh, C. P. 2014. “Elevation of Temperature and Crowding Trigger Acute Viral Nervous Necrosis in Zebra Fish, *Brachydanio Rerio* (Hamilton-Buchanan), subclinically Infected with Betanodavirus.” *Journal of Fish Diseases* 37 (3): 279–82.

Boon, Eva, Conor J. Meehan, Chris Whidden, Dennis H-J Wong, Morgan G. I. Langille, and Robert G. Beiko. 2014. “Interactions in the Microbiome: Communities of Organisms and Communities of Genes.” *FEMS Microbiology* Reviews 38 (1): 90–118.

Boutin, Sébastien, Louis Bernatchez, Céline Audet, and Nicolas Derôme. 2012. “Antagonistic Effect of Indigenous Skin Bacteria of Brook Charr (*Salvelinus Fontinalis*) against *Flavobacterium Columnare* and *F. Psychrophilum*.” *Veterinary Microbiology* 155 (2-4): 355–61.

Boutin, Sébastien, Christopher Sauvage, Louis Bernatchez, Céline Audet, and Nicolas Derome. 2014. “Inter Individual Variations of the Fish Skin Microbiota: Host Genetics Basis of Mutualism?” *PloS One* 9 (7): e102649.

Breacker, Carl, Iain Barber, William H. J. Norton, Jonathan R. McDearmid, and Ceinwen A. Tilley. 2017. “A Low-Cost Method of Skin Swabbing for the Collection of DNA Samples from Small Laboratory Fish.” *Zebrafish* 14 (1): 35–41.

Brinchmann, Monica Fengsrud. 2016. “Immune Relevant Molecules Identified in the Skin Mucus of Fish Using -Omics Technologies.” *Molecular bioSystems* 12 (7): 2056–63.

Brown, Bonnie L., Mick Watson, Samuel S. Minot, Maria C. Rivera, and Rima B. Franklin. 2017. “MinION™ Nanopore Sequencing of Environmental Metagenomes: A Synthetic Approach.” *GigaScience*.

Bruijn, Irene de, Yiying Liu, Geert F. Wiegertjes, and Jos M. Raaijmakers. 2018. “Exploring Fish Microbial Communities to Mitigate Emerging Diseases in Aquaculture.” *FEMS Microbiology Ecology* 94 (1).

Butt, Robyn Lisa, and Helene Volkoff. 2019. “Gut Microbiota and Energy Homeostasis in Fish.” *Frontiers in Endocrinology* 10 (January): 9.

Bwanika, G. N., L. J. Chapman, Y. Kizito, and J. Balirwa. 2006. “Cascading Effects of Introduced Nile Perch (*Lates Niloticus*) on the Foraging Ecology of Nile Tilapia (*Oreochromis Niloticus*).” *Ecology of Freshwater Fish* 15 (4): 470–81.

Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes. 2017. “Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis.” *The ISME Journal* 11 (12): 2639–43.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” *Nature Methods* 13 (7): 581–83.

Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK. 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic acids research*. DOI: 10.1093/nar/gkz569.

Carda-Diéguez, Miguel, Rohit Ghai, Francisco Rodríguez-Valera, and Carmen Amaro. 2017. “Wild Eel Microbiome Reveals That Skin Mucus of Fish Could Be a Natural Niche for Aquatic Mucosal Pathogen Evolution.” *Microbiome* 5 (1): 162.

Carlson, Jeanette M., Annie B. Leonard, Embriette R. Hyde, Joseph F. Petrosino, and Todd P. Primm. 2017. “Microbiome Disruption and Recovery in the Fish *Gambusia Affinis* Following Exposure to Broad-Spectrum Antibiotic.” Infection and Drug Resistance 10 (May): 143–54.

Carro, Antonia M., Diego García-Rodríguez, Paula Gonzalez-Siso, and Rosa A. Lorenzo. 2012. “Determination of Chemotherapeutic Agents in Fish and Shellfish by

Matrix Solid-Phase Dispersion and Liquid Chromatography-Tandem Mass Spectrometry." *Journal of Separation Science* 35 (21): 2866–74.

Cetron, M. S., L. Chitsulo, J. J. Sullivan, J. Pilcher, M. Wilson, J. Noh, V. C. Tsang, A. W. Hightower, and D. G. Addiss. 1996. "Schistosomiasis in Lake Malawi." *The Lancet* 348 (9037): 1274–78.

Chen, Cheng-Yu, Po-Cheng Chen, Francis Cheng-Hsuan Weng, Grace Tzun-Wen Shaw, and Daryi Wang. 2017. "Habitat and Indigenous Gut Microbes Contribute to the Plasticity of Gut Microbiome in Oriental River Prawn during Rapid Environmental Change." *PloS One* 12 (7): e0181427.

Cheung, William W. L., Vicky W. Y. Lam, Jorge L. Sarmiento, Kelly Kearney, Reg Watson, and Daniel Pauly. 2009. "Projecting Global Marine Biodiversity Impacts under Climate Change Scenarios." *Fish and Fisheries* 10 (3): 235–51.

Cho, Ilseung, and Martin J. Blaser. 2012. "The Human Microbiome: At the Interface of Health and Disease." *Nature Reviews. Genetics* 13 (4): 260–70.

Chow, Cheryl-Emiliane T., and Jed A. Fuhrman. 2012. "Seasonality and Monthly Dynamics of Marine Myovirus Communities." *Environmental Microbiology* 14 (8): 2171–83.

