Mandatory Layout of Declaration/Statements

Word Count of thesis: DECLARATION	61,457
This work has not previously been accepted in substance for any degree and is not b concurrently submitted in candidature for any degree.	
Candidate name	Álvaro García Delgado
Signature:	
Date	23/08/22

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where ***correction services** have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signature:	
Date	23/08/22

[*this refers to the extent to which the text has been corrected by others]

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signature:	
Date	

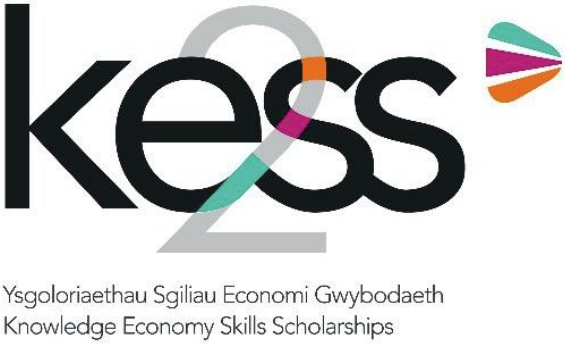
NB: *Candidates on whose behalf a bar on access (hard copy) has been approved by the University should use the following version of Statement 2:*

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by Aberystwyth University.

Signature:	
Date	29/07/22

Molecular approaches for pathogen assessment in remote aquaculture settings

Álvaro García Delgado



Fish Disease Dolau Ltd.

KESS Funding Statement:

Knowledge Economy Skills Scholarships (KESS 2) is a pan-Wales higher level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West Wales and the Valleys.

Summary

Aquaculture is a growing industry with an important role in providing a growing human population with protein. This thesis explores the use of a number of modern molecular biology techniques and technologies for detection of seven bacterial pathogens of aquaculture: *Aeromonas salmoncida*, *Flavobacterium branchiophilum*, *Flavobacterium psychrophilum*, *Moritella viscosa*, *Renibacterium salmoninarum*, *Tenacibaculum maritimum* and *Yersinia ruckeri*. It starts with a comparative genomics study, focused on *A. salmonicida*; the high genomic diversity of this bacterium highlights the value of sequencing pathogen genomes on a case-by-case basis. Nanopore sequencing was then employed to assemble the genomes of the seven bacteria, and although it has some advantages, such as low computational requirements, it performs worse several parameters than hybrid assemblies (which combine long and short read data). There was also an interest in providing a means of detecting these bacteria pre-emptively in aquaculture water by using a naïve surveillance system based on the detection of their DNA with nanopore sequencing. This led to an optimised method of obtaining DNA from river and seawater, consisting in centrifugation of the water and DNA extraction with a PowerSoil kit. Because this method did not yield enough DNA for a true metagenomic approach, other approaches to detection were considered. Loop-mediated isothermal amplification (LAMP) assays that could detect DNA of specific bacterial pathogens were developed for each of the bacteria mentioned above, with the exception of *A. salmonicida*. A proof of concept was also developed for a broad-range assay that combined LAMP and nanopore sequencing, the former to amplify DNA from a range of bacteria, the latter to discern the species of origin of the amplicons. Overall, this thesis could contribute to bring the advantages of molecular biology as close as possible to the fish farmer and improve monitoring and management of aquaculture pathogens.

Acknowledgements

The thesis you are about to read is the product of four years of working and learning. It has been a personal and academic journey which I could have never gone on without my supervisors: Drs Arwyn Edwards, Russ Morphew and John Prescott. For the opportunity to go on this journey, I must thank them. I actually wish to thank them for much more than just that: to thank John again, for his enthusiasm and motivation; Russ, for his generosity with the time he has lent to answering my queries and his good advice, and especially Arwyn, who has advised me well throughout this journey and has had more patience with me than I probably deserve.

I am also grateful for the Knowledge Economy Skills Scholarship (KESS2) program, funded by the Welsh European Funding Office, and Fish Disease Dolau Ltd., the company partner, for sponsoring this project and providing extra training. Without their participation, it would not have been possible. Specifically, I wish to thank Linda Cook, Helena Norris and Anwen Roberts, who have been the visible face of KESS for me throughout this whole period, for making the necessary bureaucracy manageable and understandable.

An important part of this thesis focuses on bioinformatics work and analyses; I wish to thank Drs Colin Sauze, Martin Swain, Kim Kenobi and André Soares for teaching me the basics of the Bash command language, Linux and the R programming language. I am also grateful to Supercomputing Wales and Aberystwyth University for enabling the analyses that relied on High Performance Computing. I am thankful to Dr Caio Ambrosio Leal Dutra, who helped me use a genome assembly program for the first time. Much appreciation goes as well to Odín Manuel Morón García, who has helped me develop my R skills.

For advice on how to sample and process water, I am grateful to Dr Sara Rassner, and to Richard Huxley, Gianmarco Sanfratello and André Soares for helping me collect those samples. For helpful discussions on how to set up LAMP experiments and discussions on experimental design, I thank Drs Chelsea Davis and Pallavee Srivastava.

A special thanks must go to Drs Gareth Griffiths and Andrew Detheridge, who have lent me many of the chemicals used in Chapter 4 and allowed me to use their lab on occasion, as well as kindly explaining how to use some of the equipment in that lab. I am especially grateful to Andrew, who has gone out of his way on a number of occasions to give me such training and explanations.

I would also like to express my gratitude towards the technicians of IBERS, Graham Brand, Rory Geoghegan, Mike Holland, Richard Huxley, Hefin Jones, Gareth Owen, Rajashree "Rosy" Swain, and Hilary Worgan, who have provided materials and training for using equipment throughout this project.

My lab colleagues also deserve acknowledgement for the general support they have given me thoughout the development of this project. They gave me a warm welcome when I arrived, helped me find my way around the new, unfamiliar lab and our lively discussions have not only been amusing and entertaining but also edifying and intellectually stimulating. So here's to Aliyah Debonnaire, André Soares, Cennydd Jones, Eleanor Furness, Guglielmo Persiani, Joe Dean, Keegan Burrows, Melanie Hay, Pallavee Srivastava, Sarah Easter and if I forget someone please forgive me, I have slept eight hours in the past three days. My sleep deprivation is not enough to make me forget to take my hat off for a second time to Drs André Soares and Pallavee Srivastava: these two people have been an inspiration as a result of their hard work and passion for science, and they have been exceptionally generous with their time and resources (thanks for the PowerSoil kit, Pallavee).

Finally, I must dedicate these following lines to my family, to whom I am grateful beyond what words can express. They have always supported me, blindly, and putting their heart into it. It is thanks to them that I am here today. It is to them that I say:

Thank you, to Pedro Antonio García Lopez, Esperanza Delgado Gonzalez, and Rodrigo García Delgado, from the bottom of my heart, for being there and having supported me all this time. I will print out a couple of copies of this thesis so you can use it as a paperweight and be reminded of me every time a breeze threatens to carry some papers away. To my extended family many, many thanks as well, although I will not force you to have this thing in your homes.

The last paragraph has been translated to Spanish for their benefit:

A Pedro Antonio García Lopez, Esperanza Delgado Gonzalez, y Rodrigo García Delgado; gracias, de corazón, por estar ahí, y haberme apoyado todo este tiempo. Imprimiré un par de copias de esta tesis para que la podáis usar de pisapapeles en casa y acordaros de mí cada vez que una brisa amenace con llevarse unos folios y este mazacote los tenga firmemente cogidos. A mi familia extendida, muchas, muchas gracias también… Aunque no os obligaré a tener esta cacharra en casa.

Table of contents

Table of figures

Table of tables

List of Abbreviations

Abbreviation	Full name
AHPND	Acute Hepatopancreatic Necrosis Disease
AMR	Anti-Microbial Resistance
ANI	Average Nucleotide Identity
ANOVA	Analysis of variance
BLAST	Basic Local Aligment Search Tool
bp	Base Pairs
BPOI	Bacterial Pathogen of Interest
CTAB	Cetyltrimethylammonium bromide
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FAO	Food and Agriculture Organisation of the United Nations
HPLC	High-Purity Liquid Chromatography
LAMP	Loop-mediated Isothermal Amplification
LOD	Limit of Detection
NCBI	National Centre for Biotechnology Information
NCIMB	National Collection of Industrial and Marine Bacteria
NEB	New England Biolabs
OIE	World Organisation for Animal Health
PATRIC	Pathosystems Resource Integration Center
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
SNP	Single Nucleotide Polymorphism
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
UV	Ultraviolet (light)
VF	Virulence Factor
WHO	World Health Organisation

Chapter 1: Main Introduction

This thesis explores several techniques that have risen to prominence in recent years in an attempt to improve disease detection in the aquaculture industry. The importance of the latter and the need for better disease detection and diagnosis systems is expounded in Chapter 1, alongside a description of the bacteria this thesis focuses on. The techniques and technologies used here were chosen for their potential to make detection simpler, faster, and more informative, with an emphasis on their ability to be carried out in the greatest possible proximity to the point of need. Chapter 2 provides examples of the value of whole-genome sequencing in the study of bacterial pathogens by considering the genomic diversity of seven bacterial species of interest, focusing on *Aeromonas salmonicida*. Chapter 3 considers the capacity of nanopore sequencing to provide useful genome assemblies that can serve to study bacterial pathogens, and whether this may be feasible in remote settings. Throughout Chapter 4, methods for the recovery of bacterial DNA from river and seawater (used as proxies for aquaculture water), are compared so as to delineate a method that can provide high yields for DNA-based pathogen detection tests. Loop-mediated isothermal amplification (LAMP), the focus of Chapter 5 and Chapter 6, offers a simpler, faster alternative to PCR for the detection of specific pathogens. Recently, multiplex assays that combine LAMP with nanopore sequencing have been used to diagnose several diseases; building a method that can be applied to an array of bacterial pathogens of aquaculture is the focus of Chapter 6. Overall, the achievements of this thesis are considered in Chapter 7, and so is the further work required to make a positive impact in aquaculture pathogen detection.

1.1. Introduction to aquaculture

Aquaculture is a practice that is growing in importance globally due to its expanding role in the provision of food for a world population in continuous growth. As of 2015 the world population was of 7,349.5 million people, up from 6,126.6 million in 2000, and yet the millennium development goal of halving hunger was almost met (United Nations Dept. of Economic Affairs, 2016). The importance of fish in human diets is substantial, as they provide 20.3 Kg of food per person and 7% of protein intake as a worldwide average (as of 2018, FAO, 2020), and fish consumption has been on the rise for decades. However, despite the rise in 31% of fish consumption between 1990 and 1997, fisheries only increased

production by 9% and there is evidence of a long-term stagnation in production (Garcia & Rosenberg, 2010; National Research Council, 2006; Tidwell & Allan, 2001).

Aquaculture, therefore, has been responsible for the majority of increases in fish supplies in recent decades and is forecast to maintain this trend until 2030 (Kobayashi et al., 2015)(see Figure 1-1). Its importance extends beyond the provision of food, as its economic and social aspects are also substantial: the industry produced 250 billion US dollars worth of fish in 2018 and the livelihood of over 20.5 million people worldwide is directly linked to the industry (FAO, 2020b).

Salmonid production is especially valuable, worth $22.8 billion in total (worldwide), and $6,415 a ton (FAO, 2020a). Salmon aquaculture, which initially started in the 1960s in Norway as a commercial enterprise, is now practiced in many regions over the world (FAO, 2021b). Norway has historically been in the lead of production, followed by Chile, the United Kingdom, Canada, the Faroe Islands and Australia (FAO, 2020a). Norway is in fact the world's second largest exporter of fish, owing to farm-reared Atlantic salmon (*Salmo salar*) in great part (FAO, 2020b). In the United Kingdom, Atlantic salmon is the major aquaculture fish, with 166,000 tonnes having been produced in 2018 (FAO, 2020b).

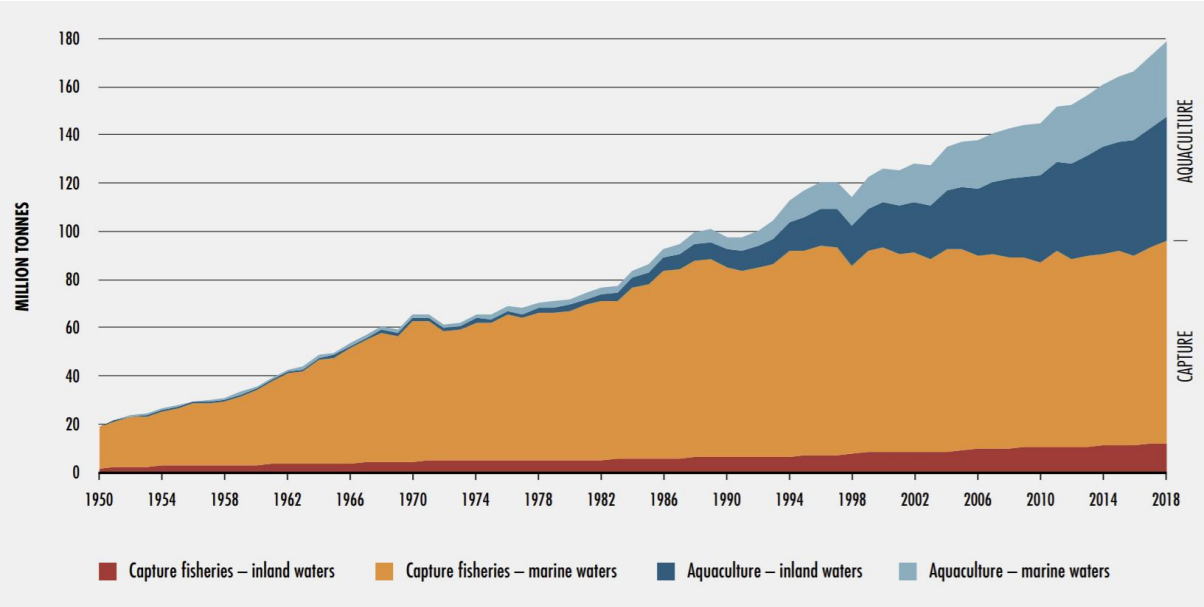


Figure 1-1: World aquatic organism production, separated into aquaculture produced or captured. From "The State of World Aquaculture and Fisheries", FAO 2020a.

1.2. Disease in aquaculture

Outbreaks can be of substantial economic importance, and especially damaging to those countries that heavily rely on this industry. Examples of serious losses due to epidemics in aquaculture can be attributed to both viruses and bacteria. One such viral outbreak was that of Infectious Salmon Anaemia Virus (ISAV) in 2007; the re-emergence of this disease caused Chile's salmon production to drop from near 400,000 tonnes in the year of 2005 to approximately 100,000 in 2010 (Asche et al., 2009). On the bacterial side, Acute Hepato-Pancreatic Necrosis Disorder (AHPND), detected for the first time in China in 2009, swept through the South-East Asian shrimp aquaculture industry during the late 2000s and early 2010s, causing severe losses. AHPND, caused by virulent *Vibrio* spp. arrived in Thailand in 2012; production decreased from a peak of 611,000 tonnes in 2011 to 280,000 in 2014 (Shinn et al, 2018). Economic losses to this country's economy were estimated at $11.6 billion in the period 2011-2016, and 100,000 jobs (Shinn et al, 2018). Of this disease, it should be noted that four years passed between the moment the disease was first noticed in 2009 and the description of the aetiological agent in 2013 (Tran et al., 2013). Within six months the first PCR assay was designed for it (Flegel & Lo, 2013), and soon afterwards the causative agent's genome sequence was announced (Yang et al., 2014).

Disease and its management in aquaculture has raised concerns beyond those directly involving the farmed fish. One such concern is that these operations serve as reservoirs for pathogens and parasites of wild fish. Experimental and phylogenetic evidence exists for the transmission of viral haemorrhagic septicaemia virus (VHSV) and salmonid alphavirus (SAV) between Atlantic salmon and wild finfish species in farm areas (Garver et al., 2013; Lovy et al., 2013; Snow et al., 2010). More attention has been paid to sea lice; infestation of wild salmonids has been related to their migratory routes past salmon farms (Krkošek et al., 2005), and infestations in farms have been correlated with declines in wild salmonid populations (Costello, 2009; Morton & Williams, 2003). This is concerning considering recent findings that have detected DNA from *Aeromonas*, *Tenacibaculum* and *Vibrio* genera, which include several pathogens of salmonid fish, in the microbiome associated to the *Caligus rogercresseyi* sea lice, which may be acting as a transmission vector or a reservoir (Gonçalves et al., 2020).

Another concern is the possibility that intensive aquaculture operations may be presenting opportunities for more dangerous pathogens to evolve. Pulkkinen and colleagues (2010) suggest that the high stocking densities of genetically homogeneous fish may provide an ideal

situation for disease transmission and that the competition of co-occurring strains favours those which are more virulent. Additionally, the demonstrated ability of aquatic bacteria to acquire genes via horizontal gene transmission (HGT) (Caro-Quintero et al., 2011; McDaniel et al., 2010) has led to the consideration of aquaculture operations as probable hotspots for the acquisition of antimicrobial resistance (AMR) (Watts et al., 2017). Bacterial strains that are resistant to antibiotics are selected for in farms where there is high antibiotic use (Higuera-Llantén et al., 2018), and, high prevalence of AMR genes in bacterial populations has been found in the proximity of fish farms (Muziasari et al., 2016).

The expansion of antibiotic resistance in aquaculture is a recognised threat for these operations. This has led some major aquaculture-producing countries, including those of the European Union, to regulate the use of antibiotics, by among other measures banning their prophylactic use (Watts et al., 2017). Norway has been credited with implementing stricter regulation and using careful management to reduce their use of antibiotics in the aquaculture industry; in 2015 the Norwegian government reported fish farming in the country used 99% less antibiotics than it did in 1987 (Høie et al., 2015). Infection prevention and control form an integral part of this careful management strategy.

1.3. Bacterial Pathogens of Interest

This thesis focuses on seven bacterial pathogens of salmonid fish (Table 1-1). These are the etiological agents of furunculosis (*Aeromonas salmonicida*), bacterial gill disease (*Flavobacterium branchiophilum*), coldwater disease, also known as rainbow trout fry syndrome (*Flavobacterium psychrophilum*), winter ulcer (*Moritella viscosa*), bacterial kidney disease (*Renibacterium salmoninarum*), tenacibaculosis (*Tenacibaculum maritimum*) and enteric redmouth disease (*Yersinia ruckeri*). They are denominated thoughout the thesis as Bacterial Pathogens of Interest (BPOI); the following section provides a description of them.

Aeromonas salmoncida, the causative agent of furunculosis, is the oldest known infectious agent of fish; it was described initially in 1896 (Lehmann & Neumann, 1896) as *Bacterium salmonicida*, although its name was revised in 1953 (Griffin et al., 1953). Currently, its distribution is global; Gulla and colleagues (2019) noted its presence in all continents except for Africa and Antarctica. The bacterium itself is a short, Gram-negative rod that produces a distinctive, diffusible brown pigment when in culture (B. Austin & Austin, 2007; Menanteau-Ledouble et al., 2016), although some strains or subspecies purportedly do not (D. A. Austin et al., 1989; Rouleau et al., 2018). The pathognomonic sign of furunculosis are furuncles, or red boils, that occur in the muscles during the chronic mode of the disease; when the disease is acute, anorexia and the presence of petechial haemorrhages also occur (B. Austin & Austin, 2007; Menanteau-Ledouble et al., 2016). In its peracute form, the disease can also lead to some necrosis of the gills and damage to the heart tissue (Bernoth, 1997; Roberts, 2012). This however, is not an extensive list of the symptoms, which are known to vary; it is also possible for infected fish to be asymptomatic (Hiney et al., 1994; Menanteau-Ledouble et al., 2016; O'Brien et al., 1994). Mortalities are variable for this disease, as they are for all those described here, often depending on the species or stock of the fish infected, its life stage, or the way the fish became infected. For a reference, challenge tests performed with a stock of Atlantic salmon (*Salmo salar*) that was expected to be relatively resistant to the disease saw mortalities of 66% after one month in this species, against the 93.8% of rainbow trout (*Oncorhynchus mykiss*) used for comparison (Holten-Andersen et al., 2012). However, as said, it can infect a variety of organisms; the "typical" strains are known as pathogens of salmon, hence the name, but there are a number of "atypical" strains that are known to affect other fish (Rouleau et al., 2018). Some strains have been isolated from warm-blooded animals, including a pied avocet (*Recurvirostra avosetta*), and humans (A. T. Vincent, Fernández-Bravo, et al., 2019; A. T. Vincent, Bernatchez, et al., 2019). During this thesis,

two strains of this species are used; the type strains of *A. salmonicida* subsp. *salmonicida* and *A. salmonicida* subsp. *achromogenes*, the latter considered "atypical".

Bacterial Gill Disease is caused by *Flavobacterium branchiophilum*, which grows on the gills of the fish, covering and damaging them; gill lamellae often fuse or become clubbed (H. Davis, 1926; Snieszko, 1981). The affected fish can die from lack of oxygen and accumulation of nitrogenous compounds, which they cannot exchange well due to the damage to the gills (Snieszko, 1981). There are no other externally visible signs, other than some behavioural changes. Mortalities, as stated before, depend on a number of factors; Ostland and colleagues carried out challenges with four different species of salmonids and a cyprinid, and observed mortalities between 0 and 75%, with the rainbow trout having the highest mortalities (Ostland et al., 1995). Although the disease was described for the first time in 1926 (H. Davis, 1926), the causative bacterium was not isolated and described until 1989 (Wakabayashi et al., 1989); it is a fastidious bacterium (Skulska, 2014). The bacteria are long Gram-negative rods that grow between 10 and 25 °C and form smooth, round colonies with a yellow pigment (Wakabayashi et al., 1989). This bacterium has been isolated throughout the northern hemisphere (Skulska, 2014).

Another *Flavobacterium* species, *F. psychrophilum*, is the causative agent of coldwater disease, also known as rainbow trout fry syndrome. This disease is characterised by necrosis of the peduncle which can be deep, sometimes reaching the spine (Bernardet & Grimont, 1989). It occurs, however, only when water temperatures are low (under 15 °C; Bernardet & Grimont, 1989). The bacterium, described for the first time in 1948 (Borg, 1948), has changed name twice (Bernardet et al., 1996; Bernardet & Grimont, 1989). Like others of its genus, it is a long, slender, Gram-negative rod, with some gliding motility but no flagella, whose colonies are yellow, circular, convex and with regular or spreading margins (Bernardet & Grimont, 1989). This species has been found around the world throughout much of the northern hemisphere, Australia and Chile (Nematollahi et al., 2003; Nicolas et al., 2008).

Moritella viscosa is among the BPOIs the most recent to be described (Lunder et al., 2000); it is the causative agent of winter ulcer, a disease that, at least upon its initial description (Lunder et al., 1995), did not appear to cause high mortalities in the Norwegian fish farms it affected (<10%). However, it is considered that the ulcerative skin lesions this pathogen causes the infected fish make them unmarketable (Colquhoun et al., 2004; Toranzo et al., 2005). It should be noted that some experimental challenges have been resulted in mortalities

of up to 33% in Atlantic salmon (*Salmo salar*) (Løvoll et al., 2009). The bacterium was described in 2000 (Lunder et al., 2000), but it was taxonomically reclassified soon after (Benediktsdóttir et al., 2000). *M. viscosa* is described as a motile, Gram-negative rod with a single polar flagellum that forms grey, translucent colonies that are round (Lunder et al., 2000). Its growth conditions are somewhat specific: the name of the disease is winter ulcer because these bacteria will not grow when water temperatures rise in the spring (over ~8 °C); additionally, they require salt water to survive, and do not grow in freshwater (Løvoll et al., 2009; Lunder et al., 1995). This species does not have a global distribution, it is mainly limited to the North Atlantic, and infects several species found in this range, as are Atlantic salmon, Atlantic cod (*Gadus morhua*) and lumpfish (*Cyclopterus lumpus*) (Karlsen et al., 2014).

Renibacterium salmoninarum is the causative agent of bacterial kidney disease (BKD), which is present throughout the northern hemisphere and Chile (Brynildsrud et al., 2014; Grayson et al., 1999). The symptoms are varied, and none of them is pathognomonic; pale gills, anaemia, exophthalmos, ascites, petechial haemorrhages on the body, vesicles with turbid liquid, etc. (Elliott, 2017). This disease can be slow to develop; Murray and colleagues experimentally infected a population of chinook salmon (*Oncorhynchus kisutch*), which developed mortalities between 15 and 50%, but this took 350 days (Murray et al., 1992). Indeed, *R. salmoninarum* is known to grow slowly; in its original description (Sanders & Fryer, 1980) the authors report it took 20-30 days from inoculation in a nutrient medium to reach its peak growth. They also described it as a Gram-positive, nonmotile, short rod that grows well between 15 and 18 °C. It is fastidious, as it requires cysteine and a means of detoxification to grow; the latter was initially provided with human blood or foetal calf serum (Sanders & Fryer, 1980), but Daly and Stevenson (1985) discovered the same effect could be achieved by including charcoal in the growth medium.

Tenacibaculum maritimum, described in 1986 (*Flexibacter maritimus*, Wakabayashi et al., 1986), changed name in 2001 (Suzuki et al., 2001); it is perhaps for this reason that the disease it causes has been known as “eroded mouth syndrome”, “black patch necrosis” or “gliding bacterial disease of fish” (Avendaño-Herrera et al., 2004). However, it is currently referred to as tenacibaculosis. It is characterised by ulcers and necroses all over the body of the infected fish, with frayed fins, tail rot and eroded mouth (Avendaño-Herrera et al., 2004; Masumura & Wakabayashi, 1977). It has been recorded throughout all the continents, with the exception of Africa and Antarctica, and is known to affect cultured Atlantic salmon,

although the disease was originally discovered in Seabream (*Pagrus major*) and Gilthead fry (*Acanthopagrus schlegeli*) (Avendaño-Herrera et al., 2004; Masumura & Wakabayashi, 1977). Mortalities were, during that original discovery, of 20 to 30%. *T. maritimum*'s cells are long, slender, Gram-negative rods similar to those of the *Flavobacterium* genus, but cells decrease in size after they are left for a long time in the same medium; colonies are flat, thin, with irregular edges and a light yellow colour (Wakabayashi et al., 1986).

Finally, the last of the BPOIs is *Yersinia ruckeri*, the etiological agent of enteric redmouth disease, which is characterised as the name indicates, by petechial haemorrhages lining the mouth of the infected fish (Pajdak-Czaus et al., 2019). It also receives the name yersinionsis, and in fact other symptoms are also common; the petechiae found around the mouth can also be found on the gut or other internal organs when dissecting the fish, the abdomen becomes distended, and exophthalmia or bloody irises also occur (Tobback et al., 2007). Mortalities can be of up to 70% in fry and juveniles of Atlantic salmon, but different serotypes have been found to have different degrees of virulence (Haig et al., 2011). *Y. ruckeri* cells are Gram-negative rods with peritrichous flagella that grow optimally at 22 to 25 °C according to Ewing and colleagues (1978) in their original description of the species.

Of the diseases caused by these bacteria, the majority of them were reported as problematic in Europe for rainbow trout (flavobacteriosis/rainbow trout fry syndrome, bacterial kidney disease, enteric redmouth and furunculosis) and salmon (flavobacteriosis, furunculosis and winter ulcer) in 2016 (Olesen & Vendramin, 2017). The reports of the Annual Workshops of the National Reference Laboratories for Fish Diseases have repeatedly included these diseases in the following years, bacterial gill disease and tenacibaculosis have also appeared throughout the 2016-2020 period (Olesen et al., 2018; Vendramin & Olesen, 2021). The Food and Agriculture Organisation of the United Nations also considers furunculosis, bacterial kidney disease, winter sores and enteric redmouth disease to be some of the main diseases in Atlantic salmon aquaculture (FAO, 2021b). Furunculosis, bacterial kidney disease and bacterial gill disease are considered have the same consideration in the case of rainbow trout (FAO, 2021a).

These diseases are problematic outside Europe as well. For instance, in Canada *A. salmoncida* accounts for 30% of the cases of fish disease recieved for analysis infections by the animal health laboratory of the Ministry of Agriculture, Fisheries and Food (RAIZO, 2018). In fact, antibiotic resistance is a recognised problem in Canada as well; certain strains

of this species are known to be resistant to all legally permitted antibiotics (Trudel et al., 2016). Recent reviews have also considered *R. salmoninarum* to be problematic in Chile, where 20.2% of samples of diseased fish registered by the National Fishery Service had an infection of this pathogen (Figueroa et al., 2019). Specifically, according to these authors, 72.6% of these cases were found in Atlantic salmon, an important species for Chilean aquaculture. Information from these sources and the suggestions of the company partner, Fish Disease Dolau Ltd., were considered in the selection of this set of pathogens as a focus of this thesis.

1.4. Current disease monitoring – diagnostic tests and surveillance

An essential element of disease control is monitoring. Monitoring involves sampling diseased animals to examine them, and even healthy ones if a disease is suspected; environmental tests are also considered to be able to provide valuable information (Adams & Thompson, 2011; OIE, 2021). Examination of these samples has traditionally involved a preliminary assessment of gross clinical signs, followed by culture of the pathogen from the locations of the infection and biochemical identification. This is known to take days or weeks in the cases of some pathogens (Avendaño-Herrera et al., 2004; Sanders & Fryer, 1980). Histology often plays a part in the classical approach to diagnosing disease.

More recently, a number of immunology-based tests have been used, which have the advantage of avoiding the culture of the pathogen. These take a variety of forms, including immunofluorescence assays, which use fluorochrome-bound antibodies to locate specific pathogens on histological sections of infected tissues (for examples see: Panangala et al., 2006; Speare et al., 1995). This technique requires a fluorescent microscope. Enzyme-Linked ImmunoSorbent Assays (ELISA), in their "sandwich" version, use an antibody to bind an antigen from a pathogen in a sample; another antibody bound to an enzyme is then applied, and finally the substrate of said enzyme is added, so that a visible reaction (ie., a colour change) occurs (for examples see: Adams & Thompson, 1990; MacPhee et al., 1995). Lateral flow assays follow a similar principle, with the analyte flowing thorough a nitrocellulose membrane that contains gold-labelled antibodies that bind to the antigen in the analyte. The antigen is also bound by another set of antibodies in a line on the nitrocellulose membrane, capturing the gold-labelled antigen-antibody complexes; this line is then easily visible. This technique, due to its simplicity, has been applied to a number of viral pathogens of shrimp, as

it is particularly useful in the low-resource settings where this crop is often grown (for examples see: P. Sithigorngul et al., 2011; W. Sithigorngul et al., 2007).

The "gold standard", however, is PCR. It is the recommended testing method by the OIE for its listed pathogens (OIE, 2021). Invented in the 1980s, the principle is well known; briefly, it uses a polymerase and two oligonucleotide primers to amplify the segment of a DNA molecule that lies between the complementary regions where the primers bind (Mullis et al., 1986). This amplification of a very specific target sequence has allowed its use as a diagnostic tool. Furthermore, it has been adapted a number of times to improve its capabilities; the addition of a reverse-transcriptase enzyme allows it to detect RNA viruses (RT-PCR); the use of nested primers to amplify samples with exceptionally low concentrations of target (nested PCR, Mullis et al., 1986); the combination of several primer sets into one assay to detect several targets simultaneously (multiplex PCR), or the use of fluorescently tagged probes to detect amplification as it takes place (quantitative PCR, or qPCR).

This technique is very sensitive, and very popular; assays have existed for the seven bacterial pathogens of interest (BPOIs) this thesis focuses on for at least ten years (see Table 1-1). The limits of detection reported in these papers are low; for instance Gustafson and colleagues (1992) detect ten colony-forming units (CFUs) of *Aeromonas salmonicida* per milligram of salmon tissue; Brown and colleagues (1994) two cells of *Renibacterium salmoninarum* per egg; and Gibello and colleagues (1999) detected down to 650 cells of *Yersinia ruckeri* per mL of sample. Often reports of qPCR assays are accompanied by a comparison to standard PCR, and demonstrate better performance (Chase & Pascho, 1998; Fringuelli et al., 2012; Powell et al., 2005). In any case, PCR has received criticism for being a complex technique that needs of specialised technicians and requires cumbersome and expensive equipment (Foo et al., 2017; Ganguli et al., 2020; W. Sithigorngul et al., 2007; Tomita et al., 2008). Therefore, despite its good performance, it is not practical in some settings, especially those where there are few resources available.

Table 1-1: List of diseases, causative agents, and PCR assays for the bacterial pathogens that are the focus of this thesis.

Disease	Species	First description	First PCR assay published
Furunculosis	*Aeromonas salmonicida*	Lehmann & Neumann, 1896	Gustafson et al., 1992
Bacterial Gill disease	*Flavobacterium branchiophilum*	Wakabayashi et al., 1989	Toyama et al., 1996
Coldwater disease	*Flavobacterium psychrophilum*	Borg, 1948	Toyama et al., 1994
Winter ulcer	*Moritella viscosa*	Lunder et al., 2000	Grove et al., 2008
Bacterial Kidney Disease	*Renibacterium salmoninarum*	Sanders & Fryer, 1980	Brown et al., 1994
Tenacibaculosis	*Tenacibaculum maritimum*	Wakabayashi et al., 1986	Toyama et al., 1996
Enteric redmouth disease	*Yersinia ruckeri*	Ewing et al., 1978	Gibello et al., 1999

More recently, a number of other techniques have arisen that serve as diagnostic tools. One of the most appealing is a nucleic acid amplification technique termed LAMP (an abbreviation of Loop-mediated isothermal AMPlification). This technique, invented by Notomi and colleagues (2000), is particularly well suited to diagnostics because it is less likely to react with non-target DNA and is usually found to detect lower quantities of its target than PCR (Kulkarni et al., 2009a, 2009b; Picón-Camacho et al., 2013; Saleh et al., 2008a, 2008b; Yu et al., 2013, etc.). It has recently come to attention as a result of the coronavirus disease (COVID-19) pandemic, as a number of researchers have proposed it as an alternative nucleic acid test to qPCR (Chow et al., 2020; Dao Thi et al., 2020; Ganguli et al., 2020). Not only does it offer better performance, but it is also faster than PCR, and it is isothermal, requiring simpler equipment. It is therefore seen as an appropriate alternative to PCR for low-resource settings (Cuadros et al., 2015; Imai et al., 2018; Tomita et al., 2008). It has already been applied to a number of viral and bacterial pathogens of cultured fish (Cai et al., 2010; Kulkarni et al., 2009a, 2009b; Soliman et al., 2009). A more complete description of the technique is given in Chapter 5; part of this thesis focuses on its potential as a means of detecting several bacterial pathogens of cultured salmonids.

1.5. DNA sequencing

DNA sequencing has advanced substantially in the last decades. New techniques have been developed that allow the massively parallel sequencing of numerous molecules of DNA; not only have these enabled new applications, but also reduced the costs, bringing them within the reach of most molecular biology laboratories. These methods, which form what is called high-throughput sequencing are diverse; two of the most commonly used, and the ones employed during this thesis, are sequencing-by-synthesis and nanopore sequencing. Sequencing by synthesis is more established, and is estimated to provide 90% of the world's sequencing data (Illumina Inc., 2021). This technology is accurate, producing highly precise short reads that have been employed for a range of projects including the reconstruction of genomes, metabarcoding, and metagenomics (Peters et al., 2018; A. T. Vincent, Fernández-Bravo, et al., 2019; Xu et al., 2020). Nanopore sequencing, on the other hand, produces longer, less accurate reads, which have been used for similar purposes (B. L. Brown et al., 2017; Hamner et al., 2019; Loman et al., 2015), but has a different set of strengths and weaknesses; these are examined in Chapter 3. One such strength is the capacity to function in field settings using minimal equipment; another is its capacity to provide data immediately from the instant the sample is loaded into it. These abilities set it as a potentially useful tool for application in resource-poor settings, remote locations and time-sensitive contexts.

Among these useful applications is genome sequencing, which has allowed insights into the diversity of life and its functioning. Increased availability of data from Whole Genome Sequencing (WGS) has underpinned the rise in digital methods of taxonomic distinction (Konstantinidis & Tiedje, 2005), informed our understanding of bacterial evolution (A. T. Vincent, Fernández-Bravo, et al., 2019; Zhong et al., 2019), or served to study the mechanisms behind pathogens' virulence (C.-T. Lee et al., 2015), just to name some of those applications. The information derived from these genomes is valuable for phylogeography studies that allow to understand how these pathogens have spread (Brynildsrud et al., 2014; Gray et al., 2011), and can serve to find the presence of genes related to AMR (Su et al., 2019). These many utilities are in part what drives the interest of several public health agencies in applying pathogen genome sequencing to human health (Grant et al., 2018; Nadon et al., 2017; Revez et al., 2017).

Another application of high-throughput sequencing is the study of environmental DNA. The use of shotgun sequencing has several qualities that make it a potentially valuable tool for the detection of pathogens in aquaculture. One of them is the fact that it is untargeted, and

therefore it sequences anything from within a DNA sample. This means that a single test based on metagenomics could gather data about presence of a large number of pathogens (Bibby et al., 2011; Rosario et al., 2009). Another beneficial characteristic of metagenomics is that it uses an environmental sample, avoiding the need for culture; as a result, detection is not limited to cultivable organisms, nor delayed by those with slow growth. Limitations are acknowledged about this approach, such as bias introduced during the extraction of DNA, preparation of metagenomic libraries, or bioinformatics analysis, but means of correcting them have been devised (McLaren et al., 2019; Walden et al., 2017). The information gained through metagenomics is not only taxonomic, but also, much like with WGS, it can serve to identify AMR and virulence genes (de Nies et al., 2021; Durso et al., 2011).

1.6. Thesis aims

This thesis focuses on the use of relatively recent molecular biology techniques and sequencing technologies to study the BPOIs. The purpose is to provide tools for their rapid, simple detection in the field, and to evaluate the potential for these sequencing technologies to provide relevant information as close as possible to the point of need. This is expected to be beneficial to disease detection and monitoring, as it can make it more accessible to farmers, especially those in remote areas, and reduce the time it takes for meaningful, targeted action to be taken.

The aims of this thesis are therefore:

1) To evaluate the benefits of nanopore sequencing on aquaculture disease detection and management
2) To ensure enough DNA can be obtained from aquaculture water for DNA-based tests
3) To produce a LAMP assay for a number of bacterial pathogens of aquaculture

Chapter 2: Comparative genomics of bacterial pathogens of interest

2.1. Introduction

The falling cost of DNA sequencing in recent decades has been followed by an increase in the publicly available sequence data. By the estimates of the National Centre for Biotechnology Information (NCBI), the contents of the GenBank databases have approximately doubled every 18 months since 1982 (information on size regularly updated in: ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt). This increase has provided opportunities to obtain insights in a number of fields, including taxonomy, epidemiology and others where abundant genetic and genomic data is of benefit.

Specifically, taxonomy has benefitted thanks to the availability of genome sequences for type strains. Average Nucleotide Identity (ANI) is a technique that compares the sequence identity between the shared genes of the genomes of two organisms; this technique was proposed in 2005 for its robustness in the comparison of strain relationships (Konstantinidis & Tiedje, 2005). Until then, it was common to use DNA-DNA hybridisation techniques, with a 70% relatedness being used as threshold (Wayne et al., 1987). This technique has limitations, one of them being the existence of a variable component to bacterial genomes even within the same species (up to 30% of gene content, according to Konstantinidis & Tiedje, 2005). Because some organisms with DNA-DNA hybridisation values above the mentioned threshold were separated into different species by virtue of some phenotypic characteristics, Konstantinidis and Tiedje (2005) considered this method inconsistent. Their proposal to use ANI for systematics set the species separation threshold at 94%, a value that approximately corresponds to the 70% DNA-DNA hybridisation, to 99% 16S rRNA gene identity (M. Kim et al., 2014), and to a dip in ANI values between bacterial genomes of different species (M. Kim et al., 2014; Rosselló-Móra & Amann, 2015). ANI continues to be used as a means of discriminating species (Colston et al., 2014; Kumru et al., 2020) and it has been advocated that the description of new prokaryotic and archaeal species should be accompanied by the publication of a high-quality genome sequence that can serve, among other things, for these comparisons (Rosselló-Móra & Amann, 2015).

Taxonomy has also benefited from the availability of genome sequences thanks to an improvement in the potential for phylogenetic analyses. As was agreed in the 1987 Workshop

for Reconciliation of Approaches to Bacterial Systematics, whole genome sequences should be the "reference standard" for phylogeny, and that phylogeny should underpin taxonomy (Wayne et al., 1987). The increased availability of genome data, thus, allows departure from the simpler, but potentially less precise, use of single loci, or a limited number of loci, in the elaboration of bacterial phylogenies. Indeed, despite its historic use in phylogenetic studies (Woese & Fox, 1977; Zhulin, 2016), 16S rRNA may be scarcely differentiated between some closely related species (Fox et al., 1992; Martínez-Murcia et al., 2005). The use of complete genomes can also yield greater phylogenetic signal when bacteria are closely related and there is little variation at any single locus (Achtman, 2008; Bayliss et al., 2018). The phylogenetic signal that complete genome sequences allows exploiting has even been useful, in combination with geographic location data, in epidemiological studies that track the spread of pathogens (Baker et al., 2008; Bayliss et al., 2018; Brynildsrud et al., 2014).

Another field of study that benefits from the availability of abundant genome data is pan-genomics. Pan-genomics is the study of the genetic composition of a group instead of a representative individual (Tettelin et al., 2005). It consists in the determination of a "core" genome, that is, a set of genes common to the entire set of organisms being examined, and a set of accessory genes, found only in groups of strains or single strains (Tettelin et al., 2005). These analyses improve our understanding of bacterial genomic diversity, and have a number of uses. For instance, core genomes are often used in phylogenetic analyses (Kumru et al., 2020; Roig et al., 2018; Touchon et al., 2009), as they provide a means of comparing the shared portion of the genome, thus ignoring the effects of horizontal gene transfer. Another use is the development of vaccines that are valid for the entire range of strains of a given pathogen, by targeting proteins encoded in the core genome (Naz et al., 2019; Shahid et al., 2021). Another facet of genomic diversity that is of considerable relevance when dealing with pathogenic bacteria of aquaculture is antimicrobial resistance (AMR) genes. Resistance can be intrinsic to a bacterium, and can become prevalent in the population as a result of selection pressure, but the frequent presence of AMR genes on mobile genetic elements, as are plasmids or bacteriophages, can accelerate their expansion (Michael et al., 2014; Opal & Pop-Vicas, 2015). Antimicrobial resistance of pathogenic bacteria is a widely recognised problem for global health (Michael et al., 2014; WHO, 2014). The genes driving this resistance are of particular interest to aquaculture, as antimicrobials are still currently extensively used in disease management (Schar et al., 2020).

Additionally, there was an interest in exploring the use of pan-genomics as a means of designing assays that were both specific to the BPOIs and that were capable of targeting the whole range of strains of each bacterium. It was expected that genes present in the core genome would constitute ideal targets for detection and diagnostic assays. Although this specific interest was later addressed using GLAPD, a software that designs specific and generic primers automatically (Jia et al., 2019, see Chapter 5), pan-genomics analyses of the BPOIs were considered intriguing for two reasons. On one hand, a number of existing comparative genomic studies for our own target bacteria incorporated less complete datasets (Barnes et al., 2016; Duchaud et al., 2018; Karlsen et al., 2017; A. T. Vincent et al., 2017). On the other, the comparative genomics study of the strains of a species would provide insights into the diversity of our bacterial pathogens that could inform as to the usefulness of targeted LAMP which is the focus of Chapter 5. Due to this, the following aim and objectives were established:

2.1.1. Aims and objectives:

Aim:

1) To study the genomic diversity of the Bacterial Pathogens Of Interest (BPOIs) of this thesis

This was divided into the following objectives:

1) To carry out pan-genomic analyses of all the BPOIs
2) To carry out comparative genomics analyses at least for one case study, in order to better understand the diversity of this bacterial pathogen

2.2. Methods

2.2.1. Data collection

Most of the publicly available genomes for each species of bacterial pathogen of interest were collected for the analyses carried out in this chapter. The list of genome accessions was obtained through the PATRIC (v3.6.10) (J. J. Davis et al., 2019) online platform, where the search term was the species name. The genome records for each of these species were grouped together, the list downloaded in table format and the accession numbers collected from this file. These were then used to download all records from the NCBI's Assembly database (NCBI Resource Coordinators et al., 2018) via the NCBI datasets download tool automatically (Sayers et al., 2021; https://www.ncbi.nlm.nih.gov/datasets/). When the software warned that it could not retrieve certain accessions, these were downloaded manually.

Besides these publicly available genomes, the genomes of the type strains of the eight BPOIs were sequenced with the ONT MinION and an Illumina HiSeq, and assembled in-house using Unicycler (R. R. Wick et al., 2017); detailed methods are given in Chapter 3. These genomes were added to the local datasets, and uploaded onto PATRIC (marked as private data). The lists of strain names and genome accessions are included in Appendix 1. Additionally, representative genomes from closely related species were included in the PATRIC datasets, so as to gain contextual information and be able to root phylogenetic trees; these are listed in Table 2-1. Because the limit for phylogenetic analyses on PATRIC is set at 100 genomes, and two representative genomes of closely related species were needed to root the tree, a total of 98 *Flavobacterium psychrophilum* genomes were gathered in the dataset. The exclusion criterion for the remaining 61 available genomes was that they were the more fragmented assemblies (greater number of contigs).

Table 2-1: Number of genomes per species included in the datasets used for the pan-genomic analyses and phylogenetic analyses; when the number used for either analysis is different, the number in brackets refers to the number of genomes used in the phylogenetic analysis. Among the outgroup genomes, those in bold were used to root the tree.

Species	Number of genomes	Outgroups
Aeromonas salmonicida	70 (74)	*Aeromonas bestiarum* GA97-22 *Aeromonas diversa* 2478-85 ***Aeromonas hydrophila* OnP3.1** *Aeromonas piscicola* LMG 24783
Flavobacterium branchiophilum	6	***Flavobacterium aquatile* LMG 4008** *Flavobacterium frigidarium* DSM 17623 *Flavobacterium limicola* DSM 15094 *Flavobacterium tiangeerense* CGMCC 1.6847 *Flavobacterium urumqiense* CGMCC 1.9230
Flavobacterium psychrophilum	99	***Flavobacterium aquatile* LMG 4008** *Flavobacterium swingsii* DSM 21789
Moritella viscosa	14	***Moritella dasanensis* ArB 0140** *Moritella marina* ATCC 15381
Renibacterium salmoninarum	4	***Psychromicrobium silvestre* DSM 102047**
Tenacibaculum maritimum	25	***Tenacibaculum adriaticum* DSM 18961** *Tenacibaculum soleae* UCD-KL19
Yersinia ruckeri	70	***Yersinia enterocolitica* FORC 002** *Yersinia kristensenii* NCTC11471

2.2.2. Pan-genomic analyses

Features of the genomes downloaded from NCBI Assembly were annotated using Prokka v1.12 (Seemann, 2014). The annotated genome files in GFF3 format were then used as input for Roary v3.13.0 (Page et al., 2015), which was used to calculate the pan-genome for each species. Prokka and Roary were both run with the default parameters, except for the option to create a multi-fasta alignment of core genes with MAFFT being selected on Roary. This

option was selected because one of the initial purposes of the pan-genomic analyses was to search for conserved regions of the genome that would serve in the development of LAMP primers that could target all strains of the species. Boxplots depicting the number of genes in the pan-genome as more genomes were built with a modified version of the create_pan_genome_plots.R script bundled with Roary.

2.2.3. Phylogenetic analyses and ANI studies

Phylogenetic and ANI analyses were carried out on each dataset in order to gain contextual information about the genomes of each bacterial species and to enable the detection of genomes incorrectly assigned to the species.

2.2.3.1. Phylogenetic analysis

The Maximum Likelihood phylogenetic analyses were carried out on the PATRIC (v3.6.10) (J. J. Davis et al., 2019) platform via the CodonTrees pipeline, using the dataset described in section 2.2.1. PATRIC allows selection of the number of genes to be used in the matrix that will serve as basis for the analysis and the number of rapid bootstraps that should be carried out to obtain branch support values; both were set at 100. The process subsequently followed by PATRIC is described in brief here. A random set of 100 protein-coding genes among those in PATRIC's global protein families database (J. J. Davis et al., 2016) that are present throughout the genomes in the dataset are selected; both the nucleotide and translated amino acid sequences are then aligned using MUSCLE (Edgar, 2004) and concatenated into a partitioned matrix. A Maximum Likelihood phylogenetic analysis is then run with RAxML, using 100 rapid bootstraps to assess clade support, with GTR+G and FLU substitution models for the nucleotide and amino acid partitions respectively (Stamatakis, 2014; Stamatakis et al., 2008). The resulting phylogenetic trees in Newick format were downloaded from the online platform and visualised and edited in FigTree v1.4.4 (Rambaut et al., 2018). Trees were rooted using the most distant relative known among the outgroups; these are written in bold in Table 2-1. The most distant relative was surmised based on a preliminary Maximum Likelihood phylogenetic analysis of the 16S rRNA genes of the type strains of all known species within each genus; a more detailed description of the procedure is given for *Tenacibaculum maritimum* in section 3.2.3.2.

2.2.3.2. ANI analysis

Average nucleotide identities (ANIs) represent the proportion of identical bases in the aligned coding regions of two genome sequences. As a measure of genomic similarity, they are often

used in bacterial taxonomy to distinguish different species. Here the ANIs between each pair of genomes were calculated using Pyani v0.2.10 (Pritchard et al., 2016); all the genomes included in the dataset described in section 2.2.1 were used in these calculations, excluding the outgroups. Pyani was set to carry out the ANI calculations using MUMmer (v3, Kurtz et al., 2004) to align coding regions of the genomes. The final ANI matrix of *Aeromonas salmonicida* was visualised with the R package ggplot2 (Wickham, 2016) and the colour scheme "viridis" (Garnier et al., 2021).

2.2.4. Search for antimicrobial resistance genes

Given the relevance of antimicrobial resistance of pathogenic bacteria in the aquaculture industry (Watts et al., 2017), all the genomes used in the phylogenetic analysis of *A. salmonicida* were searched for horizontally acquired genes responsible for antimicrobial resistance (AMR). The tool Abricate v1.0.1(Seemann, 2014/2020) was used to scan a database for features relating to antibiotic resistance. The database used was the Comprehensive Antibiotic Resistance Database (CARD, (Alcock et al., 2019), because it provides detailed information on the AMR genes it lists. To summarise the output of Abricate, a plot combining an AMR gene presence/absence matrix with the phylogenetic tree generated in section 2.2.3.1 was built. This was done on R 3.6.1 (R Core Team, 2019), using the packages ape (Paradis and Schliep, 2019), ggplot2 (Wickham, 2016), ggdendro (de Vries et al., 2016), phylogram (Wilkinson & Davy, 2018), grid (a core R package) and reshape2 (Wickham, 2007).

2.3. Results

2.3.1. Overview of phylogenetic and ANI analyses

The ANI and phylogenetic analyses of the majority of species served to identify genomes that could not be considered to be part of the species of interest according to the generally accepted criteria for ANI (identities between all members of a species > 95-96% Goris et al., 2007; M. Kim et al., 2014; Rosselló-Móra & Amann, 2015) and phylogeny (all members of the species form a clade). Across all the seven datasets only three strains were identified not to belong to the species they had been assigned. Two strains classified as *F. psychrophilum* and one of *A. salmonicida*. The phylogenetic tree of *F. psychrophilum*, rooted at *Flavobacterium aquatile*, shows that strains Z1 and Z2 are more closely related to *Flavobacterium swingsii* than to the rest of the *F. psychrophilum* clade. Additionally, the ANIs of these two strains with any other members of the *F. psychrophilum* species are below 0.83. In the case of *A. salmonicida*, the genome from strain CBA100 was found to cluster more closely with *Aeromonas bestiarum* and *Aeromonas piscicola* than with *A. salmonicida*. It's ANIs when compared with any other member of *A. salmonicida* were between 0.90397 and 0.90737; on the other hand, when compared to *A. bestiarum* this rose to 0.97204, indicating it probably belonged to this species. As a consequence, these three genomes were removed from their respective datasets for the pan-genomic analyses.

2.3.2. Pan-genomic analyses

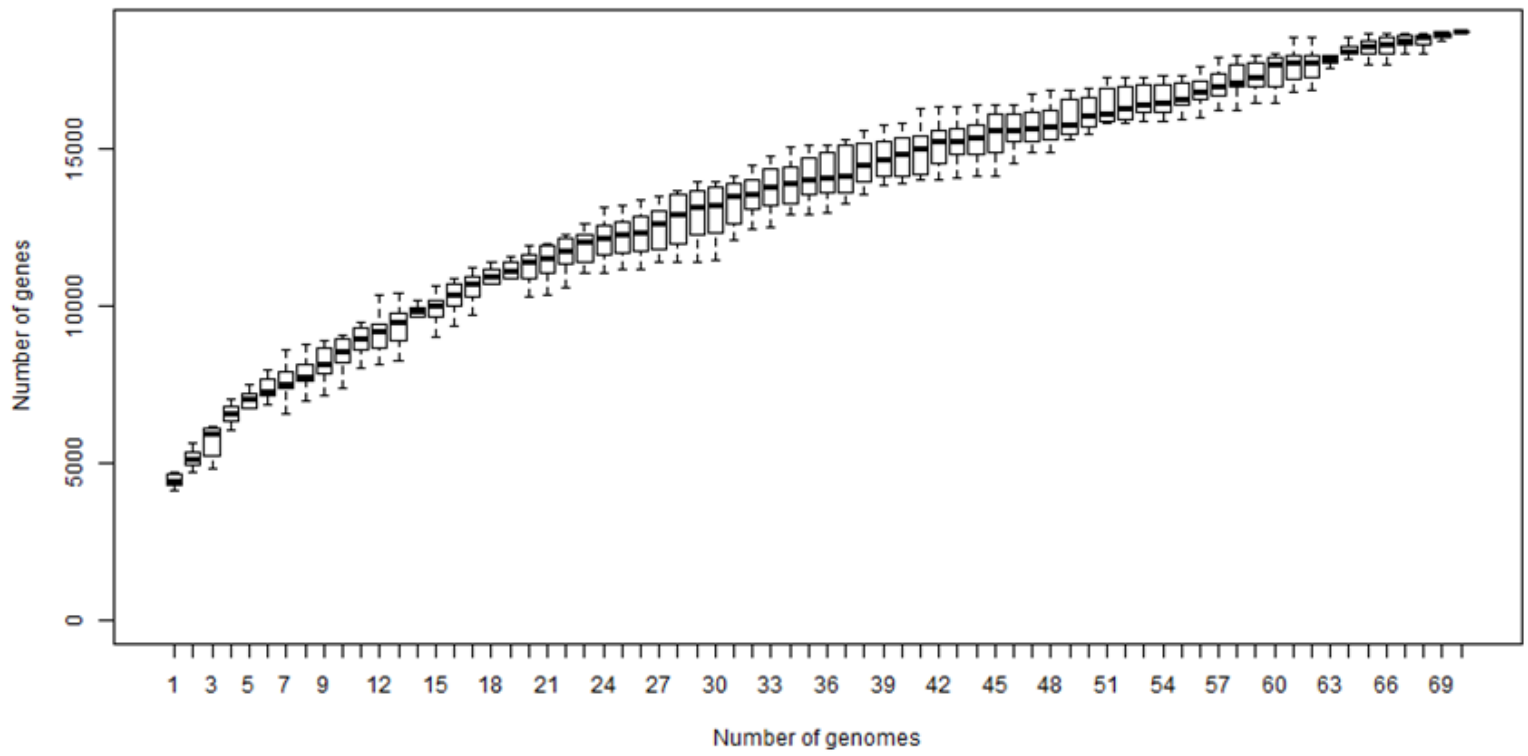
Roary identified the different genes in each genome and classified them as belonging to the core, soft core, shell and cloud categories. Genes falling into each of these categories are defined as those present in >99%, 95-99%, 15-95% and 0-15% of the genomes analysed respectively. The results are summarised in Table 2-2, where it can be seen that the species with the proportionally largest core pan-genome was *Renibacterium salmoninarum*, followed by *Moritella viscosa* and *Flavobacterium branchiophilum*. These have, respectively, 97.98% (3,486), 47.42% (3,547) and 46.51% (2,127) of their genes classified as "core", but in no cases do they have any classified as soft-core and with the exception of *M. viscosa* (28.11%, 2,103) have any classified as cloud; it is worth noting that this is a consequence of the small size of their datasets. The species with larger datasets are, unsurprisingly, the ones with the smaller core genomes (*A. salmonicida*, *F. psychrophilum*, *Tenacibaculum maritimum*, *Yersinia ruckeri*). The smallest core genome is that of *A. salmonicida*, with 8.07% (1509) of all the genes found; this species also has the highest percentage of cloud genes, 72.38% (13,528).

Table 2-2: Total number of genes found in each species' pan-genome and the percentage of them pertaining to each of the core and accessory pan-genome categories. Core genes are present in >99% of genomes, soft core genes in 95-99%, shell genes in 15-95% and cloud genes in less than 15%.

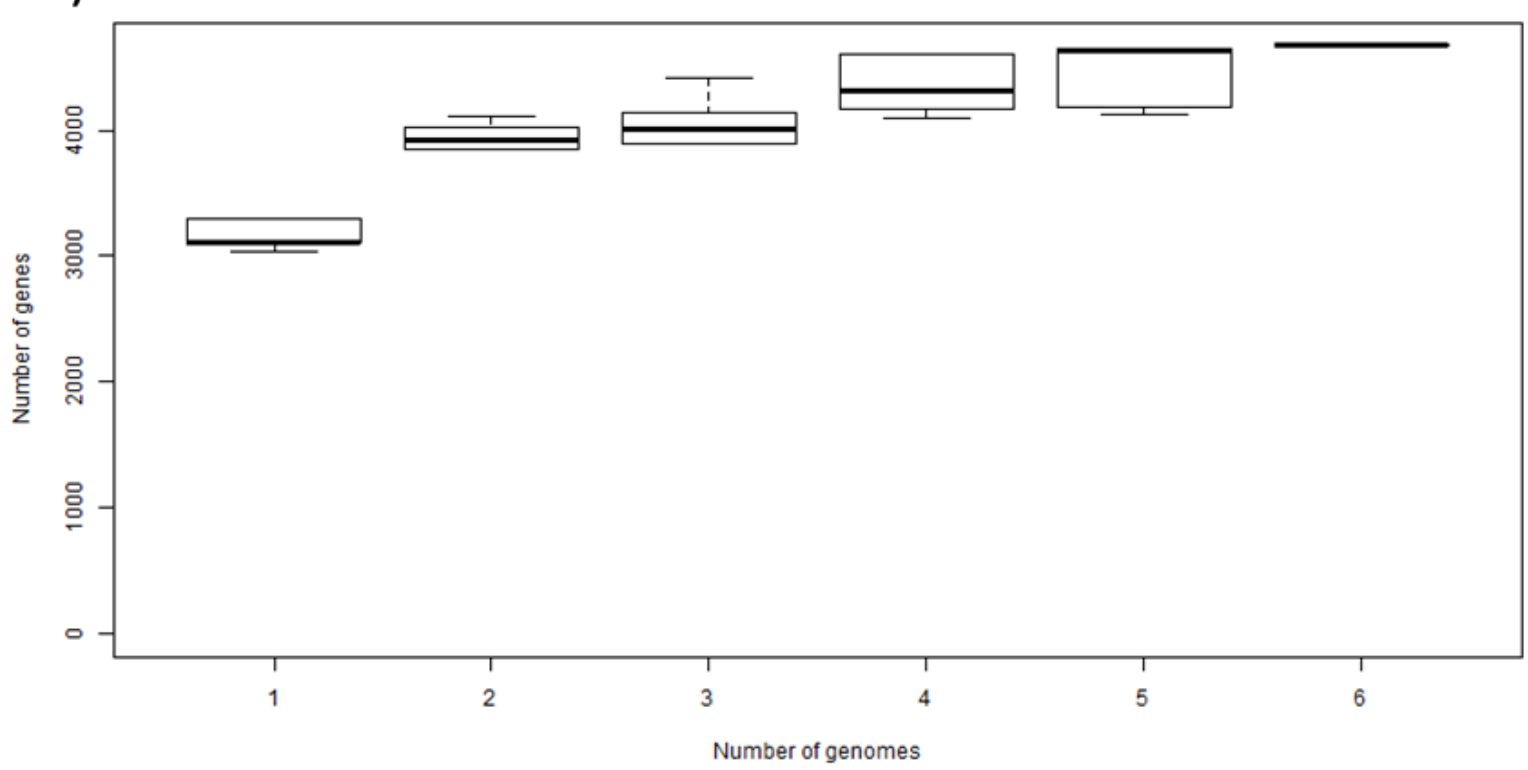
Species	% Core genes	% Soft core genes	% Shell genes	% Cloud genes	Total genes
Aeromonas salmonicida	8.07	5.93	13.61	72.38	18,689
Flavobacterium branchiophilum	46.51	0.00	53.49	0.00	4,670
Flavobacterium psychrophilum	17.83	6.45	13.67	62.05	7,175
Moritella viscosa	47.42	0.00	24.47	28.11	7,480
Renibacterium salmoninarum	97.98	0.00	2.02	0.00	3,558
Tenacibaculum maritimum	29.86	1.29	19.57	49.28	7,133
Yersinia ruckeri	41.76	5.70	14.16	38.38	6,125

The figures depicting how the number of genes changes with the addition of more genomes to the pan-genome show this data in a more intuitive manner. In Figure 2-1 it is possible to see how among the species analysed, in most cases the curves become less steep with number of new genomes added. The trend to stabilise towards an asymptote is indicative of a closed pan-genome; this is apparent in that of *R. salmoninarum*, although this may be due to the small number of genomes in the dataset. *Y. ruckeri* has, among the better documented species, a relatively closed pan-genome, with addition of new genomes to the dataset seemingly providing a smaller number of new genes. Open pan-genomes are more obvious in the cases of *F. psychrophilum* and *A. salmonicida*, the latter being, at least from visual appearance, the species with the most open pan-genome.

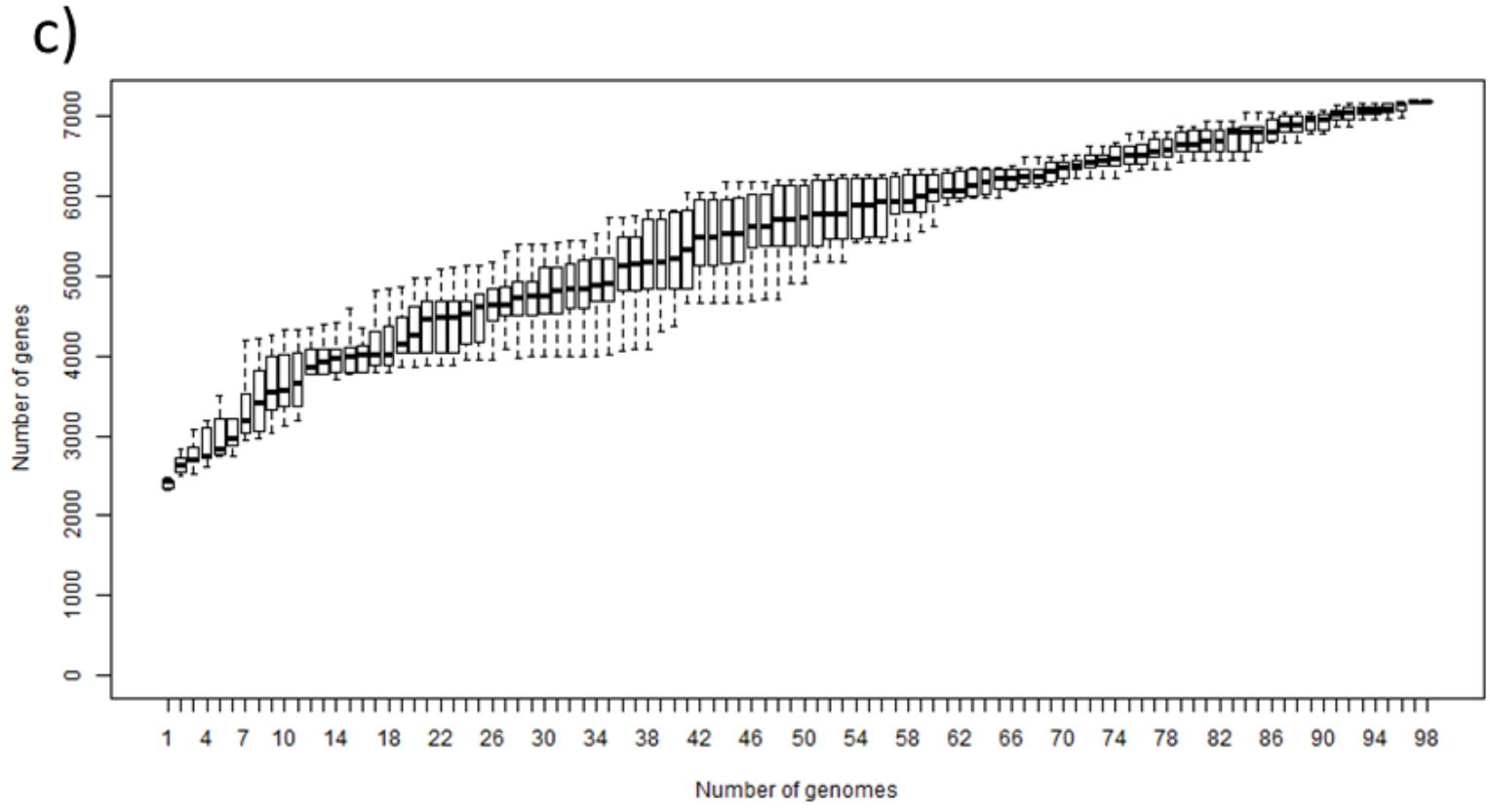
a)


Number of genes
0
5000
10000
15000
1 3 5 7 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57 60 63 66 69
Number of genomes

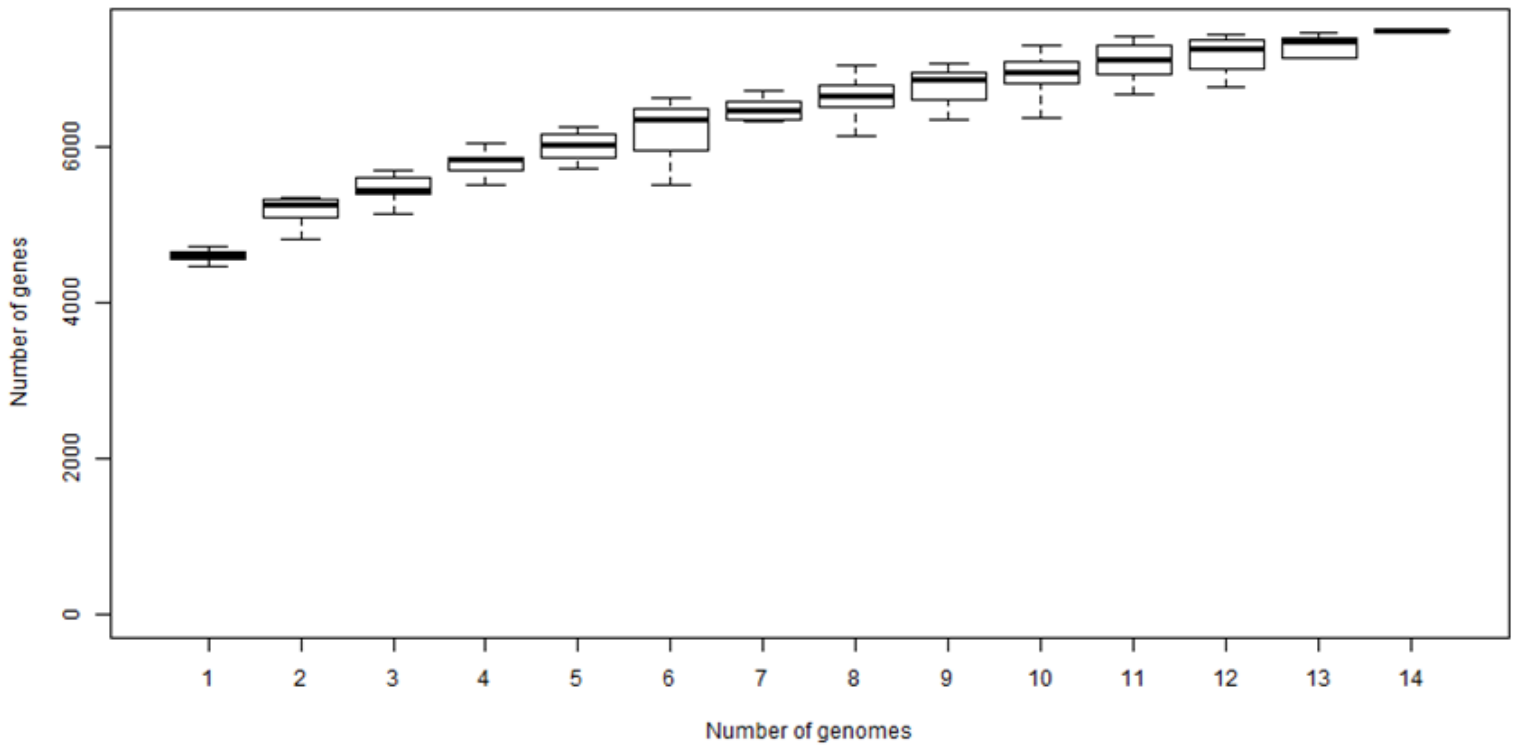
b)


Number of genes
0
1000
2000
3000
4000
1
2
3
4
5
6
Number of genomes

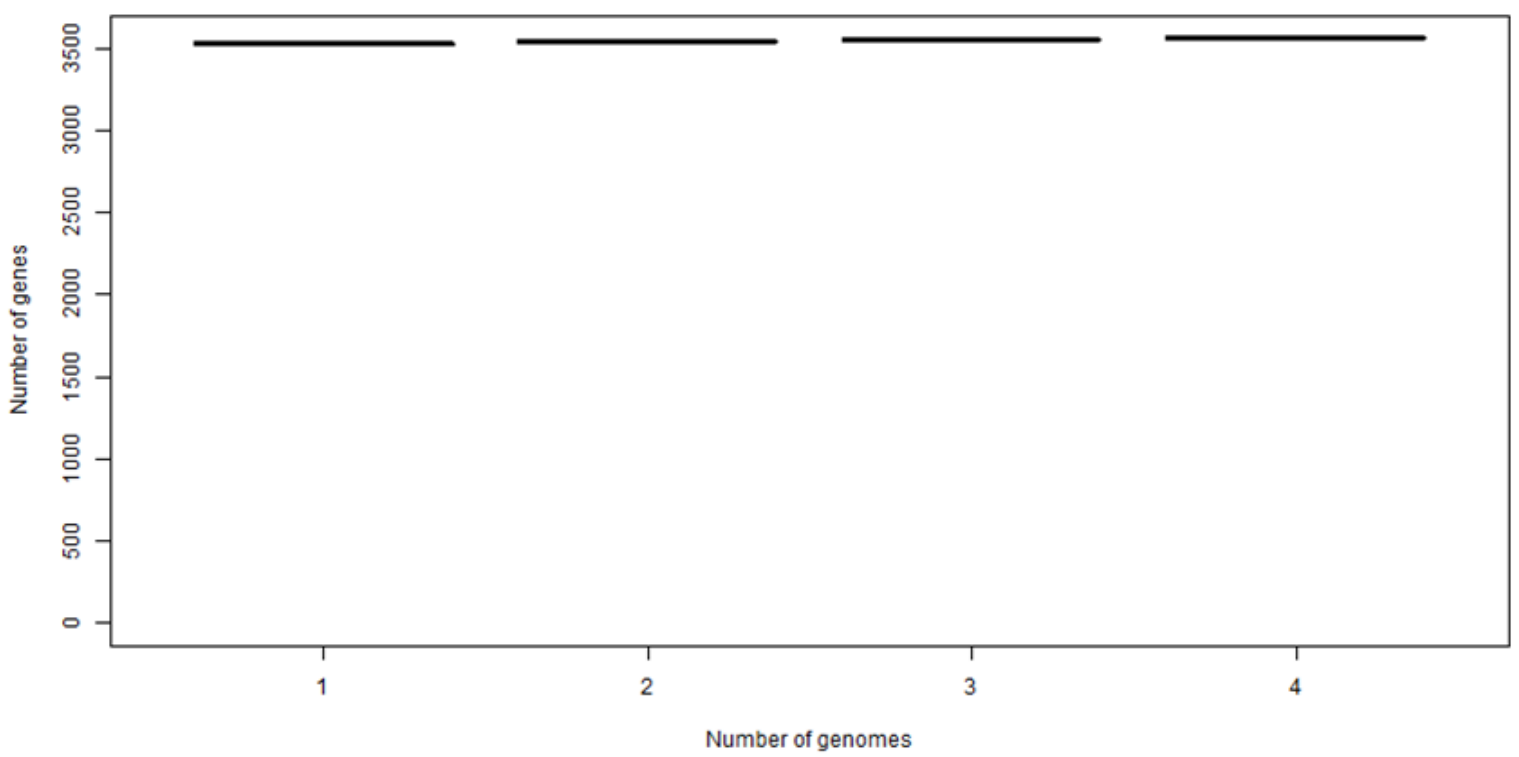
c)


Number of genes
0
1000
2000
3000
4000
5000
6000
7000
1 4 7 10 14 18 22 26 30 34 38 42 46 50 54 58 62 66 70 74 78 82 86 90 94 98
Number of genomes

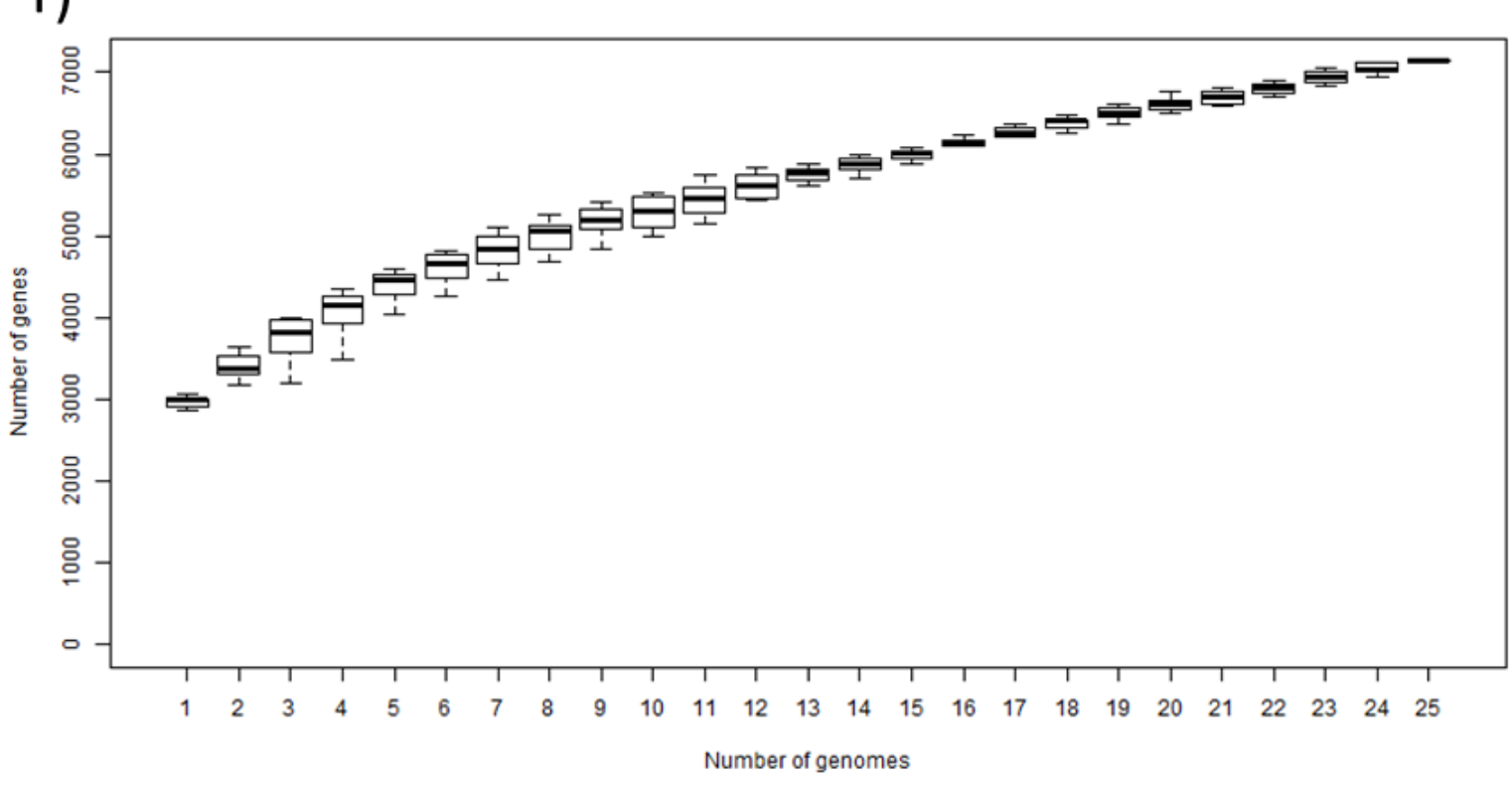
d)


Number of genes
0
2000
4000
6000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Number of genomes

e)


Number of genes
0
500
1000
1500
2000
2500
3000
3500
1
2
3
4
Number of genomes

f)


Number of genes
0
1000
2000
3000
4000
5000
6000
7000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Number of genomes

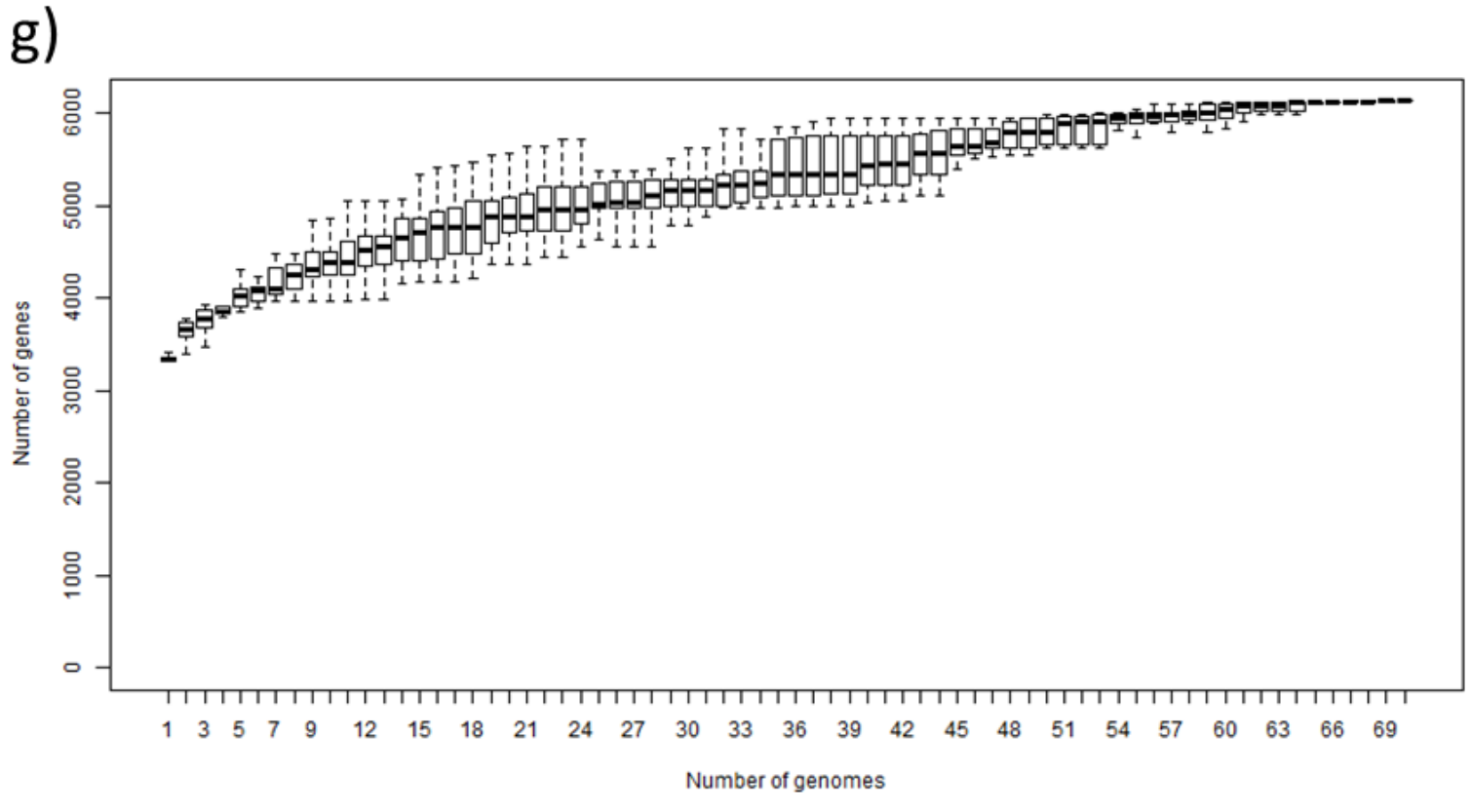


Figure 2-1 Boxplots depicting the increase in number of genes in the pan-genome as genomes are added to the calculation. The genomes are added in a random order; the variation shown by the boxplots corresponds to the different number of genes obtained over ten repetitions of the process. The following sections correspond to the following species: a) Aeromonas salmonicida, *b)* Flavobacterium branchiophilum, *c)* Flavobacterium psychrophilum, *d)* Moritella viscosa, *e)* Renibacterium salmoninarum, *f)* Tenacibaculum maritimum, *g)* Yersinia ruckeri.