Clarke, K. R. 1993. "Non-Parametric Multivariate Analyses of Changes in Community Structure." *Austral Ecology* 18 (1): 117–43.

Clarke, K. R., and R. N. Gorley. 2006. "Primer." Primer-E, *Plymouth*. *http://www.ap.smu.ca/~dclarke/home/documents/byDAC/mhd_primer.pdf*.

Colston, Timothy J., and Colin R. Jackson. 2016. "Microbiome Evolution along Divergent Branches of the Vertebrate Tree of Life: What Is Known and Unknown." *Molecular Ecology* 25 (16): 3776–3800.

Comeau, André M., and Henry M. Krisch. 2008. “The Capsid of the T4 Phage Superfamily: The Evolution, Diversity, and Structure of Some of the Most Prevalent Proteins in the Biosphere.” *Molecular Biology and Evolution* 25 (7): 1321–32.

Coon, Kerri L., Kevin J. Vogel, Mark R. Brown, and Michael R. Strand. 2014. “Mosquitoes Rely on Their Gut Microbiota for Development.” *Molecular Ecology* 23 (11): 2727–39.

Cordingley, Michael G. 2017. “Viruses: Agents of Evolutionary Invention”. Harvard University Press; 1st edition. DOI: 10.4159/9780674978638.

Cusco, A., J. Vines, S. D’Andreano, F. Riva, and J. Casellas. 2017. “Using MinION to Characterize Dog Skin Microbiota through Full-Length 16S rRNA Gene Sequencing Approach.” *bioRxiv*. doi: https://doi.org/10.1101/167015

Dash, S., S. K. Das, J. Samal, and H. N. Thatoi. 2018. “Epidermal Mucus, a Major Determinant in Fish Health: A Review.” *Iranian Journal of Veterinary Research* 19 (2): 72–81.

DeAngelis, M. M., D. G. Wang, and T. L. Hawkins. 1995. “Solid-Phase Reversible Immobilization for the Isolation of PCR Products.” *Nucleic Acids Research* 23 (22): 4742–43.

Demeke, Addisu. 2016. “Cyanobacteria Blooms and Biological Control Methods.” *Int. J. Fauna Boil. Stud* 3: 32–38.

Dong, H. T., S. Siriroob, W. Meemetta, W. Santimanawong, W. Gangnonngiw, N. Pirarat, P. Khunrae, T. Rattanarojpong, R. Vanichviriyakit, and S. Senapin. 2017. “Emergence of Tilapia Lake Virus in Thailand and an Alternative Semi-Nested RT-PCR for Detection.” *Aquaculture* 476 (July): 111–18.

Douglas, Angela E., and John H. Werren. 2016. “Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts.” *mBio* 7 (2): e02099.

Drobac, Damjana, Nada Tokodi, Jelena Lujić, Zoran Marinović, Gordana Subakov-Simić, Tamara Dulić, Tamara Važić, et al. 2016. “Cyanobacteria and Cyanotoxins in Fishponds and Their Effects on Fish Tissue.” Harmful Algae 55 (May): 66–76.

Egerton, Sian, Sarah Culloty, Jason Whooley, Catherine Stanton, and R. Paul Ross. 2018. “The Gut Microbiota of Marine Fish.” *Frontiers in Microbiology* 9 (May): 873.

Eichmiller, Jessica J., Matthew J. Hamilton, Christopher Staley, Michael J. Sadowsky, and Peter W. Sorensen. 2016. “Environment Shapes the Fecal Microbiome of Invasive Carp Species.” *Microbiome* 4 (1): 44.

Eren, A. Murat, Loïs Maignien, Woo Jun Sul, Leslie G. Murphy, Sharon L. Grim, Hilary G. Morrison, and Mitchell L. Sogin. 2013. “Oligotyping: Differentiating between Closely Related Microbial Taxa Using 16S rRNA Gene Data.” *Methods in Ecology and Evolution / British Ecological Society* 4 (12).

Eyngor, Marina, Rachel Zamostiano, Japhette Esther Kembou Tsofack, Asaf Berkowitz, Hillel Bercovier, Simon Tinman, Menachem Lev, et al. 2014. “Identification of a Novel RNA Virus Lethal to Tilapia.” *Journal of Clinical Microbiology* 52 (12): 4137–46.

Fathi, Mohamed, Cathryn Dickson, Malcolm Dickson, William Leschen, Johanna Baily, Fiona Muir, Kristina Ulrich, and Manfred Weidmann. 2017. “Identification of Tilapia Lake Virus in Egypt in Nile Tilapia Affected by ‘summer Mortality’ Syndrome.” *Aquaculture* 473 (April): 430–32.

Feiner, Ron, Tal Argov, Lev Rabinovich, Nadejda Sigal, Ilya Borovok, and Anat A. Herskovits. 2015. “A New Perspective on Lysogeny: Prophages as Active Regulatory Switches of Bacteria.” *Nature Reviews. Microbiology* 13 (10): 641–50.

Filée, Jonathan, Eric Bapteste, Edward Susko, and H. M. Krisch. 2006. “A Selective Barrier to Horizontal Gene Transfer in the T4-Type Bacteriophages That Has Preserved a Core Genome with the Viral Replication and Structural Genes.” *Molecular Biology and Evolution* 23 (9): 1688–96.

Filée, Jonathan, Françoise Tétart, Curtis A. Suttle, and H. M. Krisch. 2005. “Marine T4-Type Bacteriophages, a Ubiquitous Component of the Dark Matter of the Biosphere.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (35): 12471–76.