2.3.3. Details on the phylogenetic and ANI analyses of *A. salmonicida*

2.3.3.1. Phylogenetic analysis

The phylogenetic tree of *A. salmonicida* (Figure 2-2) shows that strain CBA100 probably does not pertain to this species, as it clusters more closely with the representative of *A. bestiarum* and *A. piscicola* than with any other strain known to be within the species. For this reason, this genome was excluded from the pan-genomic analysis.

A. salmonicida is an interesting case for several reasons. There are currently 73 publicly available genomes for the species, pertaining to several strains and all five accepted subspecies. A large proportion of the genomes sequenced for this species, however, are classified as of the subspecies *salmonicida* (33 of 73). The majority of these (30 genomes) cluster together in a single clade, which encompasses all available genome sequences for the type strain (ATCC 33658, CIP 103209, NCTC 12959 and NCIMB 1102, a product of this project's work). There are two points worth noting about this group. The first is that the bootstrap support is a relatively weak 85%, but that the clade that encompasses this one and the strains YK and BG does receive a 100% support. The second is that support values within the clade are usually very low, which is to be expected of highly similar sequences. These two points would indicate that strains YK and BG are likely part of the *salmonicida* subspecies.

However, three genomes labelled as of subspecies *salmonicida* cluster together outside this group, falling in a clade with representatives of three other subspecies (*achromogenes*, *masoucida* and *smithia*). The clade formed by the three *salmonicida* subspecies strains (J409, J410, J411) and the other three subspecies is well supported with a bootstrap value of 97. It should be noted that representatives of each of these other subspecies include the genomes of type strains. Of these, it is striking to see the type strains of subspecies *smithia* (CIP 104757) and *masoucida* (CIP 103210) being placed the closest. Though the bootstrap support is relatively low (67%) for this group, another genome of the type strain of the *masoucida* subspecies is its nearest relative (CIP 104757 = NBRC 13784), forming a clade with 100% bootstrap support. A number of strains assigned to the *masoucida* subspecies are basal to this group.

Basal to this clade is a disconcerting group that encompasses five representatives of the *achromogenes* subspecies (strains 23051, 23053, 23055, 23056 and AS03) and one of the *smithia* subspecies (strain JF4095). This is a highly supported clade, being found in 100% of

the rapid bootstraps carried out by RAxML. The acceptance that these strains are indeed correctly classified would imply that subspecies *smithia* and *achromogenes* are polyphyletic.

When going further towards the root of the tree, it is possible to see a segment where support values are considerably low. Within this section are found three genomes of subspecies *pectinolytica*; the three of them belong to the same strain, and represent either different instances when it was sequenced or different assemblies. It is surrounded by a number of genomes that have not been classified to the subspecies level. Finally, at the base of what may be considered the *A. salmonicida* group lie strains like 947C and ECFood+05, which do not cluster with any other strains.

2.3.3.2. ANI analysis

The results of the ANI analysis on *A. salmonicida* can be seen in Figure 2-3. Overall, the lowest identity of the entire *A. salmonicida* clade is the one between strains 17 and JF3517, 0.96932. Relying on Colston and colleagues' (2014) threshold of ~96% (0.96) for delimiting *Aeromonas* species, this means that, excluding that of strain CBA100, every genome available for *A. salmonicida* is within the taxonomic boundary of species. Identities with the representative genomes of other species are, at their highest of 0.91387 (*A. piscicola* and *A. salmonicida* strain A308), which is also congruent with the status of *A. salmonicida* as an independent species.

Within *A. salmonicida*, the ANI values reflect the situation described by the phylogenetic tree. All the genomes belonging to the narrower subspecies *salmonicida* clade have high identities; the minimum is 0.99912, 0.99838 when including strains YK and BG. When considering the J409, J410 and J411 strains as well, which were labelled as *salmonicida* despite not being in the main clade, the minimum identities fall to 0.99723. Considering only the smaller clade that contains type strains of three subspecies, the minimum identity is 0.99743, but then when incorporating the rest of the *salmonicida* subspecies clade it goes down to 0.99621 (between strain CIP 104001, the *achromogenes* type strain, and strain JF2507, a representative of the *salmonicida* subspecies).

Of the other subspecies excluding *salmonicida* there are few reliable representatives. In most cases these are the sequences of the type strains; it should be noted that in some cases more than one genome belongs to a type strain but is merely named differently because it comes from the sample stored at a different culture collection (for example CIP 104001 and NCIM 1110). When considering all the genomes assigned to each of the *achromogenes*, *masoucida*

and *smithia* subspecies, the minimum ANIs belonging to their most different members are respectively 0.99664 (strains NCIMB 1110 and AS03), 0.99944 (unidentified *masoucida* strain and CIP 103210) and 0.99729 (strains JF4097 and CIP 104757).

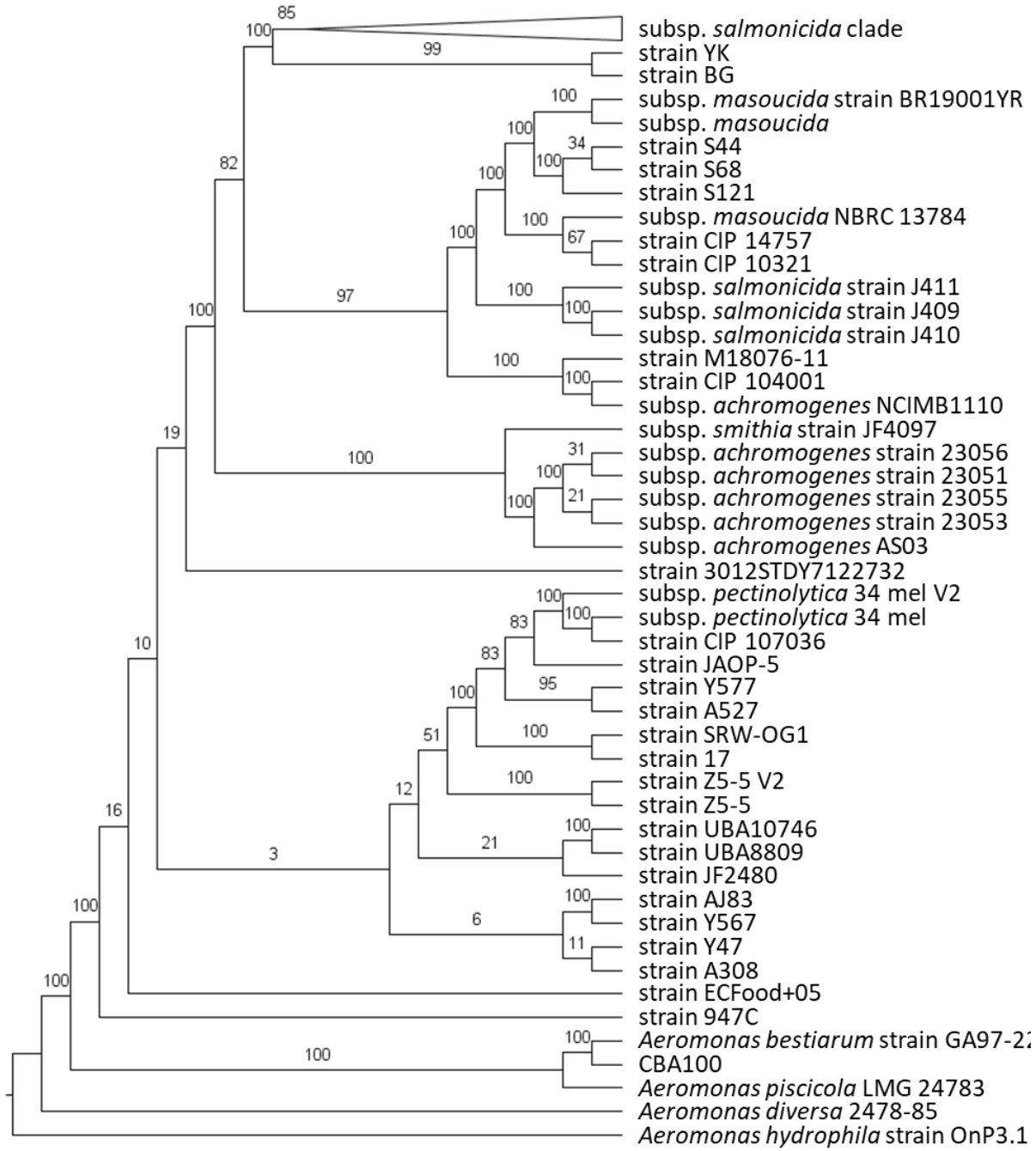


Figure 2-2: Cladogram depicting the relationships between the different genomes of the complete Aeromonas salmonicida *dataset as determined by Maximum Likelihood phylogenetic analysis on RAxML (Stamatakis, 2008). The values on the branches represent support obtained from 100 rapid bootstraps.*

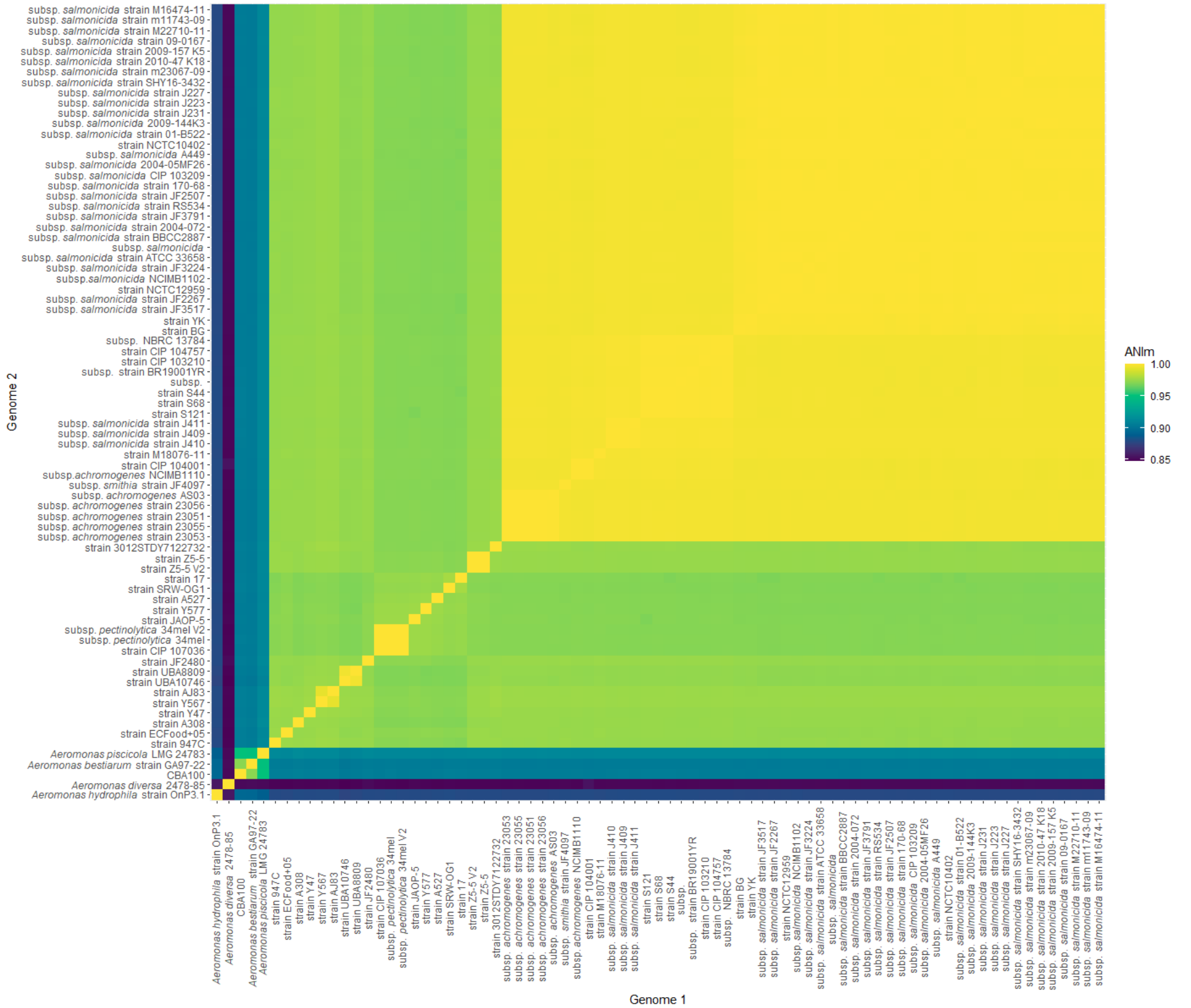


Figure 2-3: ANI matrix showing the distances between genome pairs. The colour scale represents the range of ANI values.

2.3.4. Antimicrobial resistance

Abricate detects 48 different AMR-related genes from CARD in *A. salmonicida*. In this species, each genome harboured a mean of 5.27 resistance genes, with a standard deviation of 3.47. The most commonly found AMR gene is OXA-427, coding for a class D beta-lactamase that is known to confer resistance to several cephalosporins, penicillins and carbapenems to a degree (Bogaerts et al., 2017); it is found in every genome except for that of strain UBA10746. It can also be found also in related species (*Aeromonas bestiarum* and *Aeromonas piscicola*). This is followed by cphA5, found in 56 *A. salmonicida* genomes, and which is a class B metallo-beta-lactamase that also confers resistance to carbapenems (in the Comprehensive Antibiotic Resistance Database, Alcock et al., 2019, at: https://card.mcmaster.ca/ontology/39667). The third most common, FOX-2, is present in 55 genomes, and codes for a class C beta-lactamase conferring resistance to cephamycins. The remaining resistance genes are found less frequently; see Figure 2-4 to see what combinations of antibiotic resistance genes are common.

Additionally, combining the AMR matrix with the phylogenetic tree makes apparent that there is only a moderate correlation between the two. Features like the presence of a portion of the OXA-427 gene (instead of the entire sequence), the MCR-3 and the MCR-3.12 genes seem to be characteristic of the *A. salmoncida salmonicida* subspecies, with some exceptions (see Figure 2-4). Or, for instance, the loss of the cphA5 gene in the *masoucida* clade. However, a large number of AMR genes appear and disappear in contiguous lineages, for example the FloR florfenicol resistance gene, which can be found in strains S121, S68 and S44, but none of the other representatives of the *masoucida* clade, in strain JAOP-5 alone among the cluster closest to *pectinolytica* and in six of the 32 *salmonicida* strains. Finally, it appears there are some AMR genes that are frequently or always found together, as are MCR-3 and MCR-3.12 or APH(3'')-Ib and APH(6)-Id.

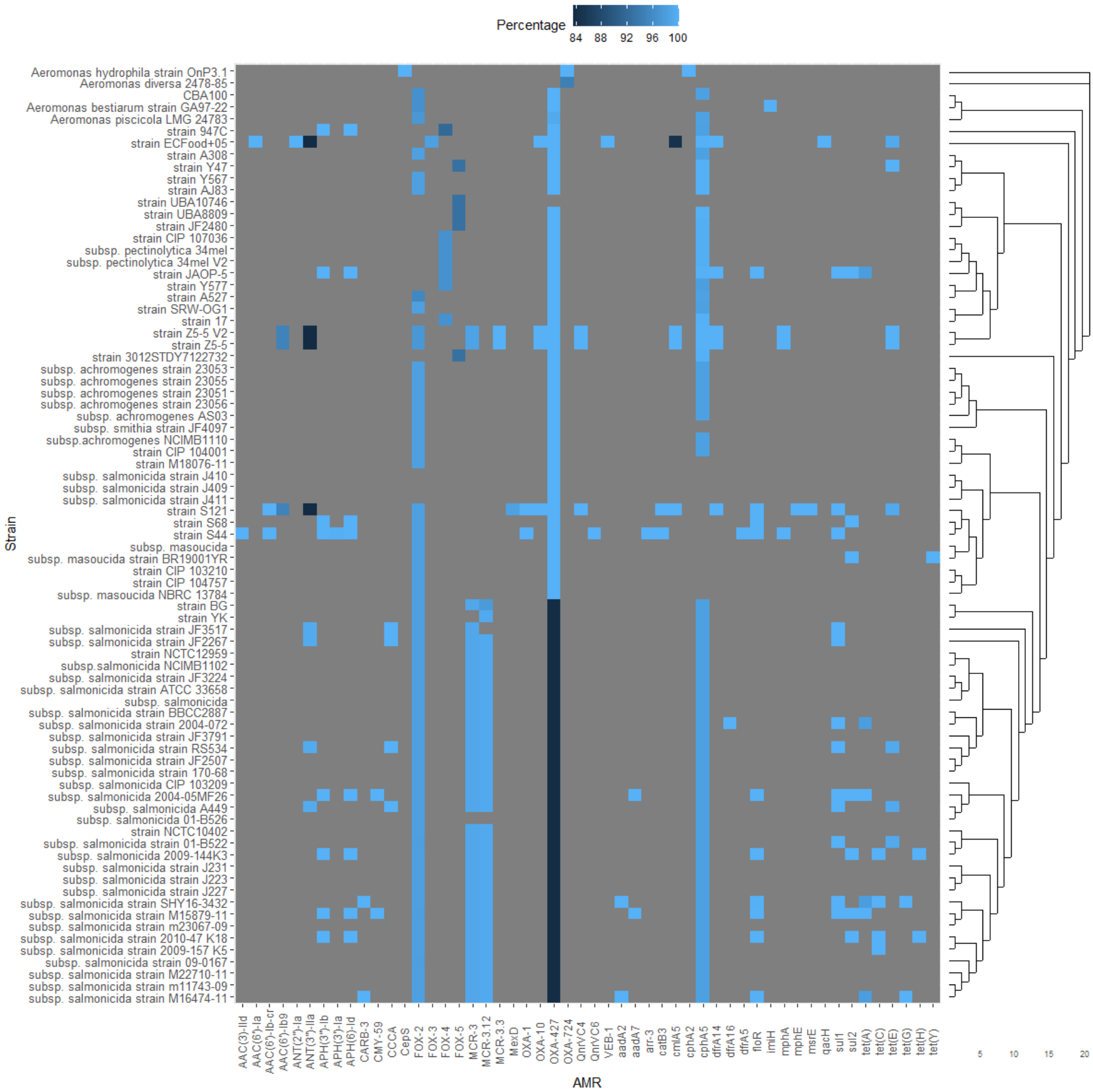


Figure 2-4: The matrix shows the AMR genes (listed on the X-axis) detected by Abricate in the genome of each strain of Aeromonas salmonicida*, plus five close relatives (listed on the left Y-axis). In this matrix, grey indicates the gene has not been found; different shades of blue indicate what percentage of the gene Abricate was able to detect. The dendrogram on the right represents the phylogenetic relationship between these strains. For all* Aeromonas salmonicida *representatives, the species name has been omitted.*

2.4. Discussion

2.4.1. Overview

Pan-genomic calculations were carried out for the eight BPOIs, and a special focus was placed on *A. salmonicida*, a particularly interesting species (objective 1.1). Although most of the bacteria included here have open pan-genomes, this is especially clear for this species. This is reflected in its AMR gene profiles, which are considerably variable, and mostly uncorrelated with phylogeny. Indeed, it seems that taxonomy is not correlated with phylogeny either, and that there are several paraphyletic/polyphyletic taxa (objective 1.2).

2.4.2. Pan-genomics

The results show that most of the BPOIs considered have open pan-genomes, where every addition of a new genome to the calculations adds new genes. The exception is *R. salmoninarum*, for which the composition of the genomes is relatively uniform throughout the different genomes available on the NCBI Assembly database. The general consensus is that this is a relatively genetically and biochemically uniform species (Bayliss et al., 2018; Bethke et al., 2017; Brynildsrud et al., 2014; Starliper, 1996), but no pan-genomic studies have been carried out before. This might be because of the poor availability of genome data for this species on public databases, despite there being substantial interest it the pathogen, and several genomes having been sequenced (Bayliss et al., 2018; Brynildsrud et al., 2014). The pan-genome analysis of this species is, therefore, of little value, given that there are only four available genomes, completely insufficient to represent the biological diversity of a pathogen with global distribution (Brynildsrud et al., 2014; Grayson et al., 1999), and yet appear concordant with previous knowledge. The situation for *F. branchiophilum* is similar; the number of genomes available is relatively small (only six), and of course, no pan-genome analyses have been published for the species (to my knowledge). In this case, however, there seems to be some more variability within the pan-genome. Kumru and colleagues (2020) do a broader analysis of pathogenic bacteria of the genus *Flavobacterium*, and they compare the virulence factor and AMR gene content of the genomes of three *F. branchiophilum* strains. They find that of 121 virulence factors and 97 AMR genes, these three strains differ in only four of the former and seven of the latter, suggesting that there is some level of variability; this agrees with this chapter's results, but also begs for the addition of more genome data for the species on public databases.

F. psychrophilum is a more well-studied species; at least two pan-genome analyses have been published (Castillo et al., 2016; Duchaud et al., 2018). Castillo and colleagues (2016) find, after analysing the genomes from 11 isolates, that this species' pan-genome had 3,373 genes of which 1,743 belonged to the core; their result coincides approximately with this chapter's for that amount of genomes (see Figure 2-1c). Duchaud and colleagues' (2018) results also approximately agree; they incorporate 41 genomes and obtain 4,805 gene families. Although this chapter's work uses more genomes in the analyses, these researchers were more cautious with the data they used. However, they then remove those genes that would translate to a peptide under 100 amino acids, as they consider these could be a result of pseudogene detection during genome annotation, or of fragmentation resulting from assemblies that are broken into numerous contigs. The latter is a recognised issue in genome annotation (Denton et al., 2014; Gabrielaite & Marvig, 2020) and it is appropriate to take precautions, such as those taken by Duchaud and colleagues (2018), or by using only complete, circularised genomes. Another recently available option is the one offered by GenAPI (Gabrielaite & Marvig, 2020), a program designed to perform pan-genome calculations accounting for fragmented genes. In any case, in any future work precautions should be implemented to account for these potential miscalculations.

In the case of *M. viscosa*, the results differ somewhat from those of Karlsen and colleagues (2017), who carried out a pan-genome analysis with 12 strains from the North Atlantic region, and find a total of 5,589 gene clusters. The analysis carried out for this chapter suggests the pan-genome of this species might be larger (see Figure 2-1d). Because all the genomes available on the NCBI Assembly database were from regions in the North Atlantic, it was considered that the use of a broader, and potentially more diverse sample of genomes, was unlikely to be behind this difference. Instead, it is possible that this difference is due to methodology: Karlsen and colleagues (2017) annotated their genomes with Glimmer v3.02 (Delcher et al., 2007) and BLASTp (Altschul et al., 1990) using the UniProt database (01 2014 release)(Apweiler, 2004). Prokka (v1.12), the annotation software used in this chapter, employs a different program to find protein-coding features (Prodigal; Hyatt et al., 2010), and uses the ISFinder (Siguier, 2006), the NCBI's Bacterial Antimicrobial Resistance Reference Gene Database (NCBI, n.d.), and Swiss-Prot (Apweiler, 2004; Uniprot Consortium, 2021) databases for protein search. A similar situation applies to the pan-genome of *T. maritimum*, for which Bridel and colleagues (2020) found a total of 5,809 gene families, 2,116 belonging to the core; this contrasts with the results of this chapter, where, using the same number of

genomes (25) 7,133 gene families are found, 2,130 of which belong to the core. The difference here could, once again, owe itself to a methodological difference; though they do not explicitly state the means of annotation, they do differ in the software used to calculate the pan-genome (PPanGGOLiN v0.1.4; Gautreau et al., 2020).

58 *Y. ruckeri* genomes are included in the analysis by Barnes and colleagues (2016). They identify 4,943 genes in total in the pan-genome, 2,745 genes in the core; this does not coincide with the analyses of this chapter, and is unlikely to be attributable to methodological differences, as they also use Prokka for annotation and Roary for pan-genome calculations. In this case it is possible that this discrepancy is owed to the 11 additional genomes used in this chapter, although it would be surprising, given that Barnes and colleagues (2016) use isolates from a variety of geographical locations and that these are all included in this chapter's analysis.

Finally, *A. salmonicida*, chosen as a case study for this chapter, is the bacterium with a most clearly open pan-genome. Although this bacterium has been extensively studied, this analysis is more comprehensive than any currently published. Vincent and Charette (2017) do calculate a pan-genome for this species using the genomes of 26 strains, collected from a variety of geographical locations and representing all five of its subspecies. Although they do not report an exact number of total genes in the pan-genome, it is below 6,500, whereas in this chapter this is over 10,000. Once again there are methodological differences. However, they agree in their assessment that the pan-genome of *A. salmonicida* is open.

2.4.3. AMR genes

The AMR gene analysis added a new dimension to the pan-genomic and phylogenetic analyses. As is visible from Figure 2-4, though there are some highly prevalent AMR genes, (OXA-427, cphA5, FOX-2), the majority are present in only a small number of strains. The majority of these less prevalent genes are present in an apparently random manner, mostly unrelated to phylogeny. It should be noted that *A. salmonicida* has been known to harbour a number of mobile genetic elements that carry antibiotic resistance genes (Attéré et al., 2017; Trudel et al., 2016; A. T. Vincent et al., 2014). In some cases there is evidence that plasmids of *A. salmonicida* are related to those carried by strains of other bacterial species (Attéré et al., 2017; A. T. Vincent et al., 2014) and as such this species has been called a potential reservoir of AMR genes (Trudel et al., 2016; A. Vincent & Charette, 2017). It is possible that the described pattern of distribution of AMR genes is a result of this mobility of genetic

elements, which may import them from other bacteria in the surroundings of fish farms (Trudel et al., 2016; A. T. Vincent et al., 2014). In fact, the waters surrounding fish farms, where antibiotic concentrations can be relatively high, are ideal settings for this to occur, as acquiring resistance to these substances could confer selective advantages (Sarmah et al., 2006; Watts et al., 2017). Trudel and colleagues (2016) have already raised the alarm about cases of *A. salmonicida* resistant to all legal antibiotics in Canada, and suggest monitoring worldwide trends. Nanopore sequencing may be uniquely placed to carry out fast and dynamic monitoring of AMR thanks to its portability and speed; this is explored in section 3.3.6.3.

2.4.4. Taxonomy

A small consideration about strain CBA100; at least until recently it was not uncommon for *Aeromonas* strains to be misclassified to the species level. Beaz-Hidalgo and colleagues (2015) found that 35.9% of the genomes they considered in their evaluation of the GenBank data on the genus were not correctly classified. As a result, they recommended the taxonomic re-evaluation of these genomes. As a positive contrast to these authors' findings, in this chapter's phylogenetic analyses and ANI calculations only one strain assigned to the *A. salmonicida* species was observed not to belong there. However, their recommendation to any submitters of genome data to carry out ANI and phylogenetic analyses before submission is still valid. This is especially true for *A. salmonicida*, which closely resembles *Aeromonas bestiarum* in biochemical assays and 16S rRNA gene sequence (Martínez-Murcia et al., 2005).

The main taxonomic observation of this chapter, however, is the incongruence of the *A. salmonicida* subspecies internal taxonomy. Whereas the entire group is mostly cohesive, and with the exception of CBA100 clearly fit well within the designated threshold of 96% ANI (Colston et al., 2014) and form a single clade, the internal taxonomy is problematic. One of these issues would result from accepting that every genome labelled to the subspecies level has been correctly classified: If we are to accept that the classification by the submitters is correct, most of the subspecies are either polyphyletic or paraphyletic. Given Beaz-Hidalgo and colleagues' (2015) warning about taxonomically misclassified submissions, it is plausible that these inaccuracies apply at the subspecies level too. Considering the genomic diversity observed for this species, it would be unsurprising that closely related strains show substantial phenotypic variation. This is indeed the case with subspecies *smithia* and *masoucida*, which despite their phenotypic differences (D. A. Austin et al., 1989), appear

clustered together in the phylogenetic tree, with high identity between their genomes (0.99973). Current recommendations are to take genomic data into account when classifying species (Arahal, 2014; Meier-Kolthoff et al., 2014), but although DDH values have been proposed for the delimitation of subspecies (Goris et al., 2007; proposal of 79-80%) the same is not true for ANI.

To compound the problem, there are a number of "orphan" strains towards the base of the *A. salmonicida* species tree. These tend to be mesophilic strains, with different hosts to *A. salmonicida salmonicida* and with certain distinctive genomic characteristics (A. T. Vincent et al., 2016, 2017; A. T. Vincent, Fernández-Bravo, et al., 2019) consider the possibility of creating new subspecies to classify these distinct strains, but are wary of doing so given the large number of strains with no close relatives. They suggest further sampling in the environments where typical *A. salmonicida* are not found, but mesophilic strains may be, and sequencing their genomes, with the hopes of better informing the delineation of the intra-species taxonomic delineation. To this suggestion I agree merely adding that future taxonomic divisions within the species should, as stated earlier, take ANI into consideration, although perhaps this should involve extending the work Colston and colleagues (2014) did to define thresholds for *A. salmonicida* into the subspecies level.

2.4.5. Conclusion

Ultimately, the pan-genomes of most of the seven BPOIs could be said to be open, but it is not reasonable to reach conclusions for *F. branchiophilum*, and especially *R. salmoninarum* when the genome data that is publicly available is so scarce. In the case of *A. salmonicida*, it is clear that the pan-genome is open, but the in-depth study of its phylogeny and ANI show that the internal taxonomy ought to be revised. Furthermore, the AMR study suggests that numerous AMR genes can be found in the species, but these are not consistent throughout, nor do they correspond with phylogeny. Studies like those of Trudel and colleagues (2016) and Vincent and Charette (2017) go further in calling *A. salmonicida* a reservoir of AMR genes. From all of this it is clear that an increased application of genome sequencing would benefit the understanding of these pathogenic organisms' pan-genome, inform the definition of their taxonomy, and improve the tracking of their evolution and surveillance of their antimicrobial resistance.

Chapter 3: Comparison of different genome assembly approaches

3.1. Introduction

Nanpore sequencing is a relatively modern nucleic acid sequencing method. Announced in 2012 and first commercialised in 2014 under the "MinION Access Program", it was the result of approximately 25 years of groundwork research (C. Brown, 2012; Deamer & Akeson, 2000; ONT, n.d.-b). The procedure is categorically different to previous sequencing techniques, which relied on DNA replication. Instead, nanopore sequencing involves a nucleic acid molecule being "read" as it crosses a pore on a membrane (C. Brown, 2012). A voltage is generated across said membrane, and the nucleic acid is forced through a pore, altering the ionic current as it transverses it. This current is measured and recorded; it constitutes a raw signal that can then be interpreted by a "basecaller", a program that infers a nucleotide (or ribonucleotide) sequence and quality scores (R. R. Wick et al., 2019).

The reads obtained through this technique are exempt from the size limitations of the Sanger method, or those of Illumina's sequencing by synthesis, depending instead on the length of the nucleic acids provided as input. This leads to them being referred to as "long reads". Their characteristics make them especially useful for genome assembly. A recognised limitation of short-read sequencing methods, which produce reads that are often only a few hundred bases long, is that they cannot resolve regions of genomes that feature repetitive sequences that are longer than this (Berlin et al., 2015; H. Lee et al., 2016). Reads that are long enough to span these repetitive regions, such as those provided by nanopore sequencing, can serve to reconstruct genomes more accurately, reducing the number of contigs an assembly is split into (H. Lee et al., 2016).

Although nanopore sequencing is not the only next generation sequencing technique that produces long reads, it offers a number of advantages. One of them is the portability of the device; although it requires some extra ancillary equipment to run, there are numerous examples of the MinION being used in remote locations or resource-limited settings. For instance, it has been used in agricultural settings in Tanzania, Uganda and Kenya to detect and classify viral pathogens of cassava (Boykin et al., 2019); it has been transported in a suitcase to a mobile field laboratory in Guinea to sequence the genome of Ebola virus samples (Quick et al., 2016); and it has been used in metabarcoding studies where sequencing took place at the sampling location on the Greenland Ice Sheet (Edwards et al., 2016).

Another advantage is the relatively low entry costs; a starter pack comprising a MinION device, a library preparation kit and a flowcell capable of generating up to 50 gigabases of read data costs £800 (ONT, n.d.-c, n.d.-d).

A field where there has been a growing level of interest for pathogen whole genome sequencing is public health. Several studies give evidence of its application in pathogen identification, phylogeny, tracking, and the study of antimicrobial resistance or virulence genes (Aanensen et al., 2016; Besser et al., 2018; McDermott et al., 2016; Pecora et al., 2015; Price et al., 2013; Walker et al., 2013). Furthermore, a number of organisations have already recognised the potential for improvement in disease monitoring and surveillance it can offer (Grant et al., 2018; Nadon et al., 2017; Revez et al., 2017). One such organisation is Public Health England, which has implemented whole-genome sequencing for *Salmonella* samples as a case study in its Colindale site. A report outlining the results of the case study noted an improvement in surveillance of this bacterium, as well as reduced reagent costs and a simplification in laboratory practices (Grant et al., 2018).

The utility of whole-genome pathogen sequencing has not been limited to human health, and there are numerous examples of studies where the genomes of agricultural bacterial pathogens have been sequenced (Condon et al., 2022; Suarez et al., 2021; Yang et al., 2014; Zankari et al., 2013). More specifically, they have demonstrated their usefulness in aquaculture. For example, Yang and colleagues (2014) sequenced the whole genomes of three pathogenic strains of *Vibrio* associated to AHPND. Their work and that of Lee and colleagues (2015) served to identify the plasmid-borne virulence genes that resulted in strains of the *Vibrio parahaemolyticus* clade becoming pathogenic. From a more practical, "on-site" perspective, Condon and colleagues (Condon et al., 2022) used data from WGS to identify a number virulence genes in *Vibrio* bacteria that were presumably causing mortality at a black tiger shrimp hatchery; the sequenced genomes served to design TaqMan probes for on-site detection of four of those virulence genes. The applications of WGS have also reached salmonid aquaculture, where it has aided the taxonomic classification of mycobacterial pathogens (Suarez et al., 2021), helped understand the origins of *Yersinia ruckeri* in Oceania (Barnes et al., 2016), and served to reconstruct the global spread of *Renibacterium salmoninarum* (Brynildsrud et al., 2014). The aforementioned advantages of the MinION sequencer, have the potential to bring the benefits of WGS to more remote locations and resource-limited settings.

Genome assembly is a crucial step in the analysis of WGS data. Considering the potential long reads have in this process, and given that the recent introduction of nanopore sequencing makes genome sequencing more widely available, it is no surprise that a large number of long-read based genome assemblers have been developed in recent years. To name just a few of them, and indeed, the ones that are the focus of this chapter, Flye (Kolmogorov et al., 2019), Raven (Vaser & Šikić, 2021), Miniasm (H. Li, 2016) and WTDBG2 ("Redbean"; Ruan & Li, 2020) have each taken on the problem of genome assembly from a different angle. They indeed often perform in a way that takes advantage of the long reads to resolve repetitive regions and obtain longer contiguous sequences. However, long read sequencing methods are error prone in terms of per-base accuracy. Specifically, nanopore sequencing recently was reported to have mean error rates per read of 14% (Workman et al., 2019), and although it has improved since its early days, it still performs worse than short-read methods. In an attempt to resolve this high error rate, most assemblers incorporate a correction step, for example in the form of the Trestle module in Flye (Kolmogorov et al., 2019) or Racon (Vaser et al., 2017), in Raven (Vaser & Šikić, 2021). Racon, for example uses a consensus of the long reads to resolve indels and mismatches (Vaser et al., 2017) to improve the accuracy of the assembly.

An alternative means of obtaining genome assemblies that are both accurate and complete, or at least less fragmented, is hybrid assembly. This approach combines short and long reads, using the short reads to produce accurate contigs and long reads to bridge unresolved gaps and areas of highly repetitive sequences (H. Lee et al., 2016; R. R. Wick et al., 2017). One software that specialises in hybrid assembly of prokaryotic genomes is Unicycler (R. R. Wick et al., 2017). It consists of a pipeline that optimises SPAdes, a common short-read assembler, using it to build contigs and a graph of their interconnections, followed by a long-read resolution step. The long-read resolution step uses minimap (H. Li, 2016) to align a relatively small number of long reads to the assembly graph and resolve the interconnections to the likeliest path. This program compares favourably to other hybrid assemblers (R. R. Wick et al., 2017) and has contributed to a number of high-quality genome publications in recent years (Humphrey et al., 2019; Nakagawa et al., 2019; Tourlousse et al., 2020).

Other hybrid assemblers have been developed recently with much lower computational requirements than Unicycler (Gatter et al., 2020; Haghshenas et al., 2020). Despite the apparent speed and low computational requirements of these latest programs, all hybrid assemblers are limited in their application by their reliance on short reads that come from

expensive, cumbersome equipment. Hybrid assembly is, therefore, despite its quality, impractical to carry out in low-resource or field settings. Therefore this chapter focuses specifically on long-read based methods. It aims to assess the different parameters that define assembly quality and the computational costs and time taken to run of a number of recently published long-read assemblers. Beyond these parameters, it also considers several others that may influence the utility of genome assembly to fish farmers.

3.1.1. Aims and objectives:

Aim(s):

1) To investigate which among a number of long-read assemblers may be the most appropriate for application in the context of aquaculture.

This aim is divided into the following objectives:

1) To sequence the genomic DNA of the eight BPOIs.
2) To assemble genomes from the data generated using a range of long-read assemblers.
3) To assess the resource requirements (time and RAM) of these programs and the quality of the assemblies they produce.

3.2. Methods

3.2.1. Bacterial cultures

Eight bacterial pathogens of farmed salmonids were used for this work; they are listed in Table 3-1; these are the seven BPOIs, including two subspecies of *Aeromonas salmonicida*. The strains selected were the type strains of each bacterial species, as they were considered representative of the species. They were obtained through the National Collection of Industrial and Marine Bacteria (NCIMB), who provided them as ampoules of lyophilised bacteria.

In preparation for the cultivation of the bacteria, appropriate broths and agars were prepared. These were in most cases those recommended by the NCIMB for the eight bacteria used; their recipes were obtained from the NCIMB website and included in Appendix 2 . However, other recipes are included in the appendix too, as in the long run some growth media were changed to improve bacterial growth. In some cases, the recipes provided by the NCIMB coincide with commercially available broth powders; this is the case for medium number 1, which contains the same ingredients in the same proportions as Nutrient Broth E, and medium number 398, which is Tryptone Soya Agar (Oxoid CM0131) plus 2% (w/v) NaCl. For the preparation of media, dry ingredients were weighed and then mixed in a glass container (bottle or flask); liquid ingredients were added posteriorly. Distilled water was then added to complete the desired volume, and pH measured (Jenway 350 pH meter, Jenway) and adjusted with HCl (5% v/v) or NaOH (1M) if necessary (see Appendix 2 for pH of each medium). Addition of certain ingredients (charcoal powder, gelatin and agar powder) was left until after the measurement of pH, to avoid any possible harm to the pH meter electrodes. Media were then sterilised at 115 °C for 15 minutes or 121 °C for 20 minutes in the case of Charcoal Salts Cytophaga Agar (Álvarez & Guijarro, 2007; Skulska, 2014). Agars were poured into vented Petri dishes (cat. Nº. 12654785, Fisher Scientific) inside the sterile flow hood (Telstar AH-100, Telstar) and lidded when solidified.

Table 3-1: Bacterial strains used in this project, with identifier given to them by the culture collection (NCIMB) and the depositor.

Bacterial species	NCIMB ID	Depositor designation
Aeromonas salmonicida achromogenes	1110	Smith6263/4/5
Aeromonas salmonicida salmonicida	1102	19291 Smith20
Flavobacterium branchiohilum	12904	BGD7721
Flavobacterium psychrophilum	1947	3068
Moritella viscosa	13584	NV1 478/88
Renibacterium salmoninarum	2235	Lea1-74
Tenacibaculum maritimum	2154	R-2
Yersinia ruckeri	2194	

Cultures were started with little change to the NCIMB's instructions: ampoules were opened by filing a circle around the centre with a diamond-tipped pen and snapping open with the provided plastic snappers. 0.5 mL of an appropriate broth were added into the snapped ampoule and mixed with the lyophilised bacteria by pipetting several times. The paper with the NCIMB ID was transferred on the pipette tip to a plate of an appropriate agar and the broth was then transferred to a test tube of an appropriate broth. Sterile equipment and technique were used throughout the process, with this taking place inside a flow hood (Telstar AH-100, Telstar, Birstall, UK) sterilised with bleach and 15 minutes of UV light.

Inoculated agar plates were then incubated at their optimum temperatures as stated in the NCIMB website, except for *Y. ruckeri* whose optimum was stated as 25 °C but was incubated at 20 °C (see Table 3-2). Bacteria were then subcultured from the agar every two weeks. A culture loop was used to pick a number of colonies and streak them on another plate of the same agar. In the case of the broths, those containing bacteria whose optimum was 15 °C were initially incubated in a room with lowered heating on a "Belly dancer" shaker (Stovall Life Science Inc., Greensboro, USA) at speed setting 1.5 and with temperature being monitored with a DS20B18 sensor connected to a Raspberry Pi computer. Those whose optimum growth temperature was 20-25 °C were placed in a shaking incubator set to 20 °C and 120 rpm. Growth of bacteria in broth was discontinued after they became established in agar, and only retaken to prepare glycerol stocks.

Table 3-2: Growth media and conditions used for bacterial cultures and their comparison with those recommended by the NCIMB. 1 – CSCA = Charcoal Salts Cytophaga Agar, 2 – KDM-C = Kidney Disease Medium - Charcoal, 3 – PYC = Peptone Yeast Cysteine broth.

Bacterial species	Recommended growth medium	Used growth medium	Recommended growth temperature	Used growth temperature
Aeromonas salmonicida achromogenes	1	1	20 °C	20 °C
Aeromonas salmonicida salmonicida	1	1	20 °C	20 °C
Flavobacterium branchiohilum	98	98/CSCA [1]	20 °C	20 °C
Flavobacterium psychrophilum	98	98/CSCA [1]	15 °C	15 °C
Moritella viscosa	398	398/TSB + 2% NaCl	15 °C	15 °C
Renibacterium salmoninarum	226	KDM-C [2]/PYC [3]	15 °C	15 °C
Tenacibaculum maritimum	223	223	20 °C	20 °C
Yersinia ruckeri	1	1	25 °C	20 °C

3.2.2. DNA extraction

DNA was extracted from bacterial biomass grown on the agar plates using the DNeasy PowerSoil kit (Qiagen Ltd., Crawley, UK). In the flow hood, a culture loop was used to scoop up some bacterial mass from each agar plate. A volume approximately sufficient to cover the wire loop end was used; this was inserted into the PowerSoil kit's extraction tube and shaken in the buffer, scraping against the quartz beads to release it. Besides extracting DNA from the eight bacteria of interest, two negative controls were included, with no bacteria added. The rest of the protocol was carried out according to the manufacturer's instructions, see Appendix 3. This procedure was carried out once (22/01/19) for bacterial identity confirmation and again at a later date (15/04/19) for genome sequencing; the purpose was to use a recent extract when sequencing.

DNA yields were measured using a Qubit™ fluorometer, and purity measurements in the form of $A_{260/280}$ ratios were also taken using the NanoDrop™ 1000. This was done following the methods detailed in section 4.2.2. The only difference was the use of buffer C6 (10 mM

Tris [pH 8.0]) from the PowerSoil kit to set the blank measurement on the NanoDrop™ 1000 instrument, as it was the elution buffer used for this DNA extraction.

3.2.3. Identity confirmation

3.2.3.1. 16S ribosomal RNA gene sequencing

Identity of the bacteria was confirmed via Sanger sequencing of the 16S ribosomal RNA gene. A PCR of this gene was used to synthesise amplicons for sequencing; the ingredients are listed in Table 3-3. Primers used were: 27F (5' – AGAGTTTGATCCTGGCTCAG – 3') (Weisburg et al., 1991) and 1389R (5' – ACGGGCGGTGTGTACAAG – 3') (Osborn et al., 2000). The reaction mix was prepared in sterile conditions and subjected to the following thermal cycles: an initial warm start step (30 s, 95 °C), 30 repetitions of denaturation (30 s, 94 °C), annealing (30 s, 57 °C) and elongation (1 m 30 s, 68 °C), and one final elongation step of 5 m at 68 °C.

Table 3-3: Reagents and quantities used to prepare the reaction mix for the 16S ribosomal RNA gene PCR.

Ingredient	Volume per sample (µL)
OneTaq Quick-Load 2X Master Mix with Standard Buffer	10
Primer 27F (75 µM)	0.08
Primer 1389R (75 µM)	0.08
DNA	2
dH_2O	7.84
Total	20

Amplification was confirmed through gel electrophoresis: a 50 mL 1.5% w/v agarose gel was prepared with 0.5X TBE (45 mM Tris-borate, 10 mM EDTA), loaded with the ten samples, including the two negative controls and the electrophoresis was run in 0.5X TBE for 45 minutes at 150 V. The gel was visualised in a GelDoc XR+ transilluminator.

The amplicons were purified with AMPure XP cleanup beads (Beckmann-Coulter) as per the manufacturer's instructions, differing on two points. A magnetic plate built in-house was used instead of a SPRIplate, and a volume of AMPure bead suspension 0.4X the volume of the sample was used instead of 1.8X. 6 µL of the bead suspension were added to 15 µL of PCR product in a 96-well plate, mixed by pipetting (ten times) and left on the magnetic plate for 10 minutes to settle. The supernatant was siphoned off, avoiding the magnetic bead pellet,

and washed twice with 200 µL of freshly prepared 70% ethanol, with a 30 second incubation before removing the ethanol. Removal of the ethanol was done carefully avoiding the magnetic bead pellet. After the last wash, the plate was left to air dry for 20 minutes, and hydrated with 40 µL of ddH_2O.

A 96-well plate was prepared with 6.1 µL of each sample in a well. For all DNA extracts but those of *Flavobacterium branchiohilum* and *R. salmoninarum*, 2.1 µL of the cleaned-up DNA, 2 µL of water and 2 µL of the 27F primer (75 µM) were included in each well. For these two cleaned-up extracts, which had less DNA, 4.1 µL of each were added to the well alongside 2 µL of 27F primer (75 µM). The DNA was sent to Aberystwyth University's in-house Sanger sequencing service, which used an ABI 3730 DNA Analyser and a BigDye Terminator kit (Life Technologies, Paisley, UK) for this purpose.

3.2.3.2. Sequence analysis

Sequences obtained from the Sanger sequencing service were then searched against the NCBI's rRNA/ITS databases (NCBI Resource Coordinators et al., 2018; NCBI Staff, 2020) using the BLASTN program and selecting the megablast option (Altschul et al., 1990). This was done through the NCBI's Standard Nucleotide BLAST portal (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

A phylogenetic analysis was carried out for the 16S rRNA gene sequence of *Tenacibaculum maritimum* to clarify the results of the BLAST search. The 16S rRNA gene sequences of the type strains of each species within the genus *Tenacibaculum* were located using the List of Prokaryotic Names with Standing in Nomenclature (maintained by the Leibniz Institute DSMZ; Euzéby, 1997; Genus *Tenacibaculum*, n.d.; Parte et al., 2020). These gene sequences in fasta format were then downloaded from the NCBI Genbank database, alongside that of a closely related *Polaribacter filamentus* (Park et al., 2017) and *Flavobacterium aquatile* to serve as outgroups (see Appendix 4 for accessions). These were then aligned with the MUSCLE program (Edgar, 2004) on the MEGA7.0.26 suite (Kumar et al., 2016). Informative blocks from the alignment were extracted with GBlocks Server v0.91 (Castresana, 2000, 2002). Phylogeny reconstruction was done through a Maximum Likelihood approach implemented through MEGA7.0.26, with the default parameters.

3.2.4. Genome sequencing

3.2.4.1. Illumina sequencing

Sequencing by synthesis was outsourced to MicrobesNG (Birmingham, UK). A sample of DNA from each bacterium was prepared from the extracts by diluting to the concentration requested by the company (15-30 ng μL^{-1}) with PowerSoil kit buffer C6 (10 mM Tris buffer). This was done on a 96-well plate which was then sealed and sent by courier to Birmingham, where the sample was processed. Libraries were prepared from the DNA extracts with a Nextera XT library Prep Kit (Illumina, San Diego, USA), deviating from the manufacturer's instruction only in the amount of input DNA (double the recommended) and increasing the duration of the elongation step of the PCR to 45 s. The process of quantifying DNA and preparing the libraries was carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). A number of libraries were pooled together, including the one prepared with our DNA and quantified with the Kapa Biosystems Library Quantification Kit for Illumina. Sequencing then took place on an Illumina HiSeq device using a 250 bp paired end protocol. Finally, the sequencer's output reads had their adapters trimmed with Trimmomatic 0.30, with the sliding window quality cutoff setting at Q15.

3.2.4.2. Nanopore sequencing

Nanopore sequencing took place in-house, at the ICEM laboratory at Aberystwyth University. Libraries were prepared for the MinION sequencer using the SQK-RBK004, which takes 400 ng of DNA as input in a volume of 7.5 µL. DNA extracts were diluted or concentrated to approximate the desired concentration of 53.3 ng μL^{-1}. For those with higher concentrations than needed a volume sufficient to reach the 400 ng of DNA was taken from the stock and placed in a 1.5 mL microcentrifuge tube, with the remaining volume until 7.5 µL being built up by adding PowerSoil kit buffer C6 (10 mM Tris buffer). Only two DNA extracts, those of *Moritella viscosa* and *R. salmoninarum*, had concentrations below the minimum necessary. The AMPure protocol for cleanup of DNA, described in section 3.2.3.1, can also be used to increase concentration of DNA, if the elution step takes places in a smaller volume than that of the initial input. This was applied to the *M. viscosa* and *R. salmoninarum* DNA extracts. The concentrated DNA was now re-quantified and adjusted to the desired concentration (53.3 ng/µL) with buffer C6 (10 mM Tris).

Library preparation followed the manufacturer's instructions using the SQK-RBK004 library preparation kit (Oxford Nanopore Technologies, Oxford); this involved adding 2.5 µL of a

different fragmentation mix (RB1-8) to each of the samples, mixing by gentle flicking, and centrifuging to ensure all the reaction mix accumulated at the bottom of the tube. The tubes containing the DNA extract-fragmentation mix combination were incubated for one minute at 30 °C and then immediately at 80 °C for another minute, followed by rapid cooling on ice. It is worth noting that the fragmentation mix fragments the DNA and tags it with one of 12 barcodes provided with the kit: these barcodes are sequences that allow reads from samples that are sequenced together to be sorted back to their original samples. To maximise the number of samples sequenced, eight samples from this project, plus four samples from a colleague's were barcoded and pooled together. An equivalent volume of AMPure XP beads were added to the pooled DNA, and the sample was incubated for 5 minutes at room temperature. Beads were pelleted on a magnetic rack, and were washed with 200 µL 70% ethanol twice, air-dried briefly, re-suspended in 10 µL buffer C6, and incubated for 2 minutes. Beads were pelleted again on a magnetic rack and 10 µL of pure barcoded DNA were removed to a new 1.5 mL microcentrifuge tube. 1 µL of the Rapid Adaptor (RAP) reagent was added to the 10 µL of the pooled, barcoded DNA and mixed by gentle flicking, then incubated in a rotisserie at room temperature for 5 minutes.

The MinION sequencer was prepared by fitting it with an appropriate flowcell (R9.4.1), which was then primed as per the manufacturer's instructions. Briefly, this involved mixing 30 µL of Flush Tether (FT) into one tube of Flush Buffer (FB) and loading 800 µL of the mix onto the flowcell via the priming port, incubating for 5 minutes. In the meantime, the loading mix was prepared by mixing 34 µL of Sequencing Buffer SQB, 25.5 µL of Loading Beads LB (recently mixed by pipetting), 4.5 µL of nuclease free water and 11 µL of DNA-adaptor mix. Priming was completed by adding a final 200 µL of Priming Mix to the flowcell via the priming port. The loading mix was mixed by gentle flicking and loaded immediately afterwards, by pipetting it onto the SpotOn™ port of the flowcell.

Sequencing was controlled by a MinIT computer, using MinKNOW (software version: minknow-core-minit-offline 3.0.0, ont-minknow-static-frontend 3.0.3) and run for a total of 16 hours, with priming mix being loaded again after 12.5 hours. Basecalling was carried out on the MinIT itself, through the program's embedded instance of Guppy (v1.8.5). Statistics of average length, quality, standard deviation of length, read length N50, number of reads and total number of bases were produced for the entirety of the reads obtained using NanoPlot 1.29.0 (De Coster et al., 2018); this program also produced histograms of read lengths by number of reads and by number of bases.

3.2.5. Comparison of demultiplexing software

Reads which passed the default quality threshold of Phred score 7 were then classified according to barcode ("demultiplexed") using Qcat v1.1.0, specifying the barcodes used were the ones in the SQK-RBK004 kit. The shortest reads were removed from the dataset corresponding to each barcode, as they are the least useful in building genome assemblies. Filtlong v0.2.0 was used to filter the 5% of the bases which formed part of the shortest reads of each dataset, and NanoPlot v1.29.0 was used to calculate statistics before and after filtering and produce plots as mentioned earlier.

The same process was carried out with Guppy_barcoder v4.4.2 using the GPU version of the software, filtering with Filtlong v0.2.0 and the same statistics were gathered as before with NanoPlot v1.29.0 both before and after filtering. After the comparison, reads demultiplexed with Qcat were used for the construction of assemblies; details are given in section 3.2.6.

A brief check for contamination of the reads was carried out with Kaiju v1.7.4 (Menzel et al., 2016), which taxonomically classifies reads to the lowest possible level, and can serve to identify reads originating from unexpected organisms. The output was visualised with KronaTools 2.7 (Ondov et al., 2011).

3.2.6. Assembler comparison

Seven genome assemblies were built for each bacterium, each using a different method. The availability of short reads from the HiSeq and long reads from the MinION allowed short-read, long-read and hybrid assemblies to be built. Unicycler v0.4.8 was used to produce short-read, long-read and hybrid assemblies. The hybrid approach was intended to produce a useful assembly, and serve as benchmark for the rest of assemblies. Long-read assemblies were constructed using Flye v2.8.3 (Kolmogorov et al., 2019), the miniasm v0.3 plus Minipolish v0.1.2 pipeline (H. Li, 2016; R. Wick & Holt, 2021), Raven v1.5.1 (Vaser & Šikić, 2021) and Redbean (WTDBG2 v0.0 + WTPOA-CNS v0.0) (Ruan & Li, 2020). These are henceforth referred to as Flye, Minipolish, Raven and Redbean respectively. The default options were used except for Flye, where polishing iterations were increased to two from the default of one. Graphical fragment assemblies (GFAs) were produced where possible. Some of these programs (Flye, Redbean) required estimates of genome size; the estimates provided were based on the size of the representative genome for each species in the NCBI Genome database (as of 05/08/21).

The assembly software was run on a High Performance Computing (HPC) cluster at Aberystwyth University ("Bert"), using the Sun Grid Engine (SGE) scheduling system to assign eight threads and 64 GB of RAM to each job. Time and maximum memory usage were recorded from the SGE accounting data.

3.2.7. Assembly evaluation

Basic assembly statistics were gathered with Quast v5.0.2 (Gurevich et al., 2013), which was also used to estimate misassemblies by comparing to a reference. The definition that Gurevich and colleagues (2013) use for misassemblies in QUAST, is the regions where two sections that are contiguous in the test assembly are separate by more than 1 kbp in the reference, or they overlap, or are present on separate chromosomes. These were calculated twice, once using the representative genome for each species in the NCBI Genome database (as of 05/08/21) as reference, and the second time using the hybrid Unicycler assembly benchmark. An annotation of the reference genome's ribosomal RNA genes was obtained using Barrnap v0.9 (Seemann, 2018; https://github.com/tseemann/barrnap), and used with Quast as well, in order to evaluate the degree to which the assemblers recovered these genes. Completeness of the assemblies was evaluated with BUSCO v5.2.2 (Manni et al., 2021; Simão et al., 2015), which ascertains the presence of a number of universal, single-copy orthologous genes that are expected to be present in every genome of a given lineage.

Additionally, the recovery of genes, especially of those relevant to aquaculture operations, as are those related to antimicrobial resistance (AMR), was assessed. Assemblies were annotated with Prokka v1.14.6 (Seemann, 2014) and searched for AMR genes with Abricate v1.0.0 (Seemann, 2020; https://github.com/tseemann/abricate) against the MEGARes 2.0 database (Doster et al., 2020).

3.3. Results

3.3.1. Identity confirmation

Bacterial cultures grew as expected based on the descriptions of the NCIMB in terms of appearance and growth times. Due to the abandonment of labwork as a result of the COVID-19 pandemic, strains had to be revived from glycerol stocks; it was impossible, however, to revive *M. viscosa*, and the DNA extracts obtained as a result of the work described in section 3.2.2 were re-used throughout the rest of the project. The first DNA extraction provided sufficient nucleic acids for 16S PCR, which in turn prvided sufficient DNA for Sanger sequencing.

Aberystwyth University's Sanger sequencing service returned eight chromatograms and their respective sequences; these were shorter than the distance between the two primers because Sanger sequencing chromatograms are often difficult to interpret near their ends. Lengths ranged from 954 to 1103 bases, which was sufficient to cover most of the region of interest and assign a species with low e-values from the BLASTN search against the NCBI's rRNA/ITS databases (Table 3-4). In all cases, the expected genus was assigned, and the expected species was assigned for seven out of eight sequences. In the case of sequence A08, which corresponded to what was purchased as *T. maritimum* NCIMB 2154, the hits obtained through BLAST correspond to a variety of different species of the genus *Tenacibaculum*, plus some to the genus *Polaromonas*.

The phylogenetic analysis run for the 16S rRNA gene of the *Tenacibaculum* genus place this project's strain as most closely related to the type strains of *T. maritimum* (*T. maritimum* NBRC 15946 and *Flexibacter maritimus*, a synonym for this species). The clade they form is backed by a 100% bootstrap support (see Figure 3-1), and the bacterium from this project groups more closely with them than with the representatives of any of the other species in the genus (Table 3-4). This suggests that the bacterium in question is indeed *T. maritimum*.

Table 3-4: Abridged results of searching the NCBI 16S database for this project's 16S sequences via BLAST.

Sample (expected species)	Sequence length	Organism	Max Score	Total Score	Query Cover	% identity	Acc. Len.	Accession
A01 (*Aeromonas salmonicida salmonicida*)	954	*Aeromonas salmonicida*	1482	1482	99%	95.28	1460	NR_119042.1
A02 (*Aeromonas salmonicida achromogenes*)	1067	*Aeromonas salmonicida* subsp. *masoucida*	1871	1871	99%	98.24	1467	NR_040829.1
A03 (*Flavobacterium branchiophilum*)	976	*Flavobacterium branchiophilum*	1773	1773	99%	99.49	1437	NR_104713.1
A04 (*Flavobacterium psychrophilum*)	1103	*Flavobacterium psychrophilum*	1978	1978	100%	98.92	1467	NR_040914.1
A05 (*Yersinia ruckeri*)	1027	*Yersinia ruckeri*	1803	1803	98%	98.63	1461	NR_115976.1
A06 (*Renibacterium salmoninarum*)	1074	*Renibacterium salmoninarum*	1927	1927	99%	98.98	1471	NR_074198.1
A07 (*Moritella viscosa*)	1029	*Moritella viscosa*	1853	1853	100%	99.04	1503	NR_028880.1
A08 (*Tenacibaculum maritimum*)	1070	*Tenacibaculum sediminilitoris*	1629	1629	100%	94.09	1442	NR_149768.1

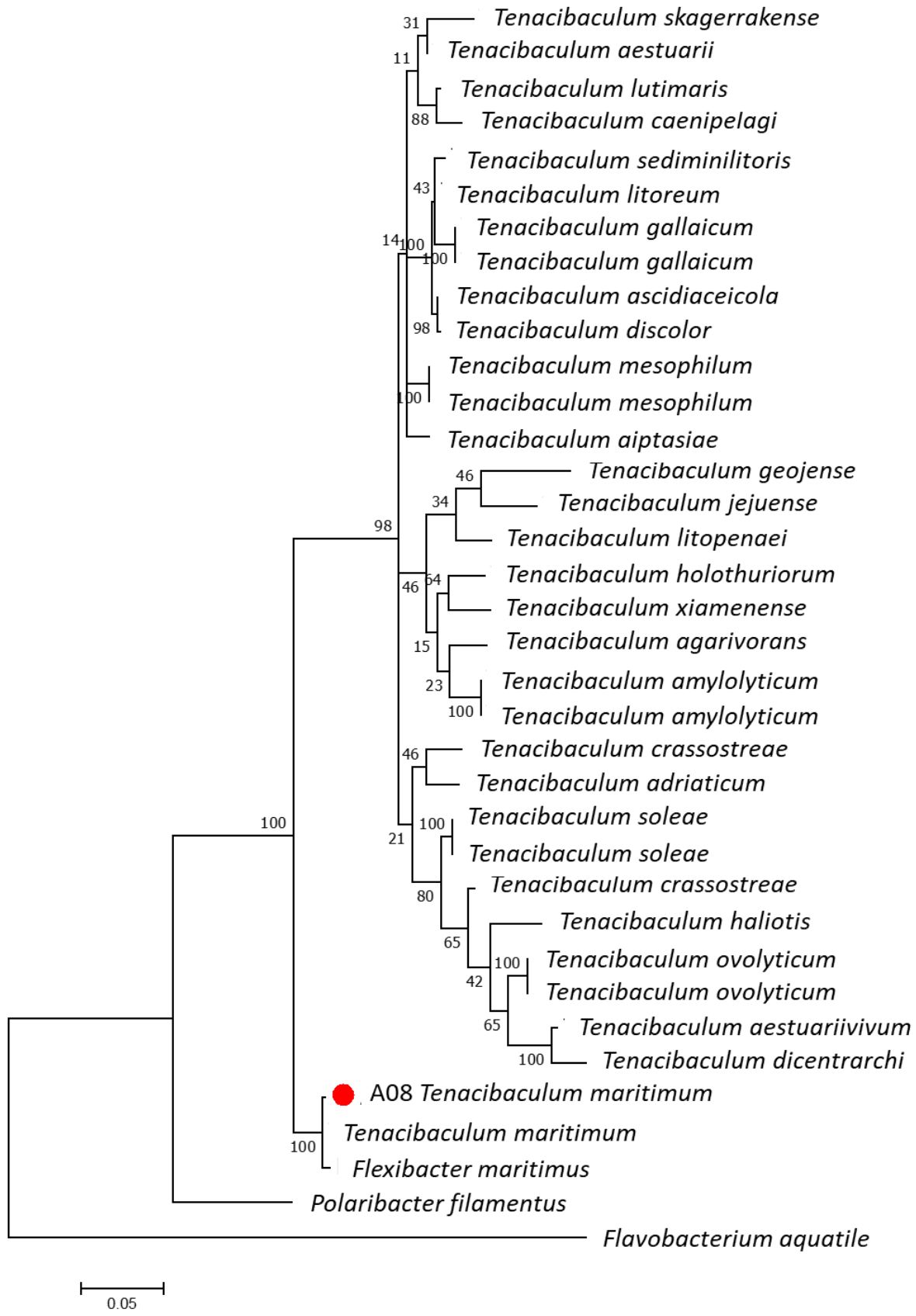


Figure 3-1: Maximum Likelihood phylogenetic tree for the 16S ribosomal RNA gene of the type strains of the Tenacibaculum *genus. It includes genes from* Polaribacter filamentus *and* Flavobacterium aquatile*, the latter is used to root the tree. The bacterium used in this project is marked by a red circle.*

3.3.2. Illumina sequencing

A total of 10,375,508 pairs of reads were obtained for this project's samples, with lengths between 36 and 251 bases. The number of reads obtained per sample ranged between 2,862,858 for *F. branchiohilum* to 224 for *T. maritimum* (Table 3-5). This low number of reads made it impossible to use them in the construction of a short-read only genome of *T. maritimum*.

Table 3-5: Overview of Illumina sequencing results; number of reads obtained per sample and coverage estimated by MicrobesNG (sequencing provider).

Sample	Read pairs	Mean coverage
Aeromonas salmonicida salmonicida	1,494,369	111.712
Aeromonas salmonicida achromogenes	1,465,998	71.616
Flavobacterium branchiophilum	2,862,858	174.346
Flavobacterium psychrophilum	416,478	54.536
Yersinia ruckeri	267,131	110.282
Renibacterium salmoninarum	1,091,429	115.239
Moritella viscosa	2,777,021	159.535
Tenacibaculum maritimum	224	102.813
Total	10,375,508	-

3.3.3. Nanopore sequencing

A total of 817,140 reads were obtained from the MinION sequencer, with 1,828,256,466 bases obtained in total. Read length was, on average (mean) 2,237.4 bases, and the N50 statistic was 7,500 bases, although read size distribution can better be appreciated in Figure 3-2. Mean read quality was 6.5 on the Phred scoring system.

Table 3-6: Results of demultiplexing with two different programs.

Barcode number	Organism of origin / expected organism	Number of reads		Number of bases	
		Qcat	Guppy barcoder	Qcat	Guppy barcoder
1	*Aeromonas salmonicida salmonicida*	37,864	35,510	202,375,415	185,654,883
2	*Aeromonas salmonicida achromogenes*	22,556	21,561	82,371,402	75,989,036
3	*Flavobacterium branchiophilum*	35,028	33,153	69,300,014	61,823,171
4	*Flavobacterium psychrophilum*	37,011	35,193	170,332,906	157,426,330
5	*Yersinia ruckeri*	58,179	54,076	307,434,092	279,600,432
6	*Renibacterium salmoninarum*	22,748	21,241	73,610,340	65,899,301
7	*Moritella viscosa*	18,498	16,789	28,666,430	24,128,778
8	*Tenacibaculum maritimum*	27,708	25,653	51,354,222	44,690,377

Table 3-6 shows the number of reads and bases per each of the barcodes after demultiplexing with either Qcat or Guppy; Qcat classified more reads and bases than Guppy in every instance. However, it is also true that the mean length and quality of the reads was higher when classified with Guppy. Despite this, Qcat was used to demultiplex reads for assembly because it classified more bases and mean read length was comparable. These are the most important parameters considered because the former directly influences the assembly coverage, and for the latter, the longer reads are, the more useful they are in resolving repetitive regions of the genome (H. Lee et al., 2016).

Filtration of the datasets removed the 5% of the bases that was present in the shortest reads of each barcode's dataset. The percentage of reads removed per Qcat-demultiplexed dataset was on average 36.5%, which served to increase the mean read length by 51.1% (see Figure 3-3).

As a collateral benefit, an increase was also seen in mean read qualities. Similar improvements were also found when filtering the Guppy-demultiplexed datasets (Figure 3-3).

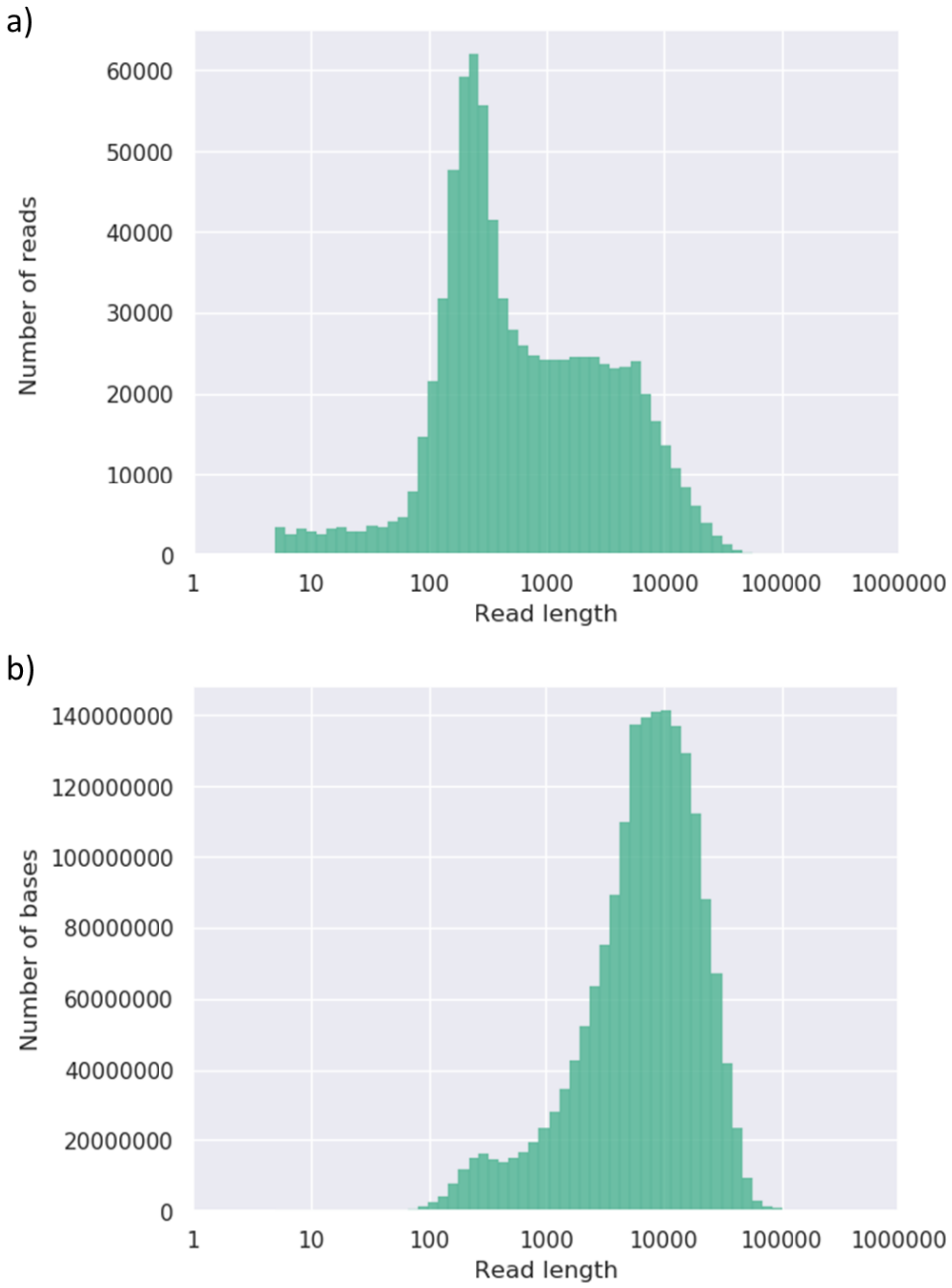


Figure 3-2: a) Size distribution of MinION long reads on a logarithmic scale b) Distribution of bases per read length on a logarithmic scale.

The checks carried out with Kaiju (Menzel et al., 2016) served to ascertain that the majority of reads for each sample corresponded to the taxa of the expected bacterium. *T. maritimum* was the exception, with no reads assigned to the expected species, although 22% were assigned to the genus *Tenacibaculum*, and 85% to the family Flavobacteriaceae, to which this species belongs (Figure 3-4).

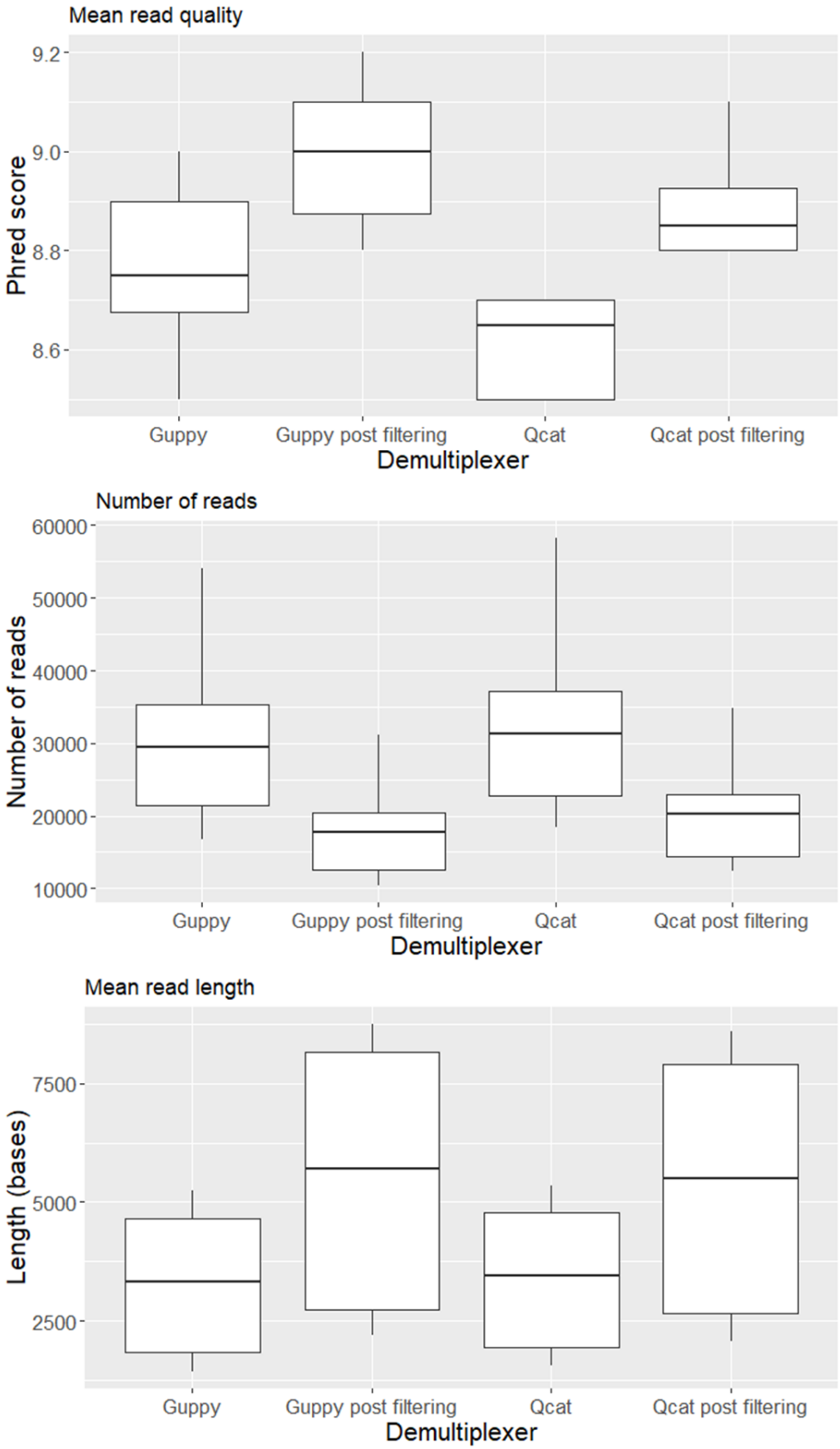


Figure 3-3: Differences in mean read quality, length and total number of reads between the datasets after demultiplexing with either Guppy barcoder or Qcat, and with or without removing the 5% shortest reads.

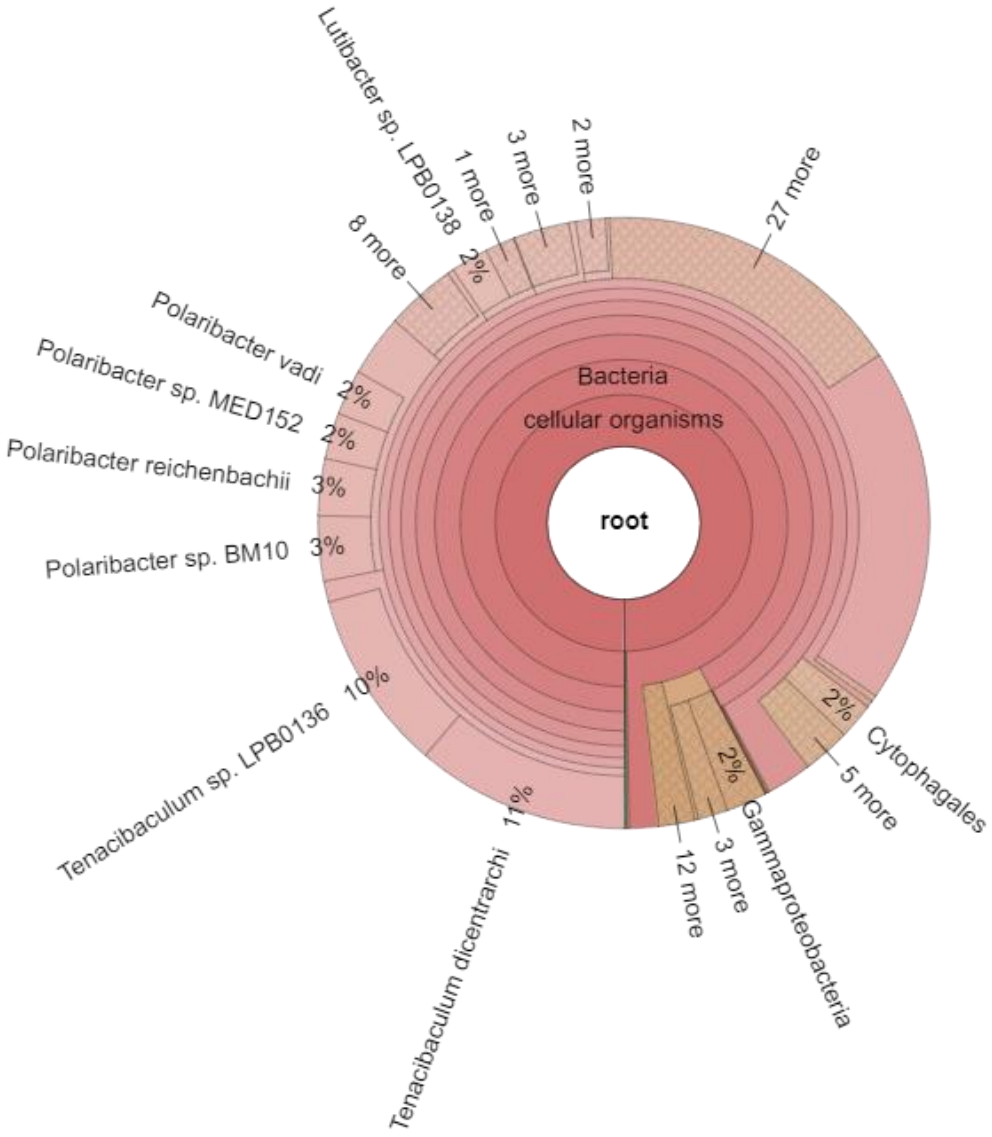


Figure 3-4: Taxonomy of the long reads obtained through nanopore sequencing for sample with barcode 8, expected to be Tenacibaculum maritimum*, as assigned by Kaiju (Menzel et al., 2016).*

3.3.4. Assembly resource requirements (time and RAM)

Assemblies ran in times ranging between 10 seconds (Raven's assembly of *M. viscosa*) and 5 hours 9 minutes (Unicycler's hybrid assembly of *F. psychrophilum*), although the maximum for a long-read-only approach was Flye's assembly of *Y. ruckeri* (25 minutes). On average, it took 6 minutes and 13 seconds for a long-read assembly (see Figure 3-5). Mean maximum memory requirements were 16.3 GB of RAM for long-read approaches, which compares favourably to the mean of short-read or hybrid approaches with Unicycler (35.9 GB and 29.0 GB respectively, see Figure 3-6). It is worth noting that in terms of maximum memory requirements, there is substantial variation around the mean; this can be appreciated in Figure 3-6.

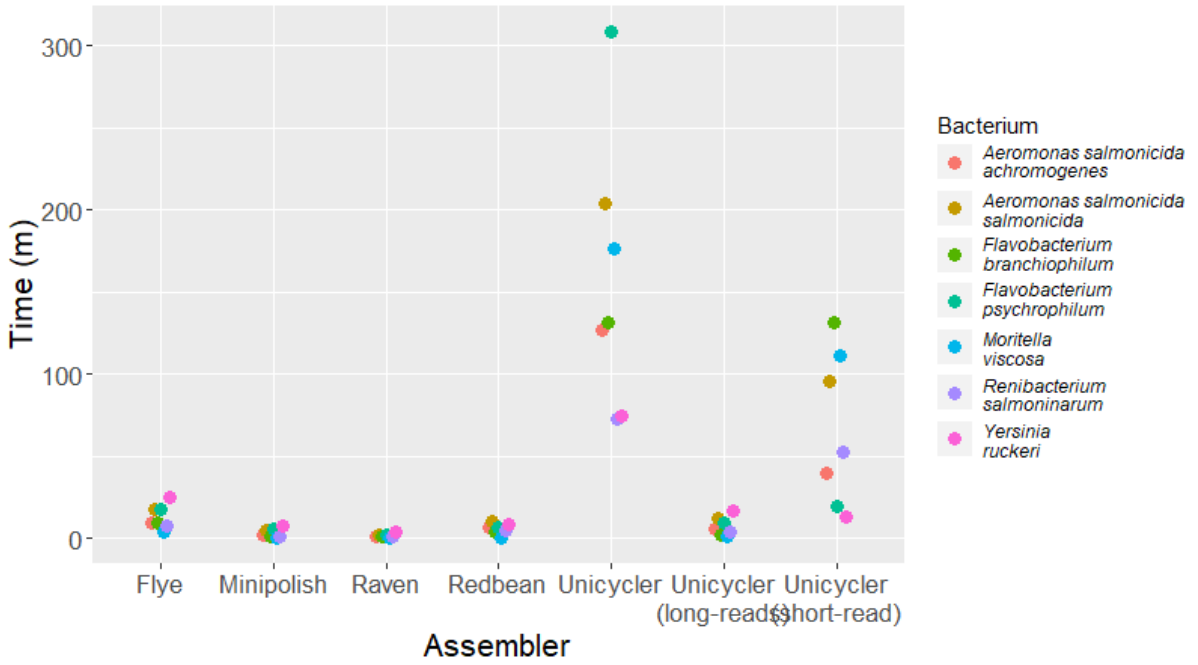


Figure 3-5: Time taken (in minutes) for the different assemblers to run to completion. Each genome is depicted in a different colour (see legend).

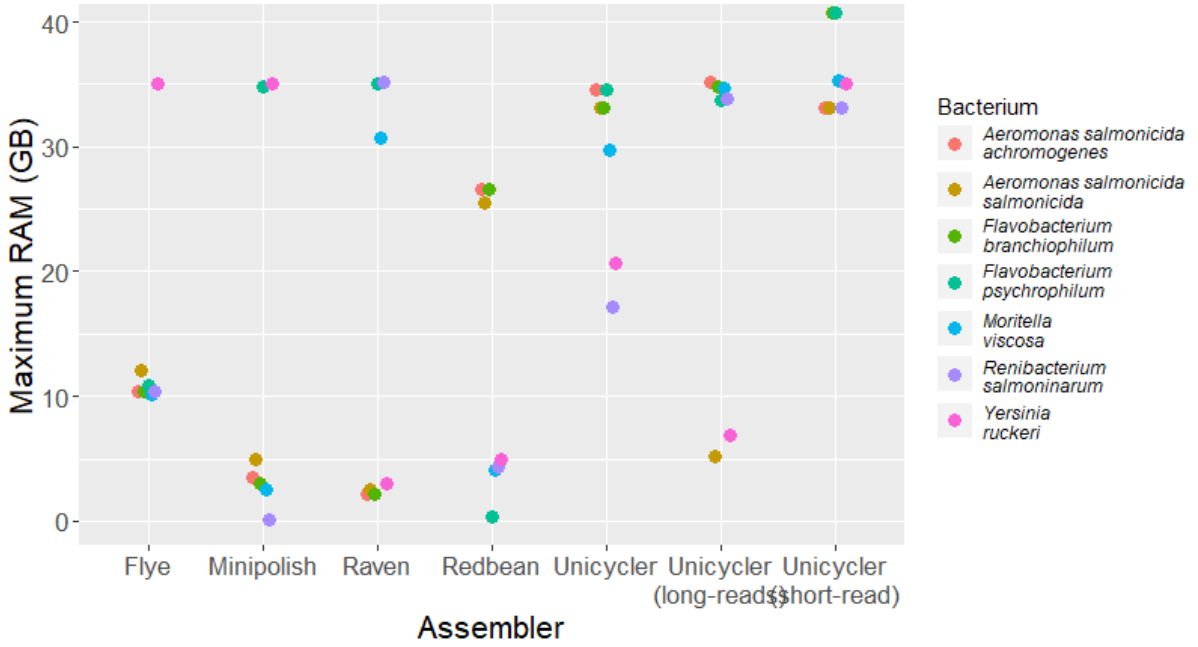


Figure 3-6: Maximum RAM used by each program during genome assembly. Each genome is depicted in a different colour (see legend).

3.3.5. Hybrid assemblies

The results of hybrid assemblies with Unicycler return genomes that are almost complete in all cases (see Table 3-7). They have low L50s, in most cases of one, which indicate that the majority of the genome sequence has been assembled into a single contig, the most striking exception being *M. viscosa*. This assembly has an L50 of 6, which might be related to the low long-read coverage; this may have impeded the resolution of bridges between contigs by the software. Visualising the graphical format assemblies for *A. salmonicida salmonicida*, and *Y. ruckeri* shows that the contigs recovered have all been circularised, that is, that the ends have been identified as overlapping and thus probably represent a complete bacterial chromosome or plasmid (Figure 3-7). Here *A. salmonicida salmonicida* is found to have one main chromosome of 4,747,635 bp, plus five plasmids of lengths ranging from 5,247 bp to 118,290 bp. For *Y. ruckeri*, it is a 3,695,287 bp long chromosome and two plasmids of 16,925 bp and 103,906 bp respectively. In the other cases, such circularised, unambiguous contigs are only retrieved for what are probably plasmids, as they are shorter and circular; there is one such contig in the assemblies of *F. branchiohilum* (3,612 bp), *F. psychrophilum* (2,602 bp), *R. salmoninarum* (62,169 bp) and three in that of *M. viscosa* (4,956 bp to 9,199 bp). In these cases, different ambiguities remain in the resolution of the bacterial chromosome. In any case, the sizes of the assemblies recovered are within 3% of the median size for their species genomes as per the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/, consulted 12/08/21) except for that of *F. branchiohilum*, which differs by 5.1%. The GC% of the genomes reconstructed in this project are also similar to their species medians, as per the NCBI, differing by less than 0.4% in all cases. See Table 3-7 for general statistics of genome size, quality and contigs assembled completeness.

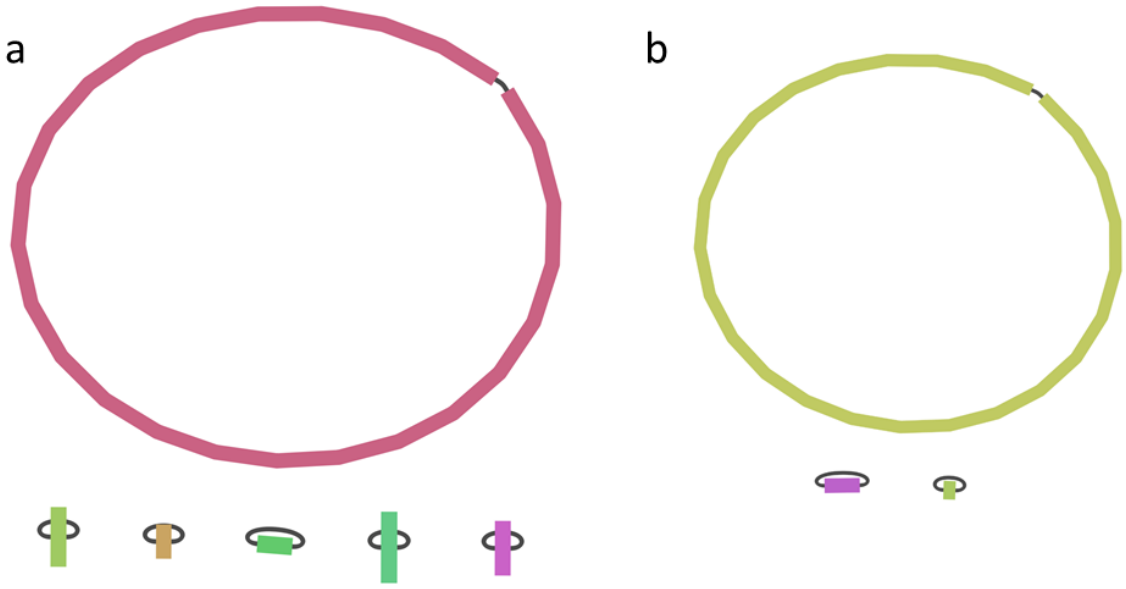


Figure 3-7: Bandage plots of the a) A. salmonicida salmonicida *and b)* Y. ruckeri *hybrid assemblies (carried out with Unicycler). These are the only two cases where the assemblies were closed (large circles, bacterial chromosome). The smaller elements represent plasmids*

Table 3-7: General statistics of genomes obtained through hybrid assembly. Length, number of contigs, long read coverage and L50 were calculated with QUAST v5.0.2, completeness was calculated with BUSCO v5.2.2.

Bacterium	Length (bp)	Contigs	GC %	Long-read coverage	L50	Completeness (%)
Aeromonas salmonicida achromogenes	4,690,209	7	58.81	16	1	99.7
Aeromonas salmonicida salmonicida	4,888,583	6	58.37	38	1	99.7
Flavobacterium branchiohilum	3,764,375	23	32.79	17	2	98.9
Flavobacterium psychrophilum	2,710,819	7	32.64	58	1	99.6
Moritella viscosa	5,207,364	52	39.31	5	6	98.1
Renibacterium salmoninarum	3,155,228	2	56.27	21	1	98.2
Yersinia ruckeri	3,816,118	3	47.5	75	1	99.5

3.3.6. Long-read assemblies

3.3.6.1. Genome size

Long-read assemblies were mostly similar to the benchmark in terms of size, although assemblies of *M. viscosa* were often much shorter (see Figure 3-8). This is possibly a consequence of the very low long-read sequencing depth, as in the case of the short read assembly, the total length is only 1.21% lower. It is important to consider this when interpreting the coverage data, as there is a coverage of 168X for Minipolish's assembly of this bacterium's genome, but its length is a meagre 71,049 bp, which is much less than the size of the benchmark (5,207,364 bp).

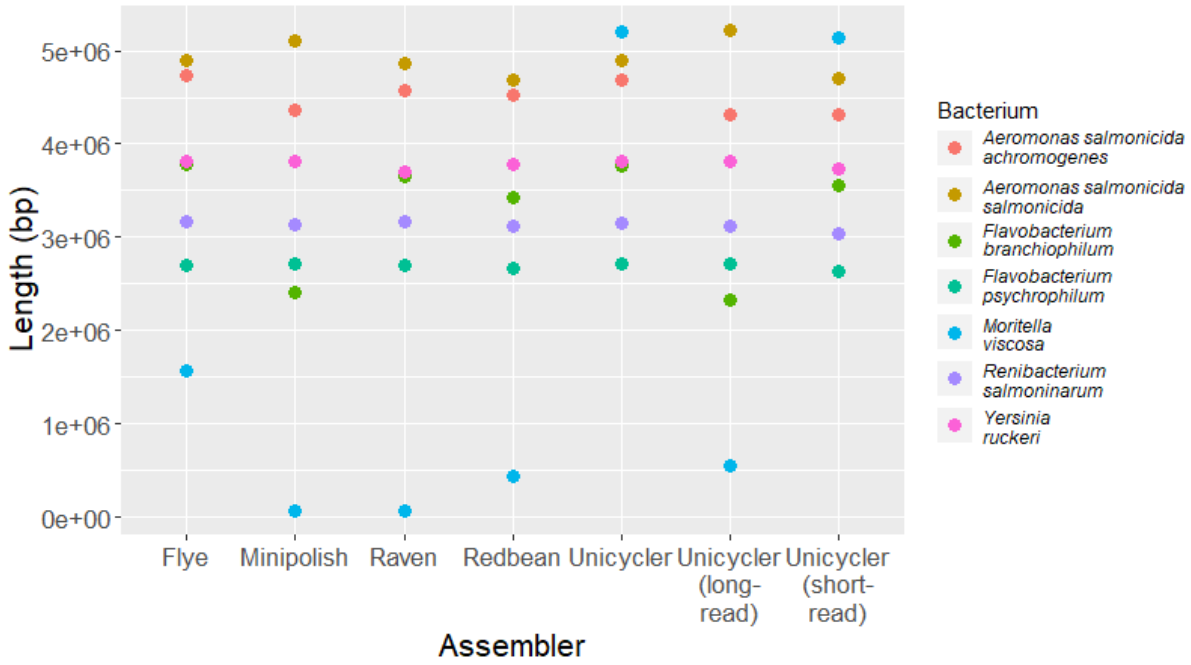


Figure 3-8: Size of assembled genomes depending on assembler and bacterial species.

3.3.6.2. Genome fragmentation

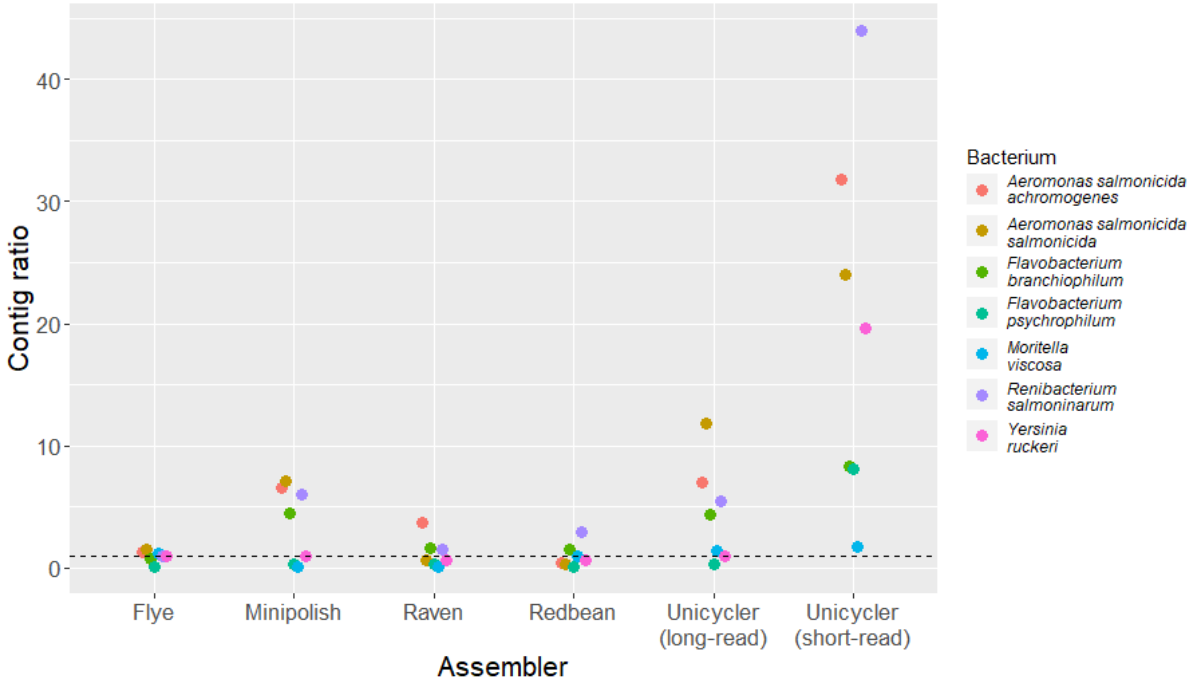


Figure 3-9: Contigs assembled per bacterium and assembler, as a ratio of contigs obtained through hybrid assembly with Unicycler. The dashed line marks a ratio of 1.

Assemblies by long-read assembly methods are less fragmented than those obtained by Unicycler's assembly of short reads. On average, when Unicycler assembles short read datasets, 19.7 times as many contigs are constructed than when it uses a hybrid dataset; the long-read approaches return between 0.989 (Flye) and 4.49 (Unicycler, long-read only) times (Figure 3-9).

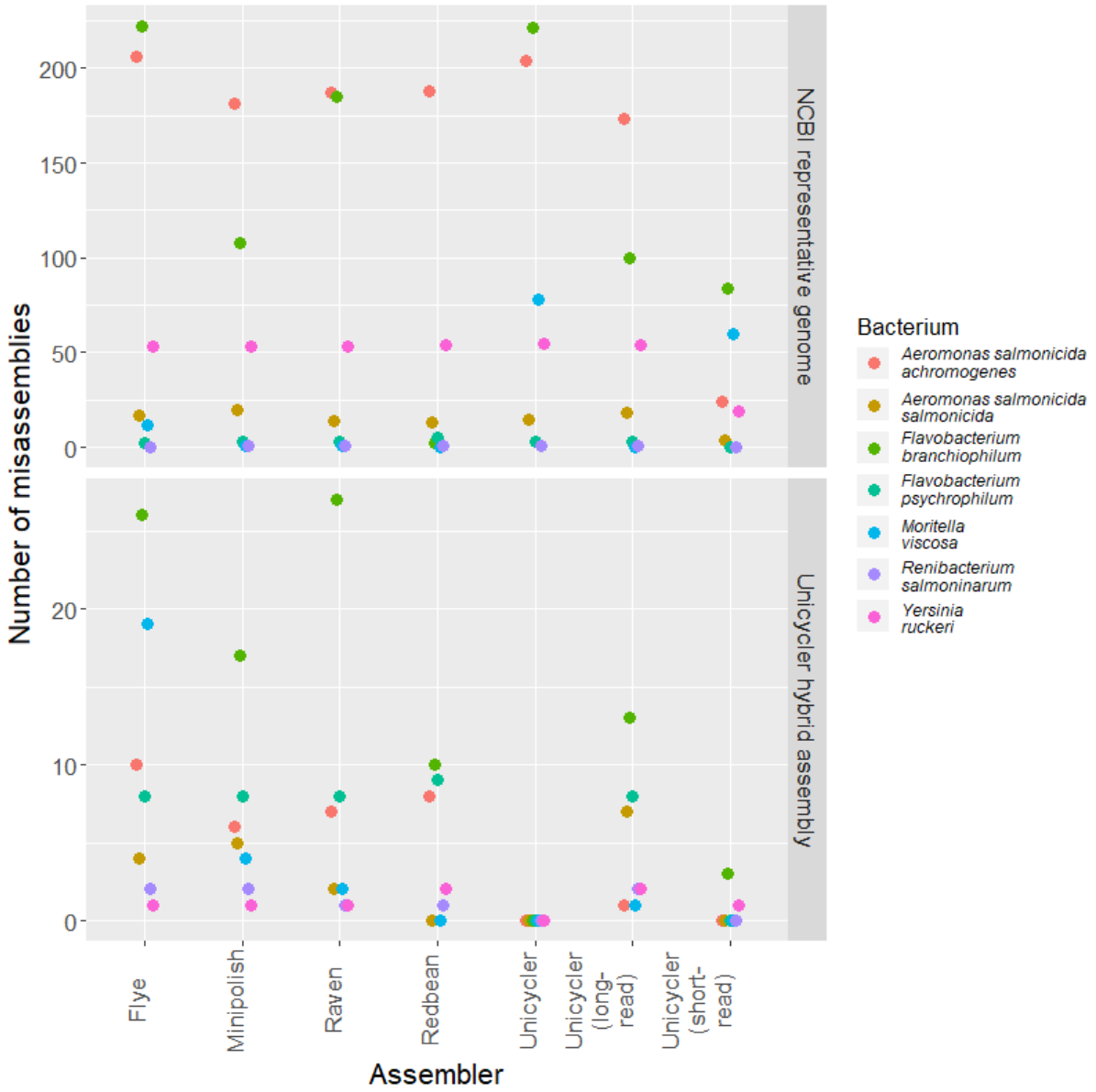


Figure 3-10: Number of misassemblies given by each assembler for each genome when compared against two different references. In one case (top), the reference is the NCBI's representative genome for the species. In the other (bottom), the hybrid genome assembled for each bacterium with Unicycler is used as reference.

Misassemblies were mostly consistent across long-read-only assemblies for each bacterium when calculated against the NCBI's representative genome, but they were much higher than when calculated against the Unicycler hybrid assembly benchmark (see Figure 3-10). When the latter was used, the mean number of misassemblies per assembler was between 4.28 (Redbean) and 10.0 (Flye). The short-read-only Unicycler approach resulted in a mean of 0.57 misassemblies; this method also resulted in the least misassmblies when calculating against the NCBI representative genome (27.3).

The long-read only genome assemblies tend to be less complete than the hybrid or short-read only ones; the mean percentage complete BUSCOs found per assembly for the long read assemblers ranged between 8.09% (Redbean) and 23.8% (Minipolish), whereas that of the hybrid and short-read methods were both 99.1%. Including fragmented BUSCOs in the count, a total of between 31.5% (Redbean) and 51.6% (Flye) of these marker genes was found in the long-read assemblies, which is still substantially lower than the assembly methods that incorporate short reads.

3.3.6.3. Genome annotations (protein-coding, ribosomal and AMR genes)

The results of the gene annotation of the assemblies are presented in Figure 3-11 and Figure 3-12. The former shows the number of protein-coding genes predicted to be found in each assembly relative to the benchmark, and the latter the number of ribosomal RNA genes (5S, 16S and 23S), both as proportions. Long-read assemblers recover a higher number of protein-coding genes than the benchmark, often more than doubling the number of genes predicted for the benchmark genome. Regarding the ribosomal RNA genes, the number recovered by the long-read methods is substantially nearer to the benchmark than that of the short-read method. Among them, the average proportion of rRNA genes recovered are 0.738 (Redbean), 0.814 (Minipolish), 0.825 (Unicycler – long-reads only), 0.900 (Raven) and 0.957 (Flye); for the short-read-only method this is 0.238. In fact, the short-read assemblies recover three or four rRNA genes in all cases; two 5S rRNA genes are recovered from the *A. salmonicida* subspecies and *Y. ruckeri*, two 16S rRNA genes are recovered for *M. viscosa*. This highlights the ability of long reads to resolve long repetitive regions of DNA sequences on assembly. It is worth noting that for all the long-read methods used here, it is for the *M. viscosa* genome that the fewest rRNA genes are recovered.

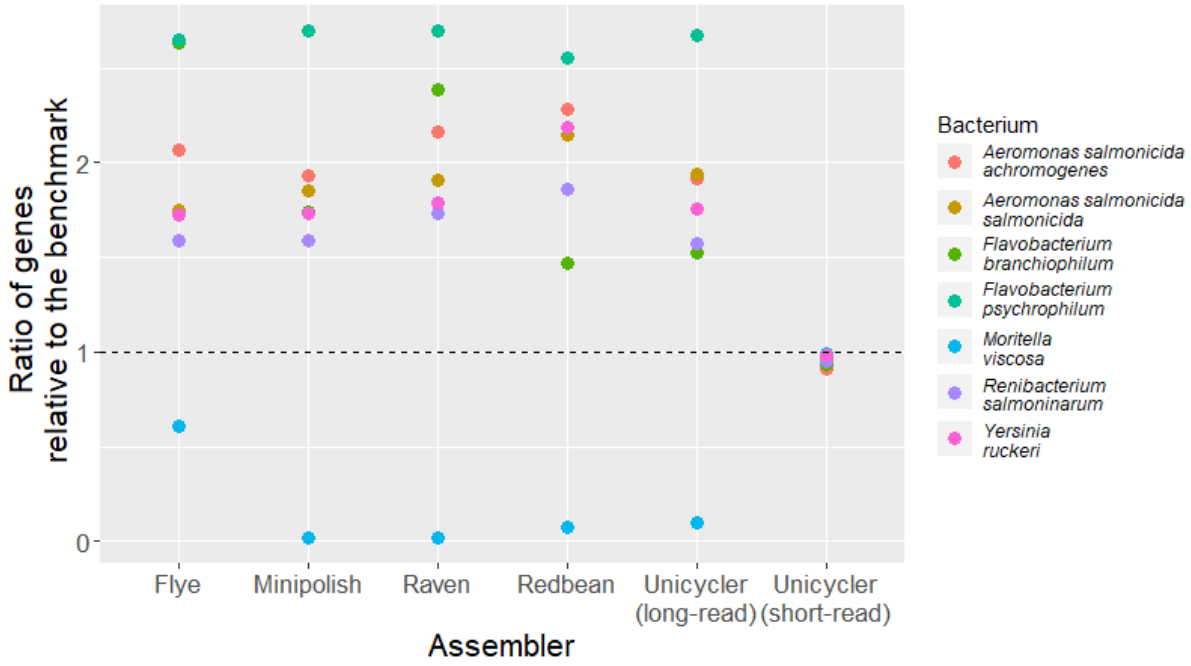


Figure 3-11: Ratio of protein coding genes identified in any given assembly as compared to the benchmark. The Unicycler hybrid assembly is used as benchmark and a ratio of 1:1 is indicated by the dashed line.

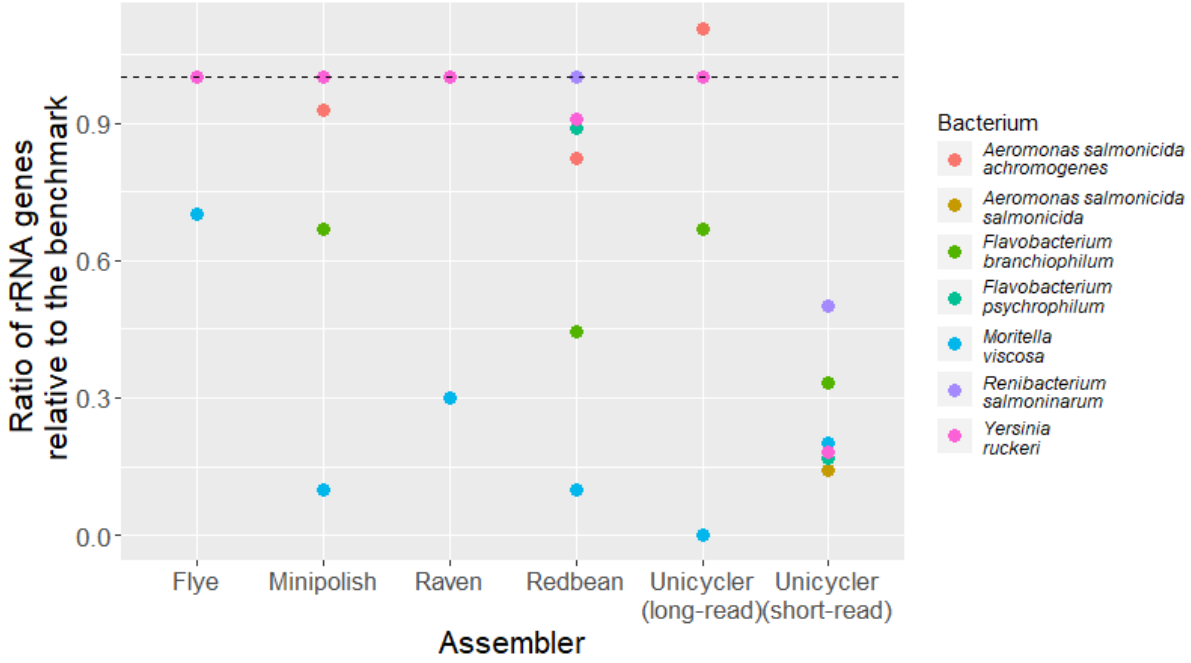


Figure 3-12: Ratio of rRNA genes identified in any given assembly as compared to the benchmark. The Unicycler hybrid assembly is used as benchmark and a ratio of 1:1 is indicated by the dashed line.

AMR genes, which are of relevance in aquaculture operations affected by any of the pathogenic bacteria considered in this study, were found through Abricate in the genomes of

three of the strains considered. Three, five and one genes were found in *A. salmonicida achromogenes*, *A. salmonicida salmonicida* and *Y. ruckeri* respectively, and none in any of the other bacteria. This number was consistent across different assembly methods, including the hybrid and short-read methods, for each bacterium, only with the exception of Redbean. For the *A. salmonicida salmonicida* genome assembled with this program, only three AMR genes were found.

3.4. Discussion

3.4.1. Overview

For this chapter, eight bacterial pathogens of interest (BPOIs) of salmonid aquaculture were cultured and their genomic DNA extracted. This DNA was sequenced both on the MinION and HiSeq platforms (objective 1.1). The reads obtained were used to assemble genomes with five long-read assemblers, one hybrid assembly method as benchmark and one short-read assembly method for comparison (objective 1.2). The time taken and maximum memory used by these programs was recorded, and the assemblies were evaluated for completeness, contiguity and coverage. These assemblies were also annotated for protein coding and ribosomal genes, and searched for antimicrobial resistance genes, and the results were compared (objective 1.3). Additionally, some consideration was given to the most appropriate demultiplexing software to process the read data.

3.4.2. A note about demultiplexing

In the experiments discussed in this chapter, Qcat was used because it retrieved more bases than Guppy_barcoder. Qcat, however, is deprecated, with no updates since July 2019 (Rescheneder et al., 2018/2019). It was considered because a comparison with an older version (v1.0.1) of the program and other basecallers (Guppy v2.3.5 and Porechop v0.2.4) had been posted on the Oxford Nanopore Technologies forum. In this comparison, Qcat was determined to be the most stringent demultiplexer, and the least prone to incorrectly place reads (Alcantara, 2019). Guppy was also considered because this software suite incorporates the most recent tested advances from ONT in demultiplexing algorithms. In fact, Guppy's results improved on those of Qcat on several metrics, such as mean read quality or length. These positive characteristics, combined with the fact that the more recent versions of the Guppy suite include a GPU version, mean it is possible that this program is more appropriate for assemblies and for low-resource settings.

3.4.3. Hybrid assemblies

The hybrid assemblies presented in this chapter were expected to serve as benchmark for the long-read assemblies, yet they are interesting in themselves. It is possible that these are the first genome assemblies that combine nanopore and Illumina sequencing-by-synthesis data for a number of our isolates; among the "complete" assemblies on NCBI Assembly, only one (*Y. ruckeri* KMM821; https://www.ncbi.nlm.nih.gov/assembly/GCF_017498685.1/) combined these two types of data. Unfortunately, only two of the assembled genomes

resulted in “complete”, circularised assemblies, with distinguishable chromosome and plasmids, those of *A. salmonicida salmonicida* and *Y. ruckeri*. This is quite possibly due to coverage, although for instance *F. psychrophilum*, for which there is more data than *A. salmonicida salmonicida*, does not assemble as well. Considering their general assembly statistics and their high degree of completeness the assemblies of *A. salmonicida salmonicida* and *Y. ruckeri* could be valuable in further studies of these organisms. This is especially interesting in the case of *A. salmonicida*, for which there were only four closed genomes available until recently (2017) on NCBI databases (A. T. Vincent et al., 2017). In fact Vincent and colleagues (2017) remarked on the importance of obtaining more genome assemblies of these characteristics to understand the evolution of *A. salmonicida*. This is especially important given the structural genomic variability of this species (see Chapter 2) and its relationship to the development of a psychrophilic lifestyle (A. T. Vincent et al., 2016).

In fact, regarding genomic rearrangements, it is possible to see that the hybrid assemblies accumulate a number of misassemblies when compared to the representative genome assemblies obtained from the NCBI (Figure 3-10). The sequencing data used in this chapter was obtained, in all cases, from strains different to those listed on the NCBI as the origin of the representative assemblies. This, together with the fact that for most bacteria every assembler that incorporates long-read data produces a similar number of misassemblies, suggests that these represent genuine genomic rearrangements.

3.4.4. Long read assemblies

When compared against the hybrid benchmark, the long-read-only assemblies often have a small number of misassemblies, no more than ten on average (mean). It is interesting to see the contrast with the short-read assemblies, though, which have a mean of 0.571. As stated in section 3.2.7, the definition for a misassembly is a region where two sections that are contiguous in the test assembly are separate by more than 1 kbp in the reference, or they overlap, or are present on separate chromosomes. Considering this definition, there may be so few of them in the short-read-only assemblies simply because whenever there is any uncertainty as to how to assemble a region, the assembly breaks. This will happen wherever there is a repetitive region that exceeds the size of the reads, which is more frequent in short read assemblies. Furthermore, the fragments that the short-read-only assembly is divided into can be expected to be very accurate, given the low error rate of the reads, which will facilitate

piling them. It is worth considering whether the improvement in contiguity that long-read methods provide compensates the higher chances of misassembly.

On the other hand, it is worth considering that the short read assemblies did not recover more than one, maximum two copies of each rRNA gene, when the genomes with the fewest copies have two of each (*R. salmoninarum*). Collapsing all the copies of these genes into a single one could be a disadvantage, given that often there is intragenomic variation in the different copies of 16S rRNA genes, which has been proposed as a means of achieving strain resolution in microbiome studies (Johnson et al., 2019). This does not happen with the long-read assemblers, which mostly recover the majority of these independent copies. The failure to assemble rRNA genes occurs mostly in the case of *M. viscosa*, so this should be taken with caution, if not entirely ignored, given the low coverage and poor assembly performance for this bacterium across the board. Excluding this bacterium from the analysis, Flye and Raven recover the same number of rRNA genes as the benchmark in every case, and the remaining assemblers recover most of them.

Mostly, long-read assemblers do not achieve remarkably different results among them in these measures of assembly quality. However, the differences when considering the universal single-copy ortholog search are more substantial. The retrieval of as few as 8.09% of these genes on average per genome in the case of Redbean, and a maximum of 23.8% with Minipolish is striking, because there is nothing else in the assembly statistics that suggests that significant portions of the genome are missing. This is most likely due to a phenomenon relating to the high error rates of long reads which is recognised among the genome-assembly community (Koren et al., 2019; Watson & Warr, 2019). The high error rates of long reads, which are especially indels, can result in frameshift mutations being introduced to the assembled sequence, and this can cause annotations of protein-coding genes to become truncated (Watson & Warr, 2019). This is in agreement with the large number of fragmented single-copy orthologs that are detected; these would represent such truncated annotations. Koren and colleagues (2019) support the use of more polishing, or better polishing algorithms, among other measures, to ameliorate this issue. It is perhaps the lack of an integrated polishing step in Redbean that causes its assemblies to have a much higher percentage of missing single-copy orthologs; the remaining tools incorporate this into their workflow, and perform substantially better on this parameter.

This characteristic of long-read based assemblies seems not to have affected the detection of AMR genes, although given that the number detected in each genome was low or null, no conclusions should be taken from this. In this chapter the default MEGARes database bundled with Abricate (Seemann, 2020) was used, and it is possible this results in not as many AMR genes being detected. For instance, (Kumru et al., 2020) detect AMR genes belonging to seven different classes in *F. branchiophilum* and *F. pscyhrophilum*, using the NCBI database and the AMR gene annotator ARG-ANNOT (Gupta et al., 2014). In general, however, it seems there is evidence for nanopore sequencing to provide sufficiently accurate assemblies for AMR gene detection (Su et al., 2019). However, Su and colleagues (2019) also warn that the relatively low consensus accuracy may not be sufficient for statistics-based methods that predict AMR based on single nucleotide polymorphisms (SNPs).

The results of the BUSCO analysis of completeness seem to contrast with those of the gene annotations; in the latter, instead of finding a reduced number of genes, a larger number than expected is found instead in the long-read assemblies. For this metric of evaluation, the closest to the benchmark is Minipolish (miniasm plus Minipolish pipeline), which still results in assemblies where a mean of 1.65 times as many genes are predicted as in the benchmark. This may be a result of several partial genes being found when a coding sequence is fragmented, as opposed to a single gene, or it could be that the inaccuracies of long reads introduce a number of start codons, although this is less likely. In any case, this could be studied further by comparing the annotations of the long-read assemblies and those of the benchmark to see what accounts for the difference. Additionally, it would be of interest to compare metabolic pathways recovered with different assemblers as compared to the benchmark.

3.4.5. Time and RAM considerations

Timewise, there is a clear advantage of all long-read assembly methods when compared to the Unicycler hybrid assembly. Indeed, it is known to be a relatively time-costly method, and it is unfeasible to run on large eukaryotic genomes (Haghshenas et al., 2020). Other faster hybrid assembly methods have been proposed (Gatter et al., 2020; Haghshenas et al., 2020), but these have not yet been used in independent, publicly available genome publications. In any case, Raven assembles genomes in the shortest time on average (1 m 37 s), followed by Minipolish (3 m 14 s); none of the long-read assemblers takes more than 26 minutes (Flye) to assemble a genome. Wick and Holt (2021), whose comparison involves several of the same

assemblers compared here, report somewhat faster mean run times for Redbean (2 m 18 s, versus the 5 m 51 s of the work in this chapter).

Regarding maximum memory (RAM) usage, there is a large variability within each assembler's results. All the assemblers take more than 30 GB of memory at some point during their execution for at least one of the assemblies except for Redbean. This is somewhat surprising, especially since Raven is specifically designed to have a low memory footprint and has performed favourably compared to other assemblers both in eukaryotic and prokaryotic datasets (Vaser & Šikić, 2021; R. Wick & Holt, 2021). In fact, the authors report using it to assemble an *Arabidopsis thaliana* genome (30x coverage) with less than 10 GB of memory (Vaser & Šikić, 2021). This indicates there might have been a methodological error in the quantification of maximum RAM used by the program in this chapter.

3.4.6. Implementation

This work carried out in this chapter suggests that the most accurate assembler, judging by the number of misassemblies, is Redbean; despite not being the fastest, it is also a time-efficient assembler. The main problem associated with it is the fact that it does not incorporate a polishing step, which, given what is argued in the previous section, could be the reason behind the high number of missing single-copy orthologs and of genes annotated. Another experiment could be run to show whether posterior polishing would enhance the number of single-copy orthologs detected.

It is entirely conceivable that Redbean could be run on a laptop, as it takes less than 30 GB RAM; it is common for higher-end laptops to have 32 GB installed. However, if results from other comparisons are taken into account (Vaser & Šikić, 2021; R. Wick & Holt, 2021), Raven would be the most memory efficient assembler. Given the estimated RAM usages of under 2 GB (R. Wick & Holt, 2021) for prokaryote genome assembly at 40x-100x depth, it could be carried out on a common laptop with 8 GB. It would be desirable to verify this prior to making any recommendations.

From the results seen in this chapter, the main obstacles to the feasibility of long-read assembly in the field would not be issues attributable to the assemblers themselves. These would lie in the nature of the nanopore reads: their high error rates limit the usefulness of assemblies obtained through this method. For instance, pan-genomic analyses would not make sense when a long-read assembly contains almost twice as many genes as a short-read based assembly, or phylogenetic analyses might be compromised if the genes used are

truncated, or the sequences accumulate mutations that are not present in the actual genome (González-Escalona et al., 2019). High error rate, however, as Koren and colleagues (2019) point out, is likely to improve with time as more accurate basecalling algorithms (for example, Bonito, released in 2019; Seymour et al., 2021) and better polishing software (for example Racon, released in 2017; Vaser et al., 2017) are developed.

It is worth noting, though, that several advantages of nanopore sequencing are already within reach, without relying on improved read quality. For instance, algorithms have been developed (Cao et al., 2016) that can type strains of bacterial pathogens within 30 minutes of commencing sequencing, and predict complete drug-resistance profiles within 10 hours. They report these can be run on a desktop computer, which suggests they can be implemented with relatively lightweight computational requirements.

3.4.7. Conclusion

From the results of this chapter it is clear that genome assembly based on long reads does not require an inordinate amount of computational resources, and it may be plausible to carry out with a personal computer. This further underscores the portability of nanopore sequencing, as it shows that not only the data acquisition is portable, but also an important element of the data processing is too. The results comparing AMR gene detection with genomes assembled through different methods suggest that at least this application of genome assemblies is not hampered by the exclusive use of this type of data. However, it is also apparent that long-read assembly still has much to mature before it becomes a substitute for hybrid, or even short-read assembly. The error-prone nature of long-read data results in incorrect annotations of protein-coding genes, which can make several other applications, such as pan-genomics or phylogenetics, unfeasible. Improvements in sequencing and basecalling accuracy are expected to improve on current error rates, but presently, hybrid assembly appears to be the most reliable means of obtaining complete, contiguous and closed bacterial genomes.

Chapter 4: Comparison of water sampling and DNA extraction methods

4.1. Introduction

Fish disease diagnostics have often relied on visual identification of symptoms or the application of a number of tests (ELISA, PCR, lateral flow assays; see section 1.3) on swabs or tissue homogenates of animals that display symptoms. There is a clear advantage, however, to detect pathogens before infection occurs. Taking water samples from the medium in which the fish grow can represent a feasible means of pre-emptively testing for organisms that infect them, without sampling every individual fish.

The advantages of this approach have been noticed in the past decades. As far back as 1994, O'Brien and colleagues used a combination of water samples from an Atlantic salmon hatchery and PCR to detect the presence of *A. salmonicida* in the rearing tanks, where they were in fact able to anticipate the onset of the disease (O'Brien et al., 1994). Similar techniques have also been applied to the detection of viruses (specific virus detection in aquaculture: Andersen et al., 2010; Otta et al., 1999; metagenomic detection in water samples: Rosario et al., 2009; Rusiñol et al., 2020). The procedures followed to recover viruses from water, however, are usually complex and time consuming (tangential flow filtration, refrigerated ultracentrifugation, prolonged incubations, etc.).

More recently there has been interest in the application of metagenomics to aquaculture (Martínez-Porchas & Vargas-Albores, 2017). Metagenomics is the study of all the DNA found in environmental samples, without enriching or searching for a specific organism. Shotgun metagenomics, frequently used in environmental studies, can serve to characterise microbial communities (Meenakshisundaram et al., 2021; Tepaamorndech et al., 2020), reconstruct genomes of organisms with important ecological functions (Haro-Moreno et al., 2020), and to detect pathogenic organisms (Hamner et al., 2019; Rusiñol et al., 2020), antimicrobial resistance genes or virulence factors (Durso et al., 2011; Higuera-Llantén et al., 2018; Muziasari et al., 2016). These capabilities have been found useful in clinical contexts, where naïve pathogen identification and characterisation can be useful to reduce number of tests needed, inform medical treatment, and simplify workflows (Grant et al., 2018; Greninger et al., 2015; Wilson et al., 2014, 2018).

Metagenomics therefore offers the possibility of studying the microbiomes of fish tanks, shrimp ponds and other culture environments, which can lead to insights into how bacterial diversity patterns can affect the cultured organism's health (He et al., 2020; Lemonnier et al., 2016). On the other, it can serve to directly identify pathogens in the farm waters by scanning the sequenced data against a database (Peters et al., 2018).

Metagenomic approaches have become more attractive with the decreasing price and increased portability brought about by more modern next-generation sequencing methods. The MinION, which was used in Chapter 2 to sequence the genomes of the seven BPOIs, is a hand-held device with a proven record in low-resource settings. Released in 2014 by Oxford Nanopore Technologies (ONT, n.d.-b), it has already been applied in a number of remote settings (Boykin et al., 2019; Edwards et al., 2017; Gowers et al., 2019; Quick et al., 2016), which highlight its portability, improved accessibility to end users and relative ease of use. Boykin and colleagues (2019) specifically apply it in remote farms, by transporting all necessary equipment to the sampling locations in Tanzania, Uganda and Kenya and carrying out the entire process, from sample collection and DNA extraction to sequencing and basic analysis on site. Additionally, entry costs are low for NGS: a MinION starter pack, comprising the sequencer, a flowcell and a library preparation kit currently cost £800 (ONT, n.d.-f). The cost of running the sequencer can be lowered by "multiplexing", the process of sequencing nucleic acids from up to 12 different samples by tagging each one with a different barcode sequence. All of these features contribute to make nanopore sequencing a very appealing option for aquaculture pathogen detection.

Despite these benefits, it also has some drawbacks. As discussed in Chapter 3, Nanopore sequencing is error-prone, with error rates per base being cited as 14% in 2019 (Workman et al., 2019, on the R9.4 flowcells and Albacore 2.1.0 basecalling), although subsequent pore structures, kit chemistries and signal-decoding algorithms have improved accuracy over the last years (ONT, 2020). This, however, can lead to issues in the attribution of metagenomic DNA sequences to the level of species (Heikema et al., 2020; Leidenfrost et al., 2020), which may difficult its application in pathogen identification. The other, which may be more of an obstacle to its implementation in the point of need, is the requirement of large amounts of DNA of high quality. Specifically, standard workflows need a minimum of 400 ng of DNA, $A_{260/280}$ ratios of 1.8 and $A_{260/230}$ ratios of 2.0-2.2 (ONT, n.d.-e). Furthermore, the DNA has to be concentrated in a volume of 7.5 µL, which can be done through certain magnetic bead-based methods (AMPure), but these incur in some loss of the genetic material (Beckman-

Coulter, n.d.). This represents an obstacle in environments where cells and DNA are widely dispersed and accompanied by a number of inhibitors.

The means for concentrating cells from water in ecological and eDNA studies are usually filtration (W.-Y. Chen et al., 2017; Deiner & Altermatt, 2014; Lu et al., 2015; Tang et al., 2016) and centrifugation (Caldwell et al., 2007; Foote et al., 2012; Williams et al., 2017). Filtration is normally carried out with 0.2 µm (W.-Y. Chen et al., 2017; Deiner & Altermatt, 2014) or 0.45 µm pore filters (Lu et al., 2015; Tang et al., 2016), and sometimes a pre-filter is included, although this is more common in eukaryotic eDNA studies (Majaneva et al., 2018; Robson et al., 2016; Takasaki et al., 2021). The purpose of the pre-filter is to remove larger particulate matter that would obstruct the smaller pores of the filter, while letting bacteria through, which are small enough to pass through the larger pores but will then be trapped by the actual filter's smaller pores. This accumulation of bacteria is then followed by a DNA extraction from the filter. With centrifugation, the procedure usually involves direct centrifugation of the sampled water and the removal of the supernatant, followed by DNA extraction from the pellet. Following this, DNA would be extracted from the pellet formed at the bottom of the centrifuged tube. These considerations prompted three questions: first, whether centrifugation or filtration would result in greater concentrations of bacterial DNA obtained from a water sample; second, whether during a filtration procedure that included a pre-filter, the pre-filter would be trapping bacteria that were attached to larger particulate matter (and therefore DNA of potential bacterial pathogens of interest); and third, whether during the DNA extraction procedure as per Griffiths and colleagues' (2000) method, the initial DNA extraction protocol considered, more DNA would be obtained if the filters were rinsed in the extraction fluid and removed, or if they were incorporated into the bead-beating process. It was considered important to answer these questions to find an optimal method of concentrating bacterial cells from water samples, which would in turn be a crucial first-step in the implementation of metagenomics-based pathogen detection in aquaculture.

This last point leads to the aim of this chapter:

4.1.1. Aims and objectives:

Aim:

1) To determine a means of concentrating bacteria from water samples and extracting their DNA that enables nanopore-based metagenomic analyses of aquaculture water.

Objectives:

1) To find out what method of concentrating bacteria from water returns the highest yields.
2) To select a method of extracting DNA from bacterial concentrates from water that returns the highest yields possible among a number of methods selected for their potential application in low-resource settings.
3) To select a method of extracting DNA from bacterial concentrates from water that returns DNA of sufficient quality for metagenomic analyses on the MinION.

4.2. Methods

4.2.1. Saturation tests

The first experiments were saturation tests carried out on the filtration setup which was to be used to process river water. Their purpose was to establish the rate of flow through the filtration system and how it changed with water volume processed, in order to decide on a cut-off volume that would not entail waiting unnecessarily for minimal returns in terms of volume filtered. For these, water was taken from the river Rheidol as it passed through Aberystwyth, coordinates 52.411039, -4.086802 and then pumped through the filtration system set up at the lab. Water was collected by immersing a bucket in the river and then pouring the water into a 25 L carboy through a funnel and a sieve (180 µm pore). All the materials had been previously sterilised by successive rinses with bleach (1,000 ppm Presept™ sodium dichloroisocyanurate, Johnson and Johnson), sterile water (autoclaved at 115 °C for 15 minutes), 5% (volume/volume) HCl and more sterile water, except for the sieve, which was autoclaved at 115 °C for 15 minutes. The bucket and carboy were pre-contaminated with river water before collecting a sample of approximately 20 L. The water samples were immediately transported to the lab, stored at 4 °C, and processed within the next three hours.

The filtration setup collected water from the carboys through 4 mm internal diameter tubing and pumped through two filters serially by a peristaltic pump at 120 rpm. The first filter was a 5.0 µm pore filter and the second was of 0.22 µm pore size; both were made of nitrocellulose, 47 mm in diameter (Whatman/GE Healthcare) and encased in airtight Swinnex-47 housings (Millipore). It is worth noting, however, that these were only sealed after purging the air from the tubing first, as once wet the filters would become impermeable to air, thus impeding the flow through the filtration system. The cumulative volume of water filtered was recorded every 2 minutes 30 seconds; it was measured with a graduated cylinder with 5 mL intervals for a total of 20 minutes. This was repeated three times. A plot of volume filtered against time was built on R 3.6.1 (R Core Team, 2019) using the ggplot2 package (Wickham, 2016).

4.2.2. Concentration method tests

To answer the three questions posed during the Introduction (section 4.1) about the comparison between filtration and centrifugation, between filters and pre-filters and between rinsing filters or including them in the bead-beating process, a single experiment was set up.

In this experiment DNA was extracted using the method mentioned above from concentrates obtained by filtering and centrifuging a river water sample. DNA was extracted from both filters and pre-filters, and both bead-beating the filters and bead beating only the extraction buffer used to rinse the filters. This was carried out with four replicates in each of these sets of conditions. The different means of obtaining/treating the bacterial concentrate and the DNA extraction process are described below.

The river water used was the one collected for the saturation experiment (section 4.2.1) at Aberystwyth South Beach; it was stored at 4 °C and processed within 32 hours of collection. Filtration was carried out with the same filtration system as the saturation experiment, with flow being stopped after 500 mL of water had been filtered; to nullify effects of storage time on the amount of DNA collected with the different treatments, the order in which they were applied to the water sample was randomised. 16 of these filtrations were carried out, in eight cases the filter was removed with sterile forceps and stored in a 2 mL screw-cap tube and in another eight the pre-filter was removed and stored in the same way. Four centrifugations were carried out by adding 50 mL of the river water to a falcon tube, centrifuging at 3,850 RCF in a Hettich Universal 320 centrifuge (Tuttlingen, Germany) for 15 minutes, decanting, and then repeating the process 10 times to process the 500 mL per sample. This concentrate was collected by resuspending the pellet formed in the remaining water (usually between 1 and 2 mL), transferring it to a 2 mL screw-cap tube, centrifuging at 10,000 RCF for 10 minutes (Heraeus Fresco 17 centrifuge; Thermo Fisher Scientific), and decanting the remaining water. Samples were stored at -20 °C until use.

As stated previously, an adaptation of Griffiths and colleagues' (2000) protocol was followed to extract the DNA. Four pre-filters and four filters were rinsed with 500 µL CTAB extraction buffer (10% w/v CTAB [hexadecyltrimethylammonium bromide], 0.7 M NaCl, 240 mM potassium phosphate buffer [pH 8.0]) and then removed from the tubes with sterile forceps. Another four pre-filters and four filters were rinsed as well, but were not removed. Finally, four bacterial concentrates from centrifugation had the same buffer added. These 20 samples then followed the same process; a scoop of glass beads was added (approximately seven beads, 3 mm diameter) to each tube, followed by 500 µL of phenol-chloroform-isoamyl alcohol mixture (25:24:1; NBS Biologicals) added in a fume hood. The tubes were then sealed tightly, vortexed (10 seconds, maximum speed on VWR analog vortex mixer) and shaken in a Mini-BeadBeater-8 (Biospec, Bartlesville, Oklahoma, USA) for 30 seconds at maximum speed.

The lysed bacterial concentrates were then centrifuged (13,000 x *g*, 10 mins, Heraeus Fresco 17 centrifuge), the aqueous phase was removed in the fume hood, carefully avoiding the interphase and placed in 1.5 mL microcentrifuge tubes (manufacturer). An approximately equivalent volume (500 µL) of chloroform-isoamyl alcohol (24:1) was added to the aqueous mixture and mixed gently by pipetting, then vortexed for 10 seconds (maximum speed on VWR analog vortex mixer) and centrifuged again (13,000 x *g*, 10 minutes). The supernatants were removed to another set of 1.5 mL microcentrifuge tubes and 1 mL PEG solution (30% w/v polyethylene glycol 6000, 1.6 M NaCl) was added. After a 2.5 hours precipitation step at room temperature, tubes were centrifuged at 15,000 x *g* at 4 °C (10 minutes) and the liquid was removed by pipetting. 1 mL 70% ethanol was added, tubes were centrifuged again at 15,000 x *g* and 4 °C (10 minutes) and as much ethanol as possible was removed without disturbing the pellet at the bottom; this wash step was then repeated. The tubes were left to dry with their lids open inside a heating block at 37 °C for as long as it took for the ethanol to evaporate, monitoring every 5 minutes. 50 µL TE buffer (1 mM Tris [pH 8.0], 1 mM EDTA) were added and the tubes were raked against the rack to elute the nucleic acids.

The DNA eluate's concentration was then measured with through a Qubit™ dsDNA HS Assay (Invitrogen/Thermo Fisher Scientific) on a Qubit™ 2.0 fluorometer (Invitrogen/Thermo Scientific, Paisley, UK); a NanoDrop™ 1000 spectrophotometer (Thermo Scientific, Paisley, UK) was used as well to measure purity in the form of $A_{260/280}$ ratios. In the case of the Qubit™ assay, 1 µL of the dsDNA HS reagent was added to 199 µL of the dsDNA HS buffer for every sample tested, then 2 µL of the DNA extract were added to 198 µL of the Qubit™ working mix, incubated for 2 minutes in the darkness and quantified with the fluorometer. In the case of the NanoDrop™ 1000, 2 µL of TE buffer (1 mM Tris [pH 8.0], 1mM EDTA) were used to set a blank measurement for the instrument and then 2 µL of each sample were added onto the pedestal to be measured, cleaning it with Optik wipes (Fisherbrand/Fisher Scientific) between samples.

4.2.2.1. Statistical analysis

Statistical analyses were carried out on R v3.6.1 (R Core Team, 2019). A linear model was fitted to the data, with concentration method as the independent variable and yield (DNA concentration of extract) as the dependent. The residuals did not fit the assumption of homoscedasticity that is required for ANOVA, and normality was questionable, so the DNA concentrations were log-transformed. An ANOVA was run on the transformed data and post-hoc Tukey HSD tests were run to compare the different pairs of treatments. The DNA

concentrations obtained after extraction of each of the different concentrates were plotted on R 3.6.1 using the ggplot2 package (Wickham, 2016).

4.2.3. Volume tests

To establish a volume of water that was sufficient to provide an adequate amount for possible downstream uses (Nanopore sequencing or LAMP assays), but offered an adequate tradeoff with time spent processing the sample, an experiment was run to find out how DNA obtained from a sample changed with volume of water processed. For this experiment, water was once again sampled from the river Rheidol as it passed through Aberystwyth, coordinates 52.411039, -4.086802, using the same technique as in section 4.2.1. Approximately 12 L were collected on this occasion; they were transported immediately to the laboratory and stored in a cold room at 4 °C; samples were processed within the next 8 hours. Sterile 50 mL falcon tubes were, as before, filled to the 50 mL mark with the water obtained from the river, centrifuged and decanted; the process was repeated to have tubes where 100, 200, 300, 400 and 500 mL had been processed, with four replicates each. The pellets were resuspended with the remaining water, transferred to 2 mL screw-cap tubes, centrifuged at 10,000 RCF and decanted again. The resulting bacterial concentrates were immediately preserved at -20 °C until use.

DNA was extracted by the same method as described in section 4.2.2, and the concentration of the extracts was measured via Qubit™ 2.0 and NanoDrop™ 1000, with purity measurements being taken with the NanoDrop™ 1000, just as in section 4.2.2 as well.

Linear regression was carried out on the by fitting a linear model to the data; the volume of water processed was considered the independent variable and the DNA concentration in the extracts was the dependent. The analysis was done in R 3.6.1 (R Core Team, 2019) and the data was plotted using the ggplot2 package (Wickham, 2016).

4.2.4. DNA extraction method tests

4.2.4.1. Sample collection

In this project the focus was on simplicity and applicability in the field; to this end a number of DNA extraction protocols were chosen for their purported simplicity, low cost or effectiveness. The next section describes briefly the main characteristics of each protocol and the adaptations that were applied to use them on the centrifuged bacterial concentrates; the full protocol instructions are included in Appendix 3. The advantages and disadvantages of each of them are detailed in Table 4-1. For these experiments, bacterial concentrates were

obtained by centrifuging 300 mL of river water collected from the same location as in the previous experiments (coordinates 52.411039, -4.086802) by the same method (see section 4.2.2, concentration method tests). Approximately 20 L were collected on this occasion, and processed over the next 36 hours, keeping the sampled river water stored at 4 °C. A total of 19.2 L were processed, yielding 64 concentrates. The centrifuged bacterial concentrates were immediately preserved at -20 °C until use.

4.2.4.2. DNA extraction protocol descriptions

Boom and colleagues' (1990) protocol involves using a sand ("silica coarse") suspension to lyse cells, bind their DNA and purify it with successive washes with ethanol and acetone.

Bordelon and colleagues' (2013) protocol is aimed at purifying DNA from samples with high amounts of inhibitors, specifically urine. The protocol uses silica-coated magnetic beads to collect DNA from a liquid sample and then allows it to be carried through several wash buffers before eluting the clean DNA in a storage buffer (TE, water). The protocol was adapted with details from posterior studies (Bitting et al., 2016), as the original was not designed specifically for the lysis of bacterial cell walls, and modified to accommodate available resources.

The protocol by Casas and colleagues (1995) involves the lysis of cells with guanidinium thiocyanate and sarkosyl, followed by washes with different alcohols.

Griffiths' protocol (Griffiths et al., 2000) is described in detail in section 4.2.2. It consists in a bead-beating lysis step followed by a phenol-chloroform extraction and ethanol precipitation.

The protocol by Martin-Platero and colleagues (2007) breaks down cell walls with lysozyme, then uses the principle of salting out to remove protein contaminants from the cell lysate and a wash with ethanol to clean the DNA of other contaminants.

The protocol designed by Orsini and Romano-Spica (2001) uses microwave lysis and polyvinylpyrrolidone to break down cells from environmental samples, and then purifies the DNA with a phenol-chloroform extraction.

Pitcher and colleagues' (1989) protocol consisted in an enzymatic lysis of cells, salting out of proteins and a chloroform based purification step.

Powersoil (see Qiagen, 2017 for manual) is a kit commonly used to extract DNA from soil. It uses a mild bead-beating step to lyse cells and then requires washing with several proprietary solutions to purify the DNA. Finally, it is trapped in a silica matrix and washed with ethanol.

Table 4-1: List of original references of the different DNA extraction protocols used during this chapter and their advantages and disadvantages.