Food and Agriculture Organization. 2014. *State of the World Fisheries and Aquaculture* 2014. FAO. *http://www.fao.org/3/a-i3720e.pdf*

Food and Agriculture Organization of the United Nations. 2016. 2016 THE STATE OF WORLD FISHERIES AND AQUACULTURE: CONTRIBUTING TO FOOD SECURITY AND NUTRITION FOR ALL. *Food & Agriculture Org*. http://www.fao.org/3/a-i5555e.pdf

Food and Agriculture Organization of the United Nations, United Nations International Children’s Emergency Fund, World Health Organization, World Food Programme, and International Fund for Agriculture Development. 2018. The State of Food Security and Nutrition in the World 2018: Building Climate Resilience for Food Security and Nutrition. *Food & Agriculture Org. http://www.fao.org/3/i9553en/i9553en.pdf*

Food and ariculture organisation of the united nations. 2017. FAO issues alert over lethal virus affecting popular tilapia fish. *http://www.fao.org/news/story/en/item/888884/icode/*

Føre, M., E. Svendsen, J. A. Alfredsen, I. Uglem, N. Bloecher, H. Sveier, L. M. Sunde, and K. Frank. 2018. “Using Acoustic Telemetry to Monitor the Effects of Crowding and Delousing Procedures on Farmed Atlantic Salmon (*Salmo Salar*).” *Aquaculture* 495 (October): 757–65.

Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, Wolfe RS, Balch WE, Tanner RS, Magrum LJ, Zablen LB, Blakemore R, Gupta R, Bonen L, Lewis BJ, Stahl DA, Luehrsen KR, Chen KN, Woese CR. 1980. The phylogeny of prokaryotes. *Science* 209:457–463.

Franke, Andrea, Catriona Clemmesen, Peter De Schryver, Linsey Garcia-Gonzalez, and Olivia Roth. 2017. “Immunostimulatory Effects of Dietary Poly-B-Hydroxybutyrate in European Sea Bass Postlarvae.” *Aquaculture Research* 48 (12): 5707-5717.

Frank, Jeremy A., Claudia I. Reich, Shobha Sharma, Jon S. Weisbaum, Brenda A. Wilson, and Gary J. Olsen. 2008. “Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes.” *Applied and Environmental Microbiology* 74 (8): 2461–70.

Fredman, David, M. Jobs, L. Strömqvist, and A. J. Brookes. 2004. “DFold: PCR Design That Minimizes Secondary Structure and Optimizes Downstream Genotyping Applications.” *Human Mutation* 24 (1): 1–8.

Fuhrman, J. A. 1999. “Marine Viruses and Their Biogeochemical and Ecological Effects.” *Nature* 399 (6736): 541–48.

Gallagher, Michael D., Iveta Matejusova, Lien Nguyen, Neil M. Ruane, Knut Falk, and Daniel J. Macqueen. 2018. “Nanopore Sequencing for Rapid Diagnostics of Salmonid RNA Viruses.” *Scientific Reports* 8 (1): 16307.

Gao, Na L., Chengwei Zhang, Zhanbing Zhang, Songnian Hu, Martin J. Lercher, Xing-Ming Zhao, Peer Bork, Zhi Liu, and Wei-Hua Chen. 2018. “MVP: A Microbe-Phage Interaction Database.” *Nucleic Acids Research* 46 (D1): D700–707.

García-Pérez, Teresa, Juan C. López, Fabiana Passos, Raquel Lebrero, Sergio Revah, and Raúl Muñoz. 2018. “Simultaneous Methane Abatement and PHB Production by *Methylocystis Hirsuta* in a Novel Gas-Recycling Bubble Column Bioreactor.” *Chemical Engineering Journal* 334 (February): 691–97.

Givens, C. E., B. Ransom, N. Bano, and J. T. Hollibaugh. 2015. “Comparison of the Gut Microbiomes of 12 Bony Fish and 3 Shark Species.” *Marine Ecology Progress* Series 518 (January): 209–23.

Gołębiewski, Marcin, Joanna Całkiewicz, Simon Creer, and Kasia Piwosz. 2017. “Tideless Estuaries in Brackish Seas as Possible Freshwater-Marine Transition Zones for Bacteria: The Case Study of the Vistula River Estuary.” *Environmental Microbiology Reports* 9 (2): 129–43.

Gomez, Daniela, J. Oriol Sunyer, and Irene Salinas. 2013. “The Mucosal Immune System of Fish: The Evolution of Tolerating Commensals While Fighting Pathogens.” *Fish & Shellfish Immunology* 35 (6): 1729–39.

Goodrich, Julia K., Sara C. Di Rienzi, Angela C. Poole, Omry Koren, William A. Walters, J. Gregory Caporaso, Rob Knight, and Ruth E. Ley. 2014. “Conducting a Microbiome Study.” *Cell* 158 (2): 250–62.

Gotelli, Nicholas J., and Robert K. Colwell. 2001. “Quantifying Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness.” *Ecology Letters* 4 (4): 379–91.

Gupta, Shruti, Adriána Fečkaninová, Jep Lokesh, Jana Koščová, Mette Sørensen, Jorge Fernandes, and Viswanath Kiron. 2018. “*Lactobacillus* Dominate in the Intestine of Atlantic Salmon Fed Dietary Probiotics.” *Frontiers in Microbiology* 9: 3247.

Hulo, Chantal, Patrick Masson, Ariane Toussaint, David Osumi-Sutherland, Edouard de Castro, Andrea H. Auchincloss, Sylvain Poux, Lydie Bougueleret, Ioannis Xenarios, and Philippe Le Mercier. 2017. “Bacterial Virus Ontology; Coordinating across Databases.” *Viruses* 9 (6).