Protocol	Advantages	Disadvantages
Boom et al., 1990	Uses relatively inexpensive reagents (ethanol, acetone, guanidinium thiocyanate, sand) which will normally keep for weeks at room temperature, and once buffers etc. are prepared the implementation is simple and moderately fast (~1 h)	Did not work with Gram+ or yeasts in the original report.
Bordelon et al., 2013	Simple, field-forward - reagents can be lyophilised and does not need pipetting. Meant to be very fast.	Potentially expensive as it uses commercial magnetic beads. Adapted version needs lysozyme (cold-chain).
Casas et al., 1995	Co-extracts DNA and RNA, so could be useful for the detection of viruses. It is also simple and fast.	Requires refrigerated centrifuge and freezer for the cold isopropanol. Not intended for bacteria originally. Adapted version needs lysozyme (cold-chain).
Griffiths et al., 2000	Co-extracts DNA and RNA, so could be useful for the detection of viruses. Validated in the EX^2EL lab; works well for inhibitor removal.	It is time consuming, uses organic solvents that require fume hood and adequate disposal. Bead-beating shears DNA.
Martín-Platero et al., 2007	Simple to carry out, reagents are not hazardous	Not intended for complex environmental samples.

Orsini & Romano-Spica, 2001	Simplifies lysis by using a common kitchen appliance (microwave).	Uses organic solvents that require fume hood and adequate disposal.
Pitcher et al., 1989	Relatively simple and validated with a variety of bacteria	Uses organic solvents that require fume hood and adequate disposal. Needs lysozyme (cold-chain).
PowerSoil (Qiagen, 2017)	Standardised, pre-packaged, requires only basic lab equipment (pipettes and centrifuge) and is straightforward. Demonstrated effectiveness in DNA extraction from wastewater and river water (Kaevska & Slana, 2015; Walden et al., 2017)	Expensive and uses large amounts of disposable plasticware.

DNA extracted by each of these methods was quantified via Qubit™ 2.0, as before (section 4.2.2, concentration method tests). Additionally, the DNA extracts were visualised by gel electrophoresis and transillumination. Two 1% w/v agarose gels were made with 75 mL 0.5X Tris-Borate-EDTA buffer (TBE, 45 mM Tris-borate, 10 mM EDTA) and 3 µL 10 000X SYBR safe each, then were loaded with a mix of 5 µL of each sample's DNA extract plus 1 µL 6X blue loading dye (0.25% w/v bromophenol blue FF, 40% w/v sucrose, 0.25% w/v xylene cyanol). 2 µL of 1kb ladder (Quick-load® Purple 1 kbp ladder, NEB) were also loaded for each row of wells. Electrophoresis was run at 60 mV for 70 minutes and then visualised in a GelDoc XR+ documentation system (Bio-Rad Laboratories Inc. Watford, UK).

An attempt was made to identify significant differences between the DNA extract concentrations between the different extraction protocols by using an ANOVA test in R v3.6.1 (R Core Team, 2019). However, when fitting a linear model to the data, the residuals did not meet the assumptions of normality or homoscedasticity, nor did they when the data was log-transformed, so an alternative non-parametric test, the Kruskal-Wallis test, was used to compare ranks. Post-hoc comparisons between the ranks of group pairs were done via a corrected Mann-Whitney U-test.

4.2.5. 16S ribosomal RNA gene PCR

To ascertain whether inhibitors of DNA amplification present in the environmental samples had been removed by the DNA extraction methods described above, a 16S ribosomal RNA (rRNA) gene PCR was run on the extracts. 16S rRNA gene PCR is a technique often used in bacterial diversity studies; it was applied here because the 16S rRNA gene is commonly found across bacteria, and would certainly be found among the DNA extracted from the bacterial concentrates from river water. Specifically, the primers used (27F and 1389R, see below) have been previously used in other studies of bacterial communities (Edwards et al., 2014; Lowe et al., 2011; Rassner et al., 2016).

16S rRNA gene PCR reaction mix was prepared with a commercial master mix and primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3′) and 1389R (5'-ACGGGCGGTGTGTACAAG-3')(Osborn et al., 2000; Weisburg et al., 1991); the components used and their exact amounts are given in Table 4-2. The reaction was run with each of the four DNA extracts from each of the eight protocols tested (non-normalised concentration, 2 µL), a positive control (high-DNA extract from *A. salmonicida salmonicida*, see section 3.2.2) and a negative control (sterile 1X TE buffer). Enough reaction mix was prepared for all the samples plus approximately 10% extra, to avoid running out of reaction mix due to pipetting error. The following program was run on a G-Storm GS0001 thermocycler: an initial warm start step (5 m, 95 °C), 30 repetitions of denaturation (35 s, 94 °C), annealing (35 s, 55 °C) and elongation (1 m, 72 °C), and one final elongation step of 5 m at 72 °C.

The amplified products were then visualised by gel electrophoresis. The 1% w/v agarose gel was made with 75 mL 0.5X TBE buffer (45 mM Tri-borate, 10 mM EDTA) and 3 µL 10,000X SYBR safe, then were loaded with a mix of 5 µL of each amplified DNA sample plus 1 µL 6X blue loading dye (0.25% bromophenol blue FF, 40% w/v sucrose, 0.25% w/v xylene cyanol). 2 µL of 100 bp ladder (Quick-load® 100 bp DNA ladder, NEB) were also loaded for each row of wells. Electrophoresis was run at 60 mV for 70 minutes and then visualised in a GelDoc XR+ transilluminator/documenter.

Table 4-2: Reagents and quantities used in the PCR reaction mixes of the 16S rRNA gene carried out for the amplification of bacterial DNA in river water concentrates.

Ingredient	Volume per sample (µL)
Thermo Fisher green hot start PCR MM	10
Primer 27F	0.08
Primer 1389R	0.08
DNA	2
dH_2O	7.84
Total	20

4.2.6. Seawater experiments

Because some farmed salmonids are anadromous and spend much of their lives at sea, where they also suffer from bacterial infections, it was considered that testing the same methods as described in the above sections would provide information about the scope of applicability of this chapter's findings. However, the experiments carried out on seawater were the first done during the entire project, and were not appropriately planned for data analysis or for comparison with further experiments. These were mostly the same as the experiments described in the above sections (4.2.1 to 4.2.5), but with variations which are described in the following paragraphs. The data collected is also presented in the results section, but it is worth noting that no statistical analyses were carried out and that the results are not always comparable to the ones of the river water experiments.

4.2.6.1. Saturation tests

Seawater was collected at Aberystwyth South Beach (coordinates 52.411780, -4.09064) by direct immersion of a 25 L carboy sterilised as described in section 4.2.1. Six saturation tests were run, filtering the water with the same setup as in section 4.2.1 with the difference that the filters had a pore size of 0.45 µm; this was done within 8 hours of sample collection.

4.2.6.2. Concentration method comparisons

The same filtration setup as for the seawater saturation tests was used for the concentration method comparisons; so was the sample. Water volume processed was 500 mL, and concentrates were frozen at -20 °C. The comparisons were as for river water; rinsed pre-filter, rinsed filter, bead-beaten pre-filter and bead-beaten filter and centrifuged water concentrates

were compared. However, in these experiments both the filter and pre-filter used to filter the same water sample were used in the comparison, making the data from the filter and pre-filter groups dependent. Measurements of nucleic acid concentration and purity were taken with a NanoDrop™ 1000 spectrophotometer as described in section 4.2.2.

4.2.6.3. Water volume experiments

Another seawater sample was collected from Aberystwyth North Beach, this time by using the method described in section 4.2.1; it was possible to do this from the Aberystwyth jetty (coordinates were: 52.416778, -4.086194).Water was processed as described in section 4.2.3 and samples were refrigerated until use. DNA extraction was also carried out as described in section 4.2.3 (river water volume tests) and extract DNA concentration and purity were measured by Qubit™ fluorometer and NanoDrop™ 1000 in the same way.

4.2.6.4. DNA extraction method comparison

These experiments were carried out in the same way as the ones described in section 4.2.4, with seawater samples collected at Aberystwyth North Beach (coordinates: 52.416778, -4.086194) in the same way as described in section 4.2.1. The water collected was processed within the next 80 hours, stored meanwhile at 4 °C. The processed samples were then refrigerated at 4 °C. DNA extraction was carried out in the same way as in section 4.2.4, and so was the measurement of DNA extract concentration. DNA extract purity was also measured on the NanoDrop™ as described for 4.2.2. A 16S rRNA gene PCR was also carried out in the same way, including visualising the result by gel electrophoresis.

4.3. Results

4.3.1. Saturation tests

The saturation tests indicated that it was possible to filter volumes of over 700 mL without filter saturation within 10 minutes (Figure 4-1). This was consistent through the three samples of river water collected on this occasion, and informed the decision to use a volume of 500 mL during the concentration method comparisons (sections 4.2.2 and 4.3.2).

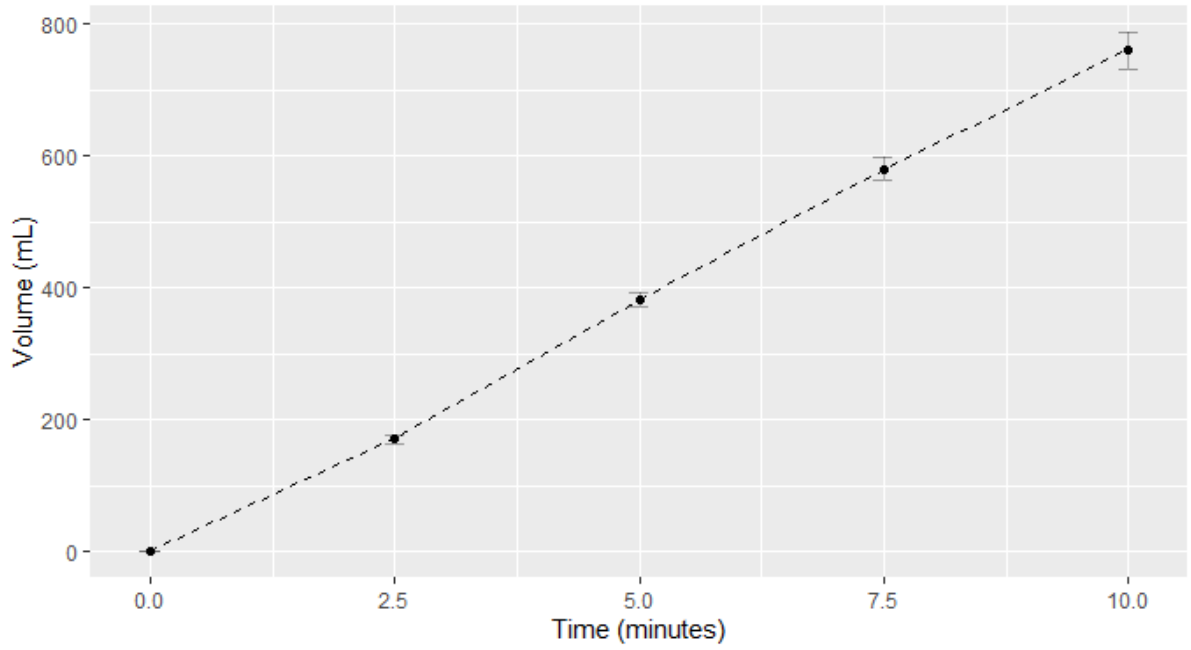


Figure 4-1: Mean volume of river water filtered over time through the filtration system (5.0 µm pore pre-filter, 0.22 µm pore filter). Bars represent one standard deviation around the mean.

4.3.2. Concentration methods comparison

Using different means of concentrating bacteria from river water resulted in very different yields after extracting DNA, as can be observed in Figure 4-2. The results of the ANOVA test on the log-transformed yield data serve to reject the null hypothesis that there is no significant difference in log-transformed yield between the different concentration methods at the 0.05 level of significance ($F_{(4,15)} = 8.31$, $p < 0.001$). The highest were obtained by centrifuging the river water, which gave an average (mean) of 11.5 ng μL^{-1} of DNA; the lowest came from bead-beading the filter in the extraction buffer, which gave 0.454 ng μL^{-1}. However, the Tukey HSD test used for post-hoc comparison between pairs of treatments did not detect any significant differences between the transformed yields from any pair of filter-based concentration methods. They did detect significant differences between centrifugation and all other treatments - bead-beating the filter ($p = 0.001$), rinsing the filter ($p = 0.004$), bead-beating the pre-filter ($p = 0.028$) and rinsing the pre-filter ($p = 0.002$).

An important consideration, given the filtration setup (two consecutive filters of decreasing pore size) is whether the sum of the DNA obtained from both filters is actually equivalent to that obtained from centrifugation. The sum of the mean DNA concentrations obtained from the rinsed filter and bead-beaten pre-filter, which is the combination of pre-filter and filter treatments that returns the highest yield, is 3.05 ng μL^{-1}, which is still substantially lower than the DNA obtained from centrifugation alone.

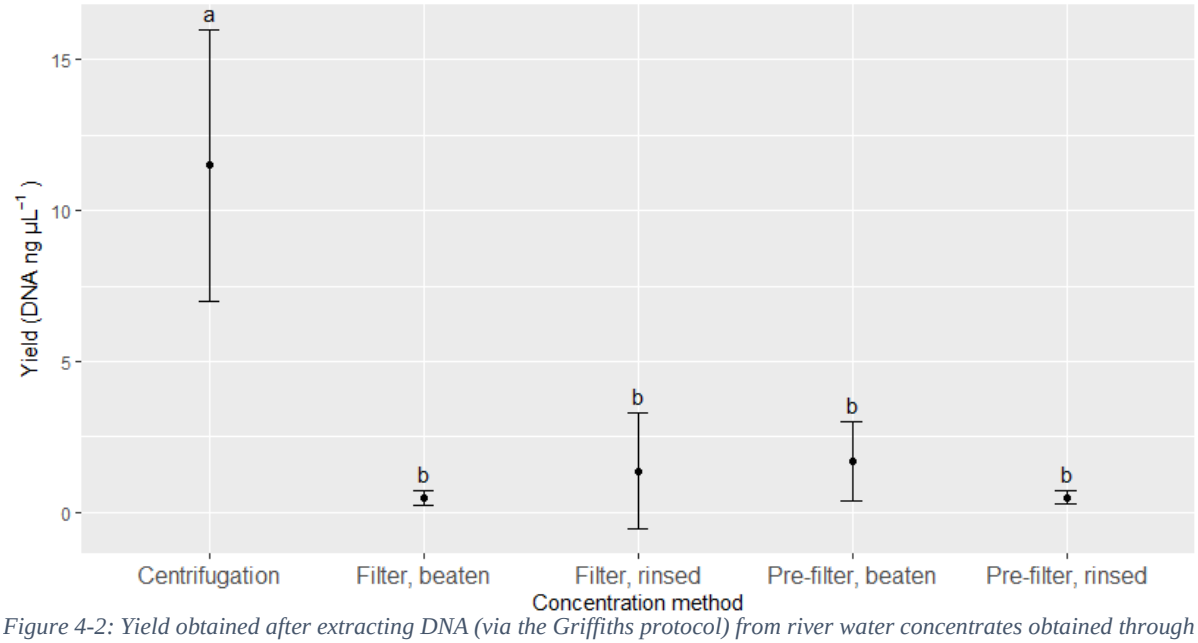


Figure 4-2: Yield obtained after extracting DNA (via the Griffiths protocol) from river water concentrates obtained through different methods. Points represent group means, bar size represents one standard deviation either side of the mean, groups marked with a different letter are significantly different in the pairwise comparisons.

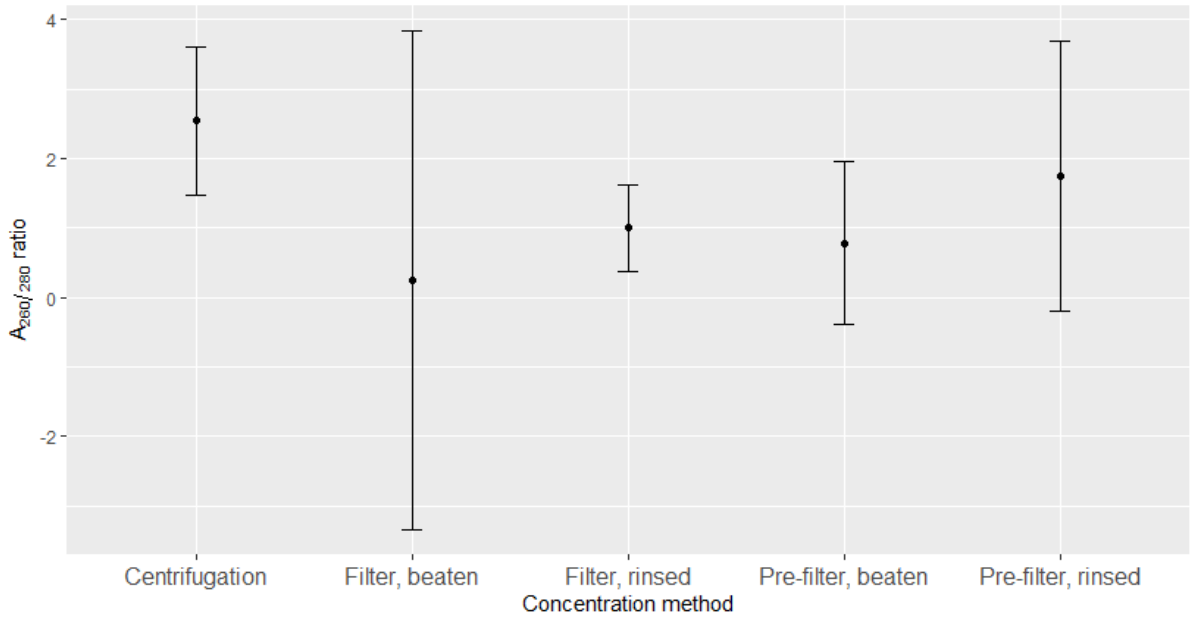


Figure 4-3: $A_{260/280}$ ratios obtained after extracting DNA (via the Griffiths protocol) from river water concentrates obtained through different means. Points represent mean, bars represent one standard deviation either side of the mean. Three extreme data points are omitted; two for the pre-filter bead-beaten samples (-40.5 and -0.527) and one for the filter, bead-beaten samples (-4.70).

Overall, it was possible to say that centrifugation concentrates the highest amounts of DNA-containing organic material from river water, and is probably the most appropriate approach to concentrating bacteria from water. An additional pilot experiment was carried out with a 15 mL falcon tube each of liquid cultures of *Yersinia ruckeri* and *Tenacibaculum maritimum* to ascertain these centrifugation forces and times served to pellet suspended bacteria. It was possible to produce a pellet and clarify the broth in both cases, and to decant the tubes without dislodgement of the pellet. These results suggest that centrifugation is an appropriate means of removing bacteria from river water and that it is more efficient, in terms of return per unit volume, than filtration, regardless of use of the filter or pre-filter and beating or rinsing of the filter.

It is worth noting that the data had to be transformed to compensate for the irregular variances across the groups. There seems to be a tendency for higher yields to be accompanied by higher standard deviations; besides giving the highest yield, centrifuging the river water also gives the highest standard deviation (4.50 ng μL^{-1}).

One other consideration is the purity of DNA obtained through the different extraction methods. Although it would have been interesting to know whether there were any differences in the abundance of contaminants depending on the bacterial concentration

method, the measurements of $A_{260/280}$ ratios obtained from these samples were not useful (see Figure 4-3). This is due to the inaccuracy of $A_{260/280}$ ratios when measuring purity of DNA extracts at low concentration; a Thermo Scientific technical bulletin for the NanoDrop™ 1000 (Thermo Scientific, 2012) points out that these are unreliable at concentrations of less than 10 ng μL^{-1}.

4.3.3. Volume-yield tests

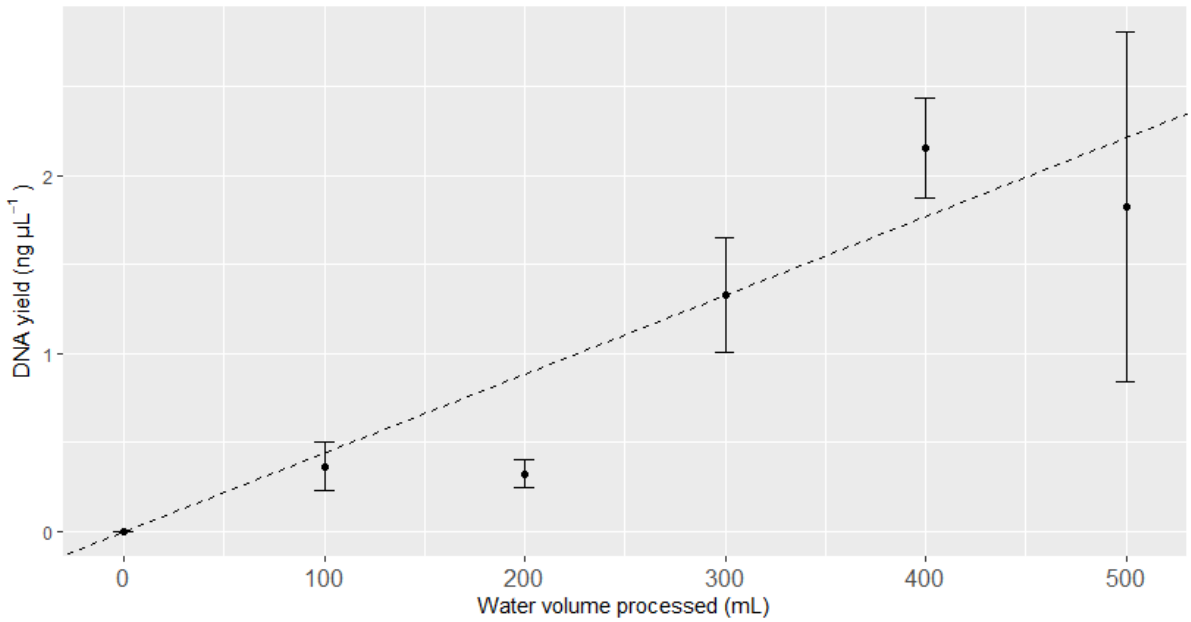


*Figure 4-4: Yield after extraction of incremental volumes of river water concentrates. Points represent mean, bars represent one standard deviation either side of the mean, and the dashed line shows the result of the linear regression (yield = 4.43 * 10^{-3} * volume).*

As a result of the tests comparing yield depending on different volumes of river water processed, a strong positive correlation was found between the DNA yield dependent on water volume processed, as expected (Pearson's r = 0.825). The regression analysis was significant at the p = 0.05 level of significance ($F_{(1,22)} = 46.73$, $p < 0.001$) with an R^2 of 0.680. It found an increase in 0.00443 ng μL^{-1} of DNA in the final extract for every extra mL of water that was included in the river water concentration, with an intercept not significantly different from zero. This coincides with the concentration obtained from negative controls. This regression curve can be seen plotted alongside the results in Figure 4-4. DNA concentrations for negative controls were zero, suggesting the method of DNA extraction did not contribute any DNA to the final extract.

4.3.4. DNA extraction method comparisons

The DNA extraction comparisons resulted in a variety of yields and variances of yields (Figure 4-5). The extraction protocol that provided the lowest yield was the Bordelon adapted protocol, which did not extract any DNA, and the one that returned the highest yield was Martin-Platero's, which gave 0.972 ng μL^{-1}. It is worth noting that five protocols (Boom, Casas, Martin-Platero, Orsini and PowerSoil) performed better than the benchmark, Griffiths' protocol, which gave 0.043 ng μL^{-1}. These fluorometry results coincide with the visual interpretation of the electrophoresis gel, where it is possible to see the extracts carried out with the Martin-Platero protocol more clearly than the rest (Figure 4-6). Despite the low amounts of DNA extracted it is also possible to see a very faint fluorescent signal coming from the DNA obtained with the PowerSoil and Orsini protocols; the next one with the highest yield, Boom, does not result in visible fluorescence. Finally, Bordelon's protocol appears to yield no DNA whatsoever.

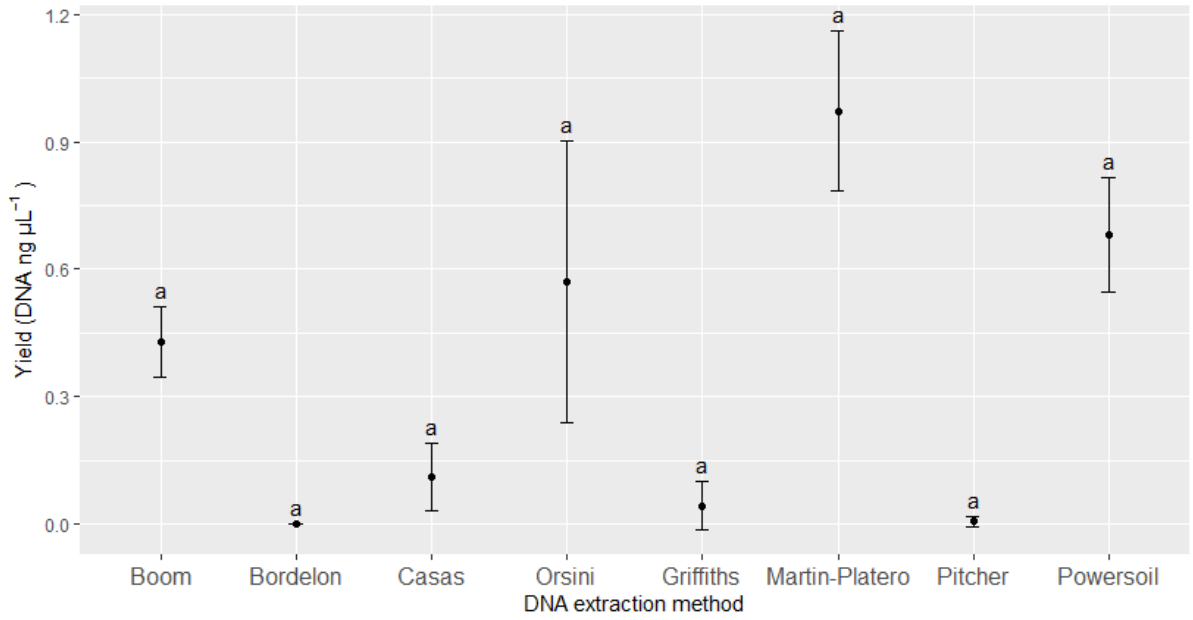


Figure 4-5: Yield (DNA ng μL^{-1}) obtained after extracting a concentrate of 300 mL river water with eight different DNA extraction protocols. Points represent group means, bar size represents one standard deviation either side of the mean and, groups marked with a different letter are significantly different in the pairwise comparisons.

A Kruskal-Wallis nonparametric test was used to evaluate whether there was a significant difference in the mean ranks of the DNA extraction methods; it was not possible to implement an ANOVA because of the heteroscedasticity of the residuals when a linear model was fitted to the data. The result of the Kruskal-Wallis test indicates there is a significant

difference in the mean ranks of the DNA yields obtained through different protocols (χ^2 = 26.818, $p < 0.001$, df = 7). Post-hoc tests, in this case Mann-Whitney U tests did not reveal any significant differences in the ranks of any pair of groups when implemented with a correction (Holm) to account for multiple comparisons.

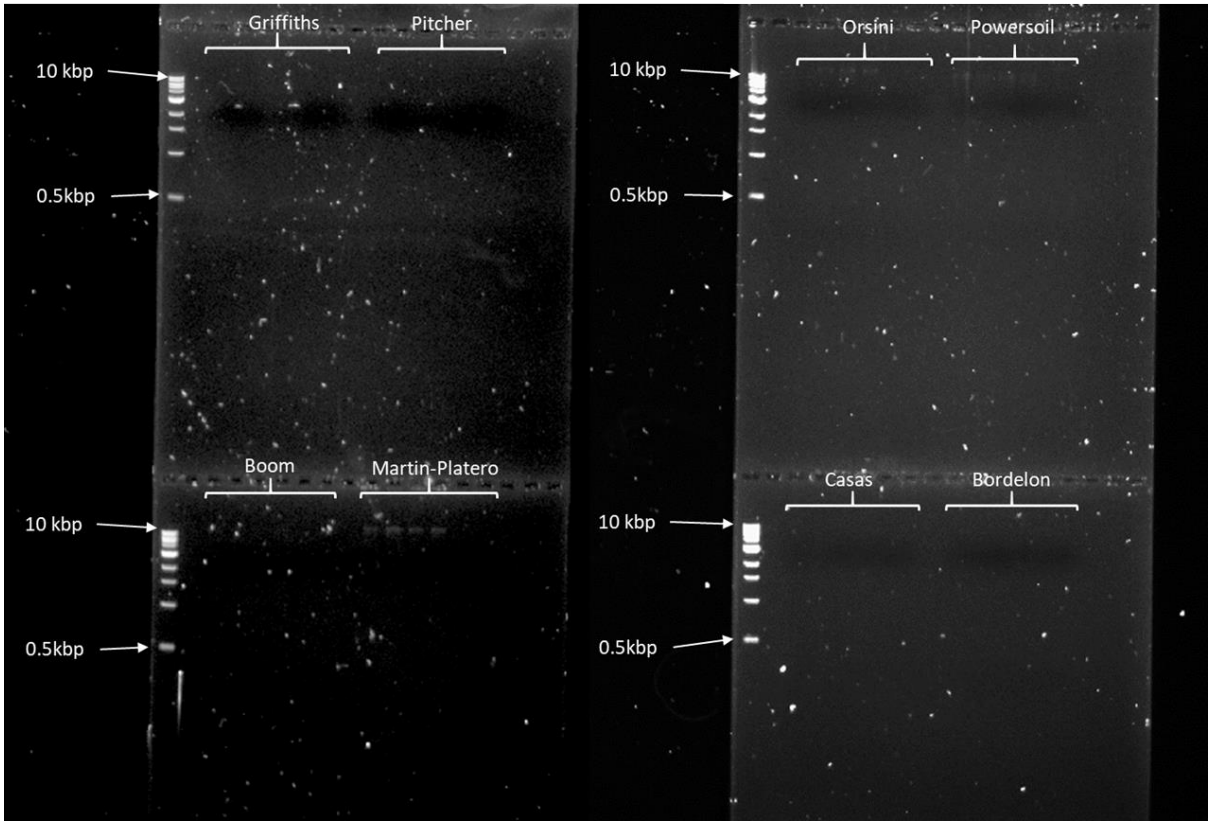


Figure 4-6: Gel electrophoresis of the DNA obtained through the different extraction protocols tested. The name of the protocol and a curly bracket mark the lanes where the extracts were placed on the gel. For each protocol there are six lanes; the two on the right hold the negative controls for the DNA extraction (no river water concentrate).

4.3.5. 16S rRNA gene PCR

The DNA amplified with all of the different extraction protocols tested was visualised in a single gel electrophoresis, see Figure 4-7. Overall, fluorescence signal of the bands corresponding to the 16S rRNA gene amplicons was greatest for the Griffiths protocol, followed by Powersoil, Casas (in three out of four replicates) and Pitcher. Boom, Martin-Platero and Orsini protocols do not result in any visible amplification after PCR, whereas the outcome is mixed for the Bordelon protocols extracts. In the latter case, one of the replicates has strong fluorescence on the gel, two of them faint, and one none. Note that though in some cases the amount of DNA a protocol extracts apparently matches the strength of the fluorescent signal obtained on the gel after PCR, this is not always the case. For example, Griffiths samples give the strongest fluorescence, but the amount of DNA this protocol extracted was relatively small. The situation is similar with another of the protocols that use

chloroform, Pitcher, although the fluorescent signal is not as strong, and with Casas, where all the samples that contained DNA in amplification signal. In the case of the Bordelon protocol it is surprising to see that, despite no DNA being detected through fluorometry, it is possible to observe some fluorescence in the gel electrophoresis.

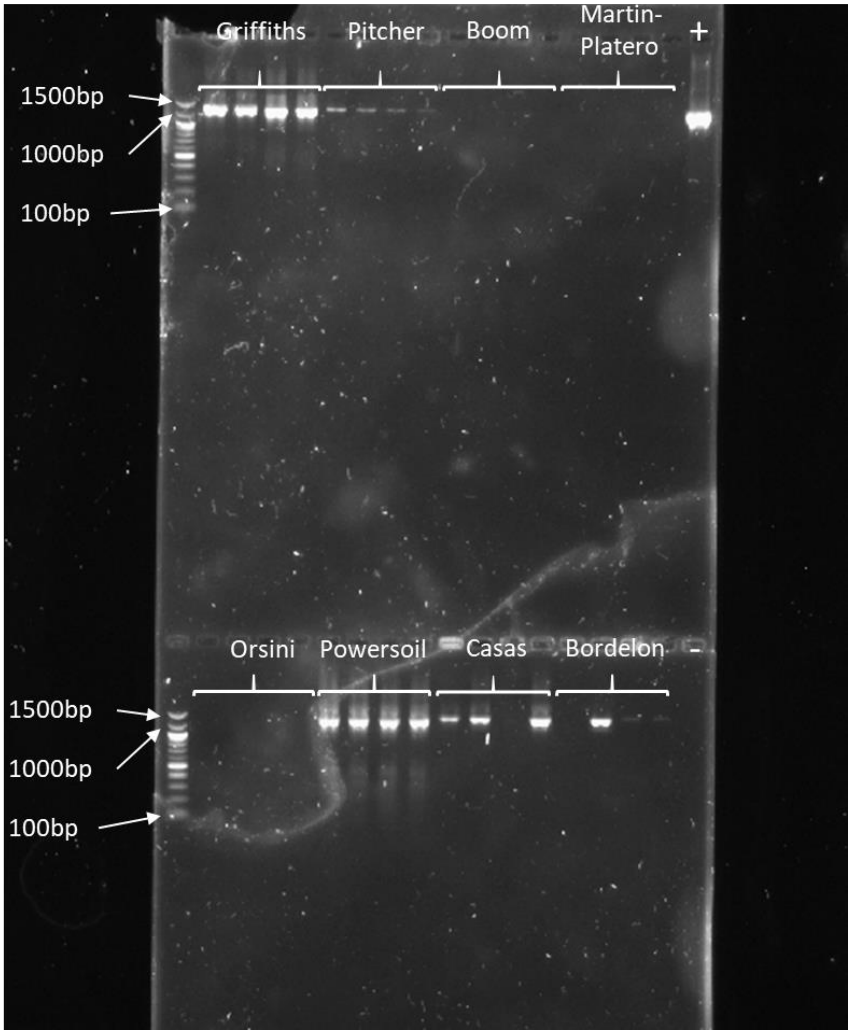


Figure 4-7: Gel electrophoresis of the results of running a 16S rRNA gene PCR on the DNA extracted by different methods from bacterial concentrates from river water.

4.3.6. Seawater results

As stated in the methods section (4.2.6), tests were also run on seawater, although with less planning; the results are summarised in the following sections.

4.3.6.1. Saturation tests

The results of the saturation tests are summarised in Figure 4-8; although after 20 minutes of filtration and an average of 983 mL the filters were approaching saturation, it was possible to consistently process 780 mL or more. This informed the choice of volume (500 mL) to be processed during the concentration method comparisons.

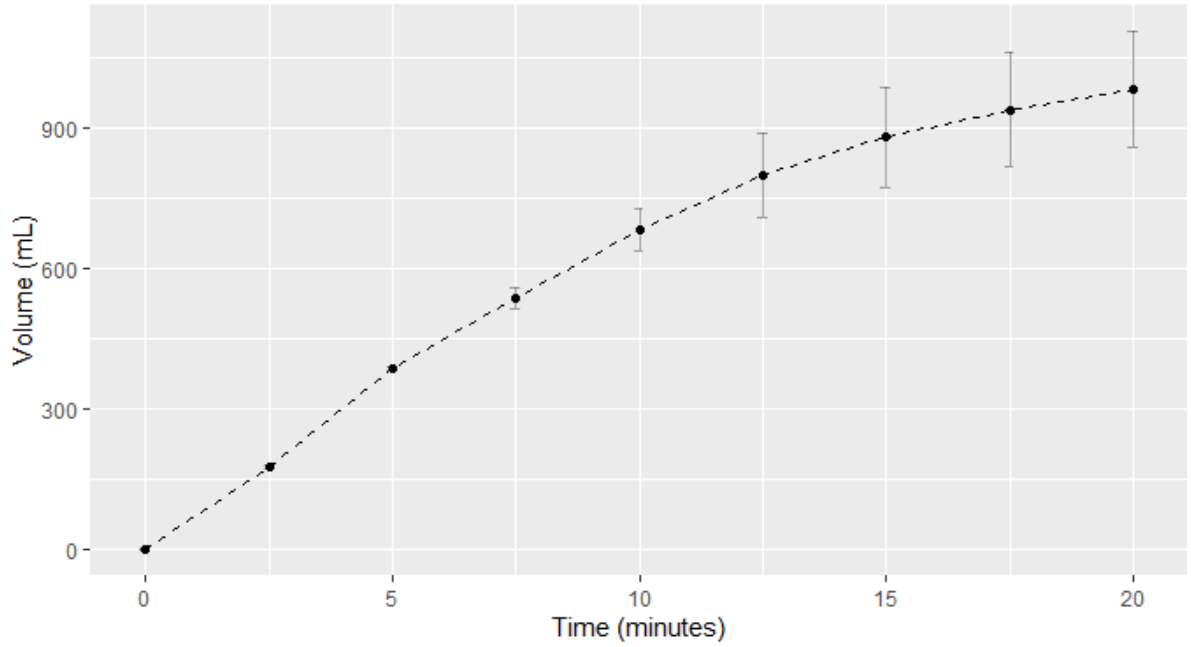


Figure 4-8: Mean volume of seawater filtered over time through the filtration system (5.0 µm pore pre-filter, 0.45 µm pore filter). Bars represent one standard deviation either side of the mean.

4.3.6.2. Concentration method comparisons

Yields from seawater were, as with river water, greater when concentrating bacteria by centrifugation, see Figure 4-9 for the means and standard deviations. The amount of DNA obtained from any combination of filter and pre-filter is still inferior to the average amount recovered from centrifugation.

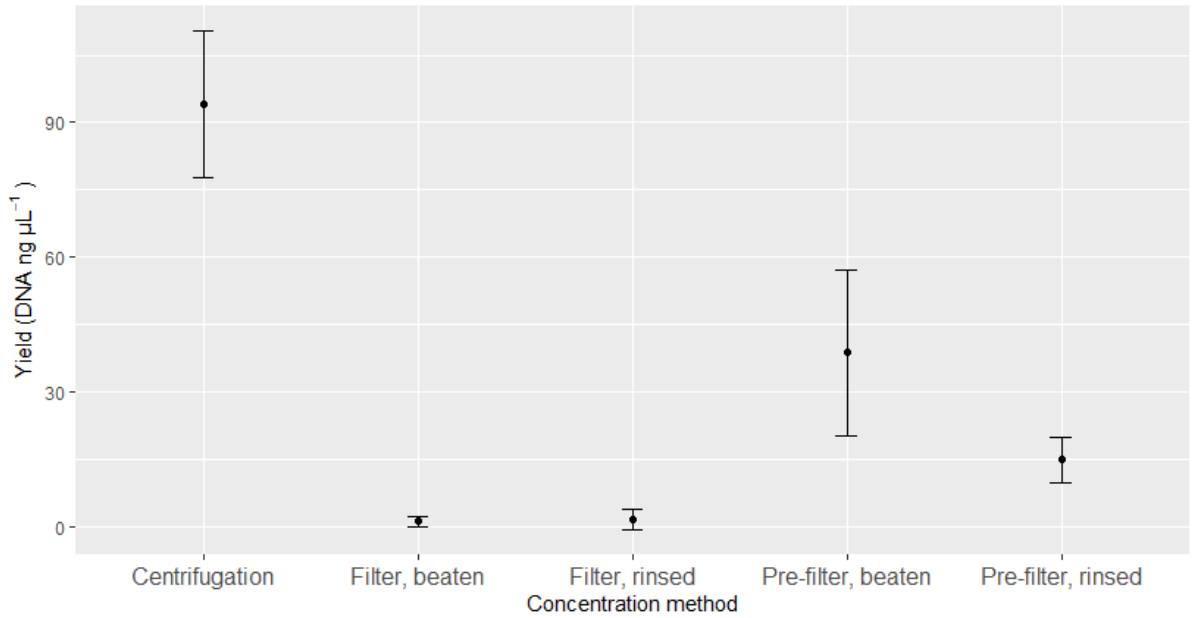


Figure 4-9: Yield obtained from extracting DNA with the Griffiths protocol from bacteria concentrated from seawater in each of the five different ways tested. Points represent group means, bar size represents one standard deviation either side of the mean.

4.3.6.3. Water volume experiments

The trend observed with volume-yield tests was an increase in DNA obtained as the volume of water processed increased. However, just as with river water, it can be seen that the trend is disturbed by discontinuities in the rate of increase; the increase in DNA yield when going from 300 mL to 400 mL water is very small, and it is surprisingly followed by a comparatively large increase when passing from 400 mL to 500 mL (Figure 4-10). Figures obtained from the NanoDrop™ 1000 readings follow a more predictable pattern, but they are unreliable at low concentrations, in fact giving non-zero estimates for the negative controls. Purity results are also variable, but the highest $A_{260/280}$ ratio was obtained when processing 300 mL seawater; it was the most consistent too. This was, in part, what motivated the selection of this volume in the subsequent DNA extraction method comparisons.

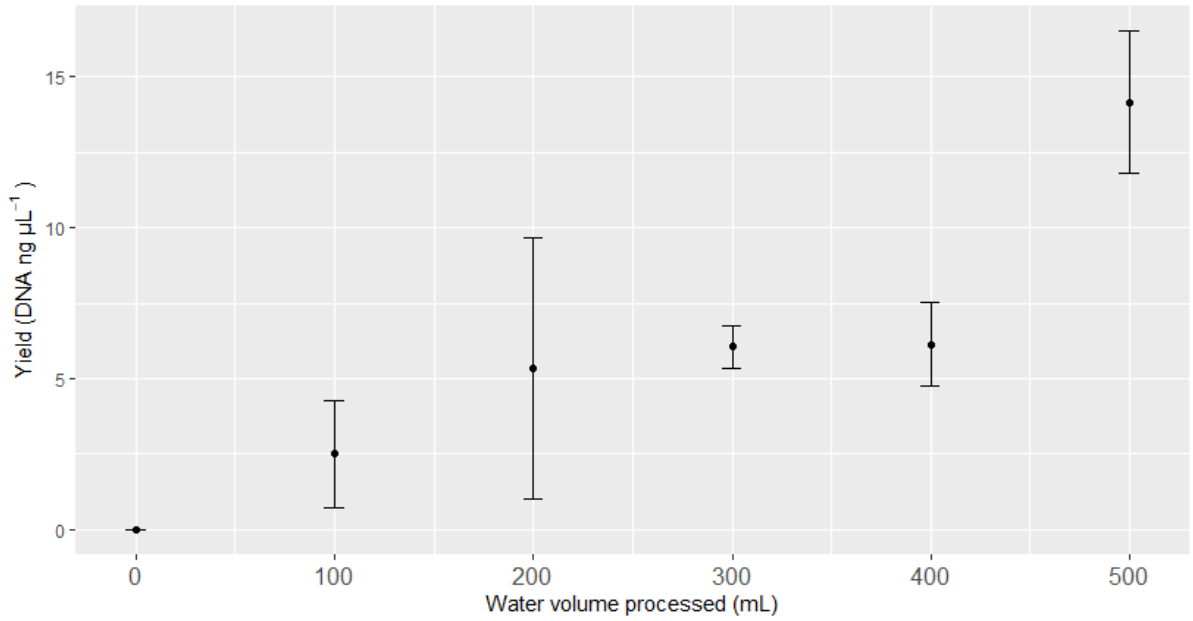


Figure 4-10: Yield obtained from extracting DNA from bacterial concentrates with each of the eight different protocols tested. Points represent group means, bar size represents one standard deviation either side of the mean.

4.3.6.4. DNA extraction method comparisons

Results from the DNA extraction protocol comparisons did not exactly coincide with those run on river water, but there were similar patterns; see Figure 4-11. Mean yield from the Martin-Platero protocol was the highest (9.86 ng μL^{-1}), followed closely by that of Boom. PowerSoil yields were also relatively high, but on this occasion were surpassed by those of the modified Bordelon protocol, which did provide measurable amounts of DNA. The Griffiths protocol, which serves as a benchmark, follows the above, returning modest yields (1.55 ng μL^{-1}) much as it did with river water. The Orsini protocol performs worse than the benchmark on this occasion. Finally come the Pitcher and Casas protocols, with the latter returning no DNA whatsoever.

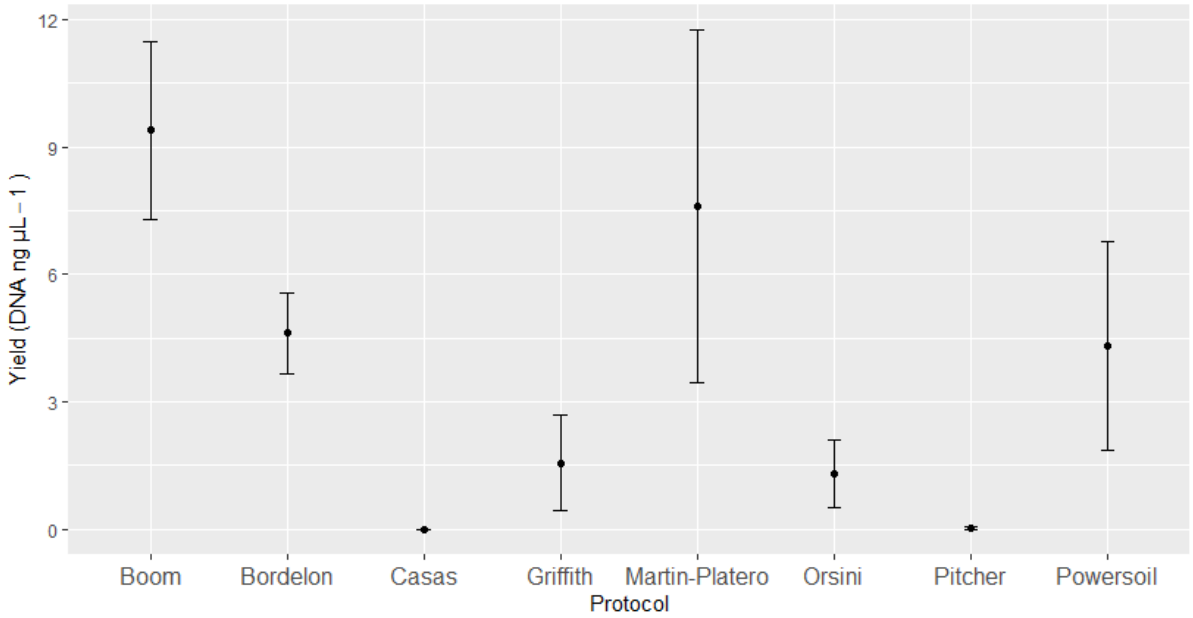


Figure 4-11: Yield obtained from extracting DNA from bacterial concentrates with each of the eight different protocols tested. Points represent group means, bar size represents one standard deviation either side of the mean.

4.3.6.5. 16S rRNA gene PCR

In the follow-up 16S rRNA gene PCR it is possible to see bands corresponding to DNA amplified from the Griffiths, PowerSoil and Bordelon protocols (Figure 4-12). In the case of the former two, this coincides with the results for river water; in the case of the latter the amplification seems to be marginally better than in river water, but also irregular across samples. No noticeable amplification is visible for the extracts obtained through the Boom, Martin-Platero, Orsini and Pitcher protocols. As with the river water DNA extracts, some of these amplification signals are low despite extract DNA concentrations being high; this is the case for the Boom and Martin-Platero protocols. For the Casas protocol, extremely faint bands can be observed, despite no DNA being detected by Qubit fluorometry in the extract used for input to the PCR.

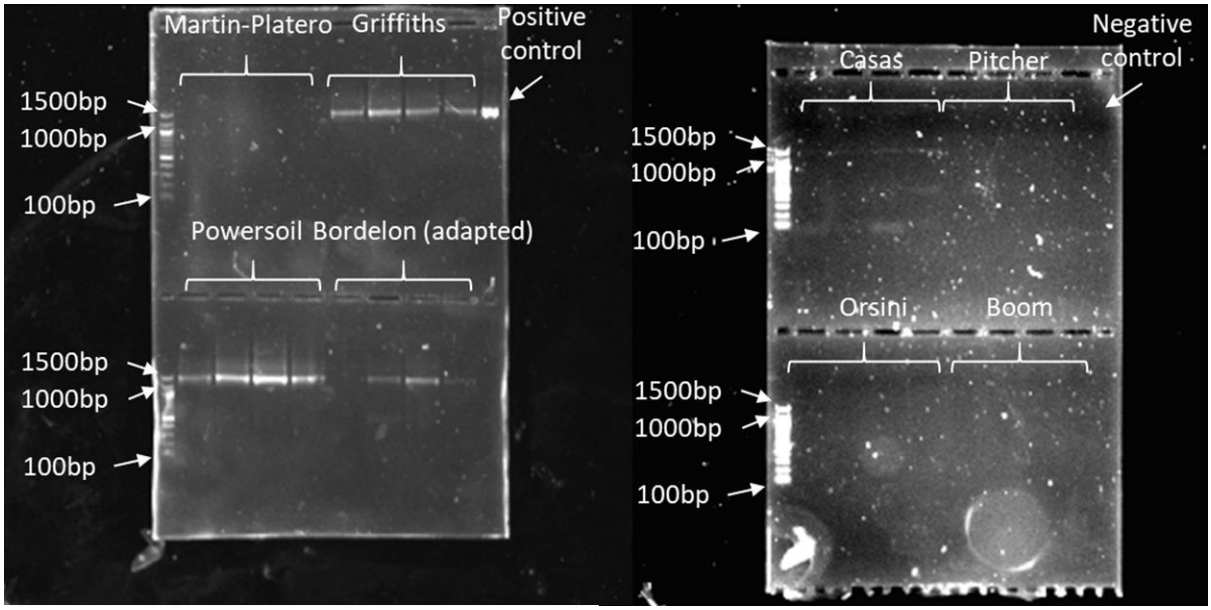


Figure 4-12: Gel electrophoresis of the results of running a 16S rRNA gene PCR on the DNA extracted by different methods from bacterial concentrates from seawater.

4.4. Discussion

4.4.1. Overview

In this chapter the effects of different means of processing sample water to collect bacteria and extract DNA from the concentrates have been assessed. The main focus has been on river water, but the methods have been applied, with some variations, to seawater as well with similar results. Overall it appears that centrifuging water, as opposed to filtering it with the common 0.2 µm pore filters, yields more DNA per unit of water volume processed (objective 1.1). The extracts resulting from this centrifugation vary in the amount of DNA recovered, but have always been above the results for filtration, even when adding the combination of DNA obtained from filter and pre-filter. The means of processing the pellets obtained through centrifugation gave similar but different results for river and seawater; the Martin-Platero protocol performed best when applied to river water and the Boom one when applied to seawater, with the former close behind (objective 1.2). However, this was only in terms of yield, and 16S rRNA gene PCR results suggest that the either the DNA extracted by these means was not bacterial, or it still contained abundant inhibitors of PCR (objective 1.3 unmet). Ultimately, the most reliable method of extracting DNA was the PowerSoil kit, which consistently recovered more nucleic acids than the benchmark method (Griffiths).

4.4.2. Implications of this chapter's findings

One of the main findings of this chapter is that the use of centrifugation is an advantageous method of concentrating bacteria from river water, as it results in higher yields of DNA after extracting the concentrates. This finding is especially interesting given that metagenomic studies often favour filtration (W.-Y. Chen et al., 2017; Hamner et al., 2019; Haro-Moreno et al., 2020). Although there is some evidence in the literature to suggest that results of centrifugation are comparable to those of filtration in the recovery of bacteria from water (Brindle et al., 1987) this is scarce, and there is also evidence to the contrary (Boulanger & Edelstein, 1995). Specifically, Boulanger and Edelstein (1995) find that filtration with 0.22 µm pore polycarbonate filters recovers significantly greater amounts of bacteria than 0.22 µm cellulose ester filters do, and suggest this could be due to the structure of the pores. The pores of the latter are described as "tortuous", and it is hypothesised that this could make release of the trapped bacterial cells more difficult. In this project such polycarbonate filters were not used or considered, which points out an opportunity for future tests to compare with centrifugation.

An important point to note regarding centrifugation is its ability to homogenise sampling efforts across different sample types. The saturation experiments shown above are only some of the ones carried out in preparation for this chapter; other experiments were carried out on water from a stream, a stagnant pond, an aquarium and on different occasions from the river and the sea. Variation on the time and volume to saturation between samples was substantial; for instance, within 20 minutes one litre of river water could be filtered, but pond water had to be diluted 1:25 in distilled water for 100 mL to pass through the same filters. Variation is also visible for the same sample type; when two samples were collected at the same location from the same river four days apart, on one occasion it was possible to pass approximately one litre through the filter within an hour; this volume was nearer to three litres on the other occasion. When aiming to compare characteristics of samples collected in different places or under different conditions it is important to maintain homogeneous sampling effort (Staley et al., 2015a). With centrifugation this is possible, as clogging does not occur. Furthermore, the volume-yield experiments carried out for this chapter indicate that it is possible to increase the amount of DNA-containing particulates collected by increasing the volume of water processed, and that it is possible to do this without using larger containers, but rather repeatedly centrifuging the same ones and decanting them until arriving at the desired volume. It may even be possible to reduce the time to process larger volumes of water by centrifuging portions of a single sample in several tubes simultaneously, and then mixing the resuspended pellets in one tube.

The results of comparing a number of different DNA extraction protocols offer a less novel insight. PowerSoil has been used on a number of occasions for the extraction of DNA from water samples (He et al., 2020; Staley et al., 2015a, 2015b; Walden et al., 2017) and has demonstrated it is better at obtaining bacterial DNA in some cases than other kits and protocols. Specifically, Kaevska and Slana (2015) extracted more DNA from filter-trapped mycobacteria than did Fisher Scientific's SurePrep Water DNA, Norgen Biotek's Water RNA/DNA, and Qiagen's PowerWater kits when homogenising the filters, although the results differed when testing other ways of treating said filters. Similarly, Riediger and colleagues (2016) find it more efficient in the recovery of *Leptospira interrogans* cells from turbid water samples (pond and sewage), although PowerWater performs better with cleaner water samples (river and ultrapure water). An additional advantage identified by Riediger and colleagues (2016) is that this kit is more reliable across sample types; this is echoed by

Walden and colleagues (2017), who find it gave them more consistent results than other kits when they used it to extract different sample types and quantify their bacterial diversity.

Despite the advantage of being verified to work on water samples efficiently and consistently, the PowerSoil kit was not chosen for these comparisons because of its field-forward qualities, or for being inexpensive. On the contrary, it is expensive, (approximately £520 per 100 samples) and generates much waste (12 pipette tips, six tubes and one silica column per sample, plus spent reagents), but it has been used in extreme conditions before, including the absence of mains electricity (Gowers et al., 2019). Additionally, the cost of a MinION flowcell is currently in the £340-£670 range, which is already expensive, making an additional £5.20p per sample relatively minor. With these considerations in mind, using PowerSoil instead of a less expensive, simpler procedure may not be a major impediment to the implementation of Nanopore sequencing in metagenomic studies for pathogen detection in aquaculture.

4.4.3. Considerations regarding DNA extraction protocols

With regards to the performance of the other DNA extraction protocols, it is not entirely surprising to see some of them perform badly on environmental samples. For instance, the Casas protocol (Casas et al., 1995) was specifically intended to obtain viral nucleic acids from cerebrospinal fluid samples, not environmental. Boom's protocol (Boom et al., 1990) was intended for urine samples and Gram negative cells, and Pitcher's protocol was for both Gram positive and Gram negative cells, but from pure cultures. These three methods involve the use of guanidinium thiocyanate to denature proteins, but vary substantially in other ways; the Casas method did not have any steps intended to lyse bacterial cell walls, so the lysozyme incubation (at 37 °C for 30 minutes) from Pitcher's protocol was incorporated. Both the Pitcher and Boom protocols have been demonstrated to work with bacteria (Boom et al., 1990; Haudiquet et al., 2021; Pitcher et al., 1989; Samarasinghe et al., 2021), and the latter has even been used in environmental studies (Samarasinghe et al., 2021; Wanigatunge et al., 2014). The principles of these guanidinium thiocyanate-based methods have been adapted into commercial kits (Fong & Lipp, 2005; Parshukov et al., 2019), but the versions used in this thesis' work were not well suited for environmental DNA extraction. The Casas and Pitcher protocols were capable of retrieving bacterial DNA of sufficient quality for PCR, but only at low concentrations; the opposite was the case with the Boom protocol, where yields were relatively high, but DNA amplification was not possible.

The Martin-Platero protocol was another one intended for bacterial isolates and not environmental samples (Martín-Platero et al., 2007). Specifically, it was aimed at extracting DNA from Gram-positive lactic acid bacteria. Most of the studies that used this method were aimed at bacterial isolates from cheese (for example Lavilla-Lerma et al., 2013; Martín-Platero et al., 2009), although it has been used in the extraction of bacterial community DNA directly from cheese homogenates (Martín-Platero et al., 2008). Additionally, it has demonstrated it is capable of extracting DNA from isolates of a diverse range of bacteria (belonging to the *Actinobacteria*, *Firmicutes* and *Proteobacteria*), beyond the original lactic acid bacteria it was intended for (Jroundi et al., 2010). It is possible that the results for using this protocol (high yield, no DNA amplification during PCR) may be due to a good ability to lyse cells but poor inhibitor removal; see section 4.4.4 for more considerations on this subject.

Other protocols were intended originally for samples with high amounts of inhibitors: Orsini's protocol (Orsini & Romano-Spica, 2001) was designed to provide a simple means of extracting DNA from sludge, soil and sediments; Griffith's was used to extract bacterial DNA from soil (Griffiths et al., 2000). These were expected to perform well in the purification of the bacterial DNA, as per their original publications, and some subsequent ones (Adelowo et al., 2018; Musa et al., 2020; Nuccio et al., 2020; Pereira et al., 2020). However, both of these methods use a phenol-chloroform step to remove proteins and lipids. This is a serious drawback for any low-cost or portable implementation, as the use of these reagents requires a fume hood to manipulate them, and an appropriate means of disposing of them, besides involving risk of injury.

The Bordelon protocol was considered particularly appealing precisely because of its portability and low footprint, and the fact that it performed well in inhibitor removal when extracting DNA from urine, known to contain PCR inhibitors (Bordelon et al., 2013). In a subsequent study that developed the method, it was coupled with lysozyme digestion of cell walls to obtain pure DNA from bacteria in human urine (*Escherichia coli*; Bitting et al., 2016). This same step to lyse cell walls was implemented during this thesis. Another deviation from the original is the form factor: Bordelon and colleagues (2013) used a loose magnet to move silica-coated magnetic beads through several buffers distributed along a length of tubing. The attempts to replicate this protocol during the work for this thesis were unsuccessful, due to the magnetic beads not travelling through the air gaps between the buffers. This may have been due to the material of the tubing not being the same as the one

used by Bordelon and colleagues (2013) (silicone versus fluorinated ethylene-propylene). The adapted implementation with microcentrifuge tubes and a magnetic rack was simple to carry out in the laboratory, and performed well in terms of purity and specificity (ie, bacterial DNA was recovered and was amplifiable by PCR). Because of its relatively high yield when extracting from seawater, good performance in inhibitor removal and field-forward form factor (the original uses lyophilised reagents), it is an interesting candidate for further tests. However, before being implemented in real scenarios, it is necessary to ensure it provides reliable results, and whether these can be extended to river water with any modifications to the method.

4.4.4. Issues with experimental design and future work

There are two issues with the experiments carried out for this chapter that should be addressed in future experiments. On one hand there is the use of samples taken from a single location for each river and seawater, which means that the conclusions that can be drawn from these experiments are limited. The use of river water from the Rheidol and seawater from the Aberystwyth coast was intended to be a proxy for aquaculture water; by collecting samples from these sources the findings are limited to the sampling locations, as they can hardly be considered representative of the overall conditions in salmonid farms of the world. Although the results from this work are valid in their own right, to be actually able to extend them to the situations where they were intended to be used it will be necessary that future experiments draw samples from a variety of farm waters, not just a single sampling point, so that the diversity of conditions found in these environments can be accounted for. Furthermore, much is to be gained from measuring a range of conditions from these waters, such as turbidity and dissolved organic carbon, as it is possible that these parameters influence the yield obtained from these waters, or predict the success of inhibitor removal with a given protocol. Any subsequent experiments should implement these changes for them to offer a real, practical value.

The other issue is the lack of an internal standard in the 16S rRNA gene PCR experiments. The fact that most bacterial life has at least one copy of the 16S rRNA gene (Woese & Fox, 1977) provides a means of knowing whether bacterial DNA is extracted, and at the same time whether there are PCR inhibitors in the extract. Because results depend on these two considerations, if amplification occurs in a PCR of a DNA extract, it can be assumed that there were both bacterial DNA and an insufficient amount of inhibitors to stop the reaction from occurring. However, if the reaction does not take place, as it did for the Martin-Platero

protocol, it is not possible to know for certain whether this is due to a lack of bacterial DNA or a presence of inhibitors. With an internal control, such as a number of copies of a known, purified gene seeded into the DNA extract, it would be possible to discern between a lack of effectiveness on bacterial DNA extraction and poor inhibitor removal. When analysing the extract, a PCR for the added gene could be carried out; if this was successful, it would be possible to conclude that inhibitor removal was accomplished. This information could be valuable for further optimisation experiments, as high-yield protocols with poor inhibitor removal could be amended to solve this specific issue.

4.4.5. Metabarcoding; a means to deal with low yields and high costs

An issue encountered during this chapter has been the unreliability of $A_{260/280}$ ratio measurements. $A_{260/280}$ ratios, which are a commonly used measure of DNA purity can serve to identify protein contaminants were in fact taken for several of the experiments. However, they were not useable due to the low concentration of DNA, at which spectrophotometers like the NanoDrop™ 1000 cannot accurately determine these values (Thermo Fisher, 2012) (This annulled objective 1.3). This represents a double issue: On one hand this ratio is a parameter used by the manufacturer of Nanopore sequencing kits to define the quality of DNA that can be used as input (ONT, n.d.-e); on the other, Nanopore sequencing kits require large amounts of DNA to work. For a rapid barcoding kit (RBK-004, see ONT, n.d.-b), one of the kits with the lowest requirement of DNA quantity, 400 ng of input DNA in a volume of 7.5 µL of nuclease-free water are needed, and most kits tend to use up to 1µg. Quality requirements are $A_{260/280}$ ratios of around 1.8 and $A_{260/230}$ of 2.0-2.2 (ONT, n.d.-e).

Considering the issues with yield and cost of DNA extraction, and that 16S ribosomal RNA gene PCRs have resulted in amplification of bacterial DNA in extracts carried out with PowerSoil, an appealing alternative to metagenomics could be metabarcoding. Metabarcoding consists in the amplification of a gene commonly present in a range of organisms, which can then be assigned to a taxonomic group once sequenced. This is the principle behind numerous ecological studies, and is especially interesting in the study of low biomass samples, such as the ones from river water (W.-Y. Chen et al., 2017; He et al., 2020; Taniguchi et al., 2016). Besides this, it also allows for multiplexing, as kits exist that allow for the use of up to 24 different identifiers (ONT, n.d.-a; SQK-16S024 kit), which could drive costs down to prices that are actually economically feasible.

On the other hand, this technique has its own disadvantages. In particular, it has come under scrutiny due to the bias introduced by barcode amplification, which can skew representation of certain taxa (Acinas et al., 2005; Brooks et al., 2015; Frank et al., 2008; Větrovský & Baldrian, 2013). This is of minor importance when searching for pathogenic organisms, as long as the biases of the amplification of their barcodes are known and corrected for if existent. A more substantial problem is the loss of functional information that metabarcoding carries, as genes associated with antimicrobial resistance or virulence are undetected by these methods. Despite these disadvantages, the ability of metabarcoding to identify different bacterial species from small amounts of DNA is in part what drives the development of a broad-range LAMP assay in Chapter 6.

4.4.6. Conclusion

In conclusion, the most appropriate means of concentrating bacteria from river and seawater is centrifugation, although to be certain that this applies to aquaculture in general it is necessary that further experiments are carried out to validate these conclusions for said environments. PowerSoil has demonstrated an ability to adequately extract DNA from these types of samples and remove inhibitors to an extent that PCR inhibition is avoided. With some more development, these methods of processing water samples could lead to the design of specific protocols for sampling and processing water from, or at, salmonid farm sites.

Chapter 5: Producing pathogen-specific LAMP assays

5.1. Introduction

Loop-mediated isothermal amplification, often abbreviated to LAMP, is technique used for the amplification of nucleic acids. Notomi and colleagues' seminal paper (2000) describes what was then a novel method, with several potential advantages with regards to PCR, the most common method for nucleic acid amplification. Their paper describes LAMP as specific, rapid and efficient, simple to set up and carry out, with limits of detection similar to PCR. Within four years this tool was first used in the detection of a fish pathogen, *Edwardsiella tarda* (Savan et al., 2004), and through the following years several more followed: *Flavobacterium columnaris* in 2006 (Yeh et al., 2006); *Yersinia ruckeri* in 2008 (Saleh et al., 2008b); *Aeromonas salmonicida* in 2009 (Kulkarni et al., 2009a), etc. Interest in this method as a tool for diagnostics and pathogen detection has continued through the past two decades, but has recently reached prominence with the development of several tests against SARS-CoV-2 (Dao Thi et al., 2020; Huang et al., 2020; Ptasinska et al., 2021).

Although the use of LAMP in practice is relatively simple, the theory behind it is more complex. It uses four sets of primers with a total of six regions of complementarity, plus, optionally, another two more primers, which serve to speed up reactions, and replication is carried out by Bst polymerase, which has strand-displacement activity (Nagamine et al., 2002; Notomi et al., 2000). Two of the primers, called the inner primers, bind to two regions of complementarity each; one of them binds, initiates amplification, and then one of the outer primers initiates replication from behind, displacing the inner primer as it goes. This frees up a region of the amplified section to which the other part of the inner primer binds, forming a hairpin structure. When this occurs on the other end of the amplified region, a dumbbell-like structure forms, which on one hand primes itself, and on the other can also be primed by any of the primers still in the solution. From here, a number of different structures form, in which the target region between the inner primers is repeated over and over. This results in exponential amplification of the target DNA. A diagram from Tomita and colleagues (2008) is reproduced to illustrate the mechanism by which LAMP works (Figure 5-1).

Despite this, LAMP is logistically simple to carry out; the strand-displacement activity of the Bst polymerase means thermal cycling is not necessary to melt the strands of the DNA apart, which means it is an isothermal reaction, normally carried out at ~65 °C. This avoids the need for a thermocycler, which is expensive and requires training to use, and makes it possible to

use a heating block or water bath, much cheaper and more straightforward to use. Isothermal occurrence of the reaction also makes amplification a continuous process, and therefore faster too; assays can be complete within an hour (Notomi et al., 2000), or even 30 minutes with the addition of loop primers (Nagamine et al., 2002).

Other advantages of LAMP are its specificity, which is greater than that of PCR, due to the six regions of complementarity it relies on to work, and its lower detection limits compared to conventional PCR. The latter is not necessarily explained by its mechanism of action, but rather a commonly reported finding in different tests that target the same gene with both techniques (Itano et al., 2006; Kulkarni et al., 2009a; Waliullah et al., 2019).

These features have made it a very useful tool in the detection of pathogens, especially with the intention of being applied in resource-limited settings, both for human pathogens (Hopkins et al., 2013) and for relevant diseases of crops (Kini et al., 2021) or farmed animals (Punati et al., 2019). It is readily combined with reverse transcription to detect RNA based viruses (Huang et al., 2020; Notomi et al., 2000), but several other augmentations have been applied to it. For instance, coupling LAMP with a lateral flow device (Punati et al., 2019), the addition of calcein, which changes colour upon reaction (Suebsing et al., 2013), or SYBR safe, which fluoresces under UV light (Carter et al., 2017) allows for better readability of results. These additions make it more accessible to non-expert audiences, more field forward and faster, as it does not require a follow-up with a gel electrophoresis of the amplified DNA, or the use of a turbidimeter to measure the phosphate precipitate the reaction normally generates.

Overall, the characteristics of colorimetric LAMP make it an appealing candidate to design assays for the detection of pathogens in aquaculture. The relatively inexpensive equipment, low detection limit and reported robustness (Francois et al., 2011; Kaneko et al., 2007), plus the simplicity of interpretation of the colorimetric assay make it highly applicable as an on-site detection tool, with potential to perform better than PCR in aquaculture settings. Despite LAMP assays not acting as naïve tests capable of detecting a variety of pathogens simultaneously, they have the advantage of requiring only a small amount of substrate to start a reaction. This could serve to overcome the challenge of low DNA yield from environmental water samples observed in Chapter 4, and provide a means of pre-emptively testing farm waters. It is for this reason that the current chapter aims to develop tests for the seven

Bacterial Pathogens Of Interest (BPOIs); see section 1.1.2 for the complete aims and objectives.

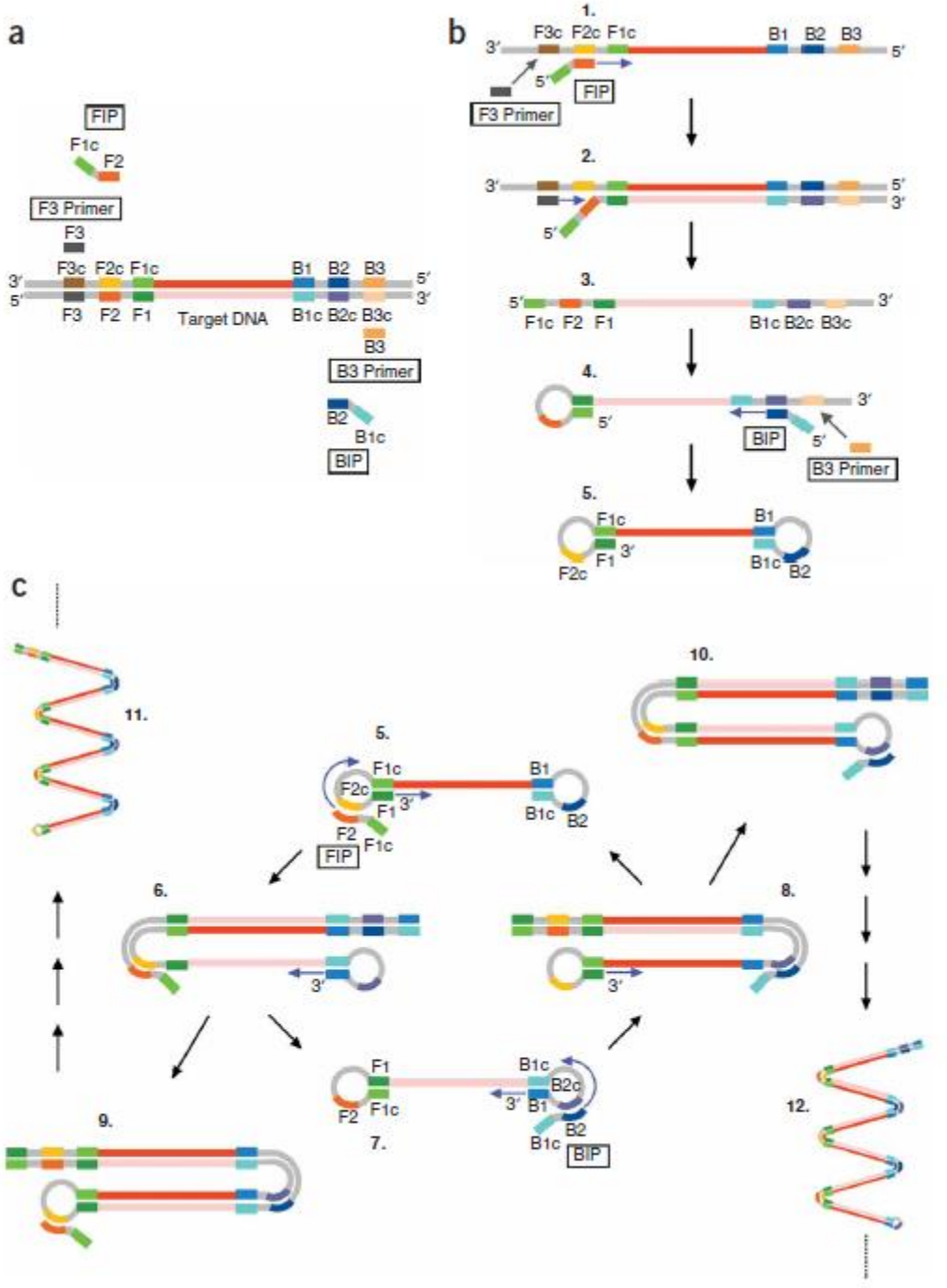


Figure 5-1: Loop-mediated isothermal amplification (LAMP) uses four primers that are complementary to six regions of the target DNA (a). Upon binding and amplifying the target region, the inner primers insert regions of self-complementarity in the sequence, which cause the DNA to take up a "dumb-bell" structure (b). This structure is self-priming, and is the starting point for an exponential amplification phase that can lead to a number of different structures (c). Diagram from Tomita and colleagues, 2008.

5.1.1. Overview of primer design software

A specialised software is needed to design LAMP primers, given their differences with common PCR primers. This is commonly done through PrimerExplorer (http://primerexplorer.jp/lampv5e/index.html), a software owned by the Eiken Chemical Corporation (Carter et al., 2017; Kulkarni et al., 2009a, 2009b; Saleh et al., 2008a, 2008b; Yeh et al., 2006, etc.). Although this program has changed with time to incorporate new features, it still is limited in its capabilities. For instance, the region for which PrimerExplorer will design primers is limited to 2000bp in size; the interface is graphical, and thus not amenable to automation or incorporation into high-throughput workflows; the code is not open source, etc. Other software has appeared through the years: LavaLAMP, published for the first time by Torres and colleagues (2011), attempted to correct the aforementioned faults and some additional ones, but despite the niche for such a software it is currently no longer available online. Other currently available options are private, like Premier Biosoft's LAMP Designer (http://www.premierbiosoft.com/isothermal/index.html), but its cost makes it less accessible and desirable. However, a recently published program called GLAPD solves the three issues mentioned earlier with PrimerExplorer (Jia et al., 2019). It is open source, is operated from the command line, and can design primers for whole genomes. Additionally, it incorporates features that make it very competitive, such as its ability to design primer sets which are commonly specific to a set of sequences input to the program, ensuring there are no regions of complementarity in another set of sequences given as input. This makes it unnecessary to make previous alignments to manually select target regions where there are mutations that separate the outgroup from the ingroup, and that are common to all members of the latter.

This last program, GLAPD (Jia et al., 2019), was the one chosen to design LAMP primers to develop assays for the BPOIs of this thesis. Given its crucial role in this chapter's workflow, the following paragraphs detail the use of this program more closely.

Broadly, the intended use of GLAPD is to take a genome of a target organism, which it will use as template to design a number of primers, and then optionally supply it with a number of other genomes or sequences, for which GLAPD can design common or exclusive primer sets. This means that it can be used to design a set of primers that will work for an entire genus, for example, or a set that will react against one strain of a bacterium, but not another similar, or closely related one. Both these characteristics are made use of during this work, in order to design primers that will on one hand, react against any strain of the target species, but on the

other, not react to any closely related species within the genus. The purpose of this is to reduce as much as possible the chances of false positives or false negatives owed to primer design when carrying out LAMP assays with these designs.

GLAPD is a complex program that works in three steps. During the first step it takes a reference DNA sequence as input and finds all possible primer sequences that could bind to it. During the second step, it searches for matches of these primer sequences among a set of DNA sequences supplied in an additional file, to find whether these primers will bind to DNA other than the reference sequence. During the third step, primers are combined into sets, and GLAPD returns a list of the sets it has found that meet the desired specifications.

The requirements of GLAPD are, therefore, the following: a reference genome sequence in fasta format; a file with all other genome sequences of relevance, comprising all target genomes besides the reference and all the non-target sequences the primers are *not* intended to bind to (in fasta format as well); a list of the genomes against which the primers must react, in text format (the genomes against which the primers must *not* react are deduced to be the ones present in the combined genome dataset, but not included in the text list). Additionally, GLAPD uses Bowtie (Langmead et al., 2009) to search the primers it has designed for the reference sequence in the additional sequences supplied to it (the common and background genomes). Because of this, it also requires a Bowtie index file for the common and background genomes before running GLAPD. Section 5.2.1 gives details on how these ancillary data files are collected or generated, and section 5.2.2 explains how the program was used to design a number of primer sets that would potentially serve to detect the seven BPOIs.

5.1.2. Aims and objectives

Aim:

1) To produce LAMP assays that serve to detect each of the bacterial pathogens of interest (BPOIs, see list in section 1.3).

This aim is divided into the following objectives:

1) To design sets of LAMP primers that will amplify the DNA of each of the BPOIs.
2) To narrow selection down to an amount small enough that makes it feasible to test them all.
3) To test designed primers to confirm they react in assays against the DNA template of the target organism.
4) To find out the limit of detection of LAMP assays carried out with these primers.
5) To find the optimal conditions to run the assays.

5.2. Method development

5.2.1. Data gathering and preparation for GLAPD

To collect the data required to run GLAPD, first the metadata corresponding to the BPOIs was obtained from the NCBI's RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/). This was done through the NCBI's own Datasets tool (Sayers et al., 2021; https://www.ncbi.nlm.nih.gov/datasets/), which provides the data in JSON format; it was then parsed with jq (https://stedolan.github.io/jq/) to obtain the relevant information. This was done for all the genomes available for each target species, and all the genomes of each target species' genus. The metadata for the latter were filtered to retain only the entries for representative RefSeq genomes, one per non-target species of the same genus. The decision to use representative genomes instead of all those available for the non-target species within the target's genus was taken to reduce the amount of redundant data input into GLAPD. Further checks would be carried out once the primers were designed to ensure specificity. However, genomes of all qualities were used for the target species because the aim was to ensure that the primers GLAPD designed would bind to sequences known to be present in every known strain of each BPOI. The metadata downloaded was specifically the genome accession, species name and genome quality (representative, draft, etc.). This metadata allowed downloading the fasta-format genome assembly files though the datasets tool for the strains of interest. The accessions of all the genome assemblies downloaded are in Appendix 5.

Seven representative genomes were downloaded, one for each BPOI, to be used as references for GLAPD (Table 5-1).

Table 5-1: Genbank accessions of genome assemblies used as reference in primer design.

Genome name	Genome accession
Aeromonas salmonicida subsp. *salmonicida* A449	GCF_000196395.1
Flavobacterium branchiophilum FL-15	GCF_000253275.1
Flavobacterium psychrophilum FPG3	GCF_000754405.1
Moritella viscosa	GCF_000953735.1
Renibacterium salmoninarum ATCC 33209	GCF_000018885.1
Tenacibaculum maritimum NCIMB 2154	GCF_900119795.1
Yersinia ruckeri	GCF_002192595.1

A total of 284 assemblies corresponding to other genomes sequenced for each target species were also downloaded (Table 5-2) and 192 assemblies corresponding to the representative genome for each non-target species remaining in each target's genus (Table 5-3). Exact

accessions for the downloaded data are given in Appendix 5. It is worth noting that no assemblies were downloaded for other species of *Renibacterium* because it is the only species of its genus.

The representative genome belonging to the target species, which was intended to be used as the reference for GLAPD, was stripped of plasmid sequences, to ensure the LAMP primers the software designed would not target these mobile genetic elements. For the rest of the genomes of each target species all the separate contigs were linked with a 30 N (unspecified nucleotide) spacer. This was done for two reasons. First, GLAPD cannot cope with genome assemblies composed of several sequences when looking for common primers; it will search for primers that are common to each of the sequences it is given, and genome assemblies are often fragmented into several contigs. Second, the designed primers should not span the gap between two contigs, which are not necessarily physically together in the genome, so a separator that is longer than the expected length of the region of complementarity of LAMP primers was used.

Table 5-2: Number of genome assemblies of the same species as the BPOIs downloaded from GenBank.

Species name	Number of assemblies downloaded
Aeromonas salmonicida	58
Flavobacterium branchiophilum	4
Flavobacterium psychrophilum	119
Moritella viscosa	12
Renibacterium salmoninarum	2
Tenacibaculum maritimum	23
Yersinia ruckeri	66
Total	284

Table 5-3: Number of representative genome assemblies downloaded for the remainder of the species in the genus of each BPOI.

Genus	Number of assemblies downloaded
Aeromonas	31
Flavobacterium	123
Moritella	3
Renibacterium	0
Tenacibaculum	17
Yersinia	18
Total	192

Finally, indexes were prepared from the combined genomes file using Bowtie (Langmead et al., 2009); these are used by GLAPD to search which genomes the primers it has designed are likely to bind to besides the reference genome.

5.2.2. Running GLAPD

Once all the data was retrieved and prepared, GLAPD was used to design primers specific to each of the BPOIs. The first step, "single.c", was run with the representative genome of each target species as reference. The second step, "par.pl", was given the corresponding Bowtie indexes and the list of common target genomes. The third step, "LAMP.c", was run with the flags "-common" and "-specific", to ensure the primer sets returned were both common to all genomes of the target species and specific to the target species.

The entire data collection, preparation and primer design process was run on the High Performance Cluster (HPC) of the Institute of Biological Environmental and Rural Sciences (IBERS) of Aberystwyth University, but GLAPD steps were run as tasks. The "Single" and "LAMP" steps were run on a single core, but "par.pl" allows parallel processing, so 12 such cores were assigned to each job; full job details can be found in Appendix 6. For each step seven jobs were run in parallel, one for each of the BPOIs, in order to reduce time spent waiting for results. The overall workflow is represented in Figure 5-2.

As a result of running GLAPD, a list of ten sets of primers for each target was produced, except for *A. salmonicida*, for which the software was run a second time to return 20 primer sets. This was because after performing follow-up checks it was seen not enough primers were specific enough to carry on to the synthesis and testing step. In all cases, the primer sets designed by GLAPD were valid for all the genomes within their target group.

The primer sets were named with an abbreviation comprising the initials of the species name, plus the number of the set (1-10, or in the case of *A. salmonicida*, 1-20); as GLAPD does not return primer sequences strictly speaking, but rather the sequence of their separate regions, the regions are referred to with the name of the primer set followed by F1c, F2, F3, B1c, B2 and B3.

5.2.3. Narrowing down selection of primers

Primer design itself is one of the main factors affecting LAMP assay optimisation (Lucigen Corporation, 2017); a good set of primers is understood as one that reacts as soon as possible with a sample containing template DNA with the longest delay until samples with no

template begin reacting. However, because the intention was to test primer sets for seven different species of bacterium, the amount of them that could be feasibly tested was limited. Therefore a sub-selection of only two primer sets for each bacterium, from the ten designed by default by GLAPD, was synthesised. These were selected on the basis of their likelihood to react with DNA from non-target organisms; though this does not address the question of whether the primers will react faster with their targets compared with no-template controls, it does serve to avoid testing primers that may ultimately be impractical because of their unspecificity.

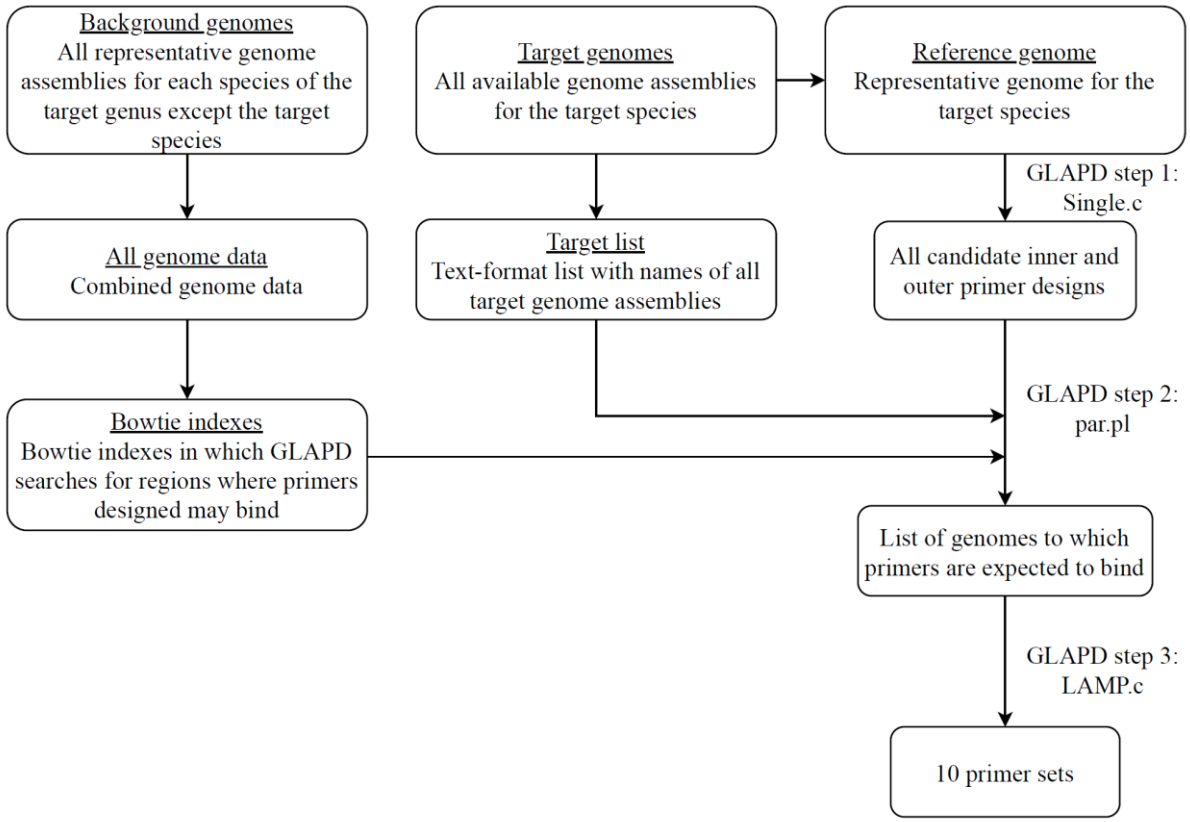


Figure 5-2: Process followed to design specific LAMP primers with GLAPD software.

The procedure to select these two primer sets consisted in running a BLAST search (Altschul et al., 1990) of each of the primer regions designed by GLAPD, recording the number of regions per set whose sequence could be found in any given genome. The BLAST was run with default parameters except a maximum e-value of 50 on the Nucleotide (nt) database. Additionally, facing pairs of regions where the primers may bind within any given contig were identified, and the distances between them calculated. Regions of non-target templates where primers were unlikely to bind because of mismatches at the primers' mismatch-

sensitive ends were excluded from the distance calculations. These are the 5' end for F1c and B1c and the 3' end for F2, F3, B2 and B3 (PrimerExplorer, Eiken Chemical Co., n.d.). A number of Bash scripts were written to carry out this job automatically (see Appendix 7), and the final results were visualised with R 3.6.1 (R Core Team, 2019) using the ggplot2 package (Wickham, 2016) (Figure 5-3).

The two sets selected for each BPOI were those which met the following conditions: first, that the six primer regions of the set could be found in the genome of its target organism; second, that they had the lowest number of primer matches possible to non-target organisms (ie., those where a maximum of two primer regions matched any given non-target genome were preferred over those where the maximum was three); and third, of those where the second condition was a tie, the set where the maximum number of primer regions matching any given non-target genome matched with the lowest number of non-target genomes.

The BLAST searches returned the names of organisms where sequences similar to those of the primer regions were found. These were of interest because of two reasons: First, ascertaining that through BLAST it was possible to find the six sequences of each of the primer sets in their target organism's genome. Second, they allowed finding out whether pairs of primer regions from within a set were present in the same organism, which could lead to amplification of the DNA of non-target organisms if it were included in the LAMP assays.

Regarding the first point, for most of the primer sets, all six primer regions matched sequences in their target organism's genome. However, if less than six matched, the primer set was excluded from the selection process; this was the case for 14 of the 20 sets designed for *A. salmonicida*. In no case did more than one organism have matches for all six primer regions (see Table 5-4).

Regarding the second point, although among the BLAST results matches that encompassed the three domains of life, plus viruses, were found, it was rarer to see matches that were likely to result in any amplification of non-target DNA. On average (mean), sequences similar to the designed primer regions of each set could be found in 122 different organisms. Far fewer of those matches posed a risk of resulting in amplification; on average, 110 of the organisms where similar sequences to the primer regions were found had only a single match, with no potential for exponential DNA amplification; only 9.26 were found to have two similar sequences, and 1.65 had three, four or five matches. As stated before, organisms besides the targets had six regions of similarity to the primer regions of any given set; this suggests that

there are few situations where any of the primer sets designed with GLAPD will bind to and amplify the DNA of a non-target organism.

The subsequent checks on distance between facing primer pairs within non-target genomes provided another reason to be confident that the primers would not amplify non-target genomes. Only situations where primer regions from within a set had matches on facing regions of a genome were considered to be liable to result in amplification; these are reflected in Figure 5-3. These show that there are a number of primer sets which have no facing primer regions in any genome in the NCBI databases, and that for many others these facing pairs only occur at distances beyond the maximum recommended distance of ~320 bp between the ends of primers F3 and B3 (Eiken Chemical Co., Ltd., n.d.). Specifically, among the selected primer sets, the closest a pair of matches on a non-target genome was at 29,940, on the genome of *Pelagornium spinosum*, pertaining to Fb3. This in-silico evaluation of the primer sets resulted in the selection, based on the criteria stated earlier in this section, of the primers listed in Table 5-5.

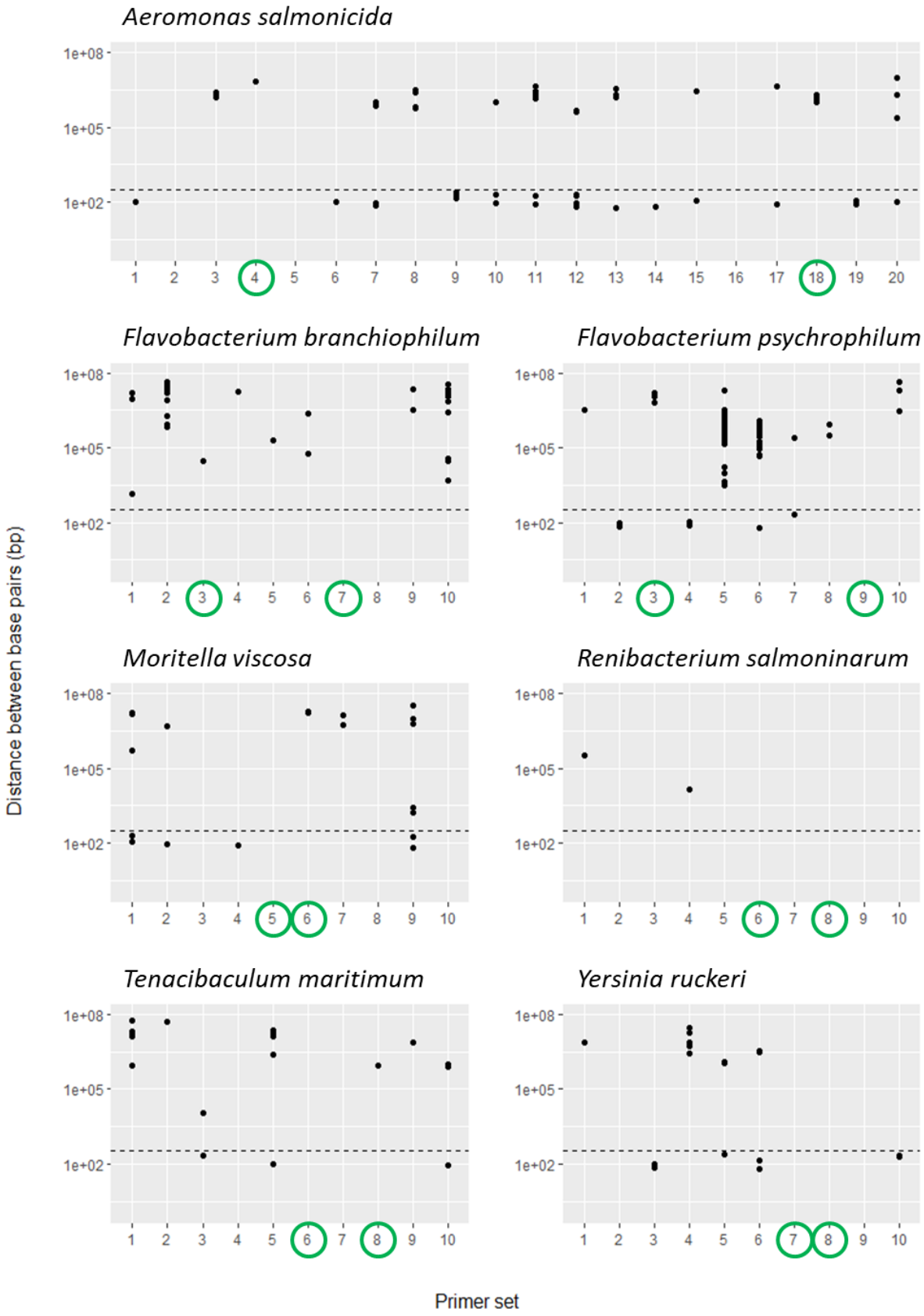


Figure 5-3: These plots depict the distances between any pair of potentially amplifying primer regions on any non-target genomes found through BLAST on the NCBI Genome database. The plots are divided according to the target organism, with the primer set design number on the X-axis. The Y axis is exponential, and the dashed line represents the maximum distance normally found between LAMP primer binding sites (320bp). Circled in green are the primer sets selected for synthesis.

Table 5-4: Number of genomes from the NCBI RefSeq database where a number of regions from the LAMP primer designs could be found via BLAST (see section 5.2.3). The number of primer regions found is indicated in the column header, and the number of genomes where these may be found are listed below. The primer sets selected for synthesis are in bold.

Target	Primer set	Total number of matches	Number of primer region matches per genome matched					
			6	5	4	3	2	1
Aeromonas salmonicida	1	201	1	0	0	3	14	183
	2	83	0	1	0	2	2	78
	3	236	0	1	0	1	7	227
	4	**137**	**1**	**0**	**0**	**1**	**9**	**126**
	5	93	0	1	0	0	8	84
	6	109	1	0	1	1	4	102
	7	101	0	0	2	3	6	90
	8	237	0	0	1	2	13	221
	9	83	1	0	0	1	13	68
	10	241	0	2	0	0	15	224
	11	141	1	0	1	2	6	131
	12	200	0	4	2	3	20	171
	13	226	0	0	2	3	15	206
	14	140	0	1	0	2	8	129
	15	137	0	1	0	0	11	125
	16	233	0	1	0	3	13	216
	17	256	0	1	1	1	20	233
	18	**86**	**1**	**0**	**0**	**0**	**2**	**83**
	19	98	0	1	0	6	4	87
	20	155	0	0	1	3	7	144
Flavobacterium branchiophilum	1	136	1	0	2	3	24	106
	2	125	1	0	0	0	8	116
	3	**64**	**1**	**0**	**0**	**0**	**3**	**60**
	4	116	1	0	0	3	12	100
	5	86	1	0	0	0	10	75
	6	173	1	0	1	4	21	146
	7	**105**	**1**	**0**	**0**	**0**	**4**	**100**
	8	96	1	0	0	2	14	79
	9	150	1	0	0	0	17	132
	10	123	1	0	0	0	10	112
Flavobacterium psychrophilum	1	226	1	0	0	1	33	191
	2	156	1	0	0	2	7	146
	3	**85**	**1**	**0**	**0**	**0**	**6**	**78**
	4	220	1	0	0	1	8	210
	5	147	1	0	1	3	8	134
	6	140	1	0	0	1	12	126
	7	108	1	0	0	1	3	103
	8	128	1	0	1	1	8	117
	9	**51**	**1**	**0**	**0**	**0**	**1**	**49**
	10	96	1	0	0	0	8	87

Moritella viscosa	1	104	1	0	0	2	14	87
	2	94	1	0	0	0	4	89
	3	51	1	0	0	0	4	46
	4	129	1	0	0	1	3	124
	5	**66**	**1**	**0**	**0**	**0**	**0**	**65**
	6	**98**	**1**	**0**	**0**	**0**	**0**	**97**
	7	102	1	0	0	3	17	81
	8	87	1	0	0	0	5	81
	9	90	1	0	0	3	12	74
	10	100	1	0	0	0	9	90
Renibacterium salmoninarum	1	134	1	0	0	0	5	128
	2	121	1	0	0	0	5	115
	3	105	1	0	0	0	6	98
	4	76	1	0	0	0	3	72
	5	43	1	0	0	0	2	40
	6	**43**	**1**	**0**	**0**	**0**	**1**	**41**
	7	53	1	0	0	0	2	50
	8	**43**	**1**	**0**	**0**	**0**	**0**	**42**
	9	52	1	0	0	0	1	50
	10	95	1	0	0	0	3	91
Tenacibaculum maritimum	1	209	1	0	1	3	24	180
	2	134	1	0	0	2	17	114
	3	180	1	0	0	1	20	158
	4	94	1	0	0	1	9	83
	5	189	1	0	1	5	18	164
	6	**131**	**1**	**0**	**0**	**0**	**16**	**114**
	7	71	1	0	0	1	8	61
	8	**86**	**1**	**0**	**0**	**0**	**6**	**79**
	9	70	1	0	0	1	6	62
	10	88	1	0	1	2	3	81
Yersinia ruckeri	1	62	1	0	0	0	8	53
	2	87	1	0	0	0	6	80
	3	49	1	0	0	1	1	46
	4	111	1	0	0	7	17	86
	5	77	1	0	0	1	7	68
	6	210	1	0	0	2	31	176
	7	**80**	**1**	**0**	**0**	**0**	**3**	**76**
	8	**75**	**1**	**0**	**0**	**0**	**5**	**69**
	9	120	1	0	0	1	14	104
	10	209	1	0	0	1	12	195
Mean		121.78	0.83	0.18	0.24	1.21	9.26	110.06

Table 5-5: Synthesised LAMP primers and their sequences. These were the primers tested in the preliminary experiments (section 5.2.5).

Target species	Primer set	Primer name (ID)	Primer sequence (5' - 3')
Aeromonas salmonicida	4	As_set4_F3	GCATCCACGTCGTTGAGG
		As_set4_FIP	GTAGTGGTGTTGACCCGTCAGGTGTCGAGCAG CTTGATGTC
		As_set4_BIP	GGCCGATAAAGCCGGTTCCTGCCAACAGAGTC AGGGAGTA
		As_set4_B3	GTTGTGCTGGCCCACTAG
	18	Ar_set18_F3	CAGCAGGCGATCATCCTCT
		Ar_set18_FIP	TTTGCCAACCGCACGTTTGTGTGGCATCTCTAT GTCCGTCT
		Ar_set18_BIP	CCCGTCCGTATCCCGACAGATCCTTGCTCACCC TGACG
		Ar_set18_B3	GTGAGTGCCTGCGTGAAT
Flavobacterium branchiophilum	3	Fb_set3_F3	TAAAGTCAATTTCAAAAAAGGCG
		Fb_set3_FIP	GCTCATAGAACGAATGGCTAAAGTAAGCTTTA AATCAATTGATGCAAG
		Fb_set3_BIP	TTGGTCATTATGGGTTAGCATTTGATCGGGATA ACGTCTTATAGG
		Fb_set3_B3	ATTGAAGCAATCTATGTACCATA
	7	Fb_set7_F3	TAGAACAAATTATCGAGGACAAC
		Fb_set7_FIP	AGCATGTCCAATATGAAGATAACCAGGTTTTTC TCAAAATCAGTTACG
		Fb_set7_BIP	AATTTGTTTGAACTTCGGTTTAGGCGTATCATC AAAACGTAAATTGACTG
		Fb_set7_B3	TTCTTCTTTTGCTGGGTTG
Flavobacterium psychrophilum	3	Fp_set3_F3	ACAGGAAGTTCTTATATAGTTTCTG
		Fp_set3_FIP	ATCCGTTTGCGACGCATAATTTACGAAATTGTA GCTGCTACT
		Fp_set3_BIP	AATGACGGAGCAATTTCGTTTGTTCTCCGTCTT GGCTTAG
		Fp_set3_B3	ACTTTAAAGAATACACCTTCTGG
	9	Fp_set9_F3	GCCATTTTCTCGCGAATTA
		Fp_set9_FIP	AACCATTTTTTAGACGCACTTCGTTTAGAAGTT GCTCCAGCT
		Fp_set9_BIP	TTACCCAAATTGAAGCGAGTTATGACGTTTACT GGCTTCGGTA
		Fp_set9_B3	ATGCAAGGTTCCAGCTAC
Moritella viscosa	5	Mv_set5_F3	GAAGGTTTGTTTCTTGCCGG
		Mv_set5_FIP	ACGAGCAGCATTTAAGCCTGCAGGTACGACGG GTTATGAGGA
		Mv_set5_BIP	CGTGGCATCCTCGTCGTGATCCGATACGGTTCT TTCGTCCC

		Mv_set5_B3	CGGTACTCAGCACGACTTG
	6	Mv_set6_F3	CGAAAAAATGGAAGCGATTG
		Mv_set6_FIP	GCCAGTAGTGCGTTGATCTTATCACGTTTACGCTCAACCTGG
		Mv_set6_BIP	AGAAGATCTTGTTCGCCGTCCCTAACTCTTCAATTGTCATCAGAG
		Mv_set6_B3	GATTCATGTTCTAACTTAGGGC
Renibacterium salmoninarum	6	Rs_set6_F3	GAAATCTCAGCCCGCCTC
		Rs_set6_FIP	CGGAGGGTAAGCTTCGGCAAATATTGCTGAAGAAGGCACCAT
		Rs_set6_BIP	CACCGGTGGAAGTAAGCACCGTGCCAAATTGAAGCGAGAGT
		Rs_set6_B3	TGGATAATCGGCCTCTGGC
	7	Rs_set7_F3	TATTTTGGCCGGCGTCAAG
		Rs_set7_FIP	CCGCTGGCTTCCAATTGAGCTAATCACCTTGTTGGCCACTG
		Rs_set7_BIP	TGCCAGGTATCTCGACCTCTGCCAAGAGTTTTCGCGACTTCG
		Rs_set7_B3	AATGTTGATCTCACCCGCG
Tenacibaculum maritimum	6	Tm_set6_F3	GCTGTACATGAAGTAGGGT
		Tm_set6_FIP	GGCAATTACAAGCTGTTCTCAAAATCCATGAAAAAAACCTTTCCCT
		Tm_set6_BIP	GGAACTATTCCTGCTGTTAGTGATGAATACCTTACATATCAACCAGA
		Tm_set6_B3	CTGGATTTGGAAAAACGGG
	8	Tm_set8_F3	TTTTTGCAACCTACTATAACTGC
		Tm_set8_FIP	GCGGACTCGTTCAAGCTTTAAATTTACCAAAAGTTATGATTCGTGC
		Tm_set8_BIP	GCCTCATCAACAATTAAGTGCACATCTATGGATGGCGACCTC
		Tm_set8_B3	GGGGAAAAAGAAATTTATGTTGT
Yersinia ruckeri	7	Yr_set7_F3	CTTGTCAGCCAACGCCTT
		Yr_set7_FIP	TGCGAGCATGTTGCATCAGGAAGGAATTCTCTGCCCCGGT
		Yr_set7_BIP	GAGTTTTCTCGCTGGGACGCGGCTGATATTTGGCGACGTTC
		Yr_set7_B3	CGGCAAAGAAAGCGGTTG
	8	Yr_set8_F3	GGAATACCGAGTTGAGCACG
		Yr_set8_FIP	GCAAAGTCAGCGTCGGCGAATCGCTAAACGGGCGTTGAA
		Yr_set8_BIP	ATGACCGATTCTCTGGCGGCTCATCACGGCAAGGCAAC
		Yr_set8_B3	TCCAACGCACATCAAAGGC

5.2.4. Primer synthesis and preparation

Two sets of primers were ordered for each one of the seven BPOIs for which primers were designed. These were synthesised by Thermo Fisher Scientific at a 25 nm scale, and provided in a dehydrated, desalted format. Primers were rehydrated on arrival to a concentration of 100 µmol L^{-1} by adding UV-treated HPLC water (Fisher Scientific). Water volumes to be added were calculated by the formula:

$$\frac{Primer\ quantity\ (\mu mol)}{Desired\ concentration\ (\frac{\mu mol}{L})} = Volume\ of\ water\ (L)$$

Aliquots were made, and from those aliquots working mixes that combined the inner (FIP, BIP) and outer primers (F3, B3) at 10X the final reaction concentration were prepared. In these, the primers were at concentrations of 2 µM in the case of the outer primers and 16 µM in the case of the inner primers, as recommended by both the master mix manufacturer (NEB, n.d.) and the Eiken Chemical Company (PrimerExplorer v5 manual, Eiken Chemical Co., Ltd., n.d.).

5.2.5. Preliminary experiments

5.2.5.1. Methods and rationale

Preliminary tests were run on each of the primer sets received from the manufacturer to assess which may be the most appropriate for further tests. These consisted in LAMP assays prepared with NEB colorimetric LAMP master mix according to the manufacturer's instructions (NEB, n.d.). However, volumes were multiplied by 0.4 (final volume 10 µL), as opposed to the default of 25 µL, in order to conserve reagents. This is because it was estimated that a large number of reactions would be run (at least 380), the reagents were relatively expensive (£827 for 6.25 mL of master mix after discount), and future supplies may not be reliable due to potential applications of LAMP in SARS-CoV-2 testing efforts. Each assay tube contained the reagents detailed in Table 5-6.

Five assays were run per primer set. Four of them contained progressively lower concentrations of template DNA, which ranged between 10 ng μL^{-1} to 10^{-5} ng μL^{-1} in 100-fold reductions in concentration, and one negative control that contained sterile UV-treated HPLC water instead of DNA solution.

Table 5-6: Reagents and quantities used in the preparation of the LAMP reaction mixes used in the preliminary experiments.

Ingredient	Volume (µL)
2X NEB colorimetric LAMP master mix	5
Sterile, UV-treated HPLC water	3.6
10X primer working mix	1
Template DNA (variable concentration)	0.4

The reaction was run for 75 minutes at 65 °C in G-Storm GS0001 thermal cycler (G-storm Ltd., Somerton, UK), pausing it after 30 minutes and every 15 minutes thereafter to visually inspect the results. The colorimetric LAMP master mix changes in colour from pink to yellow when DNA amplification occurs (Figure 5-4); assay results were classified as positive or negative accordingly. Additionally, samples where the pink colour had softened very mildly, without clearly turning yellow (or at least orange) were classified as "unclear".

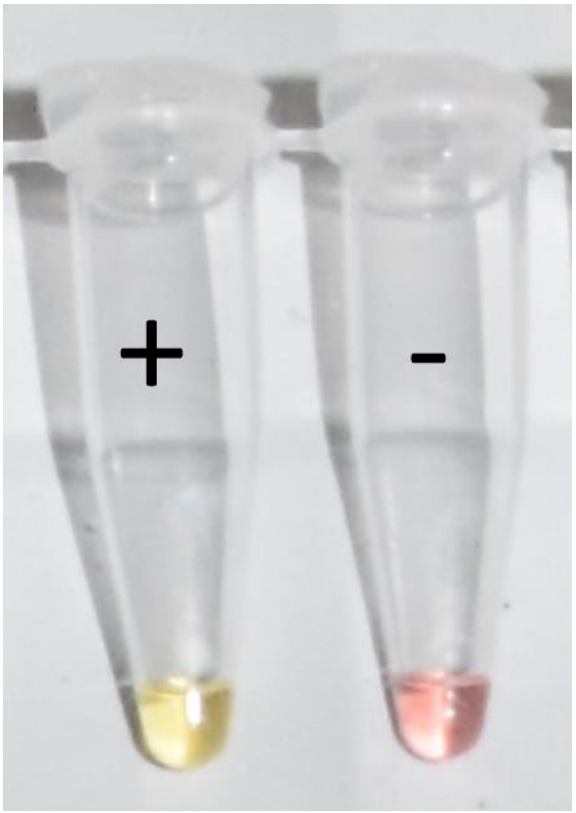


Figure 5-4: Comparison of the appearance of a positive (yellow, labelled "+" on the image) and a negative result (pink, labelled "-" on the image) of a colorimetric LAMP reaction with the NEB WarmStart Colorimetric LAMP master mix.

Because, as previously stated, primer design is such an important factor in LAMP assay optimisation (Lucigen Corporation, 2017), the objective was to select the most appropriate of the primer sets for further tests. One of the two primer sets targeting each BPOI were selected for further tests based on the following criteria, in this order: first, no reaction in no-template control; second, lower limit of detection was preferred; third, shorter times to positive reaction were preferred.

5.2.5.2. Results

Preliminary tests provided a rough estimate of which primers were more appropriate for follow-up tests. Reactions were evaluated on their colour change; if they shifted from the initial pink of the master mix to a clear yellow colour they were considered to be positive, and if they remained pink, as negative. Although those assays where colour changed very slightly were recorded too, only a clearly noticeable change in colour was considered to indicate a positive reaction.

Some of the LAMP assays reacted spuriously; this was the case for those with primer sets As18, and Rs7 (see Table 5-7), which disqualified from further tests: this prompted the selection of As4 over As18 and Rs6 over Rs7. It is worth noting that Rs6 still changed colour slightly, but as stated before, it was not considered as a positive, and that As4 did not react predictably with *A. salmonicida* subsp. *achromogenes* DNA, which could potentially limit its applicability as a diagnostic or detection test.

Another important factor considered was the minimum concentration of the template that would result in a positive reaction within the 75 minute time limit. Yr8 resulted in positives at 10^{-5} ng μL^{-1}, the lowest concentration tested, without reacting in the negative control, but it was the only one (see Table 5-7); this prompted its selection over Yr7. At this template concentration, Mv6 also had a change in colour after 75 minutes, but it was too faint to be considered as a positive. Assays with both Tm6 and Fp9 reacted at 10^{-3} ng μL^{-1} of template, which was lower than their competitors, and so were preferred to Tm8 and Fp3 respectively.

The final factor considered in the selection of primer sets for further tests was time taken to result. In the best cases assays with certain primers changed colour within 30 minutes, but this only occurred for Fp3, Mv6 and Tm6 at the highest concentrations of the template (10 ng μL^{-1}). Time to positive served as a tie-breaker between Fb3 and Fb7 and Mv5 and Mv6; those primers where assays took less time at higher target concentrations were selected. This led to Fb7 and Mv6 being used in the next tests.

Table 5-7: Time to positive in preliminary experiments, given in minutes. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive).

DNA	Primer set	DNA concentration (ng μL^{-1})				
		10	10^{-1}	10^{-3}	10^{-5}	0
Aeromonas salmonicida achromogenes	As4	-	-	60	-	-
	As18	45	60	60	60	60
Aeromonas salmonicida salmonicida	As4	60	75	-	-	-
	As18	45	45	45	30	45
Flavobacterium branchiophilum	Fb3	45	60	60	-	-
	Fb7	45	45	75	-	-
Flavobacterium psychrophilum	Fp3	30	45	-	-	-
	Fp9	45	45	60	-	-
Moritella viscosa	Mv5	45	45	45	(75)	-
	Mv6	30	45	45	-	-
Renibacterium salmoninarum	Rs6	45	45	(75)	75	(75)
	Rs7	45	45	45	45	45
Tenacibaculum maritimum	Tm6	30	45	60	-	-
	Tm8	60	-	-	-	-
Yersinia ruckeri	Yr7	45	-	-	-	-
	Yr8	45	45	60	60	-

5.2.6. Optimisation tests – exploratory phase

5.2.6.1. Methods and rationale

The reaction times observed during the preliminary tests were similar to those observed by other researchers for bacterial targets; Kulkarni and colleagues (2009a, 2009b) 45 minutes, Zhao and colleagues (Zhao et al., 2010), 45 minutes, etc. However, the manufacturer recommends running the reactions for a default of 30 minutes, and an additional ten if results are not entirely clear (NEB, n.d.). It was considered, therefore, that a reduction in reaction times was plausible. Additionally, the possibility of reducing the limits of detection was also intriguing, as it would make for a more practical tool.

Two tests were run for each primer set selected after the preliminary tests. In one, the same assay as in the preliminary tests was carried out, but at five different temperature points between 62.7 °C and 67.2 °C. In the other, three different primer concentrations (1X, 1.3X and 1.6X) were tried out whilst all other variables remained the same. All of these were carried out at two template concentrations; one just above the tentative limit of detection established in the preliminary tests, and another just under it (see Table 5-8). The criteria to decide whether there was an improvement in the assay function were the following: first, a

shorter time to positive as compared to the original conditions; second, a positive reaction with the lower concentration template. Besides these test assays, a positive control (10 ng μL^{-1} of target DNA, 1 ng μL^{-1} in the case of *Renibacterium salmoninarum*) and a negative control (no DNA), were run for every set of triplciates

Once set up, visual inspection was carried out after 30 minutes and every 15 minutes thereafter, until 75 minutes had passed, and changes in colour were logged. As before, negative reactions remain pink, positives turn yellow; unclear results were also recorded.

5.2.6.2. Results

The exploratory tests of different temperature and primer concentration levels for the LAMP assays often did not result in any positive reactions within the allotted time window of 75 minutes. This led to assays being run for longer, up to 105 minutes, in order to observe later changes. The exploratory analysis of assay conditions returned apparently random results in some cases, albeit this may be due to the lack of replication. In the case of temperature, most reactions proceeded as in the preliminary tests, with reactions taking place in assays that contained a concentration of template above the limit of detection (see Table 5-8). The assays with As4 and Mv6 were an exception, as there was no visible colour change within the first 75 minutes in the higher template concentration tested. In the case of Mv6, only faint changes were visible after 105 minutes at both the highest and the lowest temperature tested (67.2 °C and 62.7 °C respectively), and in the case of As4, a clear change was visible for some assays with lower concentrations of template after 105 minutes. In the case of Fp9 and Yr8, departing from the default 65 °C resulted in no reduction or an increase in the time to positive. In the case of Fb7, an assay temperature of 63.7 °C was followed by a shorter time to positive in comparison to the default. For Rs6 the assay with lower template concentration changed colour within 90 minutes, showing a slight colour change by 75 minutes when run at 66 °C. For Tm6 there was no reduction in the time taken to get a clear colour change, but at 62.7 °C a faint change was observed after 60 minutes.

It was found that an increase in primer concentration, was followed by a reduction in time to positive and a positive at lower concentrations of the template for Rs6 (see Table 5-9). In several cases it was found that there was no reaction even at the higher template concentration before the 75 minute expected completion time point, and when there is a change in colour it is only faint (As4, Fb7, Fp9 and Mv6). With Yr8 a colour change occurs after 75 minutes at a primer concentration of 1.3X, but no reaction at the default conditions

and the higher template concentration. Finally, in the case of Tm6 it could be seen that although at default conditions no colour change took place, at a 1.3X primer concentration a slight colour change was noticeable by 75 minutes, and a clear colour change was registered after 90 minutes.

Ultimately, it was decided to test the primer sets Rs6 and Tm6 at a combination of conditions identified through these exploratory experiments. Assays with Rs6 compared the default conditions (65 °C, 1X primer concentration) with those identified here, 66 °C and 1.6X primer concentration. For Tm6, the conditions tested were 62.7 °C (rounded to 63 °C) and 1.3X primer concentration.

Table 5-8: Time (in minutes) to positive during the exploratory temperature optimisation experiments. "X"s represent unavailable data, "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first slight colour changes were perceptible.

Primer set	Species DNA	Template DNA concentration (ng/µL)	Temperature (°C)				
			62.7	63.7	64.9	66	67.2
As4	*Aeromonas salmonicida achromogenes*	10^{-1}	X	X	(105)	(105)	(105)
		10^{-3}	-	-	-	-	-
As4	*Aeromonas salmonicida salmonicida*	10^{-1}	(105)	(105)	(105)	(105)	(105)
		10^{-3}	105 (90)	-	105	-	-
Fb7	*Flavobacterium branchiophilum*	10-3	-	60 (45)	75	-	75
		10^{-5}	-	-	-	-	-
Fp9	*Flavobacterium psychrophilum*	10^{-3}	75	75 (60)	75 (60)	75 (60)	75
		10^{-5}	-	-	-	-	-
Mv6	*Moritella viscosa*	10^{-3}	-	-	-	-	-
		10^{-5}	(105)	-	-	-	(105)
Rs6	*Renibacterium salmoninarum*	10^{-1}	60	60 (45)	60 (45)	60 (45)	60 (45)
		10^{-3}	-	(90)	-	90 (75)	-
Tm6	*Tenacibaculum maritimum*	10^{-3}	75 (60)	(90)	75		-
		10^{-5}	105 (90)	-	105 (90)	(105)	-
Yr8	*Yersinia ruckeri*	10^{-5}	105 (90)	105	75 (60)	-	-
		10^{-7}	-	-	-	-	-

Table 5-9: Time (in minutes) to positive during the exploratory primer concentration optimisation experiments. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first slight colour changes were perceptible.

Primer set	Species DNA	Template DNA concentration (ng μL^{-1})	Primer concentration		
			1X	1.3X	1.6X
As4	*Aeromonas salmonicida achromogenes*	10^{-1}	(105)	(90)	(105)
		10^{-3}	(90)	(105)	(90)
As4	*Aeromonas salmonicida salmonicida*	10^{-1}	(90)	(105)	(90)
		10^{-3}	(105)	-	(105)
Fb7	*Flavobacterium branchiophilum*	10^{-3}	-	-	(90)
		10^{-5}	-	-	-
Fp9	*Flavobacterium psychrophilum*	10^{-3}	105 (90)	(105)	105 (90)
		10^{-5}	-	-	-
Mv6	*Moritella viscosa*	10^{-3}	-	(105)	-
		10^{-5}	-	-	-
Rs6	*Renibacterium salmoninarum*	10^{-1}	90 (60)	60 (45)	60 (45)
		10^{-3}	-	(105)	90 (75)
Tm6	*Tenacibaculum maritimum*	10^{-3}	-	90 (75)	105 (90)
		10^{-5}	-	90 (75)	90 (75)
Yr8	*Yersinia ruckeri*	10^{-5}	-	75 (60)	-
		10^{-7}	-	-	-

5.2.7. Optimisation tests – assay conditions comparisons

5.2.7.1. Methods and rationale

After exploring the effects of different temperatures and primer concentrations on time to positive and limit of detection, a small number of primer sets were observed to apparently perform better at certain conditions that deviated from the default (Fb7, Rs6, Tm6 and Yr8). To ascertain that these improvements did not occur by chance, another round of experiments was carried out. LAMP assays were run at default conditions and at the new conditions identified as potentially advantageous, this time in triplicate and with positive and negative controls. The positive control had a high concentration of the template DNA (1 ng μL^{-1} for Rs6 and 10 ng μL^{-1} for Tm6) and the negative contained no template DNA (the volume was made up by adding water instead).

To identify improvements in the limits of detection, tests were carried out both at the template concentration above it, and a concentration ten times lower. Assays were set up for primer sets Rs6 and Tm6 at 66 °C and 1.6X primer concentration and 63 °C and 1.3X primer concentration respectively.

5.2.7.2. Results

Assay condition comparisons, where the temperature and primer concentrations identified as potentially advantageous for some primer sets were compared to those recommended by the manufacturer as a default, gave more robust results than the exploratory experiments. For assays with Rs6 (see Table 5-10), the new conditions tested resulted in positives at lower template concentrations, although in one of the replicates it took 90 minutes to get a clear positive. However, the negative control also reacted after 75 minutes.

For Tm6, the negative controls did not react, but neither did the assays with 10^{-3} ng μL^{-1} at the default conditions (see Table 5-11). At the newly identified conditions colour changes did take place, although in both cases it was after extending the time of the reaction beyond 75 minutes, and in one of them the change was too unclear to call a positive. Among the lower concentrations it was possible to see some colour changes after extending the reaction times, but the single assay where this happened among those run with the original conditions was faint, whereas those run with the new conditions showed clear colour changes in two out of three replicates. Again, it was only possible to see this extending reaction times to 105 minutes.

The new conditions were used in the next experiments in the case of Tm6, but not in the case of Rs6. Although the improvement in limit of detection and time to positive was more substantial in the case of Rs6, the early reaction of the negative controls discouraged their use.

Table 5-10: Time (in minutes) to positive during the assay conditions comparison for primer set Rs6. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first colour changes were perceptible. Three replicates were run for each set of conditions, plus negative (no DNA) and positive (very high target DNA concentration) controls.

Conditions	Positive control	High target concentration (10^{-1}ng/µL)			Low target concentration (10^{-3}ng/µL)			Negative control
Original conditions	45	45	45	45	-	75 (60)	-	(90)
"Improved" conditions	45	45	45	45	60	60	90 (75)	75

Table 5-11: Time (in minutes) to positive during the assay conditions comparison for primer set Tm6. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first colour changes were perceptible. Three replicates were run for each set of conditions, plus negative (no DNA) and positive (very high target DNA concentration) controls.

Conditions	Positive control	High target concentration (10^{-3}ng/µL)			Low target concentration (10^{-5}ng/µL)			Negative control
Original conditions	60 (45)	-	-	-	-	(105)	-	-
"Improved" conditions	60 (45)	(105)	75 (60)	105 (90)	-	90 (75)	105 (90)	-

5.2.8. Limit of detection tests

5.2.8.1. Methods and rationale

Once conditions for the assays with primers Rs6 and Tm6 were established, assays were carried out with all the primer sets to find the limits of detection of each set. These tests consisted in assays at different concentrations of the template DNA that ranged from 10 copies μL^{-1} to 10^5 copies μL^{-1}, plus a negative control with no DNA (the DNA solution was substituted by UV-treated HPLC water). Assays at each template concentration were run in triplicate.

LAMP assay conditions were the default as stated by the manufacturer (see above) except for Tm6, which was run at 63 °C and 1.3X primer concentration. As before, reactions were paused to observe results after 30 minutes and then every 15 minutes until 75 minutes had passed.

5.2.8.2. Results

Limit of detection tests were run with set concentrations of the target genomes, given as estimated concentrations ranging from 10^5 to 10 copies μL^{-1} at 10-fold intervals. These tests served to estimate how sensitive, in the analytical sense of the word (and not the diagnostic sense), the assays with the designed primers are. In these tests the limit of detection was considered to be the lowest concentration of template that resulted in a positive reaction (colour change from pink to yellow) in all three replicates. The lowest instance of this limit is at 10^3 copies μL^{-1} for Mv6 and Yr8, although most of the assays reacted at those template concentrations, but only in one or two of the replicates. Assays with Fb7, Fp9 and Rs6 had their limits at 10^4 genome copies μL^{-1}. None of the negative controls in these primer sets' tests reacted, indicating that they were promising candidates for practical assays.

With regards to time to positive, the time taken for these assays to change colour at their limit of detection was always at either the 45 or the 60 minute marks. This suggests that in practical situations assays should be run for 60 minutes and then stopped to establish the presence of their target organism in a sample.

In the case of As4, assay results were not consistent with their expected behaviour. As target concentration decreases, it is expected to see an increase in time to positive and a lower portion of the replicates actually resulting in amplification, with the negative control not reacting at all. For these assays clear colour changes occurred throughout the three replicates only at 10^5 copies μL^{-1}, the highest concentration tested, and only when the template was *A. salmonicida* subsp. *achromogenes* DNA. When the template was *A. salmonicida* subsp. *salmonicida* the colour change was clear in only one of three replicates, with the other two changing colour only slightly. Furthermore, a slight colour change was observed in four of the six negative controls of the assays with As4, with one actual positive. The almost simultaneous change in colour of the negative control assays and those with a higher concentration of the template suggest that As4 primers may be reacting spuriously. The apparent lack of correlation between positive results to template concentration and time to positive and the template concentration is consistent with the preliminary and optimisation experiments, but on this occasion the replication makes it less plausible that this is occurring by chance.

Initial limit of detection tests for As4 were even more unexpected, with reactions occurring at low concentrations of the target (including negative controls) when they did not occur at high ones. The results shown in Table 5-12 are the ones resulting from repeating the experiment with a new, previously unused master mix. Additionally, a fresh primer mix was prepared and filter tips were used to prepare the reaction mix, to avoid any potential source of contamination in this experiment. Furthermore, another experiment was set up to ascertain whether the old master mix could be contaminated with *A. salmonicida* DNA, or the DNA extract used could contain inhibitors. A reaction as described in methods section 5.2.7.2 was set up in triplicate at all the combinations of factors master mix (previously used tube and newly opened tube), and DNA extract (same DNA extract used and older extract), including negative controls: results were similar for both master mixes and DNA extracts in terms of time to positive, which was between the 60 and 75 minute time points. The negative controls mostly remained pink, until one of them changed colour slightly at the 60 minutes time point, and clearly after 75 minutes; this was for the new, presumably uncontaminated master mix.

Table 5-12: Time to positive (in minutes) of each limit-of-detection assay replicate at each concentration of its target DNA. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first colour changes were perceptible.

Primer set	Target	Replicate	Concentration of target DNA (copies per µL)					
			10^5	10^4	10^3	10^2	10	0
As4	*Aeromonas salmonicida achromogenes*	1	60	75 (60)	75	(75)	-	(75)
		2	75 (60)	(75)	-	(75)	(75)	(75)
		3	75 (60)	75	75 (60)	-	(75)	-
As4	*Aeromonas salmonicida salmonicida*	1	75 (60)	(75)	(75)	(75)	-	(75)
		2	(75)	-	-	-	(75)	(75)
		3	(75)	60	(75)	(75)	-	75
Fb7	*Flavobacterium branchiophilum*	1	45 (30)	45 (30)	60	-	-	-
		2	45 (30)	45	45	-	-	-
		3	45 (30)	45	-	-	-	-
Fp9	*Flavobacterium psychrophilum*	1	45	60 (45)	60	-	-	-
		2	45	60 (45)	(75)	-	-	-
		3	45	60 (45)	60 (45)	-	-	-
Mv6	*Moritella viscosa*	1	45 (30)	45	60	60 (45)	-	-
		2	45 (30)	45	60 (45)	45	-	-
		3	45 (30)	45	45	-	-	-
Rs6	*Renibacterium salmoninarum*	1	45	45	-	-	-	-
		2	45	45	-	-	75 (60)	-
		3	45	45	75	75	-	-
Tm6	*Tenacibaculum maritimum*	1	45	45	75	75	75	-
		2	45	45	45	75	60	60
		3	45	45	60	-	(75)	(75)
Yr8	*Yersinia ruckeri*	1	45	60	60	-	-	-
		2	60	60	60	-	-	-
		3	60	60	60	-	-	-

As a result of these outcomes, As4 primers did not proceed to the next tests, as they were deemed to be unreliable.

On the other hand, Tm6, which also seemed to react in the absence of target DNA, reacted sooner at higher concentrations of the template. No negative control reacted before 60 minutes, and by 45 minutes all the replicates the assays with as few as 10^4 copies μL^{-1} of the template had already reacted, besides one at 10^3 copies μL^{-1}. It was therefore decided to proceed with the tests on this primer set.

5.2.9. Cross-reactivity tests

5.2.9.1. Methods and rationale

Finally, each of the seven primer sets selected after the preliminary tests were also tested for cross-reactivity with other non-target pathogens. Due to a limited availability of non-target bacterium DNA, LAMP assays were carried out with each combination of the primer sets and DNA of the BPOIs. This would ensure, at least, that the primers do not react with any non-target organisms among the list of the BPOIs, which represent a number of common bacterial pathogens of farmed salmonids.

Each template-primer set combination assay was carried out in triplicate, with 10^5 copies of template present in each reaction. A negative control was included for each primer set, where sterile UV-treated HPLC water was added instead of the DNA solution. The reaction mix used was the default for all assays except those containing Tm6, where 1.3X primer concentration was used. Assay conditions were the default (65 °C, 75 minutes) except for assays with Tm6 (63 °C, 75 minutes), with pauses at 30 minutes and every 15 minutes thereafter to record reaction status (positive, negative, unclear).

5.2.9.2. Results

The cross reactivity tests that followed the limit-of-detection tests were aimed at finding whether the primers were accurate enough to tell each of the BPOIs apart. This would be an essential feature of any product developed as a consequence of this project. The primer sets that reached this stage of the tests were Fb7, Fp9, Mv6, Rs6, Tm6 and Yr8. In all cases the colour change expected from the reaction between the primers and their respective target templates occurred within an hour, although time to reaction was expected to be less in the case of Fp9 (see Table 5-13). The time it took for an Fp9 assay with 10^5 copies μL^{-1} of template to change colour during the limit of detection test was 45 minutes, while in this test it was 60.

In the case of Tm6, besides the expected positive reactions, which occurred within 45 minutes of experiment commencement, a number of assays that contained non-target DNA changed colour as well. One of the negative controls became a clear yellow, and two others slightly orange after 75 minutes. Much like during the limit of detection tests, it appears that assays with Tm6 can only be run for 45 minutes to remain reliable.

Table 5-13: Time to positive (in minutes) of each cross-reactivity assay replicate against any non-target DNA. "-" represent assays that did not react, bracketed times represent only faint colour change from pink (negative) to yellow (positive). When two figures share a cell, the bracketed number represents the time until the first colour changes were perceptible.

	Species DNA	Replicate	Fb7	Fp9	Mv6	Rs6	Tm6	Yr8
Template DNA, at 10^5 copies/µL	*Aeromonas salmonicida salmonicida*	1	-	-	-	-	-	-
		2	-	-	-	-	75	-
		3	-	-	-	-	60	-
	Flavobacterium branchiophilum	1	45 (30)	-	-	-	(75)	-
		2	45 (30)	-	-	-	(75)	-
		3	45	-	-	-	-	-
	Flavobacterium psychrophilum	1	-	60 (45)	-	-	(75)	-
		2	-	60 (45)	-	-	60	-
		3	-	60 (45)	-	-	(75)	-
	Moritella viscosa	1	-	-	45	-	(75)	-
		2	-	-	45	-	(75)	-
		3	-	-	45	-	-	-
	Renibacterium salmoninarum	1	-	-	-	45	(75)	-
		2	-	-	-	45	75 (60)	-
		3	-	-	-	45	75	-
	Tenacibaculum maritimum	1	-	-	-	-	45	-
		2	-	-	-	-	45	-
		3	-	-	-	-	45	-
	Yersinia ruckeri	1	-	-	-	-	-	45
		2	-	(75)	-	-	-	45
		3	-	-	-	-	-	45
	Negative control	1	-	(75)	-	-	(75)	-
		2	-	-	-	-	(75)	-
		3	-	-	-	-	75	-

5.2.10. Improved visualisation experiments

5.2.10.1. Methods and rationale

It has been reported that adding SYBR safe, a cyanine die that intercalates double-stranded DNA, to LAMP reactions allows for an easy visualisation of results (Kuan et al., 2010, Carter et al., 2017), although in pilot experiments it was found to lead to background fluorescence. Thus, the following were tested: Whether addition of SYBR safe to LAMP assays improves assay readability; and, what concentration of SYBR safe would lead to a fluorescent signal that was clearly differentiated from the background fluorescence. To test this, SYBR safe was added to LAMP assays at different concentrations and compared against a negative control without template.

SYBR safe dye (Invitrogen/Thermo Fisher Scientific) was diluted from its original concentration (10,000X) to dilutions of 100X, 50X, 40X, 30X, 20X and 10X in UV-treated HPLC water. LAMP assay reactions were then prepared with a recipe adapted from the master mix manufacturer's protocol to accommodate 1 µL of SYBR safe solution within the 10 µL of the assay. The *Flavobacterium branchiophilum* DNA template and Fb7 primer set were selected because of reliable amplification seen during the previous limit of detection and cross-reactivity experiments. Thus, a common mix was prepared containing the volumes of each reagent per assay as seen in Table 5-14.

A total volume of 189 µL reaction mix was prepared (seven test concentrations of SYBR safe x three replicates x 9 µL per assay) and 9 µL were dispensed into each of 21 0.2 mL polypropylene reaction tubes (PCR tubes, Starlab). 1 µL of the each of the SYBR safe solutions was added to three of the tubes, plus negative controls containing no SYBR safe, which received 1 µL of UV-treated HPLC water instead. This made the final concentrations of SYBR safe 10X, 5X, 4X, 3X, 2X, 1X and 0X.

Additionally, to compare background fluorescence in situations where no amplification took place, another set of 21 reactions were carried out in the same conditions, differing only in the absence of template DNA. Thus, their reaction mix was prepared with the volumes per assay noted in Table 5-14, substituting the DNA for sterile UV-treated HPLC water.

Table 5-14: Reagents and quantities used in the preparation of the common LAMP reaction mix for the improved visualisation experiments. Sterile UV-treated HPLC water was used in the negative controls.

Ingredient	Volume in test assays (µL)	Volume in control assays (µL)
Master mix	5	5
UV-treated HPLC water	2.6	3
Fb7 primer mix	1	1
DNA (*Flavobacterium branchiophilum* at 10^5 genome copies μL^{-1}) or UV-treated HPLC water	0.4	0
Total	9	9

A total volume of 189 µL was prepared, and 9µL dispensed into each of 21 0.2 mL polypropylene reaction tubes (PCR tubes). Mirroring the reactions set up with the template DNA, 1 µL of the corresponding SYBR safe dilution was added to each tube in triplicate, plus one set of triplicates that only had UV-treated HPLC water added.

Assays were then run at 65 °C for 60 minutes in a G-Storm GS0001 thermal cycler, followed by another round of 15 minutes where images were taken under white light, in a blue light transilluminator and in a GelDoc XR+ gel documentation system (Bio-Rad Laboratories Inc., Watford, UK).

5.2.10.2. Results

SYBR safe-LAMP tests were run for 60 minutes, then imaged, because this was expected to be enough time to observe a change in colour of the reaction mix. After 60 minutes the colorimetric master mix's indicator had changed colour in most of the reactions containing template, indicating positive reactions in all of the reactions containing SYBR safe and DNA; only the assays with no SYBR safe remained pink (Figure 5-5). After an additional 15 minutes at 65 °C these became yellow too. This corroborates previous reports that SYBR safe does not inhibit LAMP reactions (Carter et al., 2017).

When observed under a transilluminator's blue light and through an amber filter, it was possible to see the fluorescence emitted by the SYBR safe as a golden yellow colour (Figure 5-5). This was a brighter shade of yellow in the assays with higher concentrations of SYBR, as expected, with the 10X emitting the strongest fluorescent signal and the 0X emitting none. The difference with the no-template controls was noticeable at even the lowest concentrations of SYBR safe because the background fluorescence of SYBR safe is minimal. The signal to

background difference was greater, however, at higher concentrations of SYBR, making it easier to distinguish positives from negatives.

When viewing the results with a UV transilluminator the same pattern of increasing fluorescence with increasing SYBR safe concentrations is observed. However, in this case the background fluorescence coming from the no-template controls is more noticeable, and at the higher concentrations it is stronger than the actual signal from amplification in the assays with lower concentrations of SYBR safe (Figure 5-5). The subjective nature of interpretation of visual signals means different observers may have different preferences, but initially a final concentration of 3X SYBR safe seems to offer a clear distinction between negative and positive LAMP reaction results.

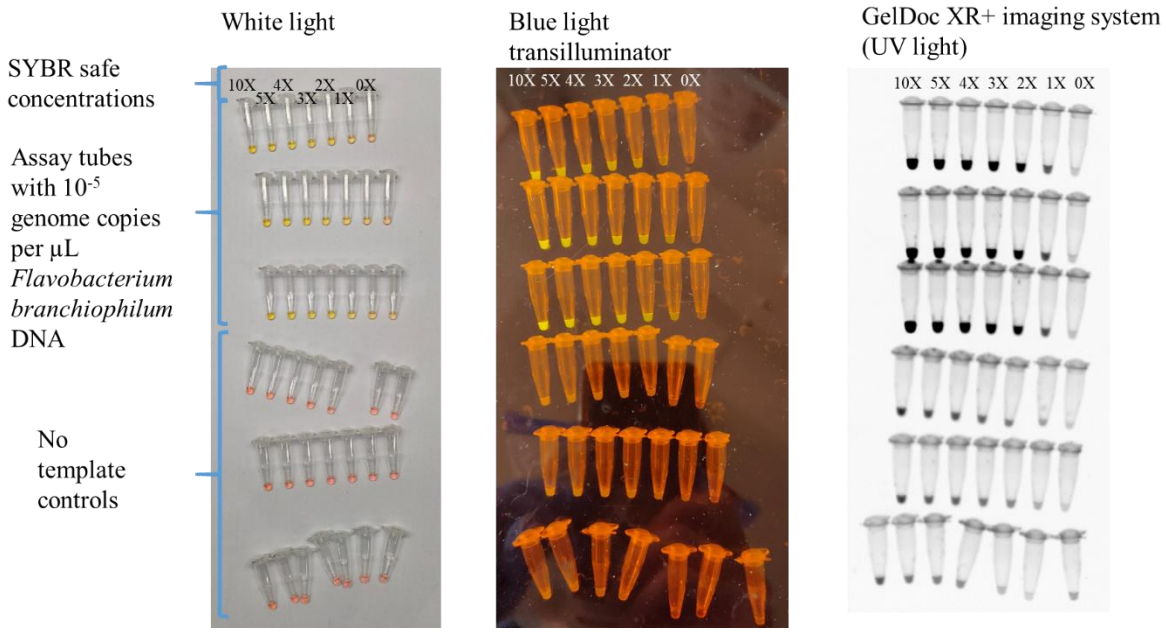


Figure 5-5: Appearance of LAMP assays with added SYBR safe visualised under different types of illumination.

5.3. Discussion

5.3.1. Overview

During this chapter LAMP primer sets were designed for seven major bacterial pathogens of salmonids (objective 1.1); a reduced number of the 80 designs, which were determined to be the least likely to react with non-target DNA, were synthesised (objective 1.2). These were then tested on DNA extracts from their targets and their limits of detection and cross-reactivity were evaluated (objectives 1.3 and 1.4). Results show that the primers designed here reacted against as few as 1,000 copies μL^{-1} of their template in the case of Mv6 and Yr8, or 400 copies per assay, and 10,000 copies μL^{-1} in the case of Fb7, Fp9, Rs6 and Tm6, or 4,000 copies per assay. The exception were the sets of primers designed to amplify *A. salmonicida* DNA, for which assays reacted spuriously, giving rise to false positives. The assays with the Tm6 primer set were modified to deviate from standard conditions based on optimisation experiments to run at 63 °C and 1.3X primer concentration (objective 1.5).

5.3.2. LAMP limits of detection

One of the advantages of LAMP is that it is sensitive to smaller amounts of its template than PCR. Although different publications report limits of detection in different ways (ie., mass per unit volume, mass per assay tube, number of template copies per unit volume, colony forming units (CFUs) per unit volume, etc.) it is usual to see limits of detection to be 10 or 100 times lower than those of equivalent conventional PCRs (Han et al., 2011; Kulkarni et al., 2009a; Saleh et al., 2008a; Savan et al., 2004; Xie et al., 2012). In some cases the improvement is even greater; Kulkarni and colleagues (2009b) reported a limit of detection 10^4 times lower for their LAMP assay against *Vibrio anguillarum* than the equivalent PCR assay.

Among the publications where mass per unit volume is reported, there is a range of values usually between femtograms per assay tube to picograms per assay tube: Saleh and colleagues (2008a) reported detection of 1 pg per assay tube when targeting the gene encoding protein p57 of *R. salmoninarum*; Zhao and colleagues (2010) report 100 fg per assay tube for their probes against the *tlh* and *tdh* genes of *Vibrio parahaemolyticus* and 1 pg per tube for the *trh* gene of *Vibrio parahaemolyticus*; and Han and colleagues (2011) reported they could detect 100 fg per assay tube of *Streptococcus iniae* DNA targeting the *pgm* gene. Others report mass per unit volume; for example, Kulkarni and colleagues (2009a) reported detection of 1 fg μL^{-1} when they targeted the *gyr*B gene of *A. salmonicida*. Finally, some, like

Xie and colleagues (2012) report detection in CFUs; in their case of ten CFUs of *Edwardsiella tarda* (presumably per assay) when targeting the *EsrB* gene.

This chapter's assays are capable of detecting an estimated 10^4 to 10^3 template copies μL^{-1} in the cases of Fb7, Fp9, Rs6 and Tm6, and Mv6 and Yr6 respectively; the equivalents in terms of mass can be seen in Table 5-15. Alongside the limit of detection of the assays developed in this chapter, Table 5-15 shows the limits of detection in published studies. This chapter's assays do not compare favourably in the cases where another set of primers already exists, as they are one to three orders of magnitude above those limits. Comparisons were more difficult with PCR assays, as studies of PCR assay detection limits for these organisms normally reported in different units; in the case, for example, of Gibello and colleagues (1999) they talk about CFUs per gram of infected tissue detected, similarly in Brown and colleagues (1994).

Table 5-15: Limits of detection determined for the primers designed in this chapter and comparison against those of LAMP assays against the same target as seen in the literature. Where possible, Limits of Detection (LOD) have been translated to pg μL^{-1}.

Target	LOD in template copies μL^{-1}	LOD in pg μL^{-1} (rounded to nearest whole pg)	LOD in literature	Publication
Aeromonas salmonicida	NA	NA	0.001 pg μL^{-1}	Kulkarni et al., 2009a
Flavobacterium branchiophilum	10^4	36.5		
Flavobacterium psychrophilum	10^4	35	20 copies μL^{-1}	Fujiwara-Nagata and Eguchi, 2009
Moritella viscosa	10^3	5.22		
Renibacterium salmoninarum	10^4	32.4	0.5 pg μL^{-1}	Saleh et al., 2008a
Tenacibaculum maritimum	10^4	34.7		
Yersinia ruckeri	10^3	3.87	0.67 pg μL^{-1}	Saleh et al., 2008b

Regardless of the existence of more sensitive LAMP assays for some of the BPOIs, the work carried out in this chapter has produced three sets of primers which target pathogens for which, to my knowledge, no LAMP primers existed previously. One of them is *Moritella viscosa*, for which this chapter has produced one of the more sensitive primer sets; this may be one of the most promising for the development of commercially applicable tests. Of the other two targets, *Tenacibaculum maritimum* and *Y. ruckeri*, tests with the former using

primer set Tm6 have given somewhat ambiguous results. The fact that these assays will reliably take longer than 45 minutes to react against non-target DNA, or in a spurious manner would need to be corroborated by further experiments before any practical application of this assay is sought. However, it might be more appropriate to try out another set from among the ten designed by GLAPD, or directly design a new set, given that primer design is one of the main factors affecting assay results (Lucigen Corporation, 2017; Tomita et al., 2008), and a greater margin between real positives and false positives would increase the confidence in the test.

It is hard to explain why the probes designed in this chapter have such high limits of detection compared to published LAMP assays. Other studies that use the same master mix (NEB WarmStart Colorimetric 2X LAMP master mix) report limits of detection that are in line with those mentioned for bacterial pathogens of fish; for example Waliullah and colleagues (2019) mention being able to detect 88 copies per assay tube of *Xyllela fastidiosa*; and Lalli and colleagues (2021) state they detect ~59 copies per assay tube of SARS-CoV-2. On the other hand, some studies mention values that are much closer to the lower detection limits of this chapter's, such as Stehlíková and colleagues' (2020), which states they can detect 1 pg μL^{-1} of *Xanthomonas gardneri*, or Sagcan and Turgut Kara's (2019), where they detect 1,360 copies or 10 pg μL^{-1} of template. One possible explanation of this is the fact that a standardised commercial kit is used in this chapter, which may be not be optimised for the specific assays carried out. In this chapter's experiments temperature and primer concentration, two major factors affecting assay results, are changed, but other relevant ones, such as the concentrations of $MgSO_4$ or enzyme, remain the same. If these assays are to be used in the future, it might be worth attempting to prepare master mixes from scratch, and testing variations on the components, as well as more thorough tests on primer concentration and temperature. To clarify, this chapter's tests did not test for any interaction between the two, but rather assumed any improvements from changing either of these parameters would be additive; more thorough tests could look at the interaction between these parameters.

Another possible explanation is stochastic factors having a more pronounced effect when using small volumes of sample, as 0.4 µL were used instead of the more common 1 µL, or 2 µL (Fujiwara-Nagata & Eguchi, 2009; Kulkarni et al., 2009a; NEB, n.d.; Saleh et al., 2008a). In fact, some studies use even greater volumes (5 µL, Sagcan & Turgut Kara, 2019; 3 µL, Stehlíková et al., 2020). Despite homogenising DNA solutions used as template in the reaction mixes by vortexing, it may be that the reduced sample size increased the chances of

getting DNA concentrations that deviated from the mean. In fact, amplification occurs, although not reliably, of template DNA in the case of Fb7, Fp9 and Mv6 at a concentration ten times smaller than what was established as their limit of detection. This suggests it is possible that upon removing the 0.4 µL of DNA solution from its container this volume contains, by chance, an amount of DNA that goes below the minimum number of copies of template that would actually be necessary for the reaction to proceed. It is also worth noting that there is anecdotal evidence of PCR assays that did not initially result in amplification of their target having reacted upon doubling the assay volume, without altering proportions of ingredients (A. Soares, pers. comm.). Once again, future experiments could attempt to clarify the effect of volume on LAMP results.

Despite their shortcomings, the primers designed in this chapter, especially those targeting bacteria for which there are no existing primer sets yet, have the potential to be developed into practical LAMP tests that can be used for detection of these pathogens in aquaculture settings or the direct diagnosis of the diseases they cause. It has been ascertained that these assays can be run within 60 minutes, which is one of the important advantages of LAMP over PCR, which tends to take longer (Kulkarni et al., 2009a). Additionally, the other main advantage of LAMP also applies: the logistical simplicity means that with a simple heating block it would be possible to run the test; it is possible to find such devices on Amazon for £230 (https://www.amazon.co.uk/FOUR-ES-Temperature-Applications-Interchangeable/dp/B08TMQ4BMJ/ref=sr_1_6?dchild=1&keywords=heating+block&qid=1625692016&sr=8-6), and their operation is straightforward. The use of a colorimetric assay adds a simplicity of result interpretation that can make it accessible to the end user, instead of requiring the intermediation of an expert. However, for this potential to be achieved it is necessary that the assays are tested with real samples; one concern is the observed increase in the limit of detection when applied in realistic scenarios. For example Kulkarni and colleagues found that running a LAMP assay with the primers they developed on bacterial DNA extracted from samples containing fish mucus resulted in a 10 (Kulkarni et al., 2009a) to 1,000 (Kulkarni et al., 2009b) fold increase in the limit of detection. Another concern is the requirement of freezers and DNA extraction for its use, although with regards to the former, LAMP assays based on lyophilised reagents have applied successfully tested (Foo et al., 2017). The question of DNA extraction is further considered in Chapter 4.