Ignacio-Espinoza, J. Cesar, and Matthew B. Sullivan. 2012. “Phylogenomics of T4 Cyanophages: Lateral Gene Transfer in the ‘Core’ and Origins of Host Genes.” *Environmental Microbiology* 14 (8): 2113–26.

Jansen, Mona Dverdal, Ha Thanh Dong, and Chadag Vishnumurthy Mohan. 2018. “Tilapia Lake Virus: A Threat to the Global Tilapia Industry?” *Reviews in Aquaculture* 46 (May): 120.

John, Seth G., Carolina B. Mendez, Li Deng, Bonnie Poulos, Anne Kathryn M. Kauffman, Suzanne Kern, Jennifer Brum, Martin F. Polz, Edward A. Boyle, and Matthew B. Sullivan. 2011. “A Simple and Efficient Method for Concentration of Ocean Viruses by Chemical Flocculation.” *Environmental Microbiology Reports* 3 (2): 195–202.

Kerkhof, Lee J., Kevin P. Dillon, Max M. Häggblom, and Lora R. McGuinness. 2017. “Profiling Bacterial Communities by MinION Sequencing of Ribosomal Operons.” *Microbiome* 5 (1): 116.

Knight, Rob, Alison Vrbanac, Bryn C. Taylor, Alexander Aksenov, Chris Callewaert, Justine Debelius, Antonio Gonzalez, et al. 2018. “Best Practices for Analysing Microbiomes.” Nature Reviews. *Microbiology* 16 (7): 410–22.

Koblitz, Rodrigo Vasconcelos, Albertina Pimentel Lima, Marcelo Menin, Diana Rojas, Luiz Henrique Condrati, and William Ernest Magnusson. 2017. “Effect of Species-Counting Protocols and the Spatial Distribution of Effort on Rarefaction Curves in Relation to Decision Making in Environmental-Impact Assessments.” *Austral Ecology* 42 (6): 723–31.

Kong, Heidi H., and Julia A. Segre. 2017. “The Molecular Revolution in Cutaneous Biology: Investigating the Skin Microbiome.” *The Journal of Investigative Dermatology* 137 (5): e119–22.

Kozich, James J., Sarah L. Westcott, Nielson T. Baxter, Sarah K. Highlander, and Patrick D. Schloss. 2013. “Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform.” *Applied and Environmental Microbiology* 79 (17): 5112–20.

Kuczynski, Justin, Christian L. Lauber, William A. Walters, Laura Wegener Parfrey, José C. Clemente, Dirk Gevers, and Rob Knight. 2011. “Experimental and Analytical Tools for Studying the Human Microbiome.” *Nature Reviews. Genetics* 13 (1): 47–58.

Kumar, Baidyanath, and Ameetha Sinha. 2014. “*Microcystis* Toxic Blooms in Fish Culture Ponds and Their Biological and Chemical Control.” *Int J Sci Technicol* Res 3 (3): 398–410.

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. “Rapid Determination of 16S Ribosomal RNA Sequences for Phylogenetic Analyses.” *Proceedings of the National Academy of Sciences of the United States of America* 82 (20): 6955–59.

Laranja, Joseph Leopoldo Q., and Peter Bossier. 2019. “Poly-Beta-Hydroxybutyrate (PHB) and Infection Reduction in Farmed Aquatic Animals.” In Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids, edited by Howard Goldfine, 1–27. *Cham: Springer International Publishing*. DOI: 10.1007/978-3-319-72473-7_35-1

Larsen, Andrea M., Stephen A. Bullard, Matthew Womble, and Covadonga R. Arias. 2015. “Community Structure of Skin Microbiome of Gulf Killifish, *Fundulus Grandis*, Is Driven by Seasonality and Not Exposure to Oiled Sediments in a Louisiana Salt Marsh.” *Microbial Ecology* 70 (2): 534–44.

Larsen, Andrea, Zhen Tao, Stephen A. Bullard, and Covadonga R. Arias. 2013. “Diversity of the Skin Microbiota of Fishes: Evidence for Host Species Specificity.” *FEMS Microbiology Ecology* 85 (3): 483–94.

Lee, Dong Geol, Jin Ju Nam, Min Ji Kim, Seung Hyun Kang, and Youn Joon Kim. 2019. Novel sporichthyaceae microorganism and use thereof. USPTO 20190040475:A1. *US Patent, filed* June 27, 2017, and issued February 7, 2019.

Legrand, Thibault P. R. A., Sarah R. Catalano, Melissa L. Wos-Oxley, Fran Stephens, Matt Landos, Matthew S. Bansemer, David A. J. Stone, Jian G. Qin, and Andrew P. A. Oxley. 2017. “The Inner Workings of the Outer Surface: Skin and Gill Microbiota as Indicators of Changing Gut Health in Yellowtail Kingfish.” *Frontiers in Microbiology* 8: 2664.

Lehodey, P., J. Alheit, M. Barange, T. Baumgartner, G. Beaugrand, K. Drinkwater, J-M Fromentin, et al. 2006. “Climate Variability, Fish, and Fisheries.” *Journal of Climate* 19 (20): 5009–30.

Letunic, Ivica, and Peer Bork. 2019. “Interactive Tree Of Life (iTOL) v4: Recent Updates and New Developments.” Nucleic Acids Research, April. https://doi.org/10.1093/nar/gkz239.