5.3.3. Primer design

The design of the primers was accomplished with GLAPD, a software tool that has only recently been released (Jia et al., 2019). This tool had only been used by the creators at the time it was used it to design this chapter's primers; it has since been used by Wang and colleagues (2021) to design probes for the detection of *Chloropillum molybdites*, a toxic fungus found in China. For this chapter's work it was preferred to PrimerExplorer, the Eiken Chemical Company's primer design tool (https://primerexplorer.jp/e/v5_manual/index.html), which is one of the few available to date for a number of reasons:

Primarily it was the conceptual simplicity primer design with GLAPD offered in terms of target selection. Throughout the literature, a number of different genes are targeted in LAMP assays for bacterial pathogens; these are often genes that have been previously targeted by PCR assays (Kulkarni et al., 2009b; Saleh et al., 2008a, 2008b), that encode virulence factors (Han et al., 2011), or ribosomal RNA genes (Yeh et al., 2006). Those targeted by PCR offer a reliable option when there is a pre-existing assay for them, and housekeeping or ribosomal genes are practically guaranteed to be present in the genomes of all strains. GLAPD on the other hand offers a naïve approach that finds regions to target that are conserved across the genomes in the entire dataset supplied to it. This means little previous knowledge about the pathogen is required to design primers for it; it can find regions conserved across the genomes of all the strains of a target species and the more data there is available means the more reliable it will be. This allows for the exploitation of public databases, such as NCBI's RefSeq, which hold more and more data every day. Additionally, the fact that the software is accessed through the command line makes it more amenable to automation and high-throughput workflows. Finally, it is open source, with the entire source code available on GitHub (https://github.com/jiqingxiaoxi/GLAPD), which makes it possible for anyone with the technical knowledge to understand how the software works and even improve on it. Among other benefits, these features have the potential to make LAMP primer design more accessible to the general public.

However, there are some clear drawbacks to the method employed. GLAPD is relatively fast for short sequences and when used with small numbers of common targets and/or background sequences, it can perform competitively. The creators of GLAPD, however, state their intended use is with genome data; although this chapter's design process corroborates that this is entirely possible, as has been demonstrated for seven different bacteria, it also makes those drawbacks clear. The resources required to run GLAPD are substantial, with large

amounts of memory being taken up, and a single core being engaged for hours on end, as unless the first and third steps of the program are run on a GPU they are not parallelisable. Its operation is also less straightforward than that of PrimerExplorer's last iteration (version 5), and the long running time makes correcting errors a long-term iterative process.

Given these points, it is worth considering using GLAPD with a more targeted approach in future projects. Since GLAPD can be used for primer design in smaller sequences much faster, it could potentially be used for the design of primers that targeted, much like other previous studies, genes that have already been targeted by PCR, housekeeping genes, or ribosomal genes (the latter option is explored in Chapter 6).

5.3.4. Visualisation

Throughout this chapter it has been possible to ascertain that colorimetric LAMP results are simple to understand, although not always straightforward to interpret. The intermediate colour between pink and yellow allows for a degree of subjectivity, which could be an impediment to its application as a diagnostic. Furthermore, the colour-based nature of result readouts can render this tool unusable to colour-blind people. In an attempt to improve result readability, SYBR safe was added to the reactions.

SYBR safe is a cyanine dye which is commonly used to stain gels containing DNA during electrophoresis. Under UV light, and when intercalating double stranded DNA, SYBR safe emits fluorescence, which allows visualising and locating the former. However, SYBR safe can also be added directly to a tube with LAMP DNA amplicons, which makes it easier to visualise the product (Kuan et al., 2010; Kulkarni et al., 2009b). Moreover, it is possible to add SYBR safe to LAMP assays during the setup stage without inhibiting the reaction (Carter et al., 2017). However, personal experience suggests this can result in background fluorescence that may make judgement more difficult. The results of the SYBR-safe experiments suggest that it is possible to clearly distinguish the fluorescence from a positive reaction from the background fluorescence of a negative one at SYBR safe concentrations of approximately 3X. Additionally, the distinction seems to be clearer with a blue-light transilluminator with an amber filter than in the gel documentation system, indicating a simple, relatively inexpensive piece of equipment could serve to read results.

Regardless, the exact ideal proportion of SYBR safe could be subject to the individual's perception. The subjectivity also extends to the improvement in comparison to the colorimetric assay. Future work could involve setting up a larger number of LAMP reactions

with different concentrations of the target, imaging the results at different time points, both under white light and blue light. A panel of observers could then be asked to classify the reactions as positive or negative; this would allow to see which type of visualisation technique is more consistent, and how they compare to each other. Currently, however, there is anecdotal evidence of a colour-blind observer being able to distinguish the fluorescent assay, as opposed to the colour assay (F. Martín, pers. comm.).

5.3.5. Conclusion

LAMP primer sets have been designed for six of the seven BPOIs; three of these sets target species for which there are currently no published LAMP primers. These colorimetric assays have the potential to improve detection of the BPOIs in aquaculture, especially of *F. branchiophilum*, *M. viscosa* and *R. salmoninarum*, but would require further testing and optimisation before practical application. One of their advantages is the relatively short time to result, and the simplicity of the readout; the interpretation of the latter can be facilitated by the addition of SYBR safe fluorescent dye. This is all possible with moderate equipment requisites.

Chapter 6: Broad-range LAMP for simultaneous multiple pathogen detection

6.1. Introduction

LAMP, invented in the year 2000 by Notomi and colleagues, is a relatively modern nucleic acid amplification technique (NAAT) (Notomi et al., 2000) (see Chapter 5). PCR, on the other hand, first published in 1986 by Mullis and colleagues, has had 14 years longer to gain popular acceptance and become a commonplace tool in molecular biology labs (Mullis et al., 1986). Both time and its characteristics, have allowed for the development of a number of derived techniques and technologies. One of these is the discovery and application of universal primers. Examples are those that target mitochondrial cytochrome oxidase c I (COI, Folmer et al., 1994), the Internal Transcribed Spacer (ITS, Schoch et al., 2012), or 16S rRNA genes (Weisburg et al., 1991), used because of the presence of these genes throughout numerous clades of the Metazoa, the fungi and bacteria respectively. Their purpose has often been "barcoding", this is, the use of the unique sequences of the amplified regions associated to each organism targeted by the primers to assign them a taxonomic label (Hebert et al., 2003).

The rise of second and third generation sequencing technologies, which reduce costs and make it possible to sequence numerous DNA molecules simultaneously, have made it more practical to apply the barcoding technique to community samples. This has led to the appearance and popularisation of "metabarcoding", which consists in the sequencing of these barcode regions of the DNA present in a mixed sample, usually environmental (Pollock et al., 2018; Taberlet et al., 2012) An advantage of this technique is the ability to study microbial communities even in low biomass samples, but it comes at a cost. It has a number of disadvantages, which include the skew of biological diversity introduced through the DNA preparation and sequencing process, which can result in some taxa within a sample being underrepresented or ignored (Brooks et al., 2015; Pollock et al., 2018; Villette et al., 2021).

LAMP, on the other hand, has not been used for these kinds of studies. This is perhaps a consequence of it being more modern, perhaps a consequence of the technical limitations deriving from the way it works, or a combination of both. Instead, LAMP studies have targeted individual organisms and demonstrated their value as diagnostic tools (for a recent example, see Dao Thi et al., 2020); they are especially useful in the detection of concrete

species because the use of six primer regions makes it less likely to amplify non-target organism DNA by chance (Notomi et al., 2000). Several studies have, however, aimed to expand LAMP's capacities to detect several organisms simultaneously. Instances of multiplex LAMP, where a different restriction site is inserted into the amplicons of different organisms through a modification in the primers, have existed since 2007 (Iseki et al., 2007). More recently, a number of studies use differential fluorescent labelling of specific probes, combined with fluorometric thermocyclers, to run assays for a number of different organisms in the same tube (J. Kim et al., 2021; Tanner et al., 2012; Y. Wang et al., 2015). The most recent development, which is to combine LAMP with Nanopore sequencing, however, has also served to discern the amplification of several different targets within a sample.

Specifically, LamPORE, developed by James and colleagues (2020), has demonstrated that nanopore sequencing can be used to measure the amount of DNA of each of several targets within the genome of SARS-CoV-2, plus a human actin gene internal standard, that is amplified in a multiplex LAMP reaction. Another approach at detecting multiple different organisms is that of Imai and colleagues (2017), who designed a primer set that targeted a region of the 18S rRNA gene of several *Plasmodium* species. They then distinguished each of the five species of *Plasmodium*, plus the two subspecies of *Plasmodium falciparum* by sequencing the amplicons and comparing them with the known sequence of this 18S rRNA gene region for each species. A similar study had been carried out by Yamagishi and colleagues (2017), for Dengue virus. In this case a region of the 3' non-coding region of the Dengue virus genome was amplified, which encompassed a number of mutations that enabled the distinction of several serotypes upon sequencing with a nanopore sequencer. Later, Hayashida and colleagues (2019) followed with a similar approach that targeted the E1 gene of Chikungunya virus; they could also distinguish several genotypes based on the genotype-specific mutations of the region.

The possibility of detecting a number of pathogens simultaneously in a single test is highly advantageous. Compared to single-target assays (such as those developed in Chapter 5) it could reduce economic and time costs, as the number of tests carried out would decrease, making it more practical for pre-emptive monitoring of fish farms. The combination of LAMP, a tool simpler and more accessible than PCR, and MinION sequencing, which is portable and also more accessible than other sequencing methods, offers a unique opportunity to do this. Additionally, it has the potential to resolve the concerns raised in Chapter 5 about low yield of DNA from water, as LAMP can amplify small amounts of DNA to meet the

requirements of nanopore sequencers. To explore the possibility of combining these two technologies, this chapter aims to develop a proof of concept of such a system. The proposed system is summarised in Figure 6-1.

6.1.1. Aims and objectives:

Aim:

1) To develop the proof of concept of a system that combines LAMP and Nanopore sequencing to detect and distinguish a broad range of bacterial pathogens of farmed fish.

This aim is divided into three objectives:

1) To design a LAMP primer set that is valid for a broad range of bacterial pathogens of fish (referred to as “broad-range LAMP” throught the chapter).
2) To test this primer set, ensuring it reacts with its targets.
3) To establish a means of assigning the sequences of amplicons to their correct species of origin.

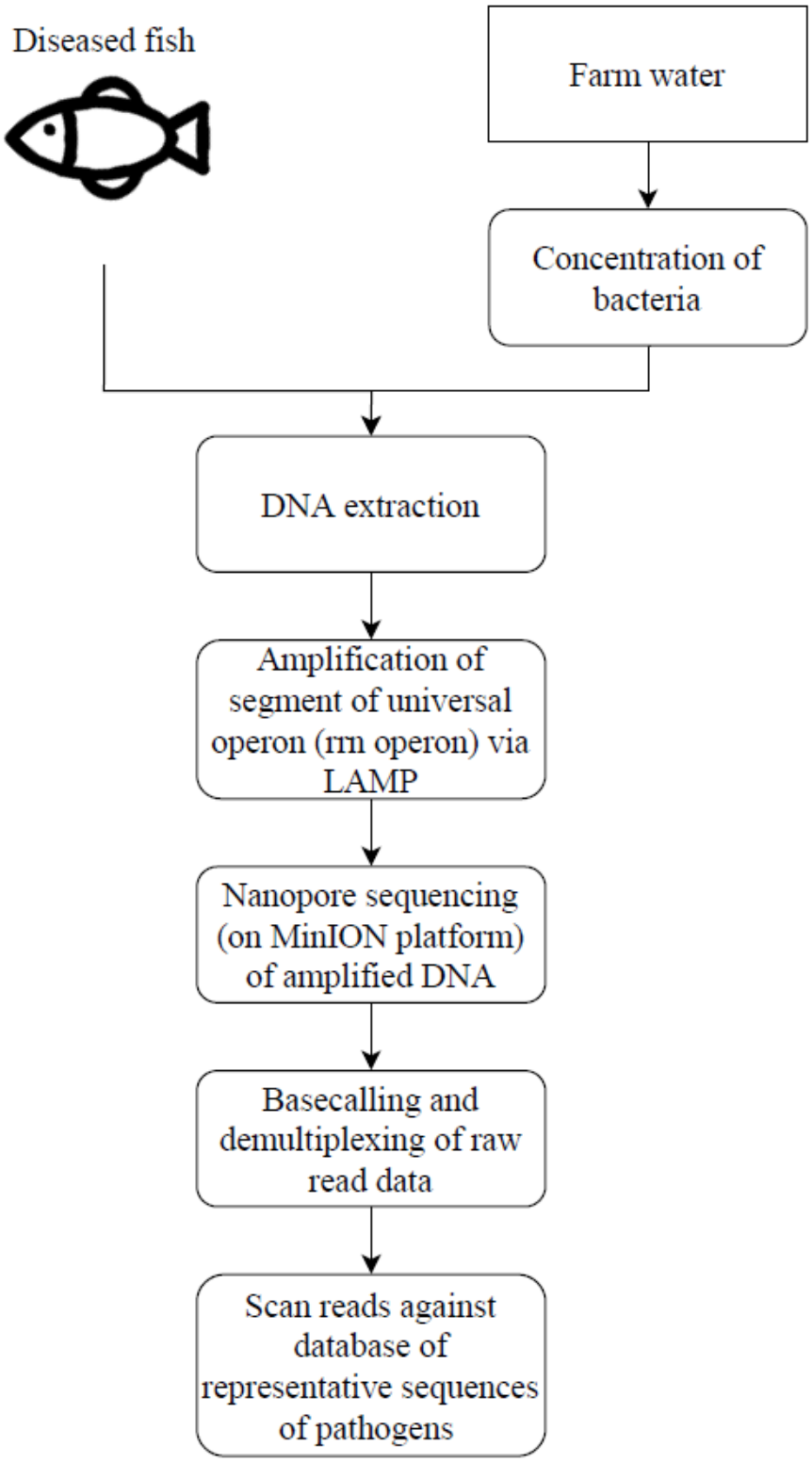


Figure 6-1: The proposed procedure would start by obtaining a DNA sample from either a swab from a diseased fish, or a bacterial concentrate from farm waters; the means of obtaining the latter is explored in Chapter 4. It would continue by using a LAMP assay to amplify a region of the rrn operon of bacteria, which comprises the genes coding for the ribosomal RNA and is thus universally present in bacteria. The amplified DNA would then be sequenced using a MinION device, and the sequence data assigned to the organism of origin by comparing to the equivalent sequences of a number of different pathogenic bacteria included in a database. Diagram created with app.diagrams.net.

6.2. Method development

6.2.1. Exploratory data gathering

6.2.1.1. Methods and rationale

From the start, the bacterial ribosomal RNA operon was considered an appropriate target, given that it is present in all bacteria, and the 16S subunit, and less frequently the 23S subunit, are commonly used in ecological studies. The focus of this chapter was originally going to be a set of 29 bacteria provided by the company partner which encompass a range of pathogens of fish and shellfish that Fish Disease Dolau Ltd. considered to be especially relevant in aquaculture (see Table 6-1). To obtain and analyse the *rrn* operon sequences of these bacteria, their genomes were downloaded from the NCBI RefSeq database. The representative genome assembly for each bacterium was searched on the assembly database, the accession retrieved, and the genome sequence in fasta format downloaded via the NCBI FTP pages.

The ribosomal RNA genes of the downloaded genomes were then annotated using Barrnap v0.9 (Seemann, 2018, https://github.com/tseemann/barrnap), and a set of custom-made Bash and R scripts was used to sort these genes into operons and save them in a separate fasta file. These scripts found any consecutive set of rRNA genes in the order 16S, 23S and 5S from 5' to 3' on the same strand of the bacterial genome; this was considered to be an *rrn* operon. The DNA sequence from the 5' end of the 16S gene to the 3' end of the 5S gene was then cut and saved in a separate file. This procedure returned *rrn* operon sequences for 24 of the 29 bacteria of interest; upon inspection, it was ascertained that the genome assemblies of the remaining bacteria did not contain any sets of contiguous 16S, 23S and 5S genes on the same strand.

Table 6-1: List of 29 species of bacterial pathogens of aquaculture provided by the company partner, and the accessions of the genome assemblies used in this chapter. The bacteria whose rrn operons were selected as cluster representatives (see section 6.2.1.2) are in bold.

Species	Accession
Aeromonas hydrophila	GCA_000014805
Aeromonas salmonicida	GCA_000196395
Aeromonas sobria	GCA_000820145
Aeromonas bestiarum	GCA_000819745
Edwardsiella ictaluri	GCA_000022885
Edwardsiella tarda	GCA_000020865
Flavobacterium branchiophilum	GCA_000253275
Flavobacterium columnare	GCA_000240075
Flavobacterium psychrophilum	**GCA_000754405**
Francisella noatunensis subsp. orientalis	GCA_000505725
Candidatus *Hepatobacter penaei*	GCA_000742475
Lactococcus garvieae	GCA_000269925
Moritella viscosa	GCA_000953735
Mycobacterium marinum	GCA_000723425
Nocardia seriolae	GCA_001865855
Photobacterium damselae	GCA_002142615
Piscirickettsia salmonis	GCA_000300295
Pseudomonas aeruginosa	GCA_001045685
Pseudomonas fluorescens	**GCA_900475215**
Renibacterium salmoninarum	**GCA_000018885**
Streptococcus iniae	**GCA_000831485**
Tenacibaculum maritimum	GCA_900119795
Vagococcus salmoninarum	GCA_003987495
Vibrio anguillarum	GCA_000217675
Vibrio alginolyticus	GCA_000354175
Vibrio ordalii	GCA_000257205
Vibrio parahaemolyticus	**GCA_001558495**
Aliivibrio salmonicida	GCA_000196495
Yersinia ruckeri	GCA_002192595

The *rrn* operons were combined into a single fasta file and aligned using MAFFT v7.450 (Katoh, 2002; Katoh & Standley, 2013). The alignment algorithm was selected automatically by MAFFT ("--auto" option), but other parameters were specified: the use of nucleic acid sequences (instead of amino acid), the direction adjustment option (as some *rrn* operons were retrieved from minus strands) and the generation of a clustal format output. The clustal format output, which identifies conserved positions within the alignment, was used to produce a plot depicting the most conserved regions and calculate the mean number of identical positions. The plot, produced with R3.6.1 (R Core Team, 2019) and the ggplot2 package (Wickham, 2016), calculated the number of conserved positions at any given point of the alignment within a window of 20 positions; this is the approximate size of a LAMP primer region.

MAFFT was also run to generate a fasta-formatted alignment, which was used to calculate a distance matrix and an unrooted Neighbour-Joining tree using Ninja (Wheeler, 2009), R v3.6.1 (R Core Team, 2019) and the R packages ape (Paradis & Schliep, 2019), phylogram (Wilkinson & Davy, 2018) and dendextend (Galili, 2015).

6.2.1.2. Results

A total of 167 *rrn* operon sequences were retrieved from the genomes of the bacteria in the list provided by the company partner. Of the 29 bacteria in the list (see Table 6-1), it was possible to obtain *rrn* operon sequences based on the contiguity criterion from 24 of them, the exceptions being *Aeromonas bestiarum*, *Aeromonas sobria*, Candidatus *Hepatobacter penaei*, *Vagococcus salmoninarum* and *Vibrio ordalii*. The genomes of these organisms all contained the three ribosomal RNA genes, but they were not placed within the same contig so they were not considered to form an operon. A mean of 6.83 operons was retrieved from each genome, with the maximum being 15 for *Photobacterium damselae* and the minimum two for three species, *Francisella noatunensis* subsp. *orientalis*, *Mycobacterium marinum* and *R. salmoninarum*.

Although the operons were 5,168 ± 124 nucleotides long on average (mean ± SD), the alignment produced with MAFFT was 7,329 positions long, owing to indels between the different sequences. In this alignment, only 29.3% of positions were conserved, ie., had the same nucleotide for all the sequences in the alignment. The plot depicting the number of conserved positions within a sliding window of 20 positions is shown in Figure 6-2. There is substantial variation throughout the alignment, with the intergenic regions showing the lowest

proportion of conserved bases (often zero), and some parts of the ribosomal genes reaching up to 20. However, there were some points in the alignment within the intragenic regions where zero positions out of 20 were conserved, although most positions were in between.

The bacteria included in the analysis encompass five different classes in four different phyla. It is therefore not surprising to see the differences in the alignment of their *rrn* operon sequences. These differences are also reflected in the Neighbour-Joining tree, where the distances calculated by Ninja are of up to 0.307. This dendrogram also makes it clear that there are differences not only between the *rrn* operon sequences of different organisms, but even within a single genome, where copies of the operon are often slightly different to the others (see Figure 6-3). In some occasions, the operons belonging to a single organism form two clusters; they are more similar to another cluster of sequences from a close relative than to the other cluster of operons originating from the same organism. This is the case for *A. salmonicida*, *Aeromonas hydrophila*, and *Edwardsiella ictaluri*.

When the tree is split into clusters where distance is under 0.200, five subtrees result. These clusters are unequal in the number of operons they contain and the number of bacterial species they encompass; 11 different species are grouped in the largest (group 1), only two in the smallest (group 4). However, distances are now reduced in each of them, and the proportion of conserved positions is higher, ranging from 52.8% in group 1 to 83.2% in group 4.

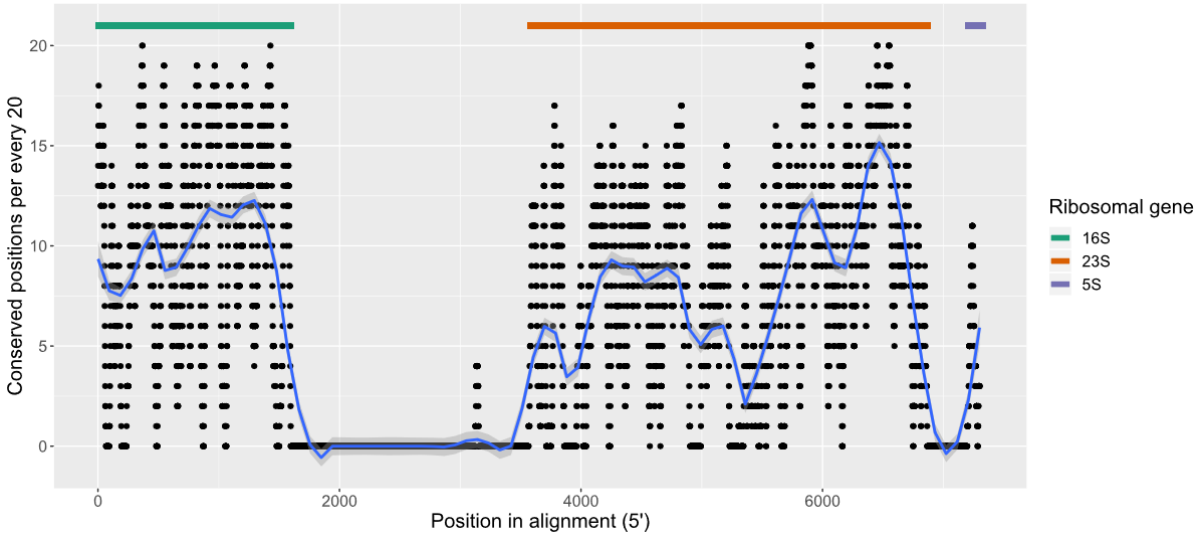


Figure 6-2: Number of conserved positions in the rrn operon alignment of 24 bacteria for a sliding window of 20 positions. The blue curve represents a smoothed mean using a Loess function. The sections of the alignment that correspond to ribosomal genes are highlighted and colour-coded.

6.2.2. Broad-range LAMP primer design

LAMP primer design was carried out with the GLAPD software (Jia et al., 2019); a more complete description of the functioning of GLAPD is provided in Chapter 5, although here I provide a brief recapitulation. GLAPD requires a reference sequence, normally intended to be a genome, but which can be any DNA sequence; an initial list of primers is generated based on it. The next step is to select primers that have complementary regions in as many members as possible among a group of "common" target sequences and not among another group of non-target sequences, although this step is optional. Finally, primers are combined into sets.

Several attempts were made at generating broad-range LAMP primers. The first one aimed to encompass the *rrn* operons of all the 24 bacteria of interest to the company partner; this failed, and a second attempt was carried out dividing the dataset into clusters where rrn sequences were more similar. This was done by cutting the tree into sub-trees at an arbitrary distance of 0.200 (see Figure 6-3). A representative sequence was chosen from these sub-trees to serve as reference in primer design (see bold species names in Table 6-1), and the GLAPD primer design procedure was run again. A third attempt used only the sequences from the seven BPOIs, which failed again, and a fourth used the same sequences, but divided into clusters, as before. These last two attempts were made because the aim was to create a "proof of concept"; a demonstration that it is possible to produce primer sets that target a range of organisms, and to distinguish them through Nanopore sequencing.

For the first attempt, the *rrn* operon sequences of all 24 bacterial species were combined in a single fasta file, with the exception of operon 1 of *Vibrio parahaemolyticus*, which was taken as the reference sequence for the primer design. A Bowtie index was generated for the fasta file containing operon sequences for the 24 species using the Bowtie-build command from Bowtie v0.12.7 (Langmead et al., 2009). GLAPD was then run with *Vibrio parahaemolyticus* operon 1 as the reference and the entire *rrn* operon dataset as "common" targets. When GLAPD was unable to design a single primer set that encompassed all the bacteria, representatives of the other clusters were used as references in successive runs of the program; see Table 6-1. For the attempts that only used the seven BPOIs, the same process was followed, but the set of "common" target sequences was reduced to contain those of the *rrn* operons of only the seven BPOIs. This excluded many of the species of the list of 24 bacteria, including those comprising two of the clusters identified in section 6.2.1, and some of the references used in the second attempt at primer design. Therefore, different reference

sequences were used for the set of seven BPOIs; these were operon 1 of *Aeromonas salmonicida*, *Flavobacterium psychrophilum* and *Renibacterium salmoninarum*.

6.2.2.1. Company partner's list of pathogens

The initial primer design attempt, which targeted the operons of all 24 bacterial species from which it was possible to retrieve sequences, did not result in a LAMP primer set that fitted operons of all bacterial species. Instead, the first primer set was predicted to bind to sequences from 13 different species (see Table 6-2), including all those belonging to cluster 1 and two of cluster 2. When designing primers based on a representative of each cluster, the sets designed for cluster 2, cluster 3, cluster 4 and cluster 5 were predicted to bind to the operons of 13, four, two and two species respectively. After this, only *Francisella noatunensis*, *Piscirickettsia salmonis* and *Mycobacterium marinum* remain without targeting. In the case of cluster 2, the predicted targets were the same as those of cluster 1. The array of species these primer sets would target are detailed in Table 6-2; for the full sequences see Appendix 8.

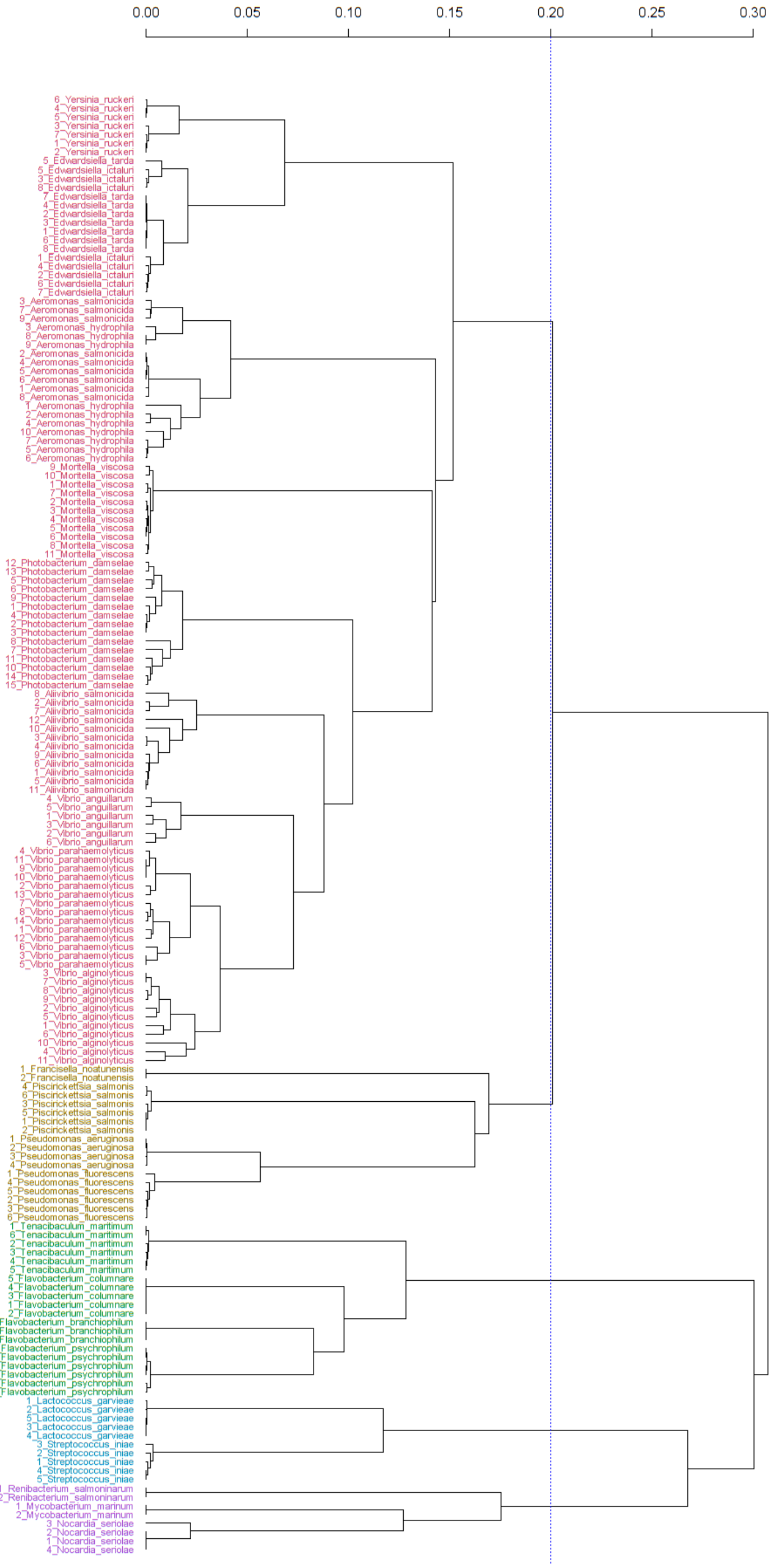


Figure 6-3: Neighbour-Joining dendrogram of rrn operon sequences of 24 species of bacteria included in the company partner's list. The dendrogram has been split into clusters at the arbitrary distance of 0.200, in order to reduce variation when designing LAMP primer sets.

Table 6-2: Primer sets designed for the bacteria of interest to the company partner and the lists of species for which they are predicted to amplify a section of the rrn operon.

Reference sequence for primer design	Target species
Vibrio parahaemolyticus operon 1	*Aeromonas hydrophila* *Aeromonas salmonicida* *Aliivibrio salmonicida* *Edwardsiella ictaluri* *Edwardsiella tarda* *Moritella viscosa* *Photobacterium damselae* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Vibrio alginolyticus* *Vibrio anguillarum* *Vibrio parahaemolyticus* *Yersinia ruckeri*
Pseudomonas fluorescens operon 1	*Aeromonas hydrophila* *Aeromonas salmonicida* *Aliivibrio salmonicida* *Edwardsiella ictaluri* *Edwardsiella tarda* *Moritella viscosa* *Photobacterium damselae* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Vibrio alginolyticus* *Vibrio anguillarum* *Vibrio parahaemolyticus* *Yersinia ruckeri*
Flavobacterium psychrophilum operon 1	*Flavobacterium branchiophilum* *Flavobacterium columnare*

	Flavobacterium psychrophilum *Tenacibaculum maritimum*
Streptococcus iniae operon 1	*Lactococcus garvieae* *Streptococcus iniae*
Renibacterium salmoninarum operon 1	*Nocardia seriolae* *Renibacterium salmoninarum*

6.2.2.2. List of seven BPOIs

The primer design attempt that targeted the seven BPOIs, as the one that targeted the whole set of 24 pathogens, failed initially. This first attempt used *F. pscyhrophilum* operon 1 as reference; the primer set designed was predicted to target rrn operons pertaining to the reference, to *F. branchiophilum* and to *T. maritimum*. The design procedure, repeated with a representative target for each cluster, gave two more primer sets; the first targeted all the BPOIs of cluster 1, *A. salmonicida*, *M. viscosa* and *Y. ruckeri*, and the second targeted the only BPOI in cluster 5, *R. salmoninarum*. The primer sequences are listed in Table 6-3.

Table 6-3: Primer sets designed for the rrn operon of the seven BPOIs.

Primer set	Primer	Sequence (5' - 3')
As1_group	F3	TAGTGATCCGGTGGTTCTGA
	FIP	TCTTGGGCGGTATCAGCCTGTGAAGGGCCATCGCTCAAC
	BIP	TATCGACGGCGGTGTTTGGCAGCCATACCCTTGGGACC
	B3	GCGTACCACTTTAAATGGCG
Fp1_group	F3	TGCAATGGCATAAGGGAGC
	FIP	ATGGCCCTTCCATGCGGAACACAGGTCGATCAGGTACGAA
	BIP	AACAGGCTGATCTCCCCCAAGATGACGAGCCGACATCGAG
	B3	TTCTCCAGCCCCAGGATG
Rs1_group	F3	ATTCTGGCTCAGGATGAACG
	FIP	ACTCACCCGTTCGCCACTAATCAACACATGCAAGTCGAACGA
	BIP	ACGTGAGTAACCTGCCCTTGACTGCGGTGATAGGTCGTATCC
	B3	ACCGCAAAAAACTTTCCACC

6.2.3. Primer testing experiments

6.2.3.1. Methods and rationale

The three LAMP primer sets that were selected for testing were named As1_group, Fp1_group and Rs1_group. As1_group targeted *A. salmonicida*, *Moritella viscosa* and *Yersinia ruckeri*; Fp1_group targeted *F. branchiophilum*, *F. psychrophilum* and *Tenacibaculum maritimum*; Rs1_group targeted *R. salmoninarum* only. They were synthesised by Life Technologies (Paisley, UK) in a 25 nmol scale, desalted, and delivered in a dehydrated format. Once received they were rehydrated to a concentration of 100 µM with UV-treated HPLC water (Fisher Scientific), a 100 µL aliquot taken, and a 10X primer mix prepared for each set. The mix included the outer primers (F3 and B3) to a concentration of 2 µM and the inner primers (FIP and BIP) to a concentration of 16 µM. The primers and primer mix were then stored at -20 °C until use.

6.2.3.1.1. Experiment 1: testing with intended templates

In order to ascertain that these primers reacted against their intended targets and that they did not react spuriously (or if they did, to measure the time this took), LAMP reactions were carried out with the primer mixes. A reaction mix was prepared with each of the reaction mixes as described in Table 6-4, and then divided into 15 0.2 mL PCR tubes (Starlab) for As1_group, 12 tubes for Fp1_group and six tubes for Rs1_group. Each tube contained 9.6 µL of the reaction mix, and 0.4 µL of the corresponding DNA solution were then added (see Table 6-5 for template DNA and primer set combinations). The DNA solutions consisted in the extracts of genomic DNA, of each BPOI, extracted from pure cultures with the PowerSoil kit (Qiagen, Manchester, UK) (see section 3.2.2) normalised to a concentration of 1 ng μL^{-1}.

The LAMP reaction was carried out at 65 °C in a G-Storm GS0001 (G-storm Ltd., Somerton, UK) thermocycler, for a period of 75 minutes. The tubes were removed and replaced after 30, 45, 60 and 75 minutes to record the progress of the reaction, as indicated by the colour change of the colorimetric master mix.

Table 6-4: Reagents and quantities for LAMP reaction mix used in experiment 1.

Ingredient	Per 10 µL reaction	As1_group: 15 x 10 µL reactions	Fp1_group: 12 x 10 µL reactions	Rs1_group: 6 x 10 µL reactions
WarmStart® Colorimetric LAMP 2X Master Mix	5	75	60	30
Water	3.6	54	43.2	21.6
Primer mix	1	15	12	6
DNA	0.4	-	-	-
Total	10	144	115.2	57.6

Table 6-5: Primer and template DNA combinations in experiment 1.

Primer set	DNA added
As1_group	*Aeromonas salmonicida achromogenes*
	Aeromonas salmonicida salmonicida
	Moritella viscosa
	Yersinia ruckeri
	Negative control
Fp1_group	*Flavobacterium branchiophilum*
	Flavobacterium psychrophilum
	Tenacibaculum maritimum
	Negative control
Rs1_group	*Renibacterium salmoninarum*
	Negative control

6.2.3.1.2. Experiment 2: Testing with non-target templates

In order to ascertain that the primer sets designed in section 6.2.1.2 do not react with non-target DNA, and, in case it occurred, to find out the time to reaction, each primer set was tested against all the available non-target DNA extracts. LAMP reaction mixes were therefore set up as in Table 6-6, dispensed into 16, 19 and 25 0.2 mL tubes for the mixes containing primer sets As1_group, Fp1_group and Rs1_group respectively (9.6 µL) and had 0.4 µL of the corresponding DNA added (see Table 6-7 for template DNA and primer set combinations). As before, this was pure genomic DNA of each BPOI obtained with the PowerSoil kit as described in section 3.2.2 and normalised to a concentration of 1 ng μL^{-1}. On this occasion, a positive control verified to react during experiment 1 was included as well, besides a no-DNA negative control. Reactions were carried out at 65 °C for 75 minutes in a G-Storm GS0001 thermocycler (G-storm Ltd., Somerton, UK). Throughout the reaction, the tubes were removed and replaced after 30, 45, 60 and 75 minutes to record the progress of the reaction according to the colour change of master mix.

Table 6-6: Reagents and quantities used in the preparation of LAMP reaction mixes of experiment 2.

Ingredient	Per 10 µL reaction	As1_group: 16 x 10 µL reactions	Fp1_group: 19 x 10 µL reactions	Rs1_group: 25 x 10 µL reactions
WarmStart® Colorimetric LAMP 2X Master Mix	5	80	95	125
Water	3.6	57.6	68.4	90
Primer mix	1	16	19	25
DNA	0.4	-	-	-
Total	10	153.6	182.4	240

Table 6-7: Primer and template DNA combinations in experiment 2.

Primer set	DNA added
As1_group	*Flavobacterium branchiophilum* *Flavobacterium psychrophilum* *Renibacterium salmoninarum* *Tenacibaculum maritimum* Negative control
Fp1_group	*Aeromonas salmonicida achromogenes* *Aeromonas salmonicida salmonicida* *Moritella viscosa* *Renibacterium salmoninarum* *Yersinia ruckeri* Negative control
Rs1_group	*Aeromonas salmonicida achromogenes* *Aeromonas salmonicida salmonicida* *Flavobacterium branchiophilum* *Flavobacterium psychrophilum* *Moritella viscosa* *Tenacibaculum maritimum* *Yersinia ruckeri* Negative control

6.2.3.1.3. Experiment 3: Limit of Detection of Rs1_group

Because the Rs1_group primer set actually targets a single organism, this primer set was considered to have potential in the specific detection of *R. salmoninarum*. An experiment was set up to determine the limit of detection of LAMP assays that use this primer set. A reaction mix was prepared with the reagents described in Table 6-8 and divided into 18 0.2 mL tubes with 9.6 µL of the mix each; 0.4 µL of *R. salmoninarum* DNA at six different concentrations was then added in triplicate. The DNA was at concentrations of 10 to 100,000 genome copies μL^{-1}, plus a negative control with no DNA (sterile, UV-treated HPLC water only). The reaction was then run at 65 °C for 75 minutes in a G-Storm GS0001 thermocycler (G-storm

Ltd., Somerton, UK). The reaction tubes were removed and replaced after 30, 45, 60 and 75 minutes to record the progress of the reaction according to the colour change of the master mix.

Table 6-8: Reagents and quantities used to prepare LAMP reaction mixes for experiment 3.

Ingredient	Per 10 µL reaction	Per 18 x 10 µL reactions
WarmStart® Colorimetric LAMP 2X Master Mix	5	90
Water	3.6	64.8
Primer mix	1	18
DNA	0.4	-
Total	10	153.6

6.2.3.2. Results

6.2.3.2.1. Experiment 1: testing with intended templates

Only the broad-range LAMP primer sets designed for the seven BPOIs were synthesised and tested; the summary of experiment 1, where they were tested against their targets, is shown in Table 6-9. The results show that the primers react against their intended targets in all cases within 45 minutes. When tested against a negative control they also react in some cases, such as two of the negative controls of the tests carried out with the As1_group set of primers, or the three negative controls of Fb1_group. When tested against the negative controls, however, the LAMP reactions take longer to reach the colour change that indicates a positive result; despite this, it was possible to notice a faint colour change in some negative controls after 45 minutes in some cases.

Table 6-9: Time taken for reactions in experiment 1 (experiments with intended targets) to change colour from pink to yellow, indicating a positive result. When two numbers share a cell, the one in brackets represents the time when a slight colour change was first perceptible.

Primer set	Tested against	Replicate		
		1	2	3
As1_group	*Aeromonas salmonicida achromogenes*	45	45	45 (30)
	Aeromonas salmonicida salmonicida	45 (30)	45 (30)	45 (30)
	Moritella viscosa	45 (30)	45 (30)	45 (30)
	Yersinia ruckeri	45 (30)	45 (30)	45 (30)
	Negative control	75 (60)	-	75 (45)
Fb1_group	*Flavobacterium branchiophilum*	45	45	45
	Flavobacterium psychrophilum	45	45	45
	Tenacibaculum maritimum	45	45	45
	Negative control	60	75 (60)	75 (45)
Rs1_group	*Renibacterium salmoninarum*	30	30	30
	Negative control	-	-	-

6.2.3.2.2. Experiment 2: Testing with non-target templates

Once the broad-range LAMP primers were verified to react against their targets, the second experiment carried out intended to verify they did not cross-react against non-target BPOIs. For the As1_group and Rs1_group primer sets, the majority of reactions prepared with non-target DNA did not react (see Table 6-10). For Rs1_group in the cases where there was a colour change it was only clear 30 minutes after the clear colour change of the positive control. In the case of As1_group, there were only clear colour changes in reactions against non-target DNA in reactions with *F. branchiophilum* and *T. maritimum* DNA. These took place 15 minutes after the colour change of the positive control, and only took place in one of the replicates. In the case of LAMP reactions prepared with Fb1_group, colour changes occurred in all but two of them within the first 75 minutes; this included negative controls. The fact that there was such a consistent reaction, including for the negative controls, indicates that these primers probably react spuriously. It is worth noting that in most cases they took longer to react than the positive control, or than during the tests against target DNA, with only the exception of one replicate with *R. salmoninarum* DNA as substrate.

Table 6-10: Time taken for reactions in experiment 2 (cross-reactivity experiments) to change colour from pink to yellow, indicating a positive result. When two numbers share a cell, the one in brackets represents the time when a slight colour change was first perceptible.

Primer set	Tested against	Replicate		
		1	2	3
As1_group	*Flavobacterium branchiophilum*	75 (60)	-	-
	Flavobacterium pscyhrophilum	(75)	-	-
	Renibacterium saloninarum	-	(75)	(75)
	Tenacibaculum maritimum	-	75 (60)	-
	Negative control	-	-	-
	Positive control	60		
Fb1_group	*Aeromonas salmonicida achromogenes*	75	75	75
	Aeromonas salmonicida salmonicida	(75)	75	75
	Moritella viscosa	75	75	(75)
	Renibacterium saloninarum	75	45	75
	Yersinia ruckeri	60	75	60
	Negative control	75 (60)	75	75
	Positive control	60		
Rs1_group	*Aeromonas salmonicida achromogenes*	-	-	-
	Aeromonas salmonicida salmonicida	-	-	-
	Flavobacterium branchiophilum	-	-	-
	Flavobacterium pscyhrophilum	-	(75)	-
	Moritella viscosa	-	-	-
	Tenacibaculum maritimum	-	-	60 (45)
	Yersinia ruckeri	(75)	-	-
	Negative control	-	(75)	-
	Positive control	30		

6.2.3.2.3. Experiment 3: Limit of Detection of Rs1_group

The speed of the LAMP reaction testing Rs1_group primers against *R. salmoninarum* DNA was faster than the rest in experiment 1 (Table 6-9), taking only 30 minutes to reach a positive. This result, along with the clear negative results of the negative controls, plus the negative results when tested against all other non-target organism DNA, indicated Rs1_group had potential in specific target detection assays. When testing this primer set against different concentrations of the target DNA, the limit of detection of assays with this primer set was found to be 1,000 genome copies μL^{-1}. For concentrations above this limit reactions were consistently positive after 45 minutes; however, two positive reactions took place at 10 genome copies μL^{-1} after 75 minutes. Regardless, no reaction was detected in the negative controls nor the 100 genome copies μL^{-1} assays.

6.2.4. Amplicon sequencing

6.2.4.1. Methods

6.2.4.1.1. DNA quantification

Concentration of the DNA amplicons generated in the experiment of section 6.2.3.1.1 was measured with the Qubit™ HS kit (Invitrogen/Thermo Fisher Scientific). Before this, the three triplicates for each primer set-organism DNA combination were pooled and a sample was diluted 1:10 in sterile UV-treated HPLC water. Qubit™ HS reagent and buffer were mixed in a 1:199 volume ratio, and 199 µL of reagent-buffer mix were combined with 1 µL of diluted sample in a thin walled Qubit™ tube (Invitrogen/Thermo Fisher Scientific). These were then incubated for 3 minutes in the dark and measured on the Qubit™ 2.0 fluorometer (Invitrogen/Thermo Fisher Scientific, Paisley, UK). Original DNA concentrations were calculated by multiplying by ten.

6.2.4.1.2. Nanopore sequencing

Library preparation was carried out with the SQK-RBK004 kit (Oxford Nanopore Technologies, Oxford, UK) as per the manufacturers' instructions with minor modifications; the process is described here. The LAMP products were diluted with nuclease free water to a concentration of 53.3 ng μL^{-1} (200 ng in 3.75 µL) and then barcoded. Barcoding was carried out by adding 1.25 µL of Fragmentation Mix RB01-08 (one separate barcode per sample), and incubating successively for 1 minute at 30 °C and then at 80 °C, followed by rapid cooling on ice. Equal volumes of these barcoded samples were then pooled, making up a total of 40 µL, and an equivalent volume of AMPure XP beads (Beckmann-Coulter) was added. The tube containing the reaction mix was inverted by hand for 5 minutes, the beads were

pelleted on a magnetic rack, and the supernatant removed. The beads were washed with 200 µL of 70% ethanol twice, air-dried briefly, re-suspended in 5 µL 10 mM Tris-HCl with 50 mM NaCl, pH 7.76 and incubated for 2 minutes. The beads were pelleted again using the magnetic rack and the eluate was removed to a clean 1.5 mL microcentrifuge tube. 0.5 µL of the Rapid Adapter (RAP) were added and the mix was incubated at room temperature for 5 minutes, then placed on ice while the MinION flowcell was primed.

A Flongle flowcell (FLO-FLG001) was used for sequencing; it was connected to the MinION device with a Flongle adapter, and primed and loaded as per the manufacturer's instructions for this specific device using reagents from both the Flongle Sequencing Expansion (EXP-FSE001) and the SQK-RBK004 kit. The reagents belonging to the EXP-FSE001 are denoted by a "II". Priming involved introducing 120 µL of priming mix, composed of 117 µL Flush Buffer II and 3 µL of Flush Tether, into the sample port. The library was prepared for loading by mixing it with 2.25 µL of nuclease free water, 17 µL of Sequencing Buffer II and 12.75 µL Loading Beads II, and mixed well by pipetting gently. The library was then loaded onto the flowcell through the sample port.

The sequencing experiment was then initiated and controlled with the MinKnow software running on a Lenovo T570 computer connected to the MinION. The experiment was run for 25.5 hours, and raw reads were stored in fast5 format.

6.2.4.1.3. Basecalling, demultiplexing and gathering of basic statistics

The sequence data was uploaded to the Swansea Sunbird system of the Supercomputing Wales network. Here, it was basecalled and demultiplexed with the GPU-based version of Guppy v4.4.2 (Oxford Nanopore Technologies). Basecalling was carried out on the Sunbird system to leverage the parallel processing capabilities of its Nvidia V100 GPU processors, which improve the speed of the process in comparison to CPU-based processing. The basecalled reads in fastq format were downloaded back onto the Lenovo T570 PC, where all further data processing took place. Relevant statistics were gathered and plots were generated with NanoPlot v1.32.1 (De Coster et al., 2018), for the combined data and for the individual demultiplexed datasets.

6.2.4.2. Results

The pooled DNA from the LAMP reactions of experiment 1 had mostly consistent concentrations across the different primers and targets tested, with an average of 789 ± 92.7 ng μL^{-1} (mean ± S.D.). This was substantially higher than the negative control, where 8.64 ng μL^{-1} were detected. Sequencing with the Flongle yielded a total of 426,501 reads, with median read lengths and Phred quality scores of 406.6 bases and 9.4 respectively.

Guppy sorted reads by barcode into eight separate categories, corresponding to each primer set/organism DNA combination. 95,569 reads remained unclassified, and a small number of reads (38 in total) were classified as being tagged with barcodes that were not used in the experiment. For the eight barcodes that were used, between 28,298 and 49,952 reads were assigned to each, and their sizes were of 303 to 335 nucleotides (see Table 6-11). The size of the reads is relevant because the expected sizes of the amplicons range between 142 and 157 nucleotides, which means that the median read length amply covers the length of the target regions.

Table 6-11: Summary statistics of the read data obtained from sequencing the amplicons of experiment 1. Reads per barcode and their quality after demultiplexing are given.

Primer set	Target	Barcode	Median read length	Median read quality	Number of reads	Total bases
As1_group	*Aeromonas salmonicida achromogenes*	01	308	9.8	35,591	13,494,910
	Aeromonas salmonicida salmonicida	02	318	9.9	37,362	14,693,304
	Moritella viscosa	03	303	9.8	28,298	10,733,879
	Yersinia ruckeri	04	323	9.9	49,736	19,910,508
Fp1_group	*Flavobacterium branchiophilum*	05	335	9.8	49,753	20,855,722
	Flavobacterium psychrophilum	06	333	9.7	42,125	17,644,452
	Tenacibaculum maritimum	07	329	9.7	49,952	21,260,998
Rs1_group	*Renibacterium salmoninarum*	08	328	9.4	38,077	16,191,582

6.2.5. Classification

6.2.5.1. Methods and rationale

The demultiplexed Nanopore read data was used as the basis for the subsequent analyses. A script was written in Bash to classify the reads according to the bacterial species of origin of the DNA amplified through the broad-range LAMP. An explanatory description of this script, which is presented in full in Appendix 9, and other ancillary work needed for it to function, is given below.

To commence, because the precise target region within the *rrn* operon of each of the sets of broad-range primers is different, one DNA sequence was collected for each of the reference operons used during primer design (those of *A. salmonicida*, operon 1, *F. psychrophilum*, operon 1 and *R. salmoninarum* operon 1). These DNA sequences corresponded to the region between the sites where the external primers bind onto the *rrn* operons of their targets. These three sequences were saved together in a single fasta file, and separately, in individual fasta files; Minimap2 v2.18 (H. Li, 2018) was used to generate indexes for all of them.

The classifier script then takes a fastq file containing the LAMP amplicon dataset and a name for a new directory; it will create a directory with this name to save analysis results and intermediate files. The fastq-format reads are aligned with Minimap2 v2.18 (H. Li, 2018) against the three sequences representative of each of the clusters of organisms. Depending on what representative sequence has the most matches, the reads are re-aligned against that specific sequence individually, with one exception. If upon the first alignment the majority of the reads match *R. salmoninarum* operon 1, they are immediately classified as such, as it was the only species targeted by the Rs1_group primer set. When matching either of the two other representative sequences, the re-alignment takes place, once again, with Minimap2 v2.18, saving the output as a SAM file. The Samtools suite v1.13 (H. Li et al., 2009) is then used to convert this to BAM format and sort the alignment; the bcftools subset of tools is then used to detect single nucleotide polymorphisms (SNPs) and record them in a VCF file. This is done because the target regions of the *rrn* operon of the different species targeted by any given primer set are often distinguished solely by a polymorphism at a single site. Standard Linux utilities are then used to retrieve the information about the nucleotide at the specific position known to differ between the target species of interest, and to communicate the result to the user. This process is summarised in Figure 6-4.

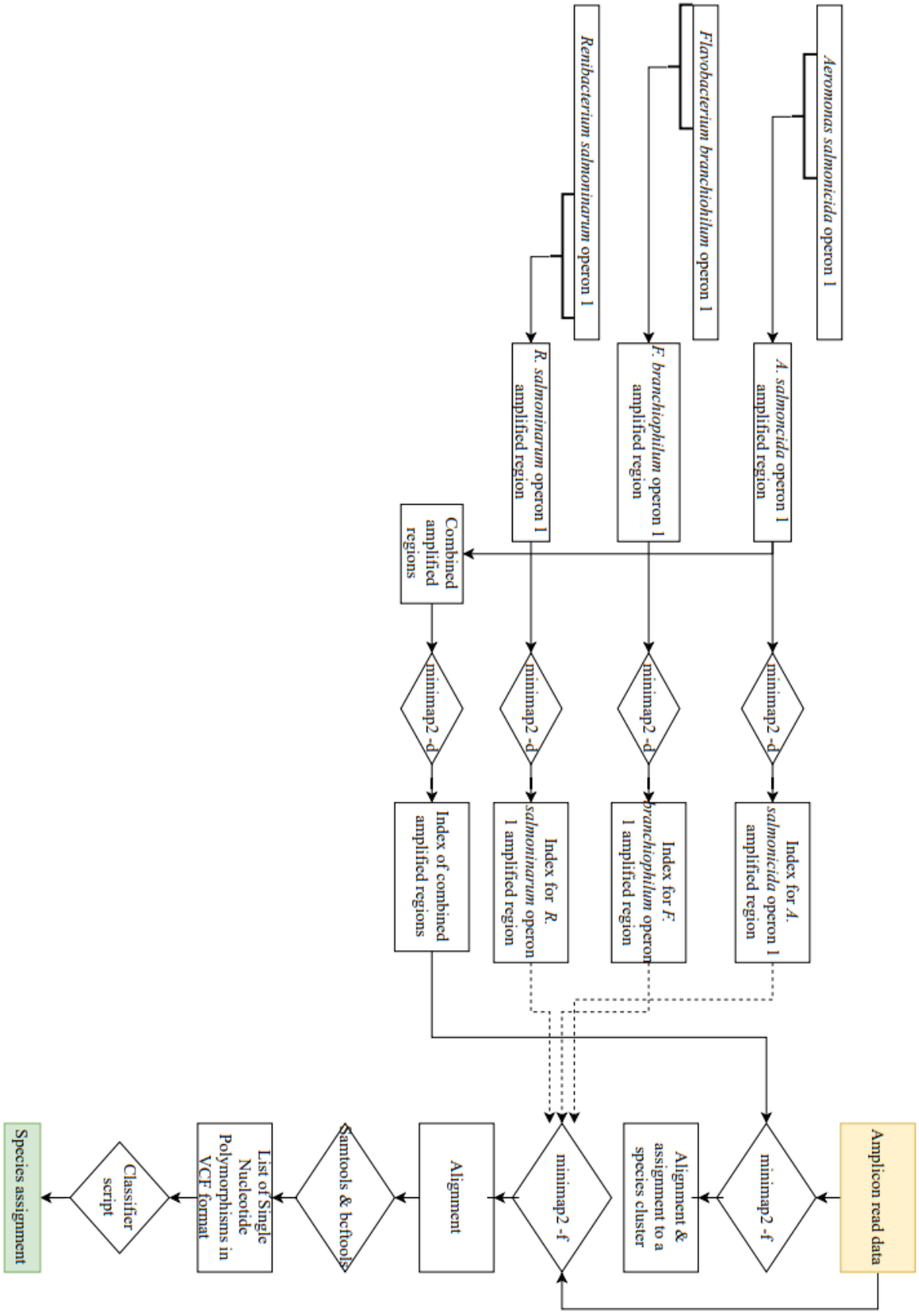


Figure 6-4: Process followed by the classifier script to assign amplicon sequences to an organism of origin (on the right; start of process in yellow box, result in green). The left hand side of the flowchart shows the preparatory process needed to enable the script to run.

6.2.5.2. Results

The classifier script classified the read datasets pertaining to each barcode to their correct genus of origin in all cases, and to species level in every case except for that of the *Flavobacterium* representatives. It took an average of 21.1 ± 9.23 seconds (mean ± SD) for it to run. Table 6-12 shows the result of the first step of the script; the assignment of a dataset's reads to their cluster (which corresponds to their primer set). It can be seen that in all cases, over 99% of the reads are assigned to the correct cluster.

Table 6-12: Results of the first alignment phase applied by the classifier script. The Renibacterium salmoninarum *group contains* Renibacterium salmoninarum, *the* Flavobacterium psychrophilum *group contains* Flavobacterium branchiophilum, Flavobacterium psychrophilum *and* Tenacibaculum maritimum, *and the* Aeromonas salmonicida *group contains* Aeromonas salmonicida, Moritella viscosa *and* Yersinia ruckeri. *The criterion for classification is the alignment to the reference sequence of each of these groups (see Figure 6-4, Table 6-2 and section 6.2.4.2).*

Barcode	Identity of input DNA	Identified by classifier script as:	Percentage (%) reads assigned to group:		
			Renibacterium salmoninarum	*Flavobacterium psychrophilum*	*Aeromonas salmonicida*
01	*Aeromonas salmonicida achromogenes*	*Aeromonas salmonicida*	0.183	0.429	99.4
02	*Aeromonas salmonicida salmonicida*	*Aeromonas salmonicida*	0.130	0.364	99.5
03	*Moritella viscosa*	*Moritella viscosa*	0.158	0.564	99.3
04	*Yersinia ruckeri*	*Yersinia ruckeri*	0.215	0.399	99.4
05	*Flavobacterium branchiophilum*	*Flavobacterium*	0.150	99.4	0.474
06	*Flavobacterium psychrophilum*	*Flavobacterium*	0.161	99.3	0.5
07	*Tenacibaculum maritimum*	*Tenacibaculum maritimum*	0.366	92.3	7.379
08	*Renibacterium salmoninarum*	*Renibacterium salmoninarum*	99.3	0.304	0.434

The second alignment step of the script, which takes place for reads determined to belong to the *A. salmonicida* or *F. psychrophilum* clusters, was accurate. Samples were assigned an identity based on an SNP at position 125 of the amplified region in the case of the *A. salmonicida* cluster and at position 82 of the amplified region in the case of the *F. psychrophilum* cluster. It should be noted that Samtools did not identify any unexpected mutations at these positions in seven out of eight cases. In the case of barcode03 (*M. viscosa*) it identified an indel in some reads, but this did not impede the detection of the SNP that distinguishes *M. viscosa* from the rest of the members of its cluster.

The limitation of this script was apparent in its inability to distinguish the two *Flavobacterium* species. This is because the amplified segment of the two *Flavobacterium* species' *rrn* operon is identical; the implications are considered in the Discussion (section 6.3.3).

6.3. Discussion

6.3.1. Overview

Throughout this chapter I have attempted to provide a proof of concept for a means of combining LAMP and Nanopore sequencing to detect and discern a number of different pathogenic bacteria of farmed fish. The work carried out during this project has not served to find a single set of LAMP primers that can amplify the *rrn* operon of the 29 pathogens that were of interest to the company partner, nor the subset of seven BPOIs. It has, however, served to design five primer sets that are predicted to bind to and amplify regions of the *rrn* operon of 21 different bacterial pathogens (objective 1.1). Additionally, it has also served to design three sets that have been demonstrated to bind to and react with the *rrn* operons of the seven BPOIs (objective 1.2). The latter have been tested to demonstrate that they can be used in combination with Nanopore sequencing and post-processing of the data to identify the organism of origin of the amplified DNA (objective 1.3). Additionally one of the primer sets designed during this chapter (Rs1_group), has the potential to detect concentrations of *R. salmoninarum* DNA ten times lower than the set designed in Chapter 5.

6.3.2. Outcomes

The main achievement of this chapter is the design of primer sets with broad ranges of targets that can be distinguished through nanopore sequencing. It is, to the best of my knowledge, the first time broad-range LAMP primers capable of reacting with the *rrn* operons of several bacteria are designed. Multiplex LAMP assays exist, but they often rely on an individual primer set to detect each of their targets (James et al., 2020; Jang et al., 2021; J. Kim et al., 2021; Y. Wang et al., 2015). In several cases, they do not actually target a set of different organisms, but rather several genes that serve to confirm the presence of the target pathogen (James et al., 2020; Jang et al., 2021). Researchers like Imai and colleagues (2017), Yamagishi and colleagues (2017), and Hayashida and colleagues (2019) do design and test LAMP primers that target a range of organisms or viruses, *Plasmodium* species in the first case and several Dengue and Chikungunya virus genotypes in the second and third. These three studies use nanopore sequencing of the amplicons to distinguish the different species or genotypes. However, their range of targets is narrower, and the wide range of targets (13 different species in four phyla) that some of the primers designed in this chapter can theoretically detect has never been reported before (again, to the best of my knowledge).

Another noteworthy outcome of this chapter is the demonstration that nanopore sequencing, and in particular a Flongle-based workflow, can serve to read the amplicons and distinguish the species of origin (relates to objective 1.3). This is not novel; preprints from summer 2020 (James et al., 2020; J. Li et al., 2020) were already announcing the use of a combination of LAMP and nanopore sequencing to detect SARS-CoV-2. Li and colleagues (2020), who use sequencing only to determine LAMP amplification, do use Flongle, but they report as few as 27,000 reads in a 12 hour sequencing run. The Flongle used in this chapter's work was received in February 2021 and produced 426,501 reads in 24 hours. This highlights the advances Oxford Nanopore Technologies has made in the design and manufacture of these devices and their kits. This improvement, alongside the demonstration that it is possible to classify samples to their correct organism with as few as ~28,000 reads, suggests it is possible that LamPORE assays could become less expensive if sequencing was carried out with Flongle, and still be viable.

The script used to classify the reads to their organism of origin is also of interest. The methodology followed by this script has several differences and similarities to those followed by Hayashida and colleagues (2019), Imai and colleagues (2017) or James and colleagues (2020), where similar classification of LAMP amplicon reads is carried out. Specifically, the software used for alignment of the reads to a reference used in this work was Minimap2 (H. Li, 2018), as opposed BWA-MEM (H. Li, 2013) or Bowtie2 (Langmead & Salzberg, 2012), used by Imai and colleagues (2017) and Hayashida and collegues (2019) respectively. The Minimap2 aligner is according to the program's author up to 50 times faster, and more accurate than BWA-MEM, and also performs better than Bowtie2 (H. Li, 2018). Classification of the reads takes less than a minute for each of the barcoded datasets of this chapter (see section 6.2.5.2), which implies this approach could be advantageous in a time-sensitive context.

On the whole, these features of the broad-range LAMP assay plus nanopore sequencing combination make it an interesting option for semi-naïve pathogen detection. The multiplicity of the assay, the relatively low resource requirements and its speed mean it could make it feasible in contexts where economic factors are key to the implementation of innovative techniques, such as private aquaculture operations or low resource settings. Its application need not be limited to identification of pathogens during an active infection, but it could be used for monitoring the health of the farm's stock through the use of environmental DNA

(see Chapter 4). However, the method has failures and drawbacks; these are discussed in the following section.

6.3.3. Drawbacks of the broad-range LAMP method

Initially, one of the objectives of this chapter was to produce a single primer set that could target a broad range of pathogenic bacteria, and that would allow for their distinction by sequencing. The high level of variation within the *rrn* operon, a consequence of having targets from four different bacterial phyla, left few sections where enough contiguous conserved positions were present to fit a primer. The workaround for this was to design a number of primer sets that could encompass the entire range of pathogens, which was only partially achieved. This attenuates the benefits of the broad-range LAMP strategy, as it implies a number of simultaneous LAMP assays are required to enable the detection or distinction of all the BPOIs. This specific drawback, however, could be solved by combining the different primer sets into a single multiplex LAMP (see section 6.1), although experiments would be necessary to ascertain they do not cross-react.

Another partial failure is in the initial choice of target. The *rrn* operon is present throughout the bacteria and archaea (Woese & Fox, 1977), making it an ideal target for broad-range assays. Although in some cases the genes that conform it are unlinked (Brewer et al., 2020), in most cases the gap can be bridged by PCR, and serve to increase the accuracy of classification of reads in microbiota studies, compared to 16S alone (Cuscó et al., 2019). However, LAMP has limits in the maximum distance between the target regions of each of the primers in a set. This is approximately 80 nucleotides between the inner primers (Eiken Chemical Co., Ltd., n.d.); it was not taken into account when choosing the entire operon as opposed to a single rRNA gene.

A third and more serious failure is a consequence of the previous point. The target regions of the designed primers do not encompass sequences that are sufficiently variable to distinguish closely related species, as it has been observed when attempting to discern the two *Flavobacterium* species. This issue is likely to repeat itself when considering any close relatives to any of the target bacteria, including non-pathogenic species. This implies that the methods developed here would not serve to reliably assess the presence of a pathogen in a sample, but only to classify samples of pathogens from a pre-existing list at best.

6.3.4. Solving methodological failures

The drawbacks encountered and described in the previous section are a consequence of the choice of target and primer design software. The *rrn* operon, as mentioned above, was chosen as target for its universality and its potential to accurately identify species. GLAPD was chosen for primer design for two reasons. First, PrimerExplorer, the main alternative design software, will only design primers for one target sequence simultaneously. Second, the way GLAPD works, by designing all primers for a reference sequence first, and then searching for the set that will theoretically amplify the largest amount of intended target sequences, simplifies the task of design in highly variable sequences.

The target choice is problematic because the *rrn* region is too long, and the ITS, which contains a large amount of variability and is therefore useful for the classification to the species level (Benítez-Páez & Sanz, 2017; Cuscó et al., 2019; Kerkhof et al., 2017) cannot be spanned by LAMP. A more appropriate choice would be one of the hypervariable regions of the 16S rRNA gene, which are often used in ecological studies (W.-Y. Chen et al., 2017; Ciric et al., 2019; García-López et al., 2020). Specifically, Chakravorty and colleagues (Chakravorty et al., 2007) found that the V6 region had substantial variability in just 58 nucleotides and that it sufficed to distinguish a number of bacterial pathogens. Given its characteristics this is a promising option for future broad-range LAMP primer design, although a preliminary, *in silico* study would be sensible.

With regards to the choice of GLAPD, it has the disadvantage of not being able to accommodate mismatches. In reality, it is possible to use degenerate primers to amplify targets where the sequence may vary slightly between organisms, as is often the case with 16S rRNA gene PCR primers (Frank et al., 2008; Higuera-Llantén et al., 2018). Additionally, mismatches between the 3' ends of primer regions F1c and B1c or the 5' ends of primer regions F2, B2, F3 and B3 and their template do not necessarily impede priming (Eiken Chemical Co., Ltd., n.d.). Future design work could be carried out by using PrimerExplorer v5 and Morphocatcher (Shirshikov et al., 2019), which permit the design of LAMP primers valid for a range of target sequences based on multiple sequence alignments.

6.3.5. Rs1_group primer set

Although this was not an initial intention of this chapter,'s work, because the Rs1_group primer set targeted the *rrn* operon of *R. salmoninarum* specifically, it made for an interesting comparison to the Rs6 primer set from Chapter 5. Specifically, the focus was on limit of

detection because of the potential improvement of analytic sensitivity when targeting a multi-copy operon. The limit of detection was established at 1,000 genome copies μL^{-1}, which is lower than the one observed for the Rs6 set (10,000 genome copies μL^{-1}). The time to positive also improved, with the reaction taking place within 30 minutes, as opposed to the 45 minutes of Rs6.

The Rs1_group LAMP also performs similarly to other published methods. Two other authors have designed and tested LAMP primers against *R. salmoninarum,* although targeting the gene that codes for the p57 Major Soluble Antigen, a known virulence factor of this species (Gahlawat et al., 2009; Saleh et al., 2008a). Gahlawat and colleagues (2009) do not give specific figures, but Saleh and colleagues (2008a) report detection of 1 pg/reaction, equivalent to 0.5 pg μL^{-1}. In the same units, the Rs1_group LAMP assays detected 3.24 pg μL^{-1}, or 1.30 pg per reaction, which is within an order of magnitude of their assays. It is worth mentioning that there are a number of studies where different variations of PCR-based assays involve designing primers against the 16S gene of *R. salmoninarum* (sequence-capture and fluorescent PCR, Königsson et al., 2005; nested RT-PCR, Magnússon et al., 1994; qPCR, Jansson et al., 2008). These often have lower limits of detection; for instance Magnusson and colleagues (1994) report detecting 1-10 bacteria per sample, or 1-10 genomes per reaction in the case of Jansson and colleagues (2008).

It is possible that the fact that each *R. salmoninarum* genome has two copies of the *rrn* operon, as opposed to one, is responsible for the improvement in limit of detection and reaction time. This may not be the case given that reactions with As1_group and *A. salmonicida*, which has nine *rrn* copies, took 45 minutes, but it is worth considering targeting this operon when designing specific primers for other bacteria in the future. Future experiments could be carried out using specific, *rrn*-based LAMP primers for the other BPOIs in order to test this hypothesis.

Regarding analytical specificity, Rs1_group was tested against the other six BPOIs without reaction. *R. salmoninarum* is one of the few representatives of the Actinobacteria phylum among the 29 bacteria listed by the company partner, and among those for which it was possible to extract *rrn* operon sequences, only three belonged to this phylum (*Mycobacterium marinum*, *Nocardia seriolae* and *R. salmoninarum*). The preliminary Neighbour-Joining tree that shaped the clusters that the operon datasets were divided into marked these three species as a distinct cluster; however, only *R. salmoninarum* was among the BPOIs, the set of seven

bacteria this thesis focuses on. This species was therefore the only representative of its cluster in the design of "broad-range" LAMP primers. Despite running GLAPD with the *rrn* operon sequences of all the seven BPOIs as intended "common" targets, the primers designed were predicted to react only with *R. salmoninarum*'s *rrn* operons. It is therefore unsurprising that the assays prepared with Rs1_group, were specific to *R. salmoninarum*. Despite this, Rs1_group has not been compared to other members of the Micrococcaceae family, which means that it is still possible that these primers could cross-react with some non-target DNA. Therefore, it is necessary that before considering this set for practical, specific LAMP assays it should be tested, at least in-silico, against other closely related bacteria.

6.3.6. Conclusion

The work carried out during this chapter provides a demonstration that it is possible to target a wide range of bacterial species with a single set of LAMP primers, and that they can be distinguished by sequencing the amplification product with a nanopore sequencer. It may be possible to combine primer sets in a multiplex assay to ensure a majority of bacterial pathogens of aquaculture are covered in a single assay. However, although this chapter serves as proof of concept, several of the characteristics of the primers make them unlikely to serve in realistic scenarios. For a functional broad-range LAMP assay, it would be necessary to re-design primers considering a more specific locus than the entire *rrn* operon, with sufficient variability to discern the different bacteria of interest. The V6 hypervariable region of the 16S rRNA gene seems like a potential target for future attempts (Chakravorty et al., 2007). If an appropriate locus is found, the technique demonstrated in this chapter has the potential to serve as a multi-pathogen test that is fast, relatively inexpensive, and relatively field-forward.

Chapter 7: Main Discussion

7.1. Overview

This thesis aimed to evaluate the benefits of nanopore sequencing on aquaculture disease detection and management (aim 1), to find means of maximising the DNA obtained from water (aim 2), in order to implement metagenomic pathogen detection, and to produce LAMP assays for the detection of seven bacterial pathogens of salmonids (aim 3). The first aim is addressed in Chapter 2 and Chapter 3. In Chapter 2 comparative genomics studies of *Aeromonas salmonicida*, and to a lesser extent other BPOIs, showed that sequencing new genomes can provide valuable data for pathogen monitoring and management; Chapter 3 showed that assembling new genomes from nanopore sequencing data is feasible with low resource requirements, although the technology needs to mature before becoming practical to use as a standalone method. Chapter 4, which addresses the second aim, devises optimal methods for the concentration of bacteria from water and extraction of their DNA, but finds that these are insufficient to reliably provide enough material for nanopore sequencing. The difficulty in implementing metagenomic pathogen detection motivates Chapter 5 and Chapter 6; here, LAMP assays which require relatively small amounts of starting genetic material, are developed for the detection of the BPOIs, addressing the third aim. In Chapter 5, assays specific to a single target are developed, while Chapter 6 provided proof of concept for a method that combines LAMP with nanopore sequencing to enable the detection of several pathogens in one assay.

7.2. Overview of proposed workflow

Overall, the different techniques and technologies explored in this thesis have the potential to greatly improve the management of disease in aquaculture. The author of this thesis would like to propose a means of integrating them in order to maximise the benefits of using them, accounting for their known limitations. This workflow is detailed in the schematic of Figure 7-1; it would use the optimised methods of DNA recovery devised in Chapter 4 to obtain DNA for LAMP assays, and implement genome sequencing only in the case of a pathogen being detected. This system would economise resources, by employing only relatively fast and inexpensive techniques at first to detect the pathogens. Then, costly but informative hybrid genome sequencing could be employed when the presence of a pathogen is established.