Le Vin, A. L., A. Adam, A. Tedder, K. E. Arnold, and B. K. Mable. 2011. “Validation of Swabs as a Non-Destructive and Relatively Non-Invasive DNA Sampling Method in Fish.” *Molecular Ecology Resources* 11 (1): 107–9.

Levy, Maayan, Aleksandra A. Kolodziejczyk, Christoph A. Thaiss, and Eran Elinav. 2017. “Dysbiosis and the Immune System.” *Nature Reviews. Immunology* 17 (4): 219–32.

Levy, Shawn E., and Richard M. Myers. 2016. “Advancements in Next-Generation Sequencing.” Annual Review of Genomics and Human Genetics 17 (August): 95–115.

Lim, Efrem S., Cynthia Rodriguez, and Lori R. Holtz. 2018. “Amniotic Fluid from Healthy Term Pregnancies Does Not Harbor a Detectable Microbial Community.” *Microbiome* 6 (1): 87.

Liu, Lu, Lanlan Cai, and Rui Zhang. 2017. “Co-Existence of Freshwater and Marine T4-like Myoviruses in a Typical Subtropical Estuary.” *FEMS Microbiology Ecology* 93 (11).

Lokesh, Jep, and Viswanath Kiron. 2016. “Transition from Freshwater to Seawater Reshapes the Skin-Associated Microbiota of Atlantic Salmon.” *Scientific Reports* 6 (January): 19707.

Lynch, A. J., and J. R. MacMillan. 2017. “The Role of Fish in a Globally Changing Food System.” In Agroclimatology: Linking Agriculture to Climate. Agronomy Monographs. Madison, WI: *American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Inc.*

https://doi.org/10.2134/agronmonogr60.2014.0059

Machimbirike, Vimbai Irene, Mona Dverdal Jansen, Saengchan Senapin, Pongsak Khunrae, Triwit Rattanarojpong, and Ha Thanh Dong. 2019. “Viral Infections in Tilapines: More than Just Tilapia Lake Virus.” *Aquaculture* 503 (March): 508–18.

Manrique, Pilar, Benjamin Bolduc, Seth T. Walk, John van der Oost, Willem M. de Vos, and Mark J. Young. 2016. “Healthy Human Gut Phageome.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (37): 10400–405.

Mao, Dan-Ping, Quan Zhou, Chong-Yu Chen, and Zhe-Xue Quan. 2012. “Coverage Evaluation of Universal Bacterial Primers Using the Metagenomic Datasets.” *BMC Microbiology* 12 (May): 66.

Marchesi, Julian R., and Jacques Ravel. 2015. “The Vocabulary of Microbiome Research: A Proposal.” *Microbiome* 3 (July): 31.

McMurdie, Paul J., and Susan Holmes. 2013. “Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.” *PloS One* 8 (4): e61217.

Merrifield, Daniel L., and Ana Rodiles. 2015. “10 - The Fish Microbiome and Its Interactions with Mucosal Tissues.” In Mucosal Health in Aquaculture, edited by Benjamin H. Beck and Eric Peatman, 273–95. *San Diego: Academic Press*.

Miller-Ensminger, Taylor, Andrea Garretto, Jonathon Brenner, Krystal Thomas-White, Adriano Zambom, Alan J. Wolfe, and Catherine Putonti. 2018. “Bacteriophages of the Urinary Microbiome.” *Journal of Bacteriology* 200 (7).

Minniti, Giusi, Live Heldal Hagen, Davide Porcellato, Sven Martin Jørgensen, Phillip B. Pope, and Gustav Vaaje-Kolstad. 2017. “The Skin-Mucus Microbial Community of Farmed Atlantic Salmon (*Salmo Salar*).” *Frontiers in Microbiology* 8 (October): 2043.

Mirzaei, Mohammadali Khan, and Corinne F. Maurice. 2017. “Ménage à Trois in the Human Gut: Interactions between Host, Bacteria and Phages.” *Nature Reviews. Microbiology* 15 (7): 397–408.

Mumby, Peter J. 2001. “Beta and Habitat Diversity in Marine Systems: A New Approach to Measurement, Scaling and Interpretation.” *Oecologia* 128 (2): 274–80.

Mutlu, Ece A., Ali Keshavarzian, John Losurdo, Garth Swanson, Basile Siewe, Christopher Forsyth, Audrey French, et al. 2014. “A Compositional Look at the Human Gastrointestinal Microbiome and Immune Activation Parameters in HIV Infected Subjects.” *PLoS Pathogens* 10 (2): e1003829.

Myhrvold, Cameron, Jonathan W. Kotula, Wade M. Hicks, Nicholas J. Conway, and Pamela A. Silver. 2015. “A Distributed Cell Division Counter Reveals Growth Dynamics in the Gut Microbiota.” *Nature Communications* 6 (November): 10039.

Needham, David M., Rohan Sachdeva, and Jed A. Fuhrman. 2017. “Ecological Dynamics and Co-Occurrence among Marine Phytoplankton, Bacteria and Myoviruses Shows Microdiversity Matters.” *The ISME Journal* 11 (7): 1614–29.

Nikouli, Eleni, Alexandra Meziti, Efthimia Antonopoulou, Eleni Mente, and Konstantinos A. Kormas. 2018. “Gut Bacterial Communities in Geographically Distant Populations of Farmed Sea Bream (*Sparus Aurata*) and Sea Bass (*Dicentrarchus Labrax*).” *Microorganisms* 6 (3).

Pratama, Akbar Adjie, Maryam Chaib De Mares, and Jan Dirk van Elsas. 2018. “Evolutionary History of Bacteriophages in the Genus Paraburkholderia.” *Frontiers in Microbiology* 9 (May): 835.

Pratte, Zoe A., Marc Besson, Rebecca D. Hollman, and Frank J. Stewart. 2018. “The Gills of Reef Fish Support a Distinct Microbiome Influenced by Host-Specific Factors.” *Applied and Environmental Microbiology* 84 (9).

Pruesse, Elmar, Jörg Peplies, and Frank Oliver Glöckner. 2012. “SINA: Accurate High-Throughput Multiple Sequence Alignment of Ribosomal RNA Genes.” *Bioinformatics* 28 (14): 1823–29.

Quick, Joshua, Nicholas J. Loman, Sophie Duraffour, Jared T. Simpson, Ettore Severi, Lauren Cowley, Joseph Akoi Bore, et al. 2016. “Real-Time, Portable Genome Sequencing for Ebola Surveillance.” *Nature* 530 (7589): 228–32.

Reid, Kristin M., Sonal Patel, Aaron J. Robinson, Lijing Bu, Jiraporn Jarungsriapisit, Lindsey J. Moore, and Irene Salinas. 2017. “Salmonid Alphavirus Infection Causes Skin Dysbiosis in Atlantic Salmon (*Salmo Salar L.)* Post-Smolts.” *PloS One* 12 (3): e0172856.

Reverter, Miriam, Pierre Sasal, N. Tapissier-Bontemps, D. Lecchini, and M. Suzuki. 2017. “Characterisation of the Gill Mucosal Bacterial Communities of Four Butterflyfish Species: A Reservoir of Bacterial Diversity in Coral Reef Ecosystems.” *FEMS Microbiology Ecology* 93 (6).

Ricci, A., A. Allende, D. Bolton, and M. Chemaly. 2017. “Scientific Opinion on the Update of the List of QPS-recommended Biological Agents Intentionally Added to Food or Feed as Notified to EFSA.” *EFSA*. *https://www.efsa.europa.eu/en/efsajournal/pub/4664*

Robinson, Christopher M., and Julie K. Pfeiffer. 2014. “Viruses and the Microbiota.” *Annual Review of Virology* 1: 55–69.

Rocap, Gabrielle, Frank W. Larimer, Jane Lamerdin, Stephanie Malfatti, Patrick Chain, Nathan A. Ahlgren, Andrae Arellano, et al. 2003. “Genome Divergence in Two *Prochlorococcus* Ecotypes Reflects Oceanic Niche Differentiation.” *Nature* 424 (6952): 1042–47.

Rodger, Hamish D. 2016. “Fish Disease Causing Economic Impact in Global Aquaculture.” In Fish Vaccines, edited by Alexandra Adams, 1–34. Basel: Springer Basel.

Roeselers, Guus, Erika K. Mittge, W. Zac Stephens, David M. Parichy, Colleen M. Cavanaugh, Karen Guillemin, and John F. Rawls. 2011. “Evidence for a Core Gut Microbiota in the Zebrafish.” *The ISME Journal* 5 (10): 1595–1608.

Rohwer, Forest, David Prangishvili, and Debbie Lindell. 2009. “Roles of Viruses in the Environment.” *Environmental Microbiology* 11 (11): 2771–74.

Rohwer, Forest, and Rebecca Vega Thurber. 2009. “Viruses Manipulate the Marine Environment.” *Nature* 459 (7244): 207–12.

Rohwer, Robin R., Joshua J. Hamilton, Ryan J. Newton, and Katherine D. McMahon. 2018. “TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution.” *mSphere* 3 (5). *https://doi.org/10.1128/mSphere.00327-18*.

Rosado, Daniela, Marcos Pérez-Losada, Ricardo Severino, Jo Cable, and Raquel Xavier. 2019. “Characterization of the Skin and Gill Microbiomes of the Farmed Seabass (*Dicentrarchus Labrax*) and Seabream (*Sparus Aurata*).” *Aquaculture* 500 (February): 57–64.

Sahu, Maloy Kumar, N. S. Swarnakumar, K. Sivakumar, T. Thangaradjou, and L. Kannan. 2008. “Probiotics in Aquaculture: Importance and Future Perspectives.” *Indian Journal of Microbiology* 48 (3): 299–308.

Salter, Susannah J., Michael J. Cox, Elena M. Turek, Szymon T. Calus, William O. Cookson, Miriam F. Moffatt, Paul Turner, Julian Parkhill, Nicholas J. Loman, and Alan W. Walker. 2014. “Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses.” *BMC Biology* 12 (November): 87.

Severins, Ivo, Malwina Szczepaniak, and Chirlmin Joo. 2018. “Multiplex Single-Molecule DNA Barcoding Using an Oligonucleotide Ligation Assay.” *Biophysical Journal* 115 (6): 957–67.

Sheik, Cody S., Brandi Kiel Reese, Katrina I. Twing, Jason B. Sylvan, Sharon L. Grim, Matthew O. Schrenk, Mitchell L. Sogin, and Frederick S. Colwell. 2018. “Identification and Removal of Contaminant Sequences From Ribosomal Gene Databases: Lessons From the Census of Deep Life.” *Frontiers in Microbiology* 9 (April): 840.

Shkoporov, Andrey N., and Colin Hill. 2019. “Bacteriophages of the Human Gut: The ‘Known Unknown’ of the Microbiome.” Cell Host & Microbe 25 (2): 195–209.