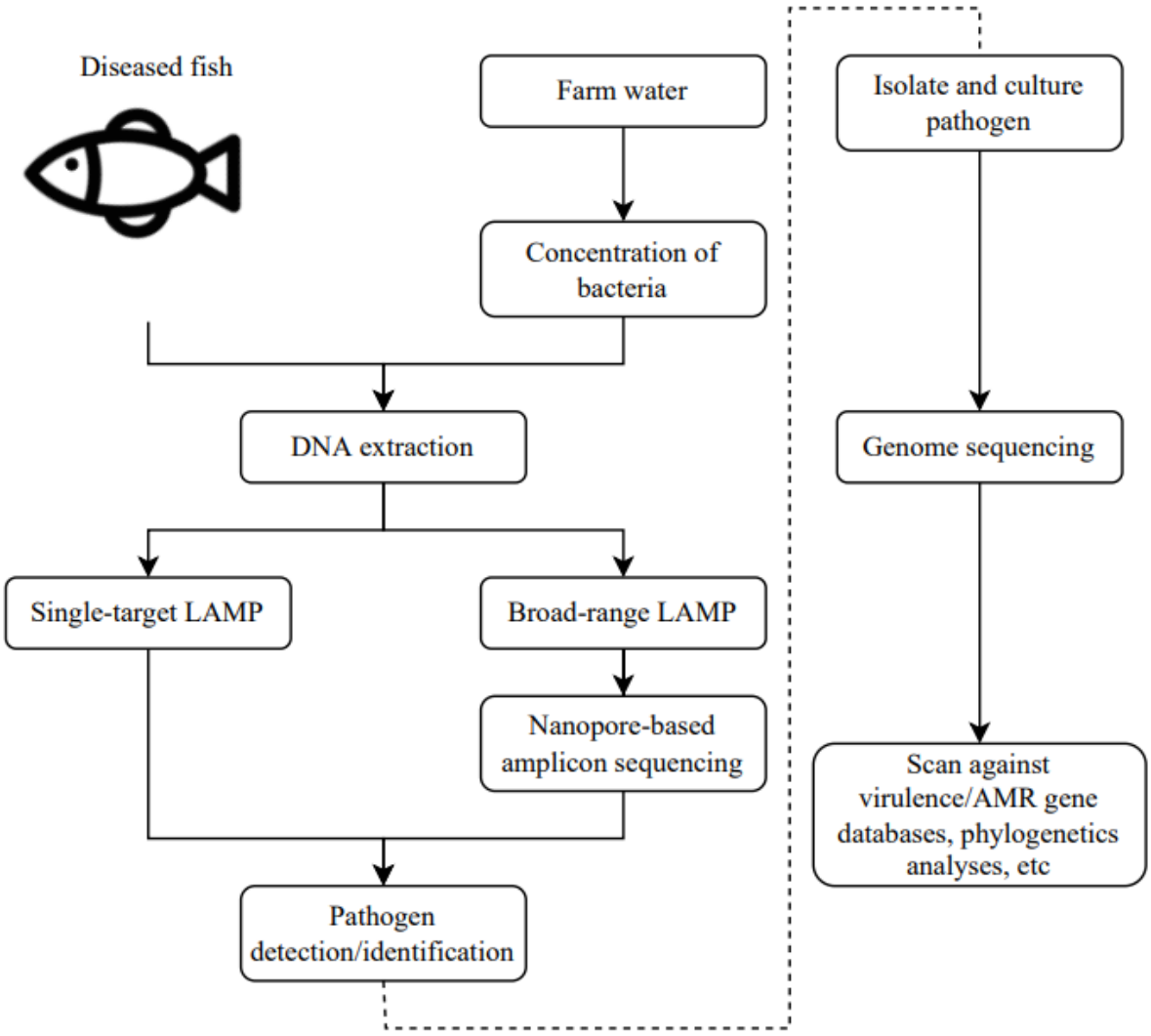


Figure 7-1: Schematic for proposed workflow integrating the techniques and technologies explored throughout this thesis.

Over the next sections the feasibility of implementation of the separate techniques of this proposal, the need for further work and the likely impact of future developments are discussed.

7.3. On genome sequencing

As discussed in Chapter 3, hybrid assembly methods have demonstrated superior performance on all metrics except for time taken and memory usage. Although long-read-only assembly methods can serve to obtain valuable information on their own, such as taxonomy (Menzel et al., 2016; Wood & Salzberg, 2014) or antibiotic resistance (Cao et al., 2016), they are insufficient on their own for accurate genome reconstruction. Techniques as those applied in Chapter 2 would be unfeasible with long-read-only assemblies; pan-genomics analyses would be skewed by the introduction of genomes annotated with more genes than they actually have, and phylogenetic reconstruction would be inaccurate if spurious mutations are introduced. As stated in said chapter, phylogenetic analyses often play a role in understanding the geographic distribution patterns of pathogens (Brynildsrud et al., 2014; Gómez-Carballa et al., 2020; Gray et al., 2011; Thapa et al., 2020), providing valuable information for outbreak tracking and control. Pan-genomics also play a role in reverse vaccinology, helping to design new vaccines (Naz et al., 2019; Shahid et al., 2021). Not least, genome sequencing can play a role in assay design for nucleic acid based tests (Chapter 5; Jia et al., 2019). Ultimately, however, the usefulness of these applications of sequencing is not limited by an inability to be applied on the field; instead it would be impeded by inaccurate assemblies. It is therefore reasonable to expect hybrid assembly, which has demonstrated an ability to reconstruct genomes that are both more complete and accurate, to take prominence in these roles.

This is not to say that nanopore sequencing does not and will not have a use and a role to play in disease detection and pathogen monitoring and surveillance. As stated earlier, it can be used to obtain information on antimicrobial resistance with nanopore reads (Cao et al., 2016), or with long-read assemblies (Chapter 3); the latter may also serve to detect virulence by scanning against databases such as VFDB (L. Chen, 2004). Furthermore, the sequencing data used in Chapter 3 was obtained in 2019; considering the speed at which innovation takes place in the field of nanopore sequencing it is possible that data obtained with the most recent sequencing chemistries and basecallers may yield better results (Koren et al., 2019). For example, the release of the R10 pore in 2020 was accompanied by reports of high accuracy (ONT, 2020). Another improvement is that of the Bonito basecaller (Seymour et al., 2019/2021), which has given high quality reads and demonstrated the ability to produce relatively high quality genome assemblies (Vereecke et al., 2020). It is therefore conceivable

that if not now, it may soon be possible to produce assemblies of comparable quality to hybrid assemblies solely with nanopore data.

To further contextualise this, the trend in human health appears to be to simplify operations by relying on pathogen Whole-Genome Sequencing (WGS). Public Health England has already tested a WGS-based scheme for typing and surveillance of gastrointestinal pathogens in their Gastrointestinal Bacteria Reference Unit and plan a transition to a genome-led service (Grant et al., 2018). PulseNet International, a network dedicated to the monitoring of pathogens in food, has already recommended transitioning to a standardised, WGS-led approach to identification and typing of foodborne bacterial pathogens (Nadon et al., 2017). WGS of pathogenic bacteria was also on the increase in the period 2015-2016 in the European Union, with three more countries implementing WGS workflows into their public health services, and those that already did expanding their capabilities (Revez et al., 2017). It would not be surprising that this trend in public health spills over to animal health as sequencing costs fall.

7.4. On effective LAMP primer design

One of the major achievements of this project is the design of seven working primer sets that can target a number of major bacterial pathogens of salmonid fish. These pathogens are: *Flavobacterium branchiophilum*, *Flavobacterium psychrophilum*, *Moritella viscosa*, *Renibacterium salmoninarum*, *Tenacibaculum maritimum* and *Yersinia ruckeri*. Of these, no published LAMP primers existed for *M. viscosa*, *T. maritimum* or *Y. ruckeri*, and in all cases it is the first time GLAPD (Jia et al., 2019) has been used to design them. This tool has proven effective in designing primers, although some of its most important characteristics, as are target-specificity and commonality have not been demonstrated in this thesis' practical experiments.

At first, the intention was to design primers based on the core genome of the seven BPOIs; this would ensure that primers would target any known strain for which there was a publicly available genome. This would make them effective even in cases like that of *Aeromonas salmonicida*, where the accessory pan-genome is diverse and only one of the 48 AMR genes is present throughout the entire species (see Chapter 2). However, this was a cumbersome task, as it involved the extraction of the core genes of the pan-genome and would still require the manual selection of one of the many core genes. Additionally, the sequences of the chosen gene should be compared throughout the homologues present in all the genomes of the dataset, in order to find conserved positions within the gene. The entire task is simplified with GLAPD, which designs all possible primers that could bind to a reference genome and then combines them into sets that are checked to be valid for an entire "common" genome dataset. As a result, the designed primers will forcibly be both in a core gene and in a conserved region; this is achieved without prior knowledge about the bacterium, merely in a systematic, easily replicable manner. Additionally, GLAPD can scan a background dataset to ensure the specificity of the primers to a desired set, so it is capable of providing primers valid for one of several closely related organisms. Indeed, if the non-target organism is closely related to the target, the primers designed by the software can be as reliably specific as the genomes in the background dataset are complete and accurate.

It is important to note, however, that two primer sets were designed for *R. salmoninarum*. The first one, Rs6, was designed based on the whole genome of *R. salmoninarum*; the second, Rs1_group, was based on the *rrn* operon. Two points are worth noting here. First, Rs1_group has a limit of detection ten times lower, and a shorter time to reaction (30 minutes as opposed to 45 minutes for the Rs6). Second, the design for Rs1_group took substantially

less time. As to whether the first point can be attributed to the use of the *rrn* operon as target, it is unclear. It would appear to make sense that a gene of which there are multiple copies in the genome would allow for detection of a lower number of cells, assuming different LAMP assays had the same sensitivity in terms of number of target copies. This comes into apparent agreement with the work of other researchers who have compared PCR probes targeting genes with multiple copies in the target organism's genome (Chern et al., 2011; Hofmann et al., 2015; Reischl et al., 2003). Specifically, Chern and colleagues (2011) demonstrated qPCR assays targeting the 23S ribosomal RNA gene of *Escherichia coli* were capable of detecting the bacterium in low-concentration samples where two qPCR assays that targeted single copy genes could not. They also demonstrated they could detect ~8 times fewer calibrator cells with the 23S rRNA-targeting assays than with the single-copy ones. However, to my knowledge, there appear to be no systematic studies on the matter, where a number of different bacteria are tested with primers targeting single-copy genes and ribosomal RNA genes. The fact that the LAMP assay with Rs1_group had a lower limit of detection will remain as anecdotal evidence until a more systematic study is carried out.

Regarding the second point, this software can take time to run. Although precise timings were not recorded, it took several hours to design LAMP primers when the input was a full genome, whereas when the single operon was used it took minutes. This is in large part due to the fact that two of its three steps can only run on a single CPU simultaneously (Jia et al., 2019). When time is pressing, this can become a serious drawback, and pending improvements of the software it is sure to limit its acceptance among the broader public, at least for its implementation as a "naïve" primer design software. To design primers that target the *rrn* operon can be a simpler approach, and potentially just as effective. One reason for this is the universal presence of ribosomal RNA genes throughout all living cells, so there will certainly be a target in any non-viral pathogen. The other is the variability of *rrn* operons; even if the 16S and 23S genes are conserved, and no positions can be found there to distinguish an organism from a close relative, the Internal Transcribed Spacer is highly variable.

A note on the issue of sensitivity: one major limitation of the LAMP assays tested here is that though their limits of detection are established, no comparisons have been carried out against previously validated methods. This means that there is uncertainty about what makes them higher than those of similar published methods. Possible reasons are considered in section 5.3.2; sub-optimal master-mix composition, stochastic factors affecting the final DNA

concentration in the assay tube and the primer designs themselves are some prominent possibilities. Validated LAMP assays already exist for *A. salmonicida* (Kulkarni et al., 2009a), *F. psychrophilum* (Fujiwara-Nagata & Eguchi, 2009), *R. salmoninarum* (Saleh et al., 2008a) and *Y. ruckeri* (Saleh et al., 2008b). Experiments could be run again comparing the performance of the primers designed in this thesis to that of those described in said publications. If done in equality of conditions, it would allow to discern whether the primers or the master mix are the cause of the high detection limit, given none of these studies uses the NEB WarmStart Colorimetric 2X LAMP master mix.

7.5. On the applicability of LAMP

The capabilities of LAMP have been demonstrated in a number of occasions. The technique is rapid, sensitive and inexpensive in comparison to PCR, and is relatively simple to carry out, even more so with the modifications that have been made to adapt it to field settings (Foo et al., 2017; Hayashida et al., 2019). There are abundant examples in the literature of LAMP assays designed for aquaculture pathogens; it is common for the aforementioned benefits to be demonstrated and remarked upon (Fujiwara-Nagata & Eguchi, 2009; Kulkarni et al., 2009a; Saleh et al., 2008a, 2008b). In recent times the interest in this technique has increased as a result of the SARS-CoV-2 pandemic; the simplicity, speed and sensitivity of this technique have led it to be proposed as a means of increasing testing capacity, especially in developing countries and other resource-limited settings (Chow et al., 2020; Dao Thi et al., 2020; Ganguli et al., 2020).

The work carried out for this thesis, which has used a commercially available kit (specifically NEB's WarmStart Colorimetric Master Mix 2X), has testified to the speed of the assays, and to an extent to their simplicity. The cost per reaction, however, was higher than that of PCR; if the manufacturer's recommendations of assay volume are followed, NEB's WarmStart® Colorimetric Master Mix 2X costs £1.65 per reaction versus £0.50 per PCR reaction using NEB's OneTaq® Quick-Load® 2X Master Mix. The simplicity of the assay was slightly hampered by the preparation step, which still involved the mixing of primer and master mixes with the test DNA sample in a sterile environment. Although this operation isn't particularly complex in itself, it is not reasonable to expect someone without any training to carry out the assays. One option would be the simplification of the assay procedure; a number of lab-on-chip implementations of LAMP have been designed (Malpartida-Cardenas et al., 2019; Tsougeni et al., 2020; S. Wang et al., 2020), which facilitate manipulation by the end-user. Another option would be the deployment of mobile laboratories, which has already been done with SARS-CoV-2 LAMP based testing (Chow et al., 2020). In either case, LAMP itself is amenable to implementation on aquaculture farms.

Specifically, three primer sets designed in this thesis could help in the control of three diseases for which there are currently no LAMP tests available (in the public literature, to the best of my knowledge). *F. branchiophilum*, *T. maritimum* and *M. viscosa*, which cause, respectively bacterial gill disease, tenacibaculosis and winter ulcer disease, are all involved in economic losses in salmonid aquaculture (FAO, 2021a, 2021b; Olesen et al., 2018; Olesen & Vendramin, 2017; Vendramin & Olesen, 2021). Although PCR assays exist for the three of

them (Avendaño-Herrera et al., 2004; Grove et al., 2008; Toyama et al., 1996), the primers designed here could provide an alternative avenue for simpler, more broadly applicable testing that required a smaller investment in equipment and was somewhat more portable. However, before these assays can be implemented in the control of these diseases it is necessary to carry out further tests to confirm they are sufficiently sensitive and specific, in the diagnostic sense. The LAMP assays must demonstrate they can serve as an appropriate alternative to PCR and other more field-forward, but often less reliable tests, such as antigen-based lateral-flow tests (Kunanopparat et al., 2011; P. Sithigorngul et al., 2011; W. Sithigorngul et al., 2007). This could be tested in experiments that used swab samples obtained from healthy and infected fish, testing the LAMP assays on DNA extracts from both types, and comparing to a similar PCR assay.

However, a more difficult task is the implementation of the methods developed in Chapter 4. As stated during that chapter, the method to obtain the highest yields of DNA would involve the concentration of material from the water by centrifugation, and DNA extraction with a PowerSoil kit. Besides the £5.20 estimated costs per sample, this would require centrifuges and pipettes. As stated in previous occasions, although this would be entirely possible using portable lab approaches as those of Quick and colleagues (2016) and Boykin and colleagues (2019), or mobile labs, as Chow and colleagues (2020) did, it is worth asking whether they would be feasible. A detailed cost-benefits analysis and an assessment of how this could be implemented in low resource settings would be necessary if these methods are to be used in practice.

7.6. On the development of a naïve monitoring system

At the start of this project there was an interest in exploring use of the MinION for metagenomics-based detection of pathogens in aquaculture waters, as explained in Chapter 4. The possibility of sampling water without any previous assumptions about presence or identity of pathogenic bacteria and carrying out a single test for a wide array of pathogens would have constituted an excellent method of active monitoring of disease. This potential has been recognised by Hamner and colleagues (2019), who tested river water downstream of a bathing location in Montana, USA, using metagenomics to detect the presence of coliform bacteria.

However, the DNA extractions from river and seawater gave very low yields (see Chapter 4), below the 400 ng needed for the library preparation kits with the lowest input requirements. Furthermore, much variability was observed between the different sampling occasions in the amount of DNA obtained, possibly as a result of changing water quality. The consideration of alternative methods that could be used to test water naïvely for a broad range of pathogens, without the constraint of requiring high amounts of DNA, was one of the reasons for the development of a broad-range LAMP assay.

The methodology developed in Chapter 6 attempted to extend the methods demonstrated by Imai and colleagues (2017) and Hayashida and colleagues (2019) to a more ambitiously broad set of targets. As discussed in said chapter, the method does serve to detect the target organisms' DNA and distinguish which among them the original sample contained. However, the broad-range LAMP assays were, as the name indicates, broad range, and not universal. This means that the number of pathogens that could potentially be detected is limited, although it was demonstrated that it was hypothetically possible to address 21 bacterial pathogens of aquaculture with four primer sets.

Additionally, although the aim of Chapter 6 was not to evaluate the limits of detection of the broad-range LAMP assay, this was done for the Rs1_group primer set. In this case the limit of detection was 1,000 genome copies per µL, or approximately 3.24 pg μL^{-1}. If this is taken as an approximate indicator of the limit of detection of As1_group and Fp1_group it is possible that these indeed react with their target bacteria at low DNA concentrations as those retrieved from river and seawater. It is evident, however, that a set of experiments must be carried out to assess the limit of detection of these LAMP assays, and the single-target assays, in terms of bacterial cells per unit of volume of water sample. These could consist in adding a

staggered number of bacterial cells into a sample of water, centrifuging and extracting DNA as per the recommendations in Chapter 6, and running the LAMP tests on these extracts. These experiments could inform as to whether the implementation of these assays on aquaculture waters as a means of monitoring the target bacteria could serve as an effective naïve detection system.

It should be noted that the metagenomic surveillance route cannot be entirely discarded. Recent research suggests that it might still be possible even in waters with low quantities of DNA. The problem with nanopore sequencing is the need of high quantities of DNA to prepare libraries: Mojarro and colleagues (2018) developed a technique to overcome this issue. It consists in adding large amounts of DNA from a known source (*Enterobacteria phage λ*) to compensate low DNA samples, then use a workflow they name CarrierSeq to filter out the reads that correspond to their added DNA. Through this method they are able to sequence low DNA samples of *Bacillus subtillis*; further experiments verify its validity in synthetic soil experiments with this same bacterium (Mojarro et al., 2019). Another option is the one proposed and tested by Maguire and colleagues (2021), who "enrich" water sampled from irrigation systems. This procedure consists in adding buffered peptone water to the samples and incubating them for a period before extracting DNA and sequencing. They demonstrate that this way they are able to detect a specific strain of E. coli at 10^3 CFUs mL^{-1}, and assemble a complete genome with $10^5 – 10^6$ CFUs mL^{-1}. These techniques could aid in the development of a practical naïve surveillance system.

7.7. On LamPORE

LAMP has, over its two decades of existence, been modified a number of times, often in ways similar to PCR. Some of these variations are reverse-transcriptase LAMP, which can be used to target RNA viruses (Huang et al., 2020), fluorescent LAMP, which allows simple visualisation of the result (Tomita et al., 2008), a quantitative LAMP that relies on the turbidity generated by phosphate precipitation during the reaction (Mori et al., 2004), etc. There have even been instances of multiplex LAMP (Iseki et al., 2007; J. Kim et al., 2021; Tanner et al., 2012; Y. Wang et al., 2015), which, as discussed in Chapter 6, have often involved the use of dye-bound probes, in a way similar to qPCR. However, in recent times the SARS-CoV-2 pandemic has sparked interest in this technique, including the innovative modification of LamPORE (James et al., 2020; Ptasinska et al., 2021). LamPORE involves the amplification of three regions of the viral genome and one internal standard (human actin gene) per sample via multiplex LAMP, which are then detected and identified by nanopore sequencing. By combining eight different barcodes with the LAMP primers and another 12 during library preparation, it is possible to process 96 samples simultaneously.

Logically, in the same way three different primer sets are used to target three different viral genes, three primer sets as the ones developed in Chapter 6 could be used to test for a broad range of bacterial pathogens. This has the potential to increase the number of simultaneous tests to 96, bringing the costs of broad-range LAMP down, making its implementation feasible in cost-sensitive aquaculture operations.

7.8. Overall Conclusion

This thesis has explored how a number of modern techniques and technologies can serve to improve the detection and characterisation of seven relevant bacterial pathogens of farmed salmonids. Specifically, the use of LAMP, nanopore sequencing and environmental DNA studies can serve to obtain relevant data to aquaculture operations and to do so in a way that is relatively inexpensive, fast, simple and de-centralised. Colorimetric LAMP offers a means of detecting specific pathogens in a fast and simple manner, which does not require complex or particularly expensive equipment to run and for readout of results. Nanopore sequencing has the potential to serve in the characterisation of pathogenic organisms' antimicrobial resistance profiles, is portable and has low entry costs compared with other sequencing technologies. Sequencing of environmental DNA (eDNA) could allow for a naïve pathogen monitoring system, reducing number of tests carried out and identifying pathogens before the onset of symptoms in the farmed fish.

Throughout this thesis, these different methods and technologies have been applied and tested. Effective LAMP assays have been developed for the detection of six bacterial pathogens of cultured salmonids, with three of them targeting organisms for which no LAMP assays existed previously (*F. branchiophilum*, *M. viscosa* and *R. salmoninarum)*. Although their limits of detection have been observed to be higher than other published LAMP assays, they still have advantages over PCR, the current "gold standard". They are faster (reactions run within one hour), the interpretation of results is simple (a colour change indicates a positive reaction), and they do not require particularly expensive or difficult to operate equipment (reactions can be run on a heating block). Similarly, nanopore sequencing has been observed to offer advantages over other methods, as Illumina's sequencing by synthesis, in its portability and low entry costs, but has important disadvantages. Assemblies based on nanopore reads could not be annotated correctly, probably as a consequence of the high error rate of the reads. Although assemblies based on these reads performed well for the purpose of identifying antimicrobial resistance genes or resolving ribosomal RNA genes, their quality in terms of completeness and annotation accuracy were low. It is also a disadvantage of nanopore sequencing that library preparation requires large amounts of DNA, which it was not possible to achieve from river water, even with the optimal methods devised in Chapter 4 (which involve centrifuging the water and extracting DNA from the pellet with a commercial PowerSoil kit). This can hamper pure metagenomic sequencing analysis of the microbiome of aquaculture waters. Because the purpose of sequencing eDNA was to establish a method for

the detection of a broad range of pathogens, a broad-range LAMP was developed to amplify the DNA of a number of them, with the intention of nanopore sequencing being used to identify the specific organism based on the amplified sequence. The proof of concept of such a broad-range LAMP method is given in Chapter 6, however, it requires refining before it can be considered viable.

All of these methods and technologies have potential to serve as valuable tools in the management of aquaculture diseases, but it is important to note that the research carried out for this thesis requires further work before it can be applied in the field. For instance, the use of the specific LAMP primers designed in Chapter 5 would require testing with samples from diseased and healthy fish, ideally in comparison against established PCR methods, to determine sensitivity and specificity of the assays. To ascertain the validity of the optimised methods of concentrating pathogenic bacteria from water and extracting their DNA for detection, experiments should be carried out where DNA of the bacterial pathogens of interest was added to a water sample, and the extract tested with a specific assay. To ensure the long-read assembly methods described in Chapter 3 can actually be run on a personal computer, assemblies should be carried out on such devices. Finally, a re-design of the primers used for broad-range LAMP should be carried out to ensure the amplified regions contain sufficient variation to discern closely related bacteria.

Finally, it is worth noting that the methods and technologies mentioned here are evolving rapidly, especially nanopore sequencing. Their continual refinement has the potential to refine or improve methods explored in this thesis, or to make them unnecessary. For instance, the issues with low quality of long-read-based assemblies may be resolved as nanopore sequencing accuracy improves; techniques such as CarrierSeq (Mojarro et al., 2018) can serve to apply pure metagenomics to low DNA samples; or broad-range LAMP may be either made unnecessary by LamPORE (James et al., 2020) or increase its capabilities enormously.

In any case, the tools developed or tested during this thesis have the potential to contribute to the improvement of pathogen detection, monitoring and management, especially in salmonid aquaculture, remote locations and low-resource settings. Ultimately, this is expected to have positive repercussions for food security by aiding in the sustainable expansion of the aquaculture industry.

References

Aanensen, D. M., Feil, E. J., Holden, M. T. G., Dordel, J., Yeats, C. A., Fedosejev, A., Goater, R., Castillo-Ramírez, S., Corander, J., Colijn, C., Chlebowicz, M. A., Schouls, L., Heck, M., Pluister, G., Ruimy, R., Kahlmeter, G., Åhman, J., Matuschek, E., Friedrich, A. W., … Grundmann, H. (2016). Whole-Genome Sequencing for Routine Pathogen Surveillance in Public Health: A Population Snapshot of Invasive *Staphylococcus aureus* in Europe. *MBio*, *7*(3). https://doi.org/10.1128/mBio.00444-16

Achtman, M. (2008). Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annual Review of Microbiology*, *62*(1), 53–70. https://doi.org/10.1146/annurev.micro.62.081307.162832

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Applied and Environmental Microbiology*, *71*(12), 8966–8969. https://doi.org/10.1128/AEM.71.12.8966-8969.2005

Adams, A., & Thompson, K. (1990). Development of an Enzyme-Linked Immunosorbent Assay (ELISA) for the Detection of *Aeromonas salmonicida* in Fish Tissue. *Journal of Aquatic Animal Health*, *2*(4), 9. https://doi.org/10.1577/1548-8667(1990)002%3C0281:DOAELI%3E2.3.CO;2

Adams, A., & Thompson, K. D. (2011). Development of diagnostics for aquaculture: Challenges and opportunities: Development of diagnostics for aquaculture. *Aquaculture Research*, *42*(s1), 93–102. https://doi.org/10.1111/j.1365-2109.2010.02663.x

Adelowo, O. O., Caucci, S., Banjo, O. A., Nnanna, O. C., Awotipe, E. O., Peters, F. B., Fagade, O. E., & Berendonk, T. U. (2018). Extended Spectrum Beta-Lactamase (ESBL)-producing bacteria isolated from hospital wastewaters, rivers and aquaculture sources in Nigeria. *Environmental Science and Pollution Research*, *25*(3), 2744–2755. https://doi.org/10.1007/s11356-017-0686-7

Alcantara, R. (2019, March 29). *Demultiplexer comparison*. Nanopore Community. https://community.nanoporetech.com/posts/demultiplexer-comparison

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., Min, S. Y., Miroshnichenko, A., Tran, H.-K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., … McArthur, A. G. (2019). CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517–D525. https://doi.org/10.1093/nar/gkz935

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Álvarez, B., & Guijarro, J. A. (2007). Recovery of *Flavobacterium psychrophilum* viable cells using a charcoal-based solid medium. *Letters in Applied Microbiology*, *44*(5), 569–572. https://doi.org/10.1111/j.1472-765X.2007.02126.x

Andersen, L., Hodneland, K., & Nylund, A. (2010). No influence of oxygen levels on pathogenesis and virus shedding in Salmonid alphavirus (SAV)-challenged Atlantic salmon (Salmo salar L.). *Virology Journal*, *7*(1), 198. https://doi.org/10.1186/1743-422X-7-198

Apweiler, R. (2004). UniProt: The Universal Protein knowledgebase. *Nucleic Acids Research*, *32*, D115-119. https://doi.org/10.1093/nar/gkh131

Arahal, D. R. (2014). Whole-Genome Analyses. In *Methods in Microbiology* (Vol. 41, pp. 103–122). Elsevier. https://doi.org/10.1016/bs.mim.2014.07.002

Asche, F., Hansen, H., & Tveteras, R. (2009). The Salmon Disease Crisis in Chile. *Marine Resource Economics*, *24*(4), 405–411.

Attéré, S. A., Vincent, A. T., Paccaud, M., Frenette, M., & Charette, S. J. (2017). The Role for the Small Cryptic Plasmids As Moldable Vectors for Genetic Innovation in *Aeromonas salmonicida* subsp. *Salmonicida*. *Frontiers in Genetics*, *8*, 211. https://doi.org/10.3389/fgene.2017.00211

Austin, B., & Austin, D. A. (2007). *Bacterial fish pathogens: Disease of farmed and wild fish* (4th ed). Springer; Published in association with Praxis Pub.

Austin, D. A., McIntosh, D., & Austin, B. (1989). Taxonomy of Fish Associated *Aeromonas* spp., with the Description of *Aeromonas salmonicida* subsp. *Smithia* subsp. Nov. *Systematic and Applied Microbiology*, *11*(3), 277–290. https://doi.org/10.1016/S0723-2020(89)80026-8

Avendaño-Herrera, R., Magariños, B., Toranzo, A., Beaz, R., & Romalde, J. (2004). Species-specific polymerase chain reaction primer sets for the diagnosis of *Tenacibaculum maritimum* infection. *Diseases of Aquatic Organisms*, *62*, 75–83. https://doi.org/10.3354/dao062075

Baker, S., Holt, K., van de Vosse, E., Roumagnac, P., Whitehead, S., King, E., Ewels, P., Keniry, A., Weill, F.-X., Lightfoot, D., van Dissel, J. T., Sanderson, K. E., Farrar, J., Achtman, M., Deloukas, P., & Dougan, G. (2008). High-Throughput Genotyping of *Salmonella enterica* Serovar Typhi Allowing Geographical Assignment of Haplotypes and Pathotypes within an Urban District of Jakarta, Indonesia. *Journal of Clinical Microbiology*, *46*(5), 1741–1746. https://doi.org/10.1128/JCM.02249-07

Barnes, A. C., Delamare-Deboutteville, J., Gudkovs, N., Brosnahan, C., Morrison, R., & Carson, J. (2016). Whole genome analysis of *Yersinia ruckeri* isolated over 27 years in Australia and New Zealand reveals geographical endemism over multiple lineages and recent evolution under host selection. *Microbial Genomics*, *2*(11), e000095. https://doi.org/10.1099/mgen.0.000095

Bayliss, S. C., Verner-Jeffreys, D. W., Ryder, D., Suarez, R., Ramirez, R., Romero, J., Pascoe, B., Sheppard, S. K., Godoy, M., & Feil, E. J. (2018). Genomic epidemiology of the commercially important pathogen *Renibacterium salmoninarum* within the

Chilean salmon industry. *Microbial Genomics*, *4*(9), e000201. https://doi.org/10.1099/mgen.0.000201

Beaz-Hidalgo, R., Hossain, M. J., Liles, M. R., & Figueras, M.-J. (2015). Strategies to Avoid Wrongly Labelled Genomes Using as Example the Detected Wrong Taxonomic Affiliation for Aeromonas Genomes in the GenBank Database. *PLOS ONE*, *10*(1), e0115813. https://doi.org/10.1371/journal.pone.0115813

Beckman-Coulter. (n.d.). *AMPure XP Performance, PCR Purification*. Beckman Coulter Life Sciences. Retrieved October 27, 2021, from https://www.mybeckman.uk/reagents/genomic/cleanup-and-size-selection/pcr/performance

Benediktsdóttir, E., Verdonck, L., Spröer, C., Helgason, S., & Swings, J. (2000). Characterization of *Vibrio viscosus* and *Vibrio wodanis* isolated at different geographical locations: A proposal for reclassification of *Vibrio viscosus* as *Moritella viscosa* comb. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *50*(2), 479–488. https://doi.org/10.1099/00207713-50-2-479

Benítez-Páez, A., & Sanz, Y. (2017). Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION™ portable nanopore sequencer. *GigaScience*, *6*(7), 1–12. https://doi.org/10.1093/gigascience/gix043

Berlin, K., Koren, S., Chin, C.-S., Drake, J. P., Landolin, J. M., & Phillippy, A. M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature Biotechnology*, *33*(6), 623–630. https://doi.org/10.1038/nbt.3238

Bernardet, J.-F., & Grimont, P. A. D. (1989). Deoxyribonucleic acid relatedness and phenotypic characterization of <i>Flexibacter columnaris<i\> sp. Nov., nom. Rev., Flexibacter psychrophilus sp. Nov., nom. Rev., and Flexibacter maritimus Wakabayashi, Hikida, and Masumura 1986. *International Journal of Systematic and Evolutionary Microbiology*, *39*(3), 346–354.

Bernardet, J.-F., Segers, P., Vancanneyt, M., Berthe, F., Kersters, K., & Vandamme, P. (1996). Cutting a Gordian knot: Emended classification and description of the genus *Flavobacterium*, emended description of the family Flavobacteriaceae, and proposal of *Flavobacterium hydatis* nom. Nov. (Basonym, *Cytophaga aquatilis* Strohl and Tait 1978). *International Journal of Systematic and Evolutionary Microbiology*, *46*(1), 128–148. https://doi.org/10.1099/00207713-46-1-128

Bernoth, E.-M. (1997). *Furunculosis: Multidisciplinary fish disease research*. Academic Press.

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, *24*(4), 335–341. https://doi.org/10.1016/j.cmi.2017.10.013

Bethke, J., Quezada, J., Poblete-Morales, M., Irgang, R., Yáñez, A., Oliver, C., Avendaño-Herrera, R., & others. (2017). Biochemical, serological, and genetic characterisation

of *Renibacterium salmoninarum* isolates recovered from salmonids in Chile. *Bull Eur Assoc Fish Pathol*, *37*, 169–180.

Bibby, K., Viau, E., & Peccia, J. (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples: Viral metagenome reveals human pathogen content. *Letters in Applied Microbiology*, *52*(4), 386–392. https://doi.org/10.1111/j.1472-765X.2011.03014.x

Bitting, A. L., Bordelon, H., Baglia, M. L., Davis, K. M., Creecy, A. E., Short, P. A., Albert, L. E., Karhade, A. V., Wright, D. W., Haselton, F. R., & Adams, N. M. (2016). Automated Device for Asynchronous Extraction of RNA, DNA, or Protein Biomarkers from Surrogate Patient Samples. *Journal of Laboratory Automation*, *21*(6), 732–742. https://doi.org/10.1177/2211068215596139

Bogaerts, P., Naas, T., Saegeman, V., Bonnin, R. A., Schuermans, A., Evrard, S., Bouchahrouf, W., Jove, T., Tande, D., de Bolle, X., Huang, T.-D., Dortet, L., & Glupczynski, Y. (2017). OXA-427, a new plasmid-borne carbapenem-hydrolysing class D β-lactamase in Enterobacteriaceae. *Journal of Antimicrobial Chemotherapy*, *72*(9), 2469–2477. https://doi.org/10.1093/jac/dkx184

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, *28*(3), 495–503. https://doi.org/10.1128/jcm.28.3.495-503.1990

Bordelon, H., Russ, P. K., Wright, D. W., & Haselton, F. R. (2013). A Magnetic Bead-Based Method for Concentrating DNA from Human Urine for Downstream Detection. *PLoS ONE*, *8*(7), e68369. https://doi.org/10.1371/journal.pone.0068369

Borg, A. F. (1948). *Studies on myxobacteria associated with diseases in salmonid fishes*. University of Washington.

Boulanger, C. A., & Edelstein, P. H. (1995). Precision and accuracy of recovery of Legionella pneumophila from seeded tap water by filtration and centrifugation. *Applied and Environmental Microbiology*, *61*(5), 1805–1809. https://doi.org/10.1128/aem.61.5.1805-1809.1995

Boykin, L. M., Sseruwagi, P., Alicai, T., Ateka, E., Mohammed, I. U., Stanton, J.-A. L., Kayuki, C., Mark, D., Fute, T., Erasto, J., Bachwenkizi, H., Muga, B., Mumo, N., Mwangi, J., Abidrabo, P., Okao-Okuja, G., Omuut, G., Akol, J., Apio, H. B., … Ndunguru, J. (2019). Tree Lab: Portable genomics for Early Detection of Plant Viruses and Pests in Sub-Saharan Africa. *Genes*, *10*(9), 632. https://doi.org/10.3390/genes10090632

Brewer, T. E., Albertsen, M., Edwards, A., Kirkegaard, R. H., Rocha, E. P. C., & Fierer, N. (2020). Unlinked rRNA genes are widespread among bacteria and archaea. *The ISME Journal*, *14*(2), 597–608. https://doi.org/10.1038/s41396-019-0552-3

Bridel, S., Bourgeon, F., Marie, A., Saulnier, D., Pasek, S., Nicolas, P., Bernardet, J.-F., & Duchaud, E. (2020). Genetic diversity and population structure of *Tenacibaculum maritimum*, a serious bacterial pathogen of marine fish: From genome comparisons to

high throughput MALDI-TOF typing. *Veterinary Research*, *51*(1), 60. https://doi.org/10.1186/s13567-020-00782-0

Brindle, R. J., Stannett, P. J., & Cunliffe, R. N. (1987). *Legionella pneumophila*: Comparison of isolation from water specimens by centrifugation and filtration. *Epidemiology and Infection*, *99*(2), 241–247. https://doi.org/10.1017/S0950268800067704

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., Buck, G. A., & Vaginal Microbiome Consortium (additional members). (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, *15*, 66. https://doi.org/10.1186/s12866-015-0351-6

Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., & Franklin, R. B. (2017). MinION™ nanopore sequencing of environmental metagenomes: A synthetic approach. *GigaScience*, *6*(3), 1–10. https://doi.org/10.1093/gigascience/gix007

Brown, C. (2012). Single molecule strand sequencing using protein nanopores and scalable electronic devices. *AGBT Conference*.

Brown, L., Iwama, G., Evelyn, T., Nelson, W., Levine, R., & others. (1994). Use of the polymerase chain reaction (PCR) to detect DNA from *Renibacterium salmoninarum* within individual salmonid eggs. *Diseases of Aquatic Organisms*, *18*(3), 165–171. https://doi.org/10.3354/dao018165

Brynildsrud, O., Feil, E. J., Bohlin, J., Castillo-Ramirez, S., Colquhoun, D., McCarthy, U., Matejusova, I. M., Rhodes, L. D., Wiens, G. D., & Verner-Jeffreys, D. W. (2014). Microevolution of *Renibacterium salmoninarum*: Evidence for intercontinental dissemination associated with fish movements. *The ISME Journal*, *8*(4), 746–756. https://doi.org/10.1038/ismej.2013.186

Cai, S. H., Lu, Y. S., Wu, Z.-H., Jian, J. C., Wang, B., & Huang, Y. C. (2010). Loop-mediated isothermal amplification method for rapid detection of *Vibrio alginolyticus* , the causative agent of vibriosis in mariculture fish. *Letters in Applied Microbiology*, *50*(5), 480–485. https://doi.org/10.1111/j.1472-765X.2010.02823.x

Caldwell, J. M., Raley, M. E., & Levine, J. F. (2007). Mitochondrial Multiplex Real-Time PCR as a Source Tracking Method in Fecal-Contaminated Effluents. *Environmental Science & Technology*, *41*(9), 3277–3283. https://doi.org/10.1021/es062912s

Cao, M. D., Ganesamoorthy, D., Elliott, A. G., Zhang, H., Cooper, M. A., & Coin, L. J. M. (2016). Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION™ sequencing. *GigaScience*, *5*(1), 32. https://doi.org/10.1186/s13742-016-0137-2

Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Höfle, M. G., Klappenbach, J., & Konstantinidis, K. T. (2011). Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *The ISME Journal*, *5*(1), 131–140. https://doi.org/10.1038/ismej.2010.93

Carter, C., Akrami, K., Hall, D., Smith, D., & Aronoff-Spencer, E. (2017). Lyophilized visually readable loop-mediated isothermal reverse transcriptase nucleic acid

amplification test for detection Ebola Zaire RNA. *Journal of Virological Methods*, *244*, 32–38. https://doi.org/10.1016/j.jviromet.2017.02.013

Casas, I., Powell, L., Klapper, P. E., & Cleator, G. M. (1995). New method for the extraction of viral RNA and DNA from cerebrospinal fluid for use in the polymerase chain reaction assay. *Journal of Virological Methods*, *53*(1), 25–36. https://doi.org/10.1016/0166-0934(94)00173-E

Castillo, D., Christiansen, R. H., Dalsgaard, I., Madsen, L., Espejo, R., & Middelboe, M. (2016). Comparative Genome Analysis Provides Insights into the Pathogenicity of *Flavobacterium psychrophilum*. *PLOS ONE*, *11*(4), e0152515. https://doi.org/10.1371/journal.pone.0152515

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Castresana, J. (2002). *Castresana Lab. Gblocks Server*. Castresana Lab: Animal Biodiversity and Evolution Program. http://molevol.cmima.csic.es/castresana/Gblocks_server.html

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, *69*(2), 330–339. https://doi.org/10.1016/j.mimet.2007.02.005

Chase, D., & Pascho, R. (1998). Development of a nested polymerase chain reaction for amplification of a sequence of the p57 gene of *Renibacterium salmoninarum* that provides a highly sensitive method for detection of the bacterium in salmonid kidney. *Diseases of Aquatic Organisms*, *34*, 223–229. https://doi.org/10.3354/dao034223

Chen, L. (2004). VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Research*, *33*(Database issue), D325–D328. https://doi.org/10.1093/nar/gki008

Chen, W.-Y., Ng, T. H., Wu, J.-H., Chen, J.-W., & Wang, H.-C. (2017). Microbiome Dynamics in a Shrimp Grow-out Pond with Possible Outbreak of Acute Hepatopancreatic Necrosis Disease. *Scientific Reports*, *7*(1), 9395. https://doi.org/10.1038/s41598-017-09923-6

Chern, E. C., Siefring, S., Paar, J., Doolittle, M., & Haugland, R. A. (2011). Comparison of quantitative PCR assays for *Escherichia coli* targeting ribosomal RNA and single copy genes: QPCR assays for *E. coli* targeting rRNA and single copy genes. *Letters in Applied Microbiology*, *52*(3), 298–306. https://doi.org/10.1111/j.1472-765X.2010.03001.x

Chow, F. W.-N., Chan, T. T.-Y., Tam, A. R., Zhao, S., Yao, W., Fung, J., Cheng, F. K.-K., Lo, G. C.-S., Chu, S., Aw-Yong, K. L., Tang, J. Y.-M., Tsang, C.-C., Luk, H. K.-H., Wong, A. C.-P., Li, K. S.-M., Zhu, L., He, Z., Tam, E. W. T., Chung, T. W.-H., … Lau, S. K.-P. (2020). A Rapid, Simple, Inexpensive, and Mobile Colorimetric Assay COVID-19-LAMP for Mass On-Site Screening of COVID-19. *International Journal of Molecular Sciences*, *21*(15), 5380. https://doi.org/10.3390/ijms21155380

Ciric, M., Waite, D., Draper, J., & Jones, J. (2019). Characterization of mid-intestinal microbiota of farmed Chinook salmon using 16S rRNA gene metabarcoding. *Archives of Biological Sciences*, *71*(4), 577–587. https://doi.org/10.2298/ABS190402040C

Colquhoun, D., Hovland, H., Hellberg, H., Haug, T., & Nilsen, H. (2004). *Moritella viscosa* isolated from farmed Atlantic cod (*Gadus morhua*). *Bulletin-European Association of Fish Pathologists*, *24*(2), 109–114.

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., & Graf, J. (2014). Bioinformatic Genome Comparisons for Taxonomic and Phylogenetic Assignments Using *Aeromonas* as a Test Case. *MBio*, *5*(6). https://doi.org/10.1128/mBio.02136-14

Condon, K., Huerlimann, R., Charles, T., Lowrey, R., Arbon, P., & Jerry, D. R. (2022). Using bacterial whole genome sequencing to identify toxin genes associated with disease outbreaks in black tiger shrimp (*Penaeus monodon*) aquaculture production. *Aquaculture*, *546*, 737255. https://doi.org/10.1016/j.aquaculture.2021.737255

Costello, M. J. (2009). How sea lice from salmon farms may cause wild salmonid declines in Europe and North America and be a threat to fishes elsewhere. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1672), 3385–3394. https://doi.org/10.1098/rspb.2009.0771

Cuadros, J., Pérez-Tanoira, R., Prieto-Pérez, L., Martin-Martin, I., Berzosa, P., González, V., Tisiano, G., Balcha, S., Ramos, J. M., & Górgolas, M. (2015). Field Evaluation of Malaria Microscopy, Rapid Malaria Tests and Loop-Mediated Isothermal Amplification in a Rural Hospital in South Western Ethiopia. *PLOS ONE*, *10*(11), e0142842. https://doi.org/10.1371/journal.pone.0142842

Cuscó, A., Catozzi, C., Viñes, J., Sanchez, A., & Francino, O. (2019). Microbiota profiling with long amplicons using Nanopore sequencing: Full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon [version 2; peer review: 2 approved, 3 approved with reservations]. *F1000Research*, *7*(1755). https://doi.org/10.12688/f1000research.16817.2

Daly, J. G., & Stevenson, R. M. (1985). Charcoal agar, a new growth medium for the fish disease bacterium *Renibacterium salmoninarum*. *Applied and Environmental Microbiology*, *50*(4), 868–871. https://doi.org/10.1128/aem.50.4.868-871.1985

Dao Thi, V. L., Herbst, K., Boerner, K., Meurer, M., Kremer, L. P., Kirrmaier, D., Freistaedter, A., Papagiannidis, D., Galmozzi, C., Stanifer, M. L., Boulant, S., Klein, S., Chlanda, P., Khalid, D., Barreto Miranda, I., Schnitzler, P., Kräusslich, H.-G., Knop, M., & Anders, S. (2020). A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science Translational Medicine*, *12*(556), eabc7075. https://doi.org/10.1126/scitranslmed.abc7075

Davis, H. (1926). A new gill disease of trout. *Transactions of the American Fisheries Society*, *56*(1), 156–160.

Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., Vonstein, V., Wattam, A. R., & Yoo, H. (2016). PATtyFams: Protein Families for the Microbial

Genomes in the PATRIC Database. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00118

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., … Stevens, R. (2019). The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Research*, *48*(D1), D606–D612. https://doi.org/10.1093/nar/gkz943

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

de Nies, L., Lopes, S., Busi, S. B., Galata, V., Heintz-Buschart, A., Laczny, C. C., May, P., & Wilmes, P. (2021). PathoFact: A pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome*, *9*(1), 49. https://doi.org/10.1186/s40168-020-00993-9

de Vries, A., PentaLibra, actions-user, Klevtsov, A., & Augustine Guilliland, J. D. (2016). *Ggdendro* (v0.1.20) [R]. https://github.com/andrie/ggdendro (Original work published 2011)

Deamer, D. W., & Akeson, M. (2000). Nanopores and nucleic acids: Prospects for ultrarapid sequencing. *Trends in Biotechnology*, *18*(4), 147–151. https://doi.org/10.1016/S0167-7799(00)01426-8

Deiner, K., & Altermatt, F. (2014). Transport Distance of Invertebrate Environmental DNA in a Natural River. *PLoS ONE*, *9*(2), e88786. https://doi.org/10.1371/journal.pone.0088786

Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, *23*(6), 673–679. https://doi.org/10.1093/bioinformatics/btm009

Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schrider, D. R., Warren, W. C., & Hahn, M. W. (2014). Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Computational Biology*, *10*(12), e1003998. https://doi.org/10.1371/journal.pcbi.1003998

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., Belk, K. E., Noyes, N. R., & Morley, P. S. (2020). MEGARes 2.0: A database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research*, *48*(D1), D561–D569. https://doi.org/10.1093/nar/gkz1010

Duchaud, E., Rochat, T., Habib, C., Barbier, P., Loux, V., Guérin, C., Dalsgaard, I., Madsen, L., Nilsen, H., Sundell, K., Wiklund, T., Strepparava, N., Wahli, T., Caburlotto, G., Manfrin, A., Wiens, G. D., Fujiwara-Nagata, E., Avendaño-Herrera, R., Bernardet, J.-F., & Nicolas, P. (2018). Genomic Diversity and Evolution of the Fish Pathogen

Flavobacterium psychrophilum. *Frontiers in Microbiology*, *9*, 138. https://doi.org/10.3389/fmicb.2018.00138

Durso, L. M., Harhay, G. P., Bono, J. L., & Smith, T. P. L. (2011). Virulence-associated and antibiotic resistance genes of microbial populations in cattle feces analyzed using a metagenomic approach. *Journal of Microbiological Methods*, *84*(2), 278–282. https://doi.org/10.1016/j.mimet.2010.12.008

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edwards, A., Debbonaire, A. R., Nicholls, S. M., Rassner, S. M. E., Sattler, B., Cook, J. M., Davy, T., Soares, A., Mur, L. A. J., & Hodson, A. J. (2016). *In-field metagenome and 16S rRNA gene amplicon nanopore sequencing robustly characterize glacier microbiota* [Preprint]. bioRxiv. https://doi.org/10.1101/073965

Edwards, A., Mur, L. A. J., Girdwood, S. E., Anesio, A. M., Stibal, M., Rassner, S. M. E., Hell, K., Pachebat, J. A., Post, B., Bussell, J. S., Cameron, S. J. S., Griffith, G. W., Hodson, A. J., & Sattler, B. (2014). Coupled cryoconite ecosystem structure-function relationships are revealed by comparing bacterial communities in alpine and Arctic glaciers. *FEMS Microbiology Ecology*, *89*(2), 222–237. https://doi.org/10.1111/1574-6941.12283

Edwards, A., Soares, A., Rassner, S. M. E., Green, P., Félix, J., & Mitchell, A. C. (2017). *Deep Sequencing: Intra-terrestrial metagenomics illustrates the potential of off-grid Nanopore DNA sequencing* [Preprint]. bioRxiv. https://doi.org/10.1101/133413

Eiken Chemical Co., Ltd. (n.d.). *A Guide to LAMP primer designing (PrimerExplorer V5)*. Retrieved October 26, 2021, from https://primerexplorer.jp/e/v5_manual/index.html

Elliott, D. (2017). *Renibacterium salmoninarum*. In P. T. K. Moo & R. C. Cipriano (Eds.), *Fish Viruses and Bacteria* (pp. 286–297). CABI.

Euzéby, J. P. (1997). List of Bacterial Names with Standing in Nomenclature: A Folder Available on the Internet. *International Journal of Systematic and Evolutionary Microbiology*, *47*(2), 590–592. https://doi.org/10.1099/00207713-47-2-590

Ewing, W., Ross, A., Brenner, D. J., & Fanning, G. (1978). *Yersinia ruckeri* sp. Nov., the redmouth (RM) bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *28*(1), 37–44. https://doi.org/10.1099/00207713-28-1-37

FAO. (2020a). *FAO Yearbook of Fishery and Aquaculture Statistics 2018*. United Nations. . https://doi.org/10.4060/cb1213t

FAO. (2020b). *The State of World Fisheries and Aquaculture 2020. Sustainability in action.* FAO. https://doi.org/10.4060/ca9229en

FAO. (2021a). *FAO Fisheries & Aquaculture—Cultured Aquatic Species Information Programme*—Oncorhynchus mykiss *(Walbaum, 1792)*. https://www.fao.org/fishery/culturedspecies/Oncorhynchus_mykiss/en

FAO. (2021b). *FAO Fisheries & Aquaculture—Cultured Aquatic Species Information Programme—*Salmo salar *(Linnaeus, 1758)*. https://www.fao.org/fishery/culturedspecies/Salmo_salar/en

Figueroa, J., Cárcamo, J., Yañez, A., Olavarria, V., Ruiz, P., Manríquez, R., Muñoz, C., Romero, A., & Avendaño-Herrera, R. (2019). Addressing viral and bacterial threats to salmon farming in Chile: Historical contexts and perspectives for management and control. *Reviews in Aquaculture*, *11*(2), 299–324. https://doi.org/10.1111/raq.12333

Flegel, T. W., & Lo, C. F. (2013). *Announcement regarding free release of primers for specific detection of bacterial isolates that cause acute hepatopancreatic necrosis disease (AHPND)*. https://enaca.org/publications/health/disease-cards/ahpnd-detection-method-announcement.pdf

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Fong, T.-T., & Lipp, E. K. (2005). Enteric Viruses of Humans and Animals in Aquatic Environments: Health Risks, Detection, and Potential Water Quality Assessment Tools. *Microbiology and Molecular Biology Reviews*, *69*(2), 357–371. https://doi.org/10.1128/MMBR.69.2.357-371.2005

Foo, P. C., Chan, Y. Y., Mohamed, M., Wong, W. K., Nurul Najian, A. B., & Lim, B. H. (2017). Development of a thermostabilised triplex LAMP assay with dry-reagent four target lateral flow dipstick for detection of *Entamoeba histolytica* and non-pathogenic *Entamoeba* spp. *Analytica Chimica Acta*, *966*, 71–80. https://doi.org/10.1016/j.aca.2017.02.019

Foote, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L. A., Salling, A. B., Galatius, A., Orlando, L., & Gilbert, M. T. P. (2012). Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals. *PLoS ONE*, *7*(8), e41781. https://doi.org/10.1371/journal.pone.0041781

Fox, G. E., Wisotzkey, J. D., & Jurtshuk, P. (1992). How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *International Journal of Systematic Bacteriology*, *42*(1), 166–170. https://doi.org/10.1099/00207713-42-1-166

Francois, P., Tangomo, M., Hibbs, J., Bonetti, E.-J., Boehme, C. C., Notomi, T., Perkins, M. D., & Schrenzel, J. (2011). Robustness of a loop-mediated isothermal amplification reaction for diagnostic applications. *FEMS Immunology & Medical Microbiology*, *62*(1), 41–48. https://doi.org/10.1111/j.1574-695X.2011.00785.x

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., & Olsen, G. J. (2008). Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. *Applied and Environmental Microbiology*, *74*(8), 2461–2470. https://doi.org/10.1128/AEM.02272-07

Fringuelli, E., Savage, P. D., Gordon, A., Baxter, E. J., Rodger, H. D., & Graham, D. A. (2012). Development of a quantitative real-time PCR for the detection of

Tenacibaculum maritimum and its application to field samples. *Journal of Fish Diseases*, *35*(8), 579–590. https://doi.org/10.1111/j.1365-2761.2012.01377.x

Fujiwara-Nagata, E., & Eguchi, M. (2009). Development and evaluation of a loop-mediated isothermal amplification assay for rapid and simple detection of *Flavobacterium psychrophilum*. *Journal of Fish Diseases*, *32*(10), 873–881. https://doi.org/10.1111/j.1365-2761.2009.01066.x

Gabrielaite, M., & Marvig, R. L. (2020). GenAPI: A tool for gene absence-presence identification in fragmented bacterial genome sequences. *BMC Bioinformatics*, *21*(1), 320. https://doi.org/10.1186/s12859-020-03657-5

Gahlawat, S. K., Ellis, A. E., & Collet, B. (2009). A sensitive loop-mediated isothermal amplification (LAMP) method for detection of *Renibacterium salmoninarum* , causative agent of bacterial kidney disease in salmonids. *Journal of Fish Diseases*, *32*(6), 491–497. https://doi.org/10.1111/j.1365-2761.2009.01005.x

Galili, T. (2015). dendextend: An R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22). https://doi.org/10.1093/bioinformatics/btv428

Ganguli, A., Mostafa, A., Berger, J., Aydin, M. Y., Sun, F., Ramirez, S. A. S. de, Valera, E., Cunningham, B. T., King, W. P., & Bashir, R. (2020). Rapid isothermal amplification and portable detection system for SARS-CoV-2. *Proceedings of the National Academy of Sciences*, *117*(37), 22727–22735. https://doi.org/10.1073/pnas.2014739117

Garcia, S. M., & Rosenberg, A. A. (2010). Food security and marine capture fisheries: Characteristics, trends, drivers and future perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2869–2880. https://doi.org/10.1098/rstb.2010.0171

García-López, R., Cornejo-Granados, F., López-Zavala, A. A., Sánchez-López, F., Cota-Huízar, A., Sotelo-Mundo, R. R., Guerrero, A., Mendoza-Vargas, A., Gómez-Gil, B., & Ochoa-Leyva, A. (2020). Doing More with Less: A Comparison of 16S Hypervariable Regions in Search of Defining the Shrimp Microbiota. *Microorganisms*, *8*(1), 134. https://doi.org/10.3390/microorganisms8010134

Garnier, S., Ross, N., Rudis, R., Camargo, A. P., Sciaini, M., & Scherer, C. (2021). *viridis—Colorblind-Friendly Color Maps for R*. https://doi.org/10.5281/zenodo.4679424

Garver, K. A., Traxler, G. S., Hawley, L. M., Richard, J., Ross, J., & Lovy, J. (2013). Molecular epidemiology of viral haemorrhagic septicaemia virus (VHSV) in British Columbia, Canada, reveals transmission from wild to farmed fish. *Diseases of Aquatic Organisms*, *104*(2), 93–104. https://doi.org/10.3354/dao02588

Gatter, T., von Löhneysen, S., Drozdova, P., Hartmann, T., & Stadler, P. F. (2020). *Economic Genome Assembly from Low Coverage Illumina and Nanopore Data* [Preprint]. bioRxiv. https://doi.org/10.1101/2020.02.07.939454

Gautreau, G., Bazin, A., Gachet, M., Planel, R., Burlot, L., Dubois, M., Perrin, A., Médigue, C., Calteau, A., Cruveiller, S., Matias, C., Ambroise, C., Rocha, E. P. C., & Vallenet,

D. (2020). PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLOS Computational Biology*, *16*(3), e1007732. https://doi.org/10.1371/journal.pcbi.1007732

Genus Tenacibaculum. (n.d.). LPSN.Dmsz.De. Retrieved March 28, 2019, from https://lpsn.dsmz.de/genus/tenacibaculum

Gibello, A., Blanco, M. M., Moreno, M. A., Cutuli, M. T., Domenech, A., Domínguez, L., & Fernández-Garayzábal, J. F. (1999). Development of a PCR Assay for Detection of *Yersinia ruckeri* in Tissues of Inoculated and Naturally Infected Trout. *Applied and Environmental Microbiology*, *65*(1), 346–350. https://doi.org/10.1128/AEM.65.1.346-350.1999

Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martinón-Torres, F., & Salas, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Research*, *30*(10), 1434–1448. https://doi.org/10.1101/gr.266221.120

Gonçalves, A. T., Collipal-Matamal, R., Valenzuela-Muñoz, V., Nuñez-Acuña, G., Valenzuela-Miranda, D., & Gallardo-Escárate, C. (2020). Nanopore sequencing of microbial communities reveals the potential role of sea lice as a reservoir for fish pathogens. *Scientific Reports*, *10*(1), 2895. https://doi.org/10.1038/s41598-020-59747-0

González-Escalona, N., Allard, M. A., Brown, E. W., Sharma, S., & Hoffmann, M. (2019). Nanopore sequencing for fast determination of plasmids, phages, virulence markers, and antimicrobial resistance genes in Shiga toxin-producing *Escherichia coli*. *PLOS ONE*, *14*(7), e0220494. https://doi.org/10.1371/journal.pone.0220494

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91. https://doi.org/10.1099/ijs.0.64483-0

Gowers, G.-Oliver. F., Vince, O., Charles, J.-H., Klarenberg, I., Ellis, T., & Edwards, A. (2019). Entirely Off-Grid and Solar-Powered DNA Sequencing of Microbial Communities during an Ice Cap Traverse Expedition. *Genes*, *10*(11), 902. https://doi.org/10.3390/genes10110902

Grant, K., Jenkins, C., Arnold, C., Green, J., & Zambon, M. (2018). *Implementing pathogen genomics: A case study*. Public Health England.

Gray, R. R., Tatem, A. J., Johnson, J. A., Alekseyenko, A. V., Pybus, O. G., Suchard, M. A., & Salemi, M. (2011). Testing Spatiotemporal Hypothesis of Bacterial Evolution Using Methicillin-Resistant *Staphylococcus aureus* ST239 Genome-wide Data within a Bayesian Framework. *Molecular Biology and Evolution*, *28*(5), 1593–1603. https://doi.org/10.1093/molbev/msq319

Grayson, T. H., Cooper, L. F., Atienzar, F. A., Knowles, M. R., & Gilpin, M. L. (1999). Molecular Differentiation of *Renibacterium salmoninarum* Isolates from Worldwide

Locations. *Applied and Environmental Microbiology*, *65*(3), 961–968. https://doi.org/10.1128/AEM.65.3.961-968.1999

Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J. M., Dodd, R., Mulembakani, P., Schneider, B. S., Muyembe-Tamfum, J.-J., Stramer, S. L., & Chiu, C. Y. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*, *7*(1), 99. https://doi.org/10.1186/s13073-015-0220-9

Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., & Bailey, M. J. (2000). Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community Composition. *Applied and Environmental Microbiology*, *66*(12), 5488–5491. https://doi.org/10.1128/AEM.66.12.5488-5491.2000

Grove, S., Reitan, L., Lunder, T., & Colquhoun, D. (2008). Real-time PCR detection of *Moritella viscosa*, the likely causal agent of winter-ulcer in Atlantic salmon *Salmo salar* and rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms*, *82*, 105–109. https://doi.org/10.3354/dao01972

Gulla, S., Bayliss, S., Björnsdóttir, B., Dalsgaard, I., Haenen, O., Jansson, E., McCarthy, U., Scholz, F., Vercauteren, M., Verner-Jeffreys, D., Welch, T., Wiklund, T., & Colquhoun, D. J. (2019). Biogeography of the fish pathogen *Aeromonas salmonicida* inferred by *vapA* genotyping. *FEMS Microbiology Letters*, *366*(7). https://doi.org/10.1093/femsle/fnz074

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., & Rolain, J.-M. (2014). ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrobial Agents and Chemotherapy*, *58*(1), 212–220. https://doi.org/10.1128/AAC.01310-13

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gustafson, C. E., Thomas, C. J., & Trust, T. J. (1992). Detection of *Aeromonas salmonicida* from fish by using polymerase chain reaction amplification of the virulence surface array protein gene. *Applied and Environmental Microbiology*, *58*(12), 3816–3825. https://doi.org/10.1128/aem.58.12.3816-3825.1992

Haghshenas, E., Asghari, H., Stoye, J., Chauve, C., & Hach, F. (2020). HASLR: Fast Hybrid Assembly of Long Reads. *IScience*, *23*(8), 101389. https://doi.org/10.1016/j.isci.2020.101389

Haig, S. J., Davies, R. L., Welch, T. J., Reese, R. Allan., & Verner-Jeffreys, D. W. (2011). Comparative susceptibility of Atlantic salmon and rainbow trout to *Yersinia ruckeri*: Relationship to O antigen serotype and resistance to serum killing. *Veterinary Microbiology*, *147*(1–2), 155–161. https://doi.org/10.1016/j.vetmic.2010.06.022

Hamner, S., Brown, B., Hasan, N., Franklin, M., Doyle, J., Eggers, M., Colwell, R., & Ford, T. (2019). Metagenomic Profiling of Microbial Pathogens in the Little Bighorn River, Montana. *International Journal of Environmental Research and Public Health*, *16*(7), 1097. https://doi.org/10.3390/ijerph16071097

Han, H.-J., Jung, S.-J., Oh, M.-J., & Kim, D.-H. (2011). Rapid and sensitive detection of *Streptococcus iniae* by loop-mediated isothermal amplification (LAMP). *Journal of Fish Diseases*, *34*(5), 395–398. https://doi.org/10.1111/j.1365-2761.2011.01242.x

Haro-Moreno, J. M., Coutinho, F. H., Zaragoza-Solas, A., Picazo, A., Almagro-Moreno, S., & López-Pérez, M. (2020). Dysbiosis in marine aquaculture revealed through microbiome analysis: Reverse ecology for environmental sustainability. *FEMS Microbiology Ecology*, *96*(12), fiaa218. https://doi.org/10.1093/femsec/fiaa218

Haudiquet, M., Buffet, A., Rendueles, O., & Rocha, E. P. C. (2021). Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of the nosocomial pathogen Klebsiella pneumoniae. *PLOS Biology*, *19*(7), e3001276. https://doi.org/10.1371/journal.pbio.3001276

Hayashida, K., Orba, Y., Sequeira, P. C., Sugimoto, C., Hall, W. W., Eshita, Y., Suzuki, Y., Runtuwene, L., Brasil, P., Calvet, G., Rodrigues, C. D. S., Santos, C. C. dos, Mares-Guia, M. A. M., Yamagishi, J., Filippis, A. M. B. de, & Sawa, H. (2019). Field diagnosis and genotyping of chikungunya virus using a dried reverse transcription loop-mediated isothermal amplification (LAMP) assay and MinION sequencing. *PLOS Neglected Tropical Diseases*, *13*(6), e0007480. https://doi.org/10.1371/journal.pntd.0007480

He, Z., Pan, L., Zhang, M., Zhang, M., Huang, F., & Gao, S. (2020). Metagenomic comparison of structure and function of microbial community between water, effluent and shrimp intestine of higher place *Litopenaeus vannamei* ponds. *Journal of Applied Microbiology*, *129*(2), 243–255. https://doi.org/10.1111/jam.14610

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Heikema, A. P., Horst-Kreft, D., Boers, S. A., Jansen, R., Hiltemann, S. D., de Koning, W., Kraaij, R., de Ridder, M. A. J., van Houten, C. B., Bont, L. J., Stubbs, A. P., & Hays, J. P. (2020). Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota. *Genes*, *11*(9), 1105. https://doi.org/10.3390/genes11091105

Higuera-Llantén, S., Vásquez-Ponce, F., Barrientos-Espinoza, B., Mardones, F. O., Marshall, S. H., & Olivares-Pacheco, J. (2018). Extended antibiotic treatment in salmon farms select multiresistant gut bacteria with a high prevalence of antibiotic resistance genes. *PLOS ONE*, *13*(9), e0203641. https://doi.org/10.1371/journal.pone.0203641

Hiney, M., Kilmartin, J., Smith, P., & others. (1994). Detection of *Aeromonas salmonicida* in Atlantic salmon with asymptomatic furunculosis infections. *Diseases of Aquatic Organisms*, *19*(3), 161–167.

Hofmann, N., Mwingira, F., Shekalaghe, S., Robinson, L. J., Mueller, I., & Felger, I. (2015). Ultra-Sensitive Detection of *Plasmodium falciparum* by Amplification of Multi-Copy Subtelomeric Targets. *PLOS Medicine*, *12*(3), e1001788. https://doi.org/10.1371/journal.pmed.1001788

Høie, B., Aspaker, E., Listhaug, S., & Sundtoft, T. (2015). *National Strategy against Antibiotic Resistance 2015-2020*. 36.

Holten-Andersen, L., Dalsgaard, I., & Buchmann, K. (2012). Baltic Salmon, *Salmo salar*, from Swedish River Lule Älv Is More Resistant to Furunculosis Compared to Rainbow Trout. *PLoS ONE*, *7*(1), e29571. https://doi.org/10.1371/journal.pone.0029571

Hopkins, H., González, I. J., Polley, S. D., Angutoko, P., Ategeka, J., Asiimwe, C., Agaba, B., Kyabayinze, D. J., Sutherland, C. J., Perkins, M. D., & Bell, D. (2013). Highly Sensitive Detection of Malaria Parasitemia in a Malaria-Endemic Setting: Performance of a New Loop-Mediated Isothermal Amplification Kit in a Remote Clinic in Uganda. *The Journal of Infectious Diseases*, *208*(4), 645–652. https://doi.org/10.1093/infdis/jit184

Huang, W. E., Lim, B., Hsu, C., Xiong, D., Wu, W., Yu, Y., Jia, H., Wang, Y., Zeng, Y., Ji, M., Chang, H., Zhang, X., Wang, H., & Cui, Z. (2020). RT-LAMP for rapid diagnosis of coronavirus SARS-CoV-2. *Microbial Biotechnology*, *13*(4), 950–961. https://doi.org/10.1111/1751-7915.13586

Humphrey, J., Seitz, T., Haan, T., Ducluzeau, A.-L., & Drown, D. M. (2019). Complete Genome Sequence of *Pantoea agglomerans* TH81, Isolated from a Permafrost Thaw Gradient. *Microbiology Resource Announcements*, *8*(1), e01486-18. https://doi.org/10.1128/MRA.01486-18

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Illumina Inc. (2021). *Sequencing Technology | Sequencing by synthesis*. https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html

Imai, K., Tarumoto, N., Amo, K., Takahashi, M., Sakamoto, N., Kosaka, A., Kato, Y., Mikita, K., Sakai, J., Murakami, T., Suzuki, Y., Maesaki, S., & Maeda, T. (2018). Non-invasive diagnosis of cutaneous leishmaniasis by the direct boil loop-mediated isothermal amplification method and MinION™ nanopore sequencing. *Parasitology International*, *67*(1), 34–37. https://doi.org/10.1016/j.parint.2017.03.001

Imai, K., Tarumoto, N., Misawa, K., Runtuwene, L. R., Sakai, J., Hayashida, K., Eshita, Y., Maeda, R., Tuda, J., Murakami, T., Maesaki, S., Suzuki, Y., Yamagishi, J., & Maeda, T. (2017). A novel diagnostic method for malaria using loop-mediated isothermal amplification (LAMP) and MinION™ nanopore sequencer. *BMC Infectious Diseases*, *17*(1), 621. https://doi.org/10.1186/s12879-017-2718-9

Iseki, H., Alhassan, A., Ohta, N., Thekisoe, O. M. M., Yokoyama, N., Inoue, N., Nambota, A., Yasuda, J., & Igarashi, I. (2007). Development of a multiplex loop-mediated isothermal amplification (mLAMP) method for the simultaneous detection of bovine *Babesia* parasites. *Journal of Microbiological Methods*, *71*(3), 281–287. https://doi.org/10.1016/j.mimet.2007.09.019

Itano, T., Kawakami, H., Kono, T., & Sakai, M. (2006). Detection of fish nocardiosis by loop-mediated isothermal amplification. *Journal of Applied Microbiology*, *100*(6), 1381–1387. https://doi.org/10.1111/j.1365-2672.2006.02872.x

James, P., Stoddart, D., Harrington, E. D., Beaulaurier, J., Ly, L., Reid, S. W., Turner, D. J., & Juul, S. (2020). *LamPORE: Rapid, accurate and highly scalable molecular screening for SARS-CoV-2 infection, based on nanopore sequencing* [Preprint]. Infectious Diseases (except HIV/AIDS). https://doi.org/10.1101/2020.08.07.20161737

Jang, W. S., Lim, D. H., Yoon, J., Kim, A., Lim, M., Nam, J., Yanagihara, R., Ryu, S.-W., Jung, B. K., Ryoo, N.-H., & Lim, C. S. (2021). Development of a multiplex Loop-Mediated Isothermal Amplification (LAMP) assay for on-site diagnosis of SARS CoV-2. *PLOS ONE*, *16*(3), e0248042. https://doi.org/10.1371/journal.pone.0248042

Jansson, E., Lindberg, L., Säker, E., & Aspán, A. (2008). Diagnosis of bacterial kidney disease by detection of *Renibacterium salmoninarum* by real-time PCR. *Journal of Fish Diseases*, *31*(10), 755–763. https://doi.org/10.1111/j.1365-2761.2008.00949.x

Jia, B., Li, X., Liu, W., Lu, C., Lu, X., Ma, L., Li, Y.-Y., & Wei, C. (2019). GLAPD: Whole Genome Based LAMP Primer Design for a Set of Target Genomes. *Frontiers in Microbiology*, *10*, 2860. https://doi.org/10.3389/fmicb.2019.02860

Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, *10*(1), 5029. https://doi.org/10.1038/s41467-019-13036-1

Jroundi, F., Fernández-Vivas, A., Rodriguez-Navarro, C., Bedmar, E. J., & González-Muñoz, M. T. (2010). Bioconservation of Deteriorated Monumental Calcarenite Stone and Identification of Bacteria with Carbonatogenic Activity. *Microbial Ecology*, *60*(1), 39–54. https://doi.org/10.1007/s00248-010-9665-y

Kaevska, M., & Slana, I. (2015). Comparison of filtering methods, filter processing and DNA extraction kits for detection of mycobacteria in water. *Annals of Agricultural and Environmental Medicine*, *22*(3), 429–432.