Simon, Jean-Christophe, Julian R. Marchesi, Christophe Mougel, and Marc-André Selosse. 2019. “Host-Microbiota Interactions: From Holobiont Theory to Analysis.” *Microbiome* 7 (1): 5.

Sinha, Rashmi, Galeb Abu-Ali, Emily Vogtmann, Anthony A. Fodor, Boyu Ren, Amnon Amir, Emma Schwager, et al. 2017. “Assessment of Variation in Microbial Community Amplicon Sequencing by the Microbiome Quality Control (MBQC) Project Consortium.” *Nature Biotechnology* 35 (11): 1077–86.

Skerra, A. 1992. “Phosphorothioate Primers Improve the Amplification of DNA Sequences by DNA Polymerases with Proofreading Activity.” Nucleic Acids Research 20 (14): 3551–54.

Steier, Liviu, Silvia Dias de Oliveira, and José Antonio Poli de Figueiredo. 2019. “Bacteriophages in Dentistry-State of the Art and Perspectives.” *Dental Journal* 7 (1).

Suguna, Ponnusamy, Chandrasekaran Binuramesh, Periyasamy Abirami, Viswanathan Saranya, Kkani Poornima, Veluchamy Rajeswari, and Rajaiah Shenbagarathai. 2014. “Immunostimulation by Poly-β Hydroxybutyrate-Hydroxyvalerate (PHB-HV) from *Bacillus Thuringiensis* in *Oreochromis Mossambicus*.” *Fish & Shellfish Immunology* 36 (1): 90–97.

Sullivan, Matthew B. 2015. “Viromes, Not Gene Markers, for Studying Double-Stranded DNA Virus Communities.” *Journal of Virology* 89 (5): 2459–61.

Sullivan, Matthew B., Katherine H. Huang, Julio C. Ignacio-Espinoza, Aaron M. Berlin, Libusha Kelly, Peter R. Weigele, Alicia S. DeFrancesco, et al. 2010. “Genomic Analysis of Oceanic Cyanobacterial Myoviruses Compared with T4-like Myoviruses from Diverse Hosts and Environments.” *Environmental Microbiology* 12 (11): 3035–56.

Tacon, Albert G. J., and Marc Metian. 2013. “Fish Matters: Importance of Aquatic Foods in Human Nutrition and Global Food Supply.” *Reviews in Fisheries Science* 21 (1): 22–38.

Tsofack, Japhette Esther Kembou, Rachel Zamostiano, Salsabeel Watted, Asaf Berkowitz, Ezra Rosenbluth, Nischay Mishra, Thomas Briese, et al. 2017. “Detection of Tilapia Lake Virus in Clinical Samples by Culturing and Nested Reverse Transcription-PCR.” *Journal of Clinical Microbiology* 55 (3): 759–67.

Tsuchiya, C., T. Sakata, and H. Sugita. 2008. “Novel Ecological Niche of Cetobacterium Somerae, an Anaerobic Bacterium in the Intestinal Tracts of Freshwater Fish.” Letters in *Applied Microbiology* 46 (1): 43–48.

Vale, Pedro F., and Tom J. Little. 2010. “CRISPR-Mediated Phage Resistance and the Ghost of Coevolution Past.” Proceedings. Biological Sciences / *The Royal Society* 277 (1691): 2097–2103.

Van Hai, Ngo. 2015. “Research Findings from the Use of Probiotics in Tilapia Aquaculture: A Review.” *Fish & Shellfish Immunology* 45 (2): 592–97.

Van Hung, Nguyen, Peter De Schryver, Nguyen Viet Dung, Nancy Nevejan, and Peter Bossier. 2019. “*Ralstonia Eutropha*, Containing High Poly-β-Hydroxybutyrate Levels, Regulates the Immune Response in Mussel Larvae Challenged with *Vibrio Coralliilyticus*.” *Fish & Shellfish Immunology* 84 (January): 196–203.

Walker, Robert J. 2016. “Population Growth and Its Implications for Global Security.” *American Journal of Economics and Sociology* 75 (4): 980–1004.

Walters, William, Embriette R. Hyde, Donna Berg-Lyons, Gail Ackermann, Greg Humphrey, Alma Parada, Jack A. Gilbert, et al. 2016. “Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys.” mSystems 1 (1).

Wang, Chun, Guoxiang Sun, Shuangshuang Li, Xian Li, and Ying Liu. 2018. “Intestinal Microbiota of Healthy and Unhealthy Atlantic Salmon *Salmo Salar* L. in a Recirculating Aquaculture System.” *Journal of Oceanology and Limnology* 36 (2): 414–26.

Wang, Miao, and Maixin Lu. 2016. “Tilapia Polyculture: A Global Review.” *Aquaculture Research* 47 (8): 2363–74.

Wang, Shaohui, Yuan Wang, Jie Ma, Yunchao Ding, and Shicui Zhang. 2011. “Phosvitin Plays a Critical Role in the Immunity of Zebrafish Embryos via Acting as a Pattern Recognition Receptor and an Antimicrobial Effector.” *The Journal of Biological Chemistry* 286 (25): 22653–64.

Wang, Xiaowei, and Brian Seed. 2003. “A PCR Primer Bank for Quantitative Gene Expression Analysis.” *Nucleic Acids Research* 31 (24): e154.

Wan, Ni, Mary Abernathy, Joseph Kuo-Hsiang Tang, Yinjie J. Tang, and Le You. 2015. “Cyanobacterial Photo-Driven Mixotrophic Metabolism and Its Advantages for Biosynthesis.” *Frontiers of Chemical Science and Engineering* 9 (3): 308–16.

Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, Temperton B. 2019. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* 7:e6800.

Webster, T. M. U., S. Consuegra, and M. Hitchings. 2018. “Inter-Population Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity.” *Applied and Environmental microbiology*. 84 (16): e00691-18.

Williams, Michael R., and Richard L. Gallo. 2015. “The Role of the Skin Microbiome in Atopic Dermatitis.” *Current Allergy and Asthma Reports* 15 (11): 65.
Willis, Amy D. 2019. “Rarefaction, Alpha Diversity, and Statistics.” *Frontiers in Microbiology* 10 (October): 2407.

Xia, Yinglin, and Jun Sun. 2017. “Hypothesis Testing and Statistical Analysis of Microbiome.” *Genes & Diseases* 4 (3): 138–48.

Yong, Ed. 2016. I Contain Multitudes: The Microbes Within Us and a Grander View of Life. *Random House*.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. “PEAR: A Fast and Accurate Illumina Paired-End reAd mergeR.” *Bioinformatics* 30 (5): 614–20.

Appendix

Rarefaction curve:

A method for estimating the number of species in any assembly, frequently used to predict the relationship between sampling effort and species to be found. (Gotelli and Colwell 2001), (Koblitz et al. 2017). Curves are created by randomly selecting an increasingly large subset of reads from a sample and counting the number of different species assigned to those reads (Gotelli and Colwell 2001). Asymptotic rarefaction curves suggest that sufficient sampling depth has been achieved to saturate sample diversity. Rapid plateauing of diversity with increasing numbers of reads is indicative of low-diversity samples. Samples which do not achieve asymptote indicate that sequencing depth was insufficient to capture the majority of community diversity. Less variation (less spread of the asymptotes) suggests similarity of community composition of the samples, while the highest variation shows dissimilarities of these communities. In our study, rearing water samples were less variant, while the highest variation was seen in the swab samples.

Phyloseq:

A package used as a tool to import, store, analyse and graphically display complex phylogenetic sequencing data that has already been clustered into Operational Taxonomic Units (OTUs) or Amplicon Sequence Variants (ASVs) (Koblitz et al. 2017).

Alpha diversity:

Alpha diversity metrics are used to measure and compare the bacterial community structure with respect to its richness (number of taxonomic groups) and evenness (distribution of abundances of the groups) within a sample. In microbial ecology, analysing the alpha diversity of amplicon sequencing data is a common first approach to assessing differences between environments (Willis 2019).

Shannon Diversity; an index commonly used to characterize species diversity in a community. To calculate Shannon diversity index (H'), this index takes both species abundance and species richness into account as seen in the equation below:

$$H' = -\sum_{i=1}^{S} p_i \, ln \, p_i$$

where S equals the number of species (species richness) and p_i equals the ratio of individuals of species i, divided by the total number of species.

To compare the alpha-diversity and the differential abundance of microbiome data, Shannon diversity of all samples was measured within **PhyloSeq** followed by the mean Shannon diversity for all sample types.

Beta Diversity:

Beta diversity refers to the measurement of the degree of difference in community membership or structure between two samples. If each taxon assigned to a sample is considered a parameter, then comparison between samples comprises calculating distances between samples, across all parameters in a high dimensional space. Multi-dimensional scaling is then used to represent the samples as points in low-dimensional space (usually two dimensions), such that the relative distance apart of all points are in the same rank order as the relative dissimilarities (or distances) of the samples in high dimensional space, as measured by some appropriate resemblance or distance matrix calculated on the abundance matrix. The interpretation of an MDS is straight forward: points that are close together represent samples that are similar in community composition, and points that are far apart correspond to communities with different structures (Clarke and Gorley 2006). Non-metric multidimensional scaling is a dimension reduction technique that is often used, in conjunction with Bray-Curtis distances, to evaluate distance between two samples and therefore their beta-diversity (Mutlu et al. 2014) (Clarke 1993).

ANOSIM

ANOSIM (Analysis of similarities) is used to compare within and between group similarity through a distance measure, to test the null hypothesis that the average rank similarity between samples within a group is the same as the average rank similarity between samples belonging to different groups. This test measures differences in community structure, and is one of the most widely used multivariate methods in microbiome studies (Clarke 1993) (Xia and Sun 2017).

First a dissimilarity matrix is constructed between samples using an appropriate measure, such as Bray-Curtis. The grouping factor is then permuted (swapped randomly) and the strength of the difference between the two groups is calculated under the null hypothesis that there is no assemblage difference between groups of samples specified by the levels of a single factor. The measured strength of difference in actual samples is then compared against the distribution of null-model permutations to determine a p-statistic for the observed value (Clarke and Gorley 2006).

SIMPER

The similarity percentages breakdown (SIMPER) procedure attempts to assess the average percent contribution of individual species to the dissimilarity between samples in a Bray-Curtis dissimilarity matrix. Its output is the most important taxa for discriminating between two pairs of groups. SIMPER calculates the mean sum of contributions from each taxon for all pairs of inter-group samples (e.g. all samples of the first group against all samples of the second group). The mean sum of contributions from each taxon between samples of the first two groups can be expressed in terms of the average contribution from each taxon. A good discriminating taxon contributes heavily to inter-group dissimilarity. These routines have been formalised in the software PRIMER. (Mumby 2001). In this study, One-way crossed SIMPER analysis was calculated, between different sample types, to estimate which taxon contribute to Bray-Curtis distances between sample types, highlighting the range of species responsible for defining the clustering pattern.