Kaneko, H., Kawana, T., Fukushima, E., & Suzutani, T. (2007). Tolerance of loop-mediated isothermal amplification to a culture medium and biological substances. *Journal of Biochemical and Biophysical Methods*, *70*(3), 499–501. https://doi.org/10.1016/j.jbbm.2006.08.008

Karlsen, C., Ellingsen, A. B., Wiik-Nielsen, C., Winther-Larsen, H. C., Colquhoun, D. J., & Sørum, H. (2014). Host specificity and clade dependent distribution of putative

virulence genes in *Moritella viscosa*. *Microbial Pathogenesis*, *77*, 53–65. https://doi.org/10.1016/j.micpath.2014.09.014

Karlsen, C., Hjerde, E., Klemetsen, T., & Willassen, N. P. (2017). Pan genome and CRISPR analyses of the bacterial fish pathogen *Moritella viscosa*. *BMC Genomics*, *18*(1), 313. https://doi.org/10.1186/s12864-017-3693-7

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kerkhof, L. J., Dillon, K. P., Häggblom, M. M., & McGuinness, L. R. (2017). Profiling bacterial communities by MinION sequencing of ribosomal operons. *Microbiome*, *5*(1), 116. https://doi.org/10.1186/s40168-017-0336-9

Kim, J., Park, B. G., Lim, D. H., Jang, W. S., Nam, J., Mihn, D.-C., & Lim, C. S. (2021). Development and evaluation of a multiplex loop-mediated isothermal amplification (LAMP) assay for differentiation of *Mycobacterium tuberculosis* and non-tuberculosis *Mycobacterium* in clinical samples. *PLOS ONE*, *16*(1), e0244753. https://doi.org/10.1371/journal.pone.0244753

Kim, M., Oh, H.-S., Park, S.-C., & Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, *64*(Pt 2), 346–351. https://doi.org/10.1099/ijs.0.059774-0

Kini, K., Wonni, I., Silué, D., & Koebnik, R. (2021). Development of two loop-mediated isothermal amplification (LAMP) genomics-informed diagnostic protocols for rapid detection of *Pantoea* species on rice. *MethodsX*, *8*, 101216. https://doi.org/10.1016/j.mex.2021.101216

Kobayashi, M., Msangi, S., Batka, M., Vannuccini, S., Dey, M. M., & Anderson, J. L. (2015). Fish to 2030: The Role and Opportunity for Aquaculture. *Aquaculture Economics & Management*, *19*(3), 282–300. https://doi.org/10.1080/13657305.2015.994240

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Königsson, M. H., Ballagi, A., Jansson, E., & Johansson, K.-E. (2005). Detection of *Renibacterium salmoninarum* in tissue samples by sequence capture and fluorescent PCR based on the 16S rRNA gene. *Veterinary Microbiology*, *105*(3), 235–243. https://doi.org/10.1016/j.vetmic.2004.11.007

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, *102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Koren, S., Phillippy, A. M., Simpson, J. T., Loman, N. J., & Loose, M. (2019). Reply to 'Errors in long-read assemblies can critically affect protein prediction.' *Nature Biotechnology*, *37*(2), 127–128. https://doi.org/10.1038/s41587-018-0005-y

Krkošek, M., Lewis, M. A., & Volpe, J. P. (2005). Transmission dynamics of parasitic sea lice from farm to wild salmon. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1564), 689–696. https://doi.org/10.1098/rspb.2004.3027

Kuan, C.-P., Wu, M.-T., Lu, Y.-L., & Huang, H.-C. (2010). Rapid detection of squash leaf curl virus by loop-mediated isothermal amplification. *Journal of Virological Methods*, *169*(1), 61–65. https://doi.org/10.1016/j.jviromet.2010.06.017

Kulkarni, A., Caipang, C. M. A., Brinchmann, M. F., Korsnes, K., & Kiron, V. (2009a). Loop-mediated isothermal amplification–an assay for the detection of atypical furunculosis caused by *Aeromonas salmonicida* in Atlantic cod, *Gadus morhua*. *Journal of Rapid Methods & Automation in Microbiology*, *17*(4), 476–489. https://doi.org/10.1111/j.1745-4581.2009.00184.x

Kulkarni, A., Caipang, C. M. A., Brinchmann, M. F., Korsnes, K., & Kiron, V. (2009b). Use of loop-mediated isothermal amplification assay for the detection of *Vibrio anguillarum* O2β, the causative agent of vibriosis in Atlantic cod, *Gadus morhua*. *Journal of Rapid Methods & Automation in Microbiology*, *17*(4), 503–518. https://doi.org/10.1111/j.1745-4581.2009.00186.x

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Kumru, S., Tekedar, H. C., Blom, J., Lawrence, M. L., & Karsi, A. (2020). Genomic diversity in flavobacterial pathogens of aquatic origin. *Microbial Pathogenesis*, *142*, 104053. https://doi.org/10.1016/j.micpath.2020.104053

Kunanopparat, A., Chaivisuthangkura, P., Senapin, S., Longyant, S., Rukpratanporn, S., Flegel, T. W., & Sithigorngul, P. (2011). Detection of infectious myonecrosis virus using monoclonal antibody specific to N and C fragments of the capsid protein expressed heterologously. *Journal of Virological Methods*, *171*(1), 141–148. https://doi.org/10.1016/j.jviromet.2010.10.015

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 9.

Lalli, M. A., Langmade, J. S., Chen, X., Fronick, C. C., Sawyer, C. S., Burcea, L. C., Wilkinson, M. N., Fulton, R. S., Heinz, M., Buchser, W. J., Head, R. D., Mitra, R. D., & Milbrandt, J. (2021). Rapid and Extraction-Free Detection of SARS-CoV-2 from Saliva by Colorimetric Reverse-Transcription Loop-Mediated Isothermal Amplification. *Clinical Chemistry*, *67*(2), 415–424. https://doi.org/10.1093/clinchem/hvaa267

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lavilla-Lerma, L., Pérez-Pulido, R., Martínez-Bueno, M., Maqueda, M., & Valdivia, E. (2013). Characterization of functional, safety, and gut survival related characteristics of *Lactobacillus* strains isolated from farmhouse goat's milk cheeses. *International Journal of Food Microbiology*, *163*(2–3), 136–145. https://doi.org/10.1016/j.ijfoodmicro.2013.02.015

Lee, C.-T., Chen, I.-T., Yang, Y.-T., Ko, T.-P., Huang, Y.-T., Huang, J.-Y., Huang, M.-F., Lin, S.-J., Chen, C.-Y., Lin, S.-S., Lightner, D. V., Wang, H.-C., Wang, A. H.-J., Wang, H.-C., Hor, L.-I., & Lo, C.-F. (2015). The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences*, *112*(34), 10798–10803. https://doi.org/10.1073/pnas.1503129112

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., Richard McCombie, W., & Schatz, M. C. (2016). *Third-generation sequencing and the future of genomics* [Preprint]. bioRxiv. https://doi.org/10.1101/048603

Lehmann, K., & Neumann, R. (1896). *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik.* J.F. Lehmann.

Leidenfrost, R. M., Pöther, D.-C., Jäckel, U., & Wünschiers, R. (2020). Benchmarking the MinION: Evaluating long reads for microbial profiling. *Scientific Reports*, *10*(1), 5125. https://doi.org/10.1038/s41598-020-61989-x

Lemonnier, H., Lantoine, F., Courties, C., Guillebault, D., Nézan, E., Chomérat, N., Escoubeyrou, K., Galinié, C., Blockmans, B., & Laugier, T. (2016). Dynamics of phytoplankton communities in eutrophying tropical shrimp ponds affected by vibriosis. *Marine Pollution Bulletin*, *110*(1), 449–459. https://doi.org/10.1016/j.marpolbul.2016.06.015

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:1303.3997*, *00*(00), 1–3. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2016). Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics*, *32*(14), 2103–2110. https://doi.org/10.1093/bioinformatics/btw152

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Quan, W., Yan, S., Wu, S., Qin, J., Yang, T., Liang, F., Wang, D., & Liang, Y. (2020). *Rapid detection of SARS-CoV-2 and other respiratory viruses by using LAMP method*

with Nanopore Flongle workflow [Preprint]. bioRxiv. https://doi.org/10.1101/2020.06.03.131474

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, *12*(8), 733–735. https://doi.org/10.1038/nmeth.3444

Løvoll, M., Wiik-Nielsen, C. R., Tunsjø, H. S., Colquhoun, D., Lunder, T., Sørum, H., & Grove, S. (2009). Atlantic salmon bath challenged with *Moritella viscosa* – Pathogen invasion and host response. *Fish & Shellfish Immunology*, *26*(6), 877–884. https://doi.org/10.1016/j.fsi.2009.03.019

Lovy, J., Piesik, P., Hershberger, P. K., & Garver, K. A. (2013). Experimental infection studies demonstrating Atlantic salmon as a host and reservoir of viral hemorrhagic septicemia virus type IVa with insights into pathology and host immunity. *Veterinary Microbiology*, *166*(1–2), 91–101. https://doi.org/10.1016/j.vetmic.2013.05.019

Lowe, B. A., Marsh, T. L., Isaacs-Cosgrove, N., Kirkwood, R. N., Kiupel, M., & Mulks, M. H. (2011). Microbial communities in the tonsils of healthy pigs. *Veterinary Microbiology*, *147*(3–4), 346–357. https://doi.org/10.1016/j.vetmic.2010.06.025

Lu, X., Zhang, X.-X., Wang, Z., Huang, K., Wang, Y., Liang, W., Tan, Y., Liu, B., & Tang, J. (2015). Bacterial Pathogens and Community Composition in Advanced Sewage Treatment Systems Revealed by Metagenomics Analysis Based on High-Throughput Sequencing. *PLOS ONE*, *10*(5), e0125549. https://doi.org/10.1371/journal.pone.0125549

Lucigen Corporation. (2017). *LavaLAMP™ DNA Master Mix*. https://www.lucigen.com/docs/manuals/MA167-LavaLAMP-DNA-Master-Mix-User-Manual.pdf

Lunder, T., Eversen, Ø., Holstad, G., & Håstein, T. (1995). "Winter ulcer" in the Atlantic salmon *Salmo salar*. Pathological and bacteriological investigations and transmission experiments. *Diseases of Aquatic Organisms*, *23*(1), 39–49. https://doi.org/10.3354/dao023039

Lunder, T., Sørum, H., Holstad, G., Steigerwalt, A. G., Mowinckel, P., & Brenner, D. J. (2000). Phenotypic and genotypic characterization of *Vibrio viscosus* sp. Nov. And *Vibrio wodanis* sp. Nov. Isolated from Atlantic salmon (*Salmo salar*) with "winter ulcer". *International Journal of Systematic and Evolutionary Microbiology*, *50*(2), 427–450. https://doi.org/10.1099/00207713-50-2-427

MacPhee, D., Ostland, V., Lumsden, J., & Ferguson, H. (1995). Development of an enzyme-linked immunosorbent assay (ELISA) to estimate the quantity of *Flavobacterium branchiophilum* on the gills of rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms*, *21*, 13–23. https://doi.org/10.3354/dao021013

Magnússon, H. B., Fridjónsson, O. H., Andrésson, O. S., Benediktsdóttir, E., Gudmundsdóttir, S., & Andrésdóttir, V. (1994). *Renibacterium salmoninarum*, the causative agent of bacterial kidney disease in salmonid fish, detected by nested

reverse transcription-PCR of 16S rRNA sequences. *Applied and Environmental Microbiology*, *60*(12), 4580–4583. https://doi.org/10.1128/aem.60.12.4580-4583.1994

Maguire, M., Kase, J. A., Roberson, D., Muruvanda, T., Brown, E. W., Allard, M., Musser, S. M., & González-Escalona, N. (2021). Precision long-read metagenomics sequencing for food safety by detection and assembly of Shiga toxin-producing *Escherichia coli* in irrigation water. *PLOS ONE*, *16*(1), e0245172. https://doi.org/10.1371/journal.pone.0245172

Majaneva, M., Diserud, O. H., Eagle, S. H. C., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, *8*(1), 4682. https://doi.org/10.1038/s41598-018-23052-8

Malpartida-Cardenas, K., Miscourides, N., Rodriguez-Manzano, J., Yu, L.-S., Moser, N., Baum, J., & Georgiou, P. (2019). Quantitative and rapid *Plasmodium falciparum* malaria diagnosis and artemisinin-resistance detection using a CMOS Lab-on-Chip platform. *Biosensors and Bioelectronics*, *145*, 111678. https://doi.org/10.1016/j.bios.2019.111678

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Martínez-Murcia, A. J., Soler, L., Saavedra, M. J., Chacón, M. R., Guarro, J., Stackebrandt, E., & Figueras, M. J. (2005). Phenotypic, genotypic, and phylogenetic discrepancies to differentiate *Aeromonas salmonicida* from *Aeromonas bestiarum*. *International Microbiology*, *8*(4), 259–269.

Martínez-Porchas, M., & Vargas-Albores, F. (2017). Microbial metagenomics in aquaculture: A potential tool for a deeper insight into the activity. *Reviews in Aquaculture*, *9*(1), 42–56. https://doi.org/10.1111/raq.12102

Martín-Platero, A. M., Valdivia, E., Maqueda, M., & Martínez-Bueno, M. (2007). Fast, convenient, and economical method for isolating genomic DNA from lactic acid bacteria using a modification of the protein "salting-out" procedure. *Analytical Biochemistry*, *366*(1), 102–104. https://doi.org/10.1016/j.ab.2007.03.010

Martín-Platero, A. M., Valdivia, E., Maqueda, M., & Martínez-Bueno, M. (2009). Characterization and safety evaluation of enterococci isolated from Spanish goats' milk cheeses. *International Journal of Food Microbiology*, *132*(1), 24–32. https://doi.org/10.1016/j.ijfoodmicro.2009.03.010

Martín-Platero, A. M., Valdivia, E., Maqueda, M., Martín-Sánchez, I., & Martínez-Bueno, M. (2008). Polyphasic Approach to Bacterial Dynamics during the Ripening of Spanish Farmhouse Cheese, Using Culture-Dependent and -Independent Methods. *Applied and Environmental Microbiology*, *74*(18), 5662–5673. https://doi.org/10.1128/AEM.00418-08

Masumura K., & Wakabayashi H. (1977). An Outbreak of Gliding Bacterial Disease in Hatchery-born Red Seabream (*Pagrus major*) and Gilthead (*Acanthopagrus schlegeli*) Fry in Hiroshima. *Fish Pathology*, *12*(3), 171–177. https://doi.org/10.3147/jsfp.12.171

McDaniel, L. D., Young, E., Delaney, J., Ruhnau, F., Ritchie, K. B., & Paul, J. H. (2010). High Frequency of Horizontal Gene Transfer in the Oceans. *Science*, *330*(6000), 50. https://doi.org/10.1126/science.1192243

McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., Ayers, S. L., Lam, C., Tate, H. P., & Zhao, S. (2016). Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal *Salmonella*. *Antimicrobial Agents and Chemotherapy*, *60*(9), 5515–5520. https://doi.org/10.1128/AAC.01030-16

McLaren, M. R., Willis, A. D., & Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *ELife*, *8*, e46923. https://doi.org/10.7554/eLife.46923

Meenakshisundaram, M., Sugantham, F., Muthukumar, C., & Chandrasekar, M. S. (2021). Metagenomic characterization of biofloc in the grow-out culture of Genetically Improved Farmed Tilapia (GIFT). *Aquaculture Research*, *52*(9), 4249–4262. https://doi.org/10.1111/are.15263

Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L. A., Chertkov, O., Reddy, T., Pati, A., Ivanova, N. N., Markowitz, V., Kyrpides, N. C., Woyke, T., Göker, M., & Klenk, H.-P. (2014). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences*, *9*(1), 2. https://doi.org/10.1186/1944-3277-9-2

Menanteau-Ledouble, S., Kumar, G., Saleh, M., & El-Matbouli, M. (2016). *Aeromonas salmonicida*: Updates on an old acquaintance. *Diseases of Aquatic Organisms*, *120*(1), 49–68. https://doi.org/10.3354/dao03006

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, *7*(1), 11257. https://doi.org/10.1038/ncomms11257

Michael, C. A., Dominey-Howes, D., & Labbate, M. (2014). The Antimicrobial Resistance Crisis: Causes, Consequences, and Management. *Frontiers in Public Health*, *2*, 145. https://doi.org/10.3389/fpubh.2014.00145

Mojarro, A., Hachey, J., Bailey, R., Brown, M., Doebler, R., Ruvkun, G., Zuber, M. T., & Carr, C. E. (2019). Nucleic Acid Extraction and Sequencing from Low-Biomass Synthetic Mars Analog Soils for *In Situ* Life Detection. *Astrobiology*, *19*(9), 1139–1152. https://doi.org/10.1089/ast.2018.1929

Mojarro, A., Hachey, J., Ruvkun, G., Zuber, M. T., & Carr, C. E. (2018). CarrierSeq: A sequence analysis workflow for low-input nanopore sequencing. *BMC Bioinformatics*, *19*(1), 108. https://doi.org/10.1186/s12859-018-2124-3

Mori, Y., Kitao, M., Tomita, N., & Notomi, T. (2004). Real-time turbidimetry of LAMP reaction for quantifying template DNA. *Journal of Biochemical and Biophysical Methods*, *59*(2), 145–157. https://doi.org/10.1016/j.jbbm.2003.12.005

Morton, A. B., & Williams, R. (2003). First Report of a Sea Louse, *Lepeophtheirus salmonis*, Infestation on Juvenile Pink Salmon, *Oncorhynchus gorbuscha*, in Nearshore Habitat. *The Canadian Field-Naturalist*, *117*(4), 634. https://doi.org/10.22621/cfn.v117i4.834

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, *51*, 263–273. https://doi.org/10.1101/sqb.1986.051.01.032

Murray, C., Evelyn, T., Beacham, T., Bamer, L., Ketcheson, J., & Prosperi-Porta, L. (1992). Experimental induction of bacterial kidney disease in Chinook salmon by immersion and cohabitation challenges. *Diseases of Aquatic Organisms*, *12*, 91–96. https://doi.org/10.3354/dao012091

Musa, Z., Ma, J., Egamberdieva, D., Abdelshafy Mohamad, O. A., Abaydulla, G., Liu, Y., Li, W.-J., & Li, L. (2020). Diversity and Antimicrobial Potential of Cultivable Endophytic Actinobacteria Associated With the Medicinal Plant *Thymus roseus*. *Frontiers in Microbiology*, *11*, 191. https://doi.org/10.3389/fmicb.2020.00191

Muziasari, W. I., Pärnänen, K., Johnson, T. A., Lyra, C., Karkman, A., Stedtfeld, R. D., Tamminen, M., Tiedje, J. M., & Virta, M. (2016). Aquaculture changes the profile of antibiotic resistance and mobile genetic element associated genes in Baltic Sea sediments. *FEMS Microbiology Ecology*, *92*(4), fiw052. https://doi.org/10.1093/femsec/fiw052

Nadon, C., Van Walle, I., Gerner-Smidt, P., Campos, J., Chinen, I., Concepcion-Acevedo, J., Gilpin, B., Smith, A. M., Kam, K. M., Perez, E., Trees, E., Kubota, K., Takkinen, J., Nielsen, E. M., Carleton, H., & FWD-NEXT Expert Panel. (2017). PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Eurosurveillance*, *22*(23), 30544. https://doi.org/10.2807/1560-7917.ES.2017.22.23.30544

Nagamine, K., Hase, T., & Notomi, T. (2002). Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Molecular and Cellular Probes*, *16*(3), 223–229. https://doi.org/10.1006/mcpr.2002.0415

Nakagawa, T., Tsuchiya, Y., & Takahashi, R. (2019). Whole-Genome Sequence of the Ammonia-Oxidizing Bacterium *Nitrosomonas stercoris* Type Strain KYUHI-S, Isolated from Composted Cattle Manure. *Microbiology Resource Announcements*, *8*(34). https://doi.org/10.1128/MRA.00742-19

National Research Council. (2006). *Dynamic Changes in Marine Ecosystems: Fishing, Food Webs, and Future Options*. The National Academies Press. https://doi.org/10.17226/11608

Naz, K., Naz, A., Ashraf, S. T., Rizwan, M., Ahmad, J., Baumbach, J., & Ali, A. (2019). PanRV: Pangenome-reverse vaccinology approach for identifications of potential

vaccine candidates in microbial pangenome. *BMC Bioinformatics*, *20*(1), 123. https://doi.org/10.1186/s12859-019-2713-9

NCBI. (n.d.). *Bacterial Antimicrobial Resistance Reference Gene ... (ID 313047)—BioProject—NCBI*. NCBI. Retrieved October 27, 2021, from https://www.ncbi.nlm.nih.gov/bioproject/313047

NCBI Resource Coordinators, Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E., Bourexis, D., Brister, J. R., Bryant, S. H., Canese, K., Cavanaugh, M., Charowhas, C., Clark, K., Dondoshansky, I., Feolo, M., Fitzpatrick, L., Funk, K., Geer, L. Y., … Zbicz, K. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

NCBI Staff. (2020, February 21). *NCBI Insights: New ribosomal RNA BLAST databases available on the web BLAST service and for download*. NCBI Insights. https://ncbiinsights.ncbi.nlm.nih.gov/2020/02/21/rrna-databases/

NEB. (n.d.). *WarmStart Colorimetric LAMP 2X Master Mix Typical LAMP Protocol (M1800)*. Retrieved October 26, 2021, from https://international.neb.com/protocols/2016/08/15/warmstart-colorimetric-lamp-2x-master-mix-typical-lamp-protocol-m1800

Nematollahi, A., Decostere, A., Pasmans, F., & Haesebrouck, F. (2003). *Flavobacterium psychrophilum* infections in salmonid fish. *Journal of Fish Diseases*, *26*(10), 563–574. https://doi.org/10.1046/j.1365-2761.2003.00488.x

Nicolas, P., Mondot, S., Achaz, G., Bouchenot, C., Bernardet, J.-F., & Duchaud, E. (2008). Population Structure of the Fish-Pathogenic Bacterium *Flavobacterium psychrophilum*. *Applied and Environmental Microbiology*, *74*(12), 3702–3709. https://doi.org/10.1128/AEM.00244-08

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, *28*(12), e63. https://doi.org/10.1093/nar/28.12.e63

Nuccio, E. E., Starr, E., Karaoz, U., Brodie, E. L., Zhou, J., Tringe, S. G., Malmstrom, R. R., Woyke, T., Banfield, J. F., Firestone, M. K., & Pett-Ridge, J. (2020). Niche differentiation is spatially and temporally regulated in the rhizosphere. *The ISME Journal*, *14*(4), 999–1014. https://doi.org/10.1038/s41396-019-0582-x

O'Brien, D., Mooney, J., Ryan, D., Powell, E., Hiney, M., Smith, P. R., & Powell, R. (1994). Detection of *Aeromonas salmonicida*, causal agent of furunculosis in salmonid fish, from the tank effluent of hatchery-reared Atlantic salmon smolts. *Applied and Environmental Microbiology*, *60*(10), 3874–3877. https://doi.org/10.1128/aem.60.10.3874-3877.1994

OIE. (2021). Chapter 2.3.0—Diseases of fish: General Information. In *Manual of Diagnostic Tests for Aquatic Animals 2021*. World Organisation for Animal Health (OIE).

Olesen, N. J., & Vendramin, N. (2017, May). *Survey & Diagnosis of Fish Diseases in 2016*. 21st Annual Workshop of the National Reference Laboratories for Fish Diseases,

Copenhagen, Denmark. https://www.eurl-fish-crustacean.eu/-/media/sites/eurl-fish-crustacean/fish/annual-workshop/22nd-aw-2018/presentations/session-1/niels-joergen-olesen_overview-of.pdf?la=da&hash=AE3E623EBA8583872A293724070F4938F784B377

Olesen, N. J., Vendramin, N., & Andersen, N. R. (2018, May). *Survey & Diagnosis of fish diseases in 2017*. 22nd Annual Workshop of the National Reference Laboratories for Fish Diseases, Copenhagen, Denmark. https://www.eurl-fish-crustacean.eu/-/media/sites/eurl-fish-crustacean/fish/annual-workshop/21st-aw-2017/presentations/session-1/niels-joergen-olesen_overview-of.pdf?la=da&hash=1C0E84451AFC14189DCDFBE57686BC66F5C0FE53

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. https://doi.org/10.1186/1471-2105-12-385

ONT. (n.d.-a). *16S Barcoding Kit 1-24: SQK-16S024*. Oxford Nanopore Technologies. Retrieved October 27, 2021, from https://store.nanoporetech.com/uk/16s-barcoding-kit-1-24.html

ONT. (n.d.-b). *Company history*. Oxford Nanopore Technologies. Retrieved October 26, 2021, from http://nanoporetech.com/about-us/history

ONT. (n.d.-c). *Flow cells and nanopores*. Oxford Nanopore Technologies. Retrieved October 27, 2021, from http://nanoporetech.com/how-it-works/flow-cells-and-nanopores

ONT. (n.d.-d). *MinION Starter Pack*. Oxford Nanopore Technologies. Retrieved October 27, 2021, from https://store.nanoporetech.com/uk/minion-basic.html

ONT. (n.d.-e). *Rapid Barcoding Sequencing (SQK-RBK004)*. Nanopore Community. Retrieved October 26, 2021, from https://community.nanoporetech.com/protocols/rapid-barcoding-sequencing-sqk-rbk004/v/rbk_9054_v2_revv_14aug2019

ONT. (n.d.-f). Oxford Nanopore Technologies. Retrieved July 17, 2022, from https://store.nanoporetech.com/uk/minion.html

ONT. (2020, January 13). *R10.3: The newest nanopore for high accuracy nanopore sequencing – now available in store*. Oxford Nanopore Technologies. http://nanoporetech.com/about-us/news/r103-newest-nanopore-high-accuracy-nanopore-sequencing-now-available-store

Opal, S. M., & Pop-Vicas, A. (2015). Molecular mechanisms of antibiotic resistance in bacteria Chapter 18. In *Mandell, Douglas, and Bennett's principles and practice of infectious diseases* (8th ed., Vol. 1, pp. 235–251). Saunders.

Orsini, M., & Romano-Spica, V. (2001). A microwave-based method for nucleic acid isolation from environmental samples. *Letters in Applied Microbiology*, *33*(1), 17–20. https://doi.org/10.1046/j.1472-765X.2001.00938.x

Osborn, A. M., Moore, E. R. B., & Timmis, K. N. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of

microbial community structure and dynamics. *Environmental Microbiology*, *2*(1), 39–50. https://doi.org/10.1046/j.1462-2920.2000.00081.x

Ostland, V. E., MacPhee, D. D., Lumsden, J. S., & Ferguson, H. W. (1995). Virulence of *Flavobacterium branchiophilum* in experimentally infected salmonids. *Journal of Fish Diseases*, *18*(3), 249–262. https://doi.org/10.1111/j.1365-2761.1995.tb00300.x

Otta, S., Shubha, G., Joseph, B., Chakraborty, A., Karunasagar, I., & Karunasagar, I. (1999). Polymerase chain reaction (PCR) detection of white spot syndrome virus (WSSV) in cultured and wild crustaceans in India. *Diseases of Aquatic Organisms*, *38*, 67–70. https://doi.org/10.3354/dao038067

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Pajdak-Czaus, J., Platt-Samoraj, A., Szweda, W., Siwicki, A. K., & Terech-Majewska, E. (2019). *Yersinia ruckeri*—A threat not only to rainbow trout. *Aquaculture Research*, *50*(11), 3083–3096. https://doi.org/10.1111/are.14274

Panangala, V., Shelby, R., Shoemaker, C., Klesius, P., Mitra, A., & Morrison, E. (2006). Immunofluorescent test for simultaneous detection of *Edwardsiella ictaluri* and *Flavobacterium columnare*. *Diseases of Aquatic Organisms*, *68*, 197–207. https://doi.org/10.3354/dao068197

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Park, S., Choi, S. J., Won, S.-M., & Yoon, J.-H. (2017). *Tenacibaculum aestuariivivum* sp. Nov., isolated from a tidal flat. *International Journal of Systematic and Evolutionary Microbiology*, *67*(11), 4612–4618. https://doi.org/10.1099/ijsem.0.002343

Parshukov, A. N., Kashinskaya, E. N., Simonov, E. P., Hlunov, O. V., Izvekova, G. I., Andree, K. B., & Solovyev, M. M. (2019). Variations of the intestinal gut microbiota of farmed rainbow trout, *Oncorhynchus mykiss* (Walbaum), depending on the infection status of the fish. *Journal of Applied Microbiology*, *127*(2), 379–395. https://doi.org/10.1111/jam.14302

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., & Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, *70*(11), 5607–5612. https://doi.org/10.1099/ijsem.0.004332

Pecora, N. D., Li, N., Allard, M., Li, C., Albano, E., Delaney, M., Dubois, A., Onderdonk, A. B., & Bry, L. (2015). Genomically Informed Surveillance for Carbapenem-Resistant Enterobacteriaceae in a Health Care System. *MBio*, *6*(4), e01030. https://doi.org/10.1128/mBio.01030-15

Pereira, F. C., Wasmund, K., Cobankovic, I., Jehmlich, N., Herbold, C. W., Lee, K. S., Sziranyi, B., Vesely, C., Decker, T., Stocker, R., Warth, B., von Bergen, M., Wagner, M., & Berry, D. (2020). Rational design of a microbial consortium of mucosal sugar

utilizers reduces Clostridiodes difficile colonization. *Nature Communications*, *11*(1), 5104. https://doi.org/10.1038/s41467-020-18928-1

Peters, L., Spatharis, S., Dario, M. A., Dwyer, T., Roca, I. J. T., Kintner, A., Kanstad-Hanssen, Ø., Llewellyn, M. S., & Praebel, K. (2018). Environmental DNA: A New Low-Cost Monitoring Tool for Pathogens in Salmonid Aquaculture. *Frontiers in Microbiology*, *9*, 3009. https://doi.org/10.3389/fmicb.2018.03009

Picón-Camacho, S. M., Thompson, W. P., Blaylock, R. B., & Lotz, J. M. (2013). Development of a rapid assay to detect the dinoflagellate *Amyloodinium ocellatum* using loop-mediated isothermal amplification (LAMP). *Veterinary Parasitology*, *196*(3–4), 265–271. https://doi.org/10.1016/j.vetpar.2013.04.010

Pitcher, D. G., Saunders, N. A., & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology*, *8*(4), 151–156. https://doi.org/10.1111/j.1472-765X.1989.tb00262.x

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Applied and Environmental Microbiology*, *84*(7), e02627-17. https://doi.org/10.1128/AEM.02627-17

Powell, M., Overturf, K., Hogge, C., & Johnson, K. (2005). Detection of *Renibacterium salmoninarum* in chinook salmon, *Oncorhynchus tshawytscha* (Walbaum), using quantitative PCR. *Journal of Fish Diseases*, *28*(10), 615–622. https://doi.org/10.1111/j.1365-2761.2005.00669.x

Price, J., Gordon, N. C., Crook, D., Llewelyn, M., & Paul, J. (2013). The usefulness of whole genome sequencing in the management of *Staphylococcus aureus* infections. *Clinical Microbiology and Infection*, *19*(9), 784–789. https://doi.org/10.1111/1469-0691.12109

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods*, *8*(1), 12–24. https://doi.org/10.1039/C5AY02550H

Ptasinska, A., Whalley, C., Bosworth, A., Poxon, C., Bryer, C., Machin, N., Grippon, S., Wise, E. L., Armson, B., Howson, E. L. A., Goring, A., Snell, G., Forster, J., Mattocks, C., Frampton, S., Anderson, R., Cleary, D., Parker, J., Boukas, K., … Beggs, A. D. (2021). Diagnostic accuracy of loop-mediated isothermal amplification coupled to nanopore sequencing (LamPORE) for the detection of SARS-CoV-2 infection at scale in symptomatic and asymptomatic populations. *Clinical Microbiology and Infection*, *27*(9), 1348.e1-1348.e7. https://doi.org/10.1016/j.cmi.2021.04.008

Pulkkinen, K., Suomalainen, L.-R., Read, A. F., Ebert, D., Rintamäki, P., & Valtonen, E. T. (2010). Intensive fish farming and the evolution of pathogen virulence: The case of columnaris disease in Finland. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 593–600. https://doi.org/10.1098/rspb.2009.1659

Punati, R. D., Mallepaddi, P. C., Poonati, R., Maity, S. N., Sohal, J. S., Polavarapu, K. K. B., & Polavarapu, R. (2019). Development and evaluation of LAMP-coupled lateral flow device for the detection of MAP in livestock at point of care resource-limited areas. *Brazilian Journal of Microbiology*, *50*(4), 1105–1114. https://doi.org/10.1007/s42770-019-00116-z

Qiagen. (2017). *DNeasy® PowerSoil® Kit Handbook*. Qiagen.

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., … Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228–232. https://doi.org/10.1038/nature16996

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

RAIZO. (2018). *La Furonculose: Défis et Importance de la Peévention* (Bulletin Zoosanitaire). Ministère de l'Agriculture, des Pêcheries et de l'Alimentation. https://www.mapaq.gouv.qc.ca/SiteCollectionDocuments/Santeanimale/Bulletins/Bulletinzoosan_Furonculose.pdf

Rambaut, A., Suchard, M., Nenarokov, S., Klötzl, F., & Flood, S. (2018). *Figtree* (v1.4.4) [Java]. https://github.com/rambaut/figtree (Original work published 2015)

Rassner, S. M. E., Anesio, A. M., Girdwood, S. E., Hell, K., Gokul, J. K., Whitworth, D. E., & Edwards, A. (2016). Can the Bacterial Community of a High Arctic Glacier Surface Escape Viral Control? *Frontiers in Microbiology*, *7*, 956. https://doi.org/10.3389/fmicb.2016.00956

Reischl, U., Bretagne, S., Krüger, D., Ernault, P., & Costa, J.-M. (2003). Comparison of two DNA targets for the diagnosis of Toxoplasmosis by real-time PCR using fluorescence resonance energy transfer hybridization probes. *BMC Infectious Diseases*, *3*(1), 7. https://doi.org/10.1186/1471-2334-3-7

Rescheneder, P., cjw65, Pettett, R., & Amery, M. (2019). *Qcat* (v1.1.0) [Python]. Oxford Nanopore Technologies. https://github.com/nanoporetech/qcat (Original work published 2018)

Revez, J., Espinosa, L., Albiger, B., Leitmeyer, K. C., Struelens, M. J., & ECDC National Microbiology Focal Points and Experts Group. (2017). Survey on the Use of Whole-Genome Sequencing for Infectious Diseases Surveillance: Rapid Expansion of European National Capacities, 2015–2016. *Frontiers in Public Health*, *5*, 347. https://doi.org/10.3389/fpubh.2017.00347

Riediger, I. N., Hoffmaster, A. R., Casanovas-Massana, A., Biondo, A. W., Ko, A. I., & Stoddard, R. A. (2016). An Optimized Method for Quantification of Pathogenic *Leptospira* in Environmental Water Samples. *PLOS ONE*, *11*(8), e0160523. https://doi.org/10.1371/journal.pone.0160523

Roberts, R. J. (2012). *Fish pathology*. John Wiley & Sons.

Robson, H. L. A., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., & Jerry, D. R. (2016). Fine-tuning for the tropics: Application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, *16*(4), 922–932. https://doi.org/10.1111/1755-0998.12505

Roig, F. J., González-Candelas, F., Sanjuán, E., Fouz, B., Feil, E. J., Llorens, C., Baker-Austin, C., Oliver, J. D., Danin-Poleg, Y., Gibas, C. J., Kashi, Y., Gulig, P. A., Morrison, S. S., & Amaro, C. (2018). Phylogeny of *Vibrio vulnificus* from the Analysis of the Core-Genome: Implications for Intra-Species Taxonomy. *Frontiers in Microbiology*, *8*, 2613. https://doi.org/10.3389/fmicb.2017.02613

Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., & Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*, *11*(11), 2806–2820. https://doi.org/10.1111/j.1462-2920.2009.01964.x

Rosselló-Móra, R., & Amann, R. (2015). Past and future species definitions for Bacteria and Archaea. *Systematic and Applied Microbiology*, *38*(4), 209–216. https://doi.org/10.1016/j.syapm.2015.02.001

Rouleau, F. D., Vincent, A. T., & Charette, S. J. (2018). Genomic and phenotypic characterization of an atypical *Aeromonas salmonicida* strain isolated from a lumpfish and producing unusual granular structures. *Journal of Fish Diseases*, *41*(4), 673–681. https://doi.org/10.1111/jfd.12769

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, *17*(2), 155–158. https://doi.org/10.1038/s41592-019-0669-3

Rusiñol, M., Martínez-Puchol, S., Timoneda, N., Fernández-Cassi, X., Pérez-Cataluña, A., Fernández-Bravo, A., Moreno-Mesonero, L., Moreno, Y., Alonso, J. L., Figueras, M. J., Abril, J. F., Bofill-Mas, S., & Girones, R. (2020). Metagenomic analysis of viruses, bacteria and protozoa in irrigation water. *International Journal of Hygiene and Environmental Health*, *224*, 113440. https://doi.org/10.1016/j.ijheh.2019.113440

Sagcan, H., & Turgut Kara, N. (2019). Detection of Potato ring rot Pathogen *Clavibacter michiganensis* subsp. *Sepedonicus* by Loop-mediated isothermal amplification (LAMP) assay. *Scientific Reports*, *9*(1), 20393. https://doi.org/10.1038/s41598-019-56680-9

Saleh, M., Soliman, H., & El-Matbouli, M. (2008a). Loop-mediated isothermal amplification (LAMP) for rapid detection of *Renibacterium salmoninarum*, the causative agent of bacterial kidney disease. *Diseases of Aquatic Organisms*, *81*(2), 143–151. https://doi.org/10.3354/dao01945

Saleh, M., Soliman, H., & El-Matbouli, M. (2008b). Loop-mediated isothermal amplification as an emerging technology for detection of *Yersinia ruckeri* the causative agent of enteric red mouth disease in fish. *BMC Veterinary Research*, *4*(1), 31. https://doi.org/10.1186/1746-6148-4-31

Samarasinghe, S. N., Wanigatunge, R. P., & Magana-Arachchi, D. N. (2021). Bacterial Diversity in a Sri Lankan Geothermal Spring Assessed by Culture-Dependent and

Culture-Independent Approaches. *Current Microbiology*, *78*(9), 3439–3452. https://doi.org/10.1007/s00284-021-02608-4

Sanders, J. E., & Fryer, J. L. (1980). *Renibacterium salmoninarum* gen. Nov., sp. Nov., the Causative Agent of Bacterial Kidney Disease in Salmonid Fishes. *International Journal of Systematic Bacteriology*, *30*(2), 496–502. https://doi.org/10.1099/00207713-30-2-496

Sarmah, A. K., Meyer, M. T., & Boxall, A. B. A. (2006). A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. *Chemosphere*, *65*(5), 725–759. https://doi.org/10.1016/j.chemosphere.2006.03.026

Savan, R., Igarashi, A., Matsuoka, S., & Sakai, M. (2004). Sensitive and Rapid Detection of Edwardsiellosis in Fish by a Loop-Mediated Isothermal Amplification Method. *Applied and Environmental Microbiology*, *70*(1), 621–624. https://doi.org/10.1128/AEM.70.1.621-624.2004

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S., Klimke, W., Marchler-Bauer, A., Landrum, M., Lathrop, S., Lu, Z., Madden, T. L., O'Leary, N., Phan, L., Rangwala, S. H., Schneider, V. A., … Sherry, S. T. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *49*(D1), D10–D17. https://doi.org/10.1093/nar/gkaa892

Schar, D., Klein, E. Y., Laxminarayan, R., Gilbert, M., & Van Boeckel, T. P. (2020). Global trends in antimicrobial use in aquaculture. *Scientific Reports*, *10*(1), 21878. https://doi.org/10.1038/s41598-020-78849-3

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., Chen, W., Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List, Bolchacova, E., Voigt, K., Crous, P. W., Miller, A. N., Wingfield, M. J., Aime, M. C., An, K.-D., Bai, F.-Y., Barreto, R. W., Begerow, D., … Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241–6246. https://doi.org/10.1073/pnas.1117018109

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seemann, T. (2018). *Barrnap* (v0.9) [Perl]. https://github.com/tseemann/barrnap (Original work published 2013)

Seemann, T. (2020). *ABRicate* (v1.0.0) [Perl]. https://github.com/tseemann/abricate (Original work published 2014)

Seymour, C., Page, D. C., Slim, L., Rana, S., Pettett, R., Vella, M., Young, S., sirelkhatim, hoho2b, & Lloyd-Palmer, M. (2021). *Bonito* (v0.4.0) [Python]. Oxford Nanopore Technologies. https://github.com/nanoporetech/bonito (Original work published 2019)

Shahid, F., Zaheer, T., Ashraf, S. T., Shehroz, M., Anwer, F., Naz, A., & Ali, A. (2021). Chimeric vaccine designs against *Acinetobacter baumannii* using pan genome and

reverse vaccinology approaches. *Scientific Reports*, *11*(1), 13213. https://doi.org/10.1038/s41598-021-92501-8

Shinn, A. P., Pratoomyot, J., Griffiths, D., Trong, T. Q., Vu, N. T., Jiravanichpaisal, P., & Briggs, M. (2018). Asian Shrimp Production and the Economic Costs of Disease. *Asian Fisheries Science*, *31S*, 29–58. https://doi.org/10.33997/j.afs.2018.31.S1.003

Shirshikov, F. V., Pekov, Y. A., & Miroshnikov, K. A. (2019). MorphoCatcher: A multiple-alignment based web tool for target selection and designing taxon-specific primers in the loop-mediated isothermal amplification method. *PeerJ*, *7*, e6801. https://doi.org/10.7717/peerj.6801

Siguier, P. (2006). ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Research*, *34*(90001), D32–D36. https://doi.org/10.1093/nar/gkj014

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sithigorngul, P., Rukpratanporn, S., Chaivisuthangkura, P., Sridulyakul, P., & Longyant, S. (2011). Simultaneous and rapid detection of white spot syndrome virus and yellow head virus infection in shrimp with a dual immunochromatographic strip test. *Journal of Virological Methods*, *173*(1), 85–91. https://doi.org/10.1016/j.jviromet.2011.01.011

Sithigorngul, W., Rukpratanporn, S., Sittidilokratna, N., Pecharaburanin, N., Longyant, S., Chaivisuthangkura, P., & Sithigorngul, P. (2007). A convenient immunochromatographic test strip for rapid diagnosis of yellow head virus infection in shrimp. *Journal of Virological Methods*, *140*(1–2), 193–199. https://doi.org/10.1016/j.jviromet.2006.11.034

Skulska, I. (2014). *Culture of the Bacterial Gill Disease Organism, Flavobacterium branchiophilum and Strain Differences Relevant to Epizoology* [Masters Thesis, University of Guelph]. https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/8523/Skulska_Iwona_201410_Msc.pdf?sequence=1&isAllowed=y

Snieszko, S. F. (1981). *Bacterial gill disease of freshwater fishes* (Vol. 62). US Department of the Interior, Fish and Wildlife Service, Division of Fishery Research.

Snow, M., Black, J., Matejusova, I., McIntosh, R., Baretto, E., Wallace, I., & Bruno, D. (2010). Detection of salmonid alphavirus RNA in wild marine fish: Implications for the origins of salmon pancreas disease in aquaculture. *Diseases of Aquatic Organisms*, *91*(3), 177–188. https://doi.org/10.3354/dao02265

Soliman, H., Midtlyng, P. J., & El-Matbouli, M. (2009). Sensitive and rapid detection of infectious pancreatic necrosis virus by reverse transcription loop mediated isothermal amplification. *Journal of Virological Methods*, *158*(1–2), 77–83. https://doi.org/10.1016/j.jviromet.2009.01.018

Speare, D. J., Markham, R. J. F., Despres, B., Whitman, K., & MacNair, N. (1995). Examination of Gills from Salmonids with Bacterial Gill Disease using Monoclonal

Antibody Probes for *Flavobacterium branchiophilum* and *Cytophaga columnaris*. *Journal of Veterinary Diagnostic Investigation*, *7*(4), 500–505. https://doi.org/10.1177/104063879500700413

Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2015a). Evaluation of water sampling methodologies for amplicon-based characterization of bacterial community structure. *Journal of Microbiological Methods*, *114*, 43–50. https://doi.org/10.1016/j.mimet.2015.05.003

Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2015b). High-throughput functional screening reveals low frequency of antibiotic resistance genes in DNA recovered from the Upper Mississippi River. *Journal of Water and Health*, *13*(3), 693–703. https://doi.org/10.2166/wh.2014.215

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, *57*(5), 758–771. https://doi.org/10.1080/10635150802429642

Starliper, C. (1996). Genetic diversity of North American isolates of *Renibacterium salmoninarum*. *Diseases of Aquatic Organisms*, *27*, 207–213. https://doi.org/10.3354/dao027207

Stehlíková, D., Beran, P., Cohen, S. P., & Čurn, V. (2020). Development of Real-Time and Colorimetric Loop Mediated Isothermal Amplification Assay for Detection of *Xanthomonas gardneri*. *Microorganisms*, *8*(9), 1301. https://doi.org/10.3390/microorganisms8091301

Su, M., Satola, S. W., & Read, T. D. (2019). Genome-Based Prediction of Bacterial Antibiotic Resistance. *Journal of Clinical Microbiology*, *57*(3), e01405-18. https://doi.org/10.1128/JCM.01405-18

Suarez, R., Kusch, K., Miranda, C. D., Li, T., Campanini, J., Behra, P. R. K., Aro, L., Martínez, A., Godoy, M., & Medina, D. A. (2021). Whole-Genome sequencing and comparative genomics of *Mycobacterium* spp. From farmed Atlantic and coho salmon in Chile. *Antonie van Leeuwenhoek*, *114*(9), 1323–1336. https://doi.org/10.1007/s10482-021-01592-w

Suebsing, R., Kampeera, J., Tookdee, B., Withyachumnarnkul, B., Turner, W., & Kiatpathomchai, W. (2013). Evaluation of colorimetric loop-mediated isothermal amplification assay for visual detection of *Streptococcus agalactiae* and *Streptococcus iniae* in tilapia. *Letters in Applied Microbiology*, *57*(4), 317–324. https://doi.org/10.1111/lam.12114

Suzuki, M., Nakagawa, Y., Harayama, S., & Yamamoto, S. (2001). Phylogenetic analysis and taxonomic study of marine *Cytophaga*-like bacteria: Proposal for *Tenacibaculum* gen. Nov. With *Tenacibaculum maritimum* comb. Nov. And *Tenacibaculum ovolyticum* comb. Nov., and description of *Tenacibaculum mesophilum* sp. Nov. And

Tenacibaculum amylolyticum sp. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*(5), 1639–1652. https://doi.org/10.1099/00207713-51-5-1639

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Takasaki, K., Aihara, H., Imanaka, T., Matsudaira, T., Tsukahara, K., Usui, A., Osaki, S., & Doi, H. (2021). Water pre-filtration methods to improve environmental DNA detection by real-time PCR and metabarcoding. *PLOS ONE*, *16*(5), e0250162. https://doi.org/10.1371/journal.pone.0250162

Tang, J., Bu, Y., Zhang, X.-X., Huang, K., He, X., Ye, L., Shan, Z., & Ren, H. (2016). Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water. *Ecotoxicology and Environmental Safety*, *132*, 260–269. https://doi.org/10.1016/j.ecoenv.2016.06.016

Taniguchi, A., Aoki, R., & Eguchi, M. (2016). Microbial communities in various waters used for fish larval rearing. *Aquaculture Research*, *47*(2), 370–378. https://doi.org/10.1111/are.12495

Tanner, N. A., Zhang, Y., & Evans, T. C. (2012). Simultaneous multiple target detection in real-time loop-mediated isothermal amplification. *BioTechniques*, *53*(2), 81–89. https://doi.org/10.2144/0000113902

Tepaamorndech, S., Nookaew, I., Higdon, S. M., Santiyanont, P., Phromson, M., Chantarasakha, K., Mhuantong, W., Plengvidhya, V., & Visessanguan, W. (2020). Metagenomics in bioflocs and their effects on gut microbiome and immune responses in Pacific white shrimp. *Fish & Shellfish Immunology*, *106*, 733–741. https://doi.org/10.1016/j.fsi.2020.08.042

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., … Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome." *Proceedings of the National Academy of Sciences*, *102*(39), 13950–13955. https://doi.org/10.1073/pnas.0506758102

Thapa, S. P., De Francesco, A., Trinh, J., Gurung, F. B., Pang, Z., Vidalakis, G., Wang, N., Ancona, V., Ma, W., & Coaker, G. (2020). Genome-wide analyses of *Liberibacter* species provides insights into evolution, phylogenetic relationships, and virulence factors. *Molecular Plant Pathology*, *21*(5), 716–731. https://doi.org/10.1111/mpp.12925

Thermo Scientific. (2012). *T123—Technical Bulletin, NanoDrop Lite: Interpretation of Nucleic Acid 260/280 Ratios*. Thermo Scientific. https://tools.thermofisher.com/content/sfs/brochures/T123-NanoDrop-Lite-Interpretation-of-Nucleic-Acid-260-280-Ratios.pdf

Tidwell, J. H., & Allan, G. L. (2001). Fish as food: Aquaculture's contribution: Ecological and economic impacts and contributions of fish farming and capture fisheries. *EMBO Reports*, *2*(11), 958–963. https://doi.org/10.1093/embo-reports/kve236

Tobback, E., Decostere, A., Hermans, K., Haesebrouck, F., & Chiers, K. (2007). *Yersinia ruckeri* infections in salmonid fish. *Journal of Fish Diseases*, *30*(5), 257–268. https://doi.org/10.1111/j.1365-2761.2007.00816.x

Tomita, N., Mori, Y., Kanda, H., & Notomi, T. (2008). Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nature Protocols*, *3*(5), 877–882. https://doi.org/10.1038/nprot.2008.57

Toranzo, A. E., Magariños, B., & Romalde, J. L. (2005). A review of the main bacterial fish diseases in mariculture systems. *Aquaculture*, *246*(1–4), 37–61. https://doi.org/10.1016/j.aquaculture.2005.01.002

Torres, C., Vitalis, E. A., Baker, B. R., Gardner, S. N., Torres, M. W., & Dzenitis, J. M. (2011). LAVA: An Open-Source Approach To Designing LAMP (Loop-Mediated Isothermal Amplification) DNA Signatures. *BMC Bioinformatics*, *12*(1), 240. https://doi.org/10.1186/1471-2105-12-240

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M. E., Frapy, E., … Denamur, E. (2009). Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths. *PLoS Genetics*, *5*(1), e1000344. https://doi.org/10.1371/journal.pgen.1000344

Tourlousse, D. M., Sakamoto, M., Miura, T., Narita, K., Ohashi, A., Uchino, Y., Yamazoe, A., Kameyama, K., Terauchi, J., Ohkuma, M., Kawasaki, H., & Sekiguchi, Y. (2020). Complete Genome Sequence of *Flavonifractor plautii* JCM 32125 ^{T}. *Microbiology Resource Announcements*, *9*(17), e00135-20. https://doi.org/10.1128/MRA.00135-20

Toyama, T., Kita-Tsukamoto, K., & Wakabayashi, H. (1994). Identification of *Cytophaga psychrophila* by PCR targeted 16S ribosomal RNA. *Fish Pathology*, *29*(4), 271–275.

Toyama, T., Kita-Tsukamoto, K., & Wakabayashi, H. (1996). Identification of *Flexibacter maritimus*, *Flavobacterium branchiophilum* and *Cytophaga columnaris* by PCR targeted 16S ribosomal DNA. *Fish Pathology*, *31*(1), 25–31.

Tran, L., Nunan, L., Redman, R., Mohney, L., Pantoja, C., Fitzsimmons, K., & Lightner, D. (2013). Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Diseases of Aquatic Organisms*, *105*(1), 45–55. https://doi.org/10.3354/dao02621

Trudel, M. V., Vincent, A. T., Attéré, S. A., Labbé, M., Derome, N., Culley, A. I., & Charette, S. J. (2016). Diversity of antibiotic-resistance genes in Canadian isolates of *Aeromonas salmonicida* subsp. *Salmonicida*: Dominance of pSN254b and discovery of pAsa8. *Scientific Reports*, *6*(1), 35617. https://doi.org/10.1038/srep35617

Tsougeni, K., Kaprou, G., Loukas, C. M., Papadakis, G., Hamiot, A., Eck, M., Rabus, D., Kokkoris, G., Chatzandroulis, S., Papadopoulos, V., Dupuy, B., Jobst, G., Gizeli, E.,

Tserepi, A., & Gogolides, E. (2020). Lab-on-Chip platform and protocol for rapid foodborne pathogen detection comprising on-chip cell capture, lysis, DNA amplification and surface-acoustic-wave detection. *Sensors and Actuators B: Chemical*, *320*, 128345. https://doi.org/10.1016/j.snb.2020.128345

Uniprot Consortium. (2021). *UniProtKB 2021_03*. UniProt. https://www.uniprot.org/uniprot/?query=reviewed:yes

United Nations Dept. of Economic Affairs. (2016). *2015 Demographic Yearbook* (66th ed.). United Nations.

Vaser, R., & Šikić, M. (2021). Time- and memory-efficient genome assembly with Raven. *Nature Computational Science*, *1*(5), 332–336. https://doi.org/10.1038/s43588-021-00073-4

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Vendramin, N., & Olesen, N. J. (2021, June). *Survey & Diagnosis of Fish Diseases in 2020*. 25th Annual Workshop of the National Reference Laboratories for Fish Diseases, Copenhagen, Denmark. https://www.eurl-fish-crustacean.eu/-/media/sites/eurl-fish-crustacean/fish/annual-workshop/25th-aw-2021/2-s-d-2020.pdf?la=da&hash=4F9DDC416182072F9836ECD25BD07E133039BA66

Vereecke, N., Bokma, J., Haesebrouck, F., Nauwynck, H., Boyen, F., Pardon, B., & Theuns, S. (2020). High quality genome assemblies of *Mycoplasma bovis* using a taxon-specific Bonito basecaller for MinION and Flongle long-read nanopore sequencing. *BMC Bioinformatics*, *21*(1), 517. https://doi.org/10.1186/s12859-020-03856-0

Větrovský, T., & Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS ONE*, *8*(2), e57923. https://doi.org/10.1371/journal.pone.0057923

Villette, R., Autaa, G., Hind, S., Holm, J. B., Moreno-Sabater, A., & Larsen, M. (2021). Refinement of 16S rRNA gene analysis for low biomass biospecimens. *Scientific Reports*, *11*(1), 10741. https://doi.org/10.1038/s41598-021-90226-2

Vincent, A., & Charette, S. (2017). Phylogenetic analysis of the fish pathogen *Aeromonas salmonicida* underlines the dichotomy between European and Canadian strains for the *salmonicida* subspecies. *Journal of Fish Diseases*, *40*(9), 1241–1247. https://doi.org/10.1111/jfd.12595

Vincent, A. T., Bernatchez, A., Frey, J., & Charette, S. J. (2019). A Mesophilic *Aeromonas salmonicida* Strain Isolated from an Unsuspected Host, the Migratory Bird Pied Avocet. *Microorganisms*, *7*(12), 592. https://doi.org/10.3390/microorganisms7120592

Vincent, A. T., Fernández-Bravo, A., Sanchis, M., Mayayo, E., Figueras, M. J., & Charette, S. J. (2019). Investigation of the virulence and genomics of *Aeromonas salmonicida* strains isolated from human patients. *Infection, Genetics and Evolution*, *68*, 1–9. https://doi.org/10.1016/j.meegid.2018.11.019

Vincent, A. T., Rouleau, F. D., Moineau, S., & Charette, S. J. (2017). Study of mesophilic *Aeromonas salmonicida* A527 strain sheds light on the species' lifestyles and taxonomic dilemma. *FEMS Microbiology Letters*, *364*(23). https://doi.org/10.1093/femsle/fnx239

Vincent, A. T., Trudel, M. V., Freschi, L., Nagar, V., Gagné-Thivierge, C., Levesque, R. C., & Charette, S. J. (2016). Increasing genomic diversity and evidence of constrained lifestyle evolution due to insertion sequences in *Aeromonas salmonicida*. *BMC Genomics*, *17*(1), 44. https://doi.org/10.1186/s12864-016-2381-3

Vincent, A. T., Trudel, M. V., Paquet, V. E., Boyle, B., Tanaka, K. H., Dallaire-Dufresne, S., Daher, R. K., Frenette, M., Derome, N., & Charette, S. J. (2014). Detection of Variants of the pRAS3, pAB5S9, and pSN254 Plasmids in *Aeromonas salmonicida* subsp. *Salmonicida*: Multidrug Resistance, Interspecies Exchanges, and Plasmid Reshaping. *Antimicrobial Agents and Chemotherapy*, *58*(12), 7367–7374. https://doi.org/10.1128/AAC.03730-14

Wakabayashi, H., Hikida, M., & Masumura, K. (1986). Flexibacter maritimus sp. Nov., a pathogen of marine fishes. *International Journal of Systematic and Evolutionary Microbiology*, *36*(3), 396–398.

Wakabayashi, H., Huh, G. J., & Kimura, N. (1989). *Flavobacterium branchiophila* sp. Nov., a Causative Agent of Bacterial Gill Disease of Freshwater Fishes. *International Journal of Systematic Bacteriology*, *39*(3), 213–216. https://doi.org/10.1099/00207713-39-3-213

Walden, C., Carbonero, F., & Zhang, W. (2017). Assessing impacts of DNA extraction methods on next generation sequencing of water and wastewater samples. *Journal of Microbiological Methods*, *141*, 10–16. https://doi.org/10.1016/j.mimet.2017.07.007

Waliullah, S., Hudson, O., Oliver, J. E., Brannen, P. M., Ji, P., & Ali, M. E. (2019). Comparative analysis of different molecular and serological methods for detection of *Xylella fastidiosa* in blueberry. *PLOS ONE*, *14*(9), e0221903. https://doi.org/10.1371/journal.pone.0221903

Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., Eyre, D. W., Wilson, D. J., Hawkey, P. M., Crook, D. W., Parkhill, J., Harris, D., Walker, A. S., Bowden, R., Monk, P., Smith, E. G., & Peto, T. E. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: A retrospective observational study. *The Lancet Infectious Diseases*, *13*(2), 137–146. https://doi.org/10.1016/S1473-3099(12)70277-3

Wang, N., Zhao, Z., Gao, J., Tian, E., Yu, W., Li, H., Zhang, J., Xie, R., Zhao, X., & Chen, A. (2021). Rapid and Visual Identification of *Chlorophyllum molybdites* With Loop-Mediated Isothermal Amplification Method. *Frontiers in Microbiology*, *12*, 638315. https://doi.org/10.3389/fmicb.2021.638315

Wang, S., Liu, N., Zheng, L., Cai, G., & Lin, J. (2020). A lab-on-chip device for the sample-in-result-out detection of viable *Salmonella* using loop-mediated isothermal amplification and real-time turbidity monitoring. *Lab on a Chip*, *20*(13), 2296–2305. https://doi.org/10.1039/D0LC00290A

Wang, Y., Wang, Y., Lan, R., Xu, H., Ma, A., Li, D., Dai, H., Yuan, X., Xu, J., & Ye, C. (2015). Multiple Endonuclease Restriction Real-Time Loop-Mediated Isothermal Amplification. *The Journal of Molecular Diagnostics*, *17*(4), 392–401. https://doi.org/10.1016/j.jmoldx.2015.03.002

Wanigatunge, R. P., Magana-Arachchi, D. N., Chandrasekharan, N. V., & Kulasooriya, S. A. (2014). Genetic Diversity and Molecular Phylogeny of Cyanobacteria from Sri Lanka Based on 16S rRNA Gene. *Environmental Engineering Research*, *19*(4), 317–329. https://doi.org/10.4491/eer.2014.035

Watson, M., & Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction. *Nature Biotechnology*, *37*(2), 124–126. https://doi.org/10.1038/s41587-018-0004-z

Watts, J., Schreier, H., Lanska, L., & Hale, M. (2017). The Rising Tide of Antimicrobial Resistance in Aquaculture: Sources, Sinks and Solutions. *Marine Drugs*, *15*(6), 158. https://doi.org/10.3390/md15060158

Wayne, L., Brenner, D., Colwell, R., Grimont, P., Kandler, O., Krichevsky, M., Moore, L., Moore, W., Murray, R., Stackebrandt, E., & others. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic and Evolutionary Microbiology*, *37*(4), 463–464. https://doi.org/10.1099/00207713-37-4-463

Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, *173*(2), 697–703. https://doi.org/10.1128/jb.173.2.697-703.1991

Wheeler, T. J. (2009). Large-scale neighbor-joining with NINJA. *International Workshop on Algorithms in Bioinformatics*, 375–389. https://doi.org/10.1007/978-3-642-04241-6_31

WHO. (2014). *Antimicrobial Resistance: Global Report of Surveillance; 2014 Summary*.

Wick, R., & Holt, K. (2021). Benchmarking of long-read assemblers for prokaryote whole genome sequencing [version 4; peer review: 4 approved]. *F1000Research*, *8*(2138). https://doi.org/10.12688/f1000research.21782.4

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, *13*(6), e1005595. https://doi.org/10.1371/journal.pcbi.1005595

Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, *20*(1), 129. https://doi.org/10.1186/s13059-019-1727-y

Wickham, H. (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, *21*(12), 1–20. https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilkinson, S. P., & Davy, S. K. (2018). phylogram: An R package for phylogenetic analysis with nested lists. *Journal of Open Source Software*, *3*(26), 790. https://doi.org/10.21105/joss.00790

Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLOS ONE*, *12*(7), e0179282. https://doi.org/10.1371/journal.pone.0179282

Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., Salamat, S. M., Somasekar, S., Federman, S., Miller, S., Sokolic, R., Garabedian, E., Candotti, F., Buckley, R. H., Reed, K. D., Meyer, T. L., Seroogy, C. M., Galloway, R., Henderson, S. L., … Chiu, C. Y. (2014). Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing. *New England Journal of Medicine*, *370*(25), 2408–2417. https://doi.org/10.1056/NEJMoa1401268

Wilson, M. R., O'Donovan, B. D., Gelfand, J. M., Sample, H. A., Chow, F. C., Betjemann, J. P., Shah, M. P., Richie, M. B., Gorman, M. P., Hajj-Ali, R. A., Calabrese, L. H., Zorn, K. C., Chow, E. D., Greenlee, J. E., Blum, J. H., Green, G., Khan, L. M., Banerji, D., Langelier, C., … DeRisi, J. L. (2018). Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. *JAMA Neurology*, *75*(8), 947. https://doi.org/10.1001/jamaneurol.2018.0463

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, *74*(11), 5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46. https://doi.org/10.1186/gb-2014-15-3-r46

Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., Zuzarte, P. C., Gilpatrick, T., Payne, A., Quick, J., Sadowski, N., Holmes, N., de Jesus, J. G., Jones, K. L., Soulette, C. M., Snutch, T. P., Loman, N., Paten, B., Loose, M., … Timp, W. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods*, *16*(12), 1297–1305. https://doi.org/10.1038/s41592-019-0617-2

Xie, G., Zhang, Q., Han, N., Shi, C., Wang, X., Liu, Q., & Huang, J. (2012). An improved method for detection of *Edwardsiella tarda* by loop-mediated isothermal amplification by targeting the EsrB gene. *Chinese Journal of Oceanology and Limnology*, *30*(4), 595–603. https://doi.org/10.1007/s00343-012-1293-6

Xu, C., Lv, Z., Shen, Y., Liu, D., Fu, Y., Zhou, L., Liu, W., Chen, K., Ye, H., Xia, X., Xia, J., Wang, Y., Ke, Y., & Shen, J. (2020). Metagenomic insights into differences in environmental resistome profiles between integrated and monoculture aquaculture farms in China. *Environment International*, *144*, 106005. https://doi.org/10.1016/j.envint.2020.106005

Yamagishi, J., Runtuwene, L. R., Hayashida, K., Mongan, A. E., Thi, L. A. N., Thuy, L. N., Nhat, C. N., Limkittikul, K., Sirivichayakul, C., Sathirapongsasuti, N., Frith, M., Makalowski, W., Eshita, Y., Sugano, S., & Suzuki, Y. (2017). Serotyping dengue

virus with isothermal amplification and a portable sequencer. *Scientific Reports*, *7*(1), 3510. https://doi.org/10.1038/s41598-017-03734-5

Yang, Y.-T., Chen, I.-T., Lee, C.-T., Chen, C.-Y., Lin, S.-S., Hor, L.-I., Tseng, T.-C., Huang, Y.-T., Sritunyalucksana, K., Thitamadee, S., Wang, H.-C., & Lo, C.-F. (2014). Draft Genome Sequences of Four Strains of *Vibrio parahaemolyticus*, Three of Which Cause Early Mortality Syndrome/Acute Hepatopancreatic Necrosis Disease in Shrimp in China and Thailand. *Genome Announcements*, *2*(5). https://doi.org/10.1128/genomeA.00816-14

Yeh, H.-Y., Shoemaker, C. A., & Klesius, P. H. (2006). Sensitive and rapid detection of *Flavobacterium columnare* in channel catfish *Ictalurus punctatus* by a loop-mediated isothermal amplification method. *Journal of Applied Microbiology*, *100*(5), 919–925. https://doi.org/10.1111/j.1365-2672.2006.02853.x

Yu, L., Hu, Y., Zhang, X., & Sun, B. (2013). Development of a triplex loop-mediated isothermal amplification method for rapid on-site detection of three *Vibrio* species associated with fish diseases. *Aquaculture*, *414–415*, 267–273. https://doi.org/10.1016/j.aquaculture.2013.08.016

Zankari, E., Hasman, H., Kaas, R. S., Seyfarth, A. M., Agerso, Y., Lund, O., Larsen, M. V., & Aarestrup, F. M. (2013). Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *Journal of Antimicrobial Chemotherapy*, *68*(4), 771–777. https://doi.org/10.1093/jac/dks496

Zhao, X., Wang, L., Chu, J., Li, Y., Li, Y., Xu, Z., Li, L., Shirtliff, M. E., He, X., Liu, Y., Wang, J., & Yang, L. (2010). Rapid detection of *Vibrio parahaemolyticus* strains and virulent factors by loop-mediated isothermal amplification assays. *Food Science and Biotechnology*, *19*(5), 1191–1197. https://doi.org/10.1007/s10068-010-0170-3

Zhong, C., Han, M., Yang, P., Chen, C., Yu, H., Wang, L., & Ning, K. (2019). Comprehensive Analysis Reveals the Evolution and Pathogenicity of *Aeromonas*, Viewed from Both Single Isolated Species and Microbial Communities. *MSystems*, *4*(5). https://doi.org/10.1128/mSystems.00252-19

Zhulin, I. B. (2016). Classic Spotlight: 16S rRNA Redefines Microbiology. *Journal of Bacteriology*, *198*(20), 2764–2765. https://doi.org/10.1128/JB.00616-16

Appendices

Appendix 1

Strain names and GenBank accessions used in the comparative genomics analyses of Chapter 2. The names of *Aeromonas salmonicida* strains in bold were used in the phylogenetic and antimicrobial resistance analyses but not in the pan-genomics or ANI analyses.

Aeromonas salmonicida	
Strain Name	Assembly Accession
CBA100	**GCA_000746985.1**
strain 17	GCA_003259515.1
strain 3012STDY7122732	GCA_900683655.1
strain 947C	GCA_003947395.1
strain A308	GCA_003947375.1
strain A527	GCA_002764135.1
strain AJ83	GCA_003947355.1
strain BG	GCA_002093695.1
strain CIP 103210	GCA_006246135.1
strain CIP 104001	GCA_006246315.1
strain CIP 104757	GCA_006246305.1
strain CIP 107036	GCA_006243325.1
strain ECFood+05	GCA_002317045.1
strain JAOP-5	GCA_018274765.1
strain JF2480	GCA_009725185.1
strain M18076-11	GCA_002883135.1
strain NCTC10402	GCA_900445125.1
strain NCTC12959	GCA_900445115.1
strain S121	GCA_002214245.1
strain S44	GCA_002214305.1
strain S68	GCA_002214265.1
strain SRW-OG1	GCA_012931585.1
strain UBA10746	GCA_003483825.1
strain UBA8809	GCA_003505635.1
strain Y47	GCF_001481535.1
strain Y567	GCF_001466435.1
strain Y577	GCF_001481545.1
strain YK	GCA_002093675.1
strain Z5-5	GCA_003265515.1
strain Z5-5	**GCA_900491655.1**
subsp. achromogenes AS03	GCA_000315855.2
subsp. achromogenes NCIMB1110	Sequenced in this project
subsp. achromogenes strain 23051	GCA_006043955.1
subsp. achromogenes strain 23053	GCA_006044015.1
subsp. achromogenes strain 23055	GCA_006044075.1
subsp. achromogenes strain 23056	GCA_006044035.1
subsp. masoucida	GCA_002313065.1

subsp. masoucida NBRC 13784	GCA_000647955.1
subsp. masoucida strain BR19001YR	GCA_014872735.1
subsp. pectinolytica 34mel	GCA_000447435.1
subsp. pectinolytica 34mel	**GCA_002735225.1**
subsp. salmonicida	GCF_001902065.1
subsp. salmonicida 01-B526 1076135.3	**GCA_000234845.2**
subsp. salmonicida 2004-05MF26	GCA_000786805.1
subsp. salmonicida 2009-144K3	GCA_000786795.1
subsp. salmonicida A449	GCA_000196395.1
subsp. salmonicida CIP 103209	GCA_000820065.1
subsp. salmonicida strain 01-B522	GCF_001901975.1
subsp. salmonicida strain 09-0167	GCF_001902165.1
subsp. salmonicida strain 170-68	GCF_001901965.1
subsp. salmonicida strain 2004-072	GCA_016811055.1
subsp. salmonicida strain 2009-157 K5	GCF_001902045.1
subsp. salmonicida strain 2010-47 K18	GCF_001902055.1
subsp. salmonicida strain ATCC 33658	GCF_001643305.1
subsp. salmonicida strain BBCC2887	GCF_016918885.1
subsp. salmonicida strain J223	GCF_001643275.1
subsp. salmonicida strain J227	GCF_001643355.1
subsp. salmonicida strain J231	GCF_001643285.1
subsp. salmonicida strain J409	GCA_009858115.1
subsp. salmonicida strain J410	GCA_009858135.1
subsp. salmonicida strain J411	GCA_012933685.1
subsp. salmonicida strain JF2267	GCA_001901985.1
subsp. salmonicida strain JF2507	GCF_001902105.1
subsp. salmonicida strain JF3224	GCF_000931985.1
subsp. salmonicida strain JF3517	GCF_001902125.1
subsp. salmonicida strain JF3791	GCF_001597895.1
subsp. salmonicida strain m11743-09	GCA_004151085.1
subsp. salmonicida strain M15879-11 29491.36	**GCA_002110585.1**
subsp. salmonicida strain M16474-11	GCA_005476635.1
subsp. salmonicida strain M22710-11	GCA_002811185.1
subsp. salmonicida strain m23067-09	GCA_001902025.1
subsp. salmonicida strain RS534	GCF_001499805.1
subsp. salmonicida strain SHY16-3432	GCA_008370735.1
subsp. smithia strain JF4097	GCF_001466445.1
subsp. salmonicida NCIMB1102	Sequenced in this project

Flavobacterium branchiophilum	
Strain name	Genome accession
NBRC 15030 = ATCC 35035	GCA_007990815.1
NCIMB12904	Sequenced in this project
strain ATCC 35035	GCA_002217205.1
strain ATCC 35036	GCA_002530755.1

strain DSM 24789	GCA_006716585.1
strain FL-15	GCA_000253275.1

Flavobacterium psychrophilum	
Strain name	Genome accession
10	GCA_003732575.1
950106-1/1	GCA_000767095.1
CN	GCA_000801645.1
CR	GCA_000801625.1
CSF-259-93	GCA_000801635.1
DSM 3660	GCA_900101925.1
FPG101	GCA_000754365.1
FPG3	GCA_000754405.1
JIP02/86	GCA_000064305.2
NCIMB1947	Sequenced in this project
strain 010418-2/1	GCA_016803555.1
strain 030522-1/1	GCA_016803635.1
strain 13-SE026_S4_R1	GCA_900610015.1
strain 141127-1/2N	GCA_016803575.1
strain 160401-1/5M	GCA_016803615.1
strain 160401-1/5N	GCA_013343195.1
strain 21-SE080_S8_R1	GCA_900610065.1
strain 23-SE091_S9_R1	UXES01000000 (from GenBank)
strain 27-SE164_S11_R1	GCA_900610105.1
strain 29-SE328_S12_R1	GCA_900610085.1
strain 3	GCF_000971735.1
strain 31-SE409_S13_R1	GCA_900610095.1
strain 3-M2.99_S8_R1	GCA_900609985.1
strain 4	GCA_000971785.1
strain 43-SE552_S2_R1	GCA_900610175.1
strain 45-SE553_S11_R1	GCA_900610185.1
strain 47-SE554_S13_R1	GCA_900610165.1
strain 5	GCF_000971575.1
strain 51-SE556_S6_R1	GCA_900610215.1
strain 71-SE566_S4_R1	GCA_900610295.1
strain 75-SE568_S7_R1	GCA_900610335.1
strain 77-SE569_S9_R1	GCA_900610325.1
strain 7-SE009_S1_R1	GCA_900609995.1
strain 990512-1/2A	GCA_016803595.1
strain 9-SE010_S2_R1	GCA_900609975.1
strain ATCC 49418	GCA_002217405.1
strain CH1895	GCA_900186435.1
strain CH8	GCA_900186345.1
strain CSF259-93	GCF_000739395.1
strain DK001	GCA_900186405.1

strain DK002	GCA_900186425.1
strain DK095	GCA_900186385.1
strain DK150	GCA_900186365.1
strain F164	GCA_016803535.1
strain FI055	GCA_900186395.1
strain FI056	GCA_900186375.1
strain FI070	GCA_900186445.1
strain FI146	GCA_900186455.1
strain FI166	GCA_900186565.1
strain FPCH6	GCA_009730675.1
strain FPG10	GCA_002150255.1
strain FPG103	GCF_001887145.1
strain FPG1W08	GCA_002150245.1
strain FPG48	GCA_002150145.1
strain FPG87	GCF_001887135.1
strain FPG92	GCA_002150215.1
strain FPRT1	GCA_013426055.1
strain FPS-D10	GCA_016804085.1
strain FPS-D15	GCA_016804105.1
strain FPS-F15	GCA_016803495.1
strain FPS-F16	GCA_016803695.1
strain FPS-F21	GCA_016804225.1
strain FPS-F22	GCA_016804205.1
strain FPS-F27	GCA_016803675.1
strain FPS-F30	GCA_016804145.1
strain FPS-F32	GCA_016804165.1
strain FPS-F33	GCA_016804185.1
strain FPS-G1	GCA_016803515.1
strain FPS-R7	GCA_016803475.1
strain FPS-S10	GCA_016803735.1
strain FPS-S11A	GCA_016803715.1
strain FPS-S11B	GCA_016803775.1
strain FPS-S6	GCA_013343175.1
strain FPS-S9	GCA_016803655.1
strain FRGDSA 1882/11	GCA_900186415.1
strain IT02	GCA_900186545.1
strain IT09	GCA_900186525.1
strain IWL08	GCA_003788665.1
strain K9/00	GCA_016803755.1
strain LM-01-Fp	GCA_900186685.1
strain LM-02-Fp	GCA_900186665.1
strain MH1	GCF_000971605.1
strain NO004	GCA_900186755.1
strain NO014	GCA_900186785.1
strain NO042	GCA_900186825.1
strain NO083	GCA_900186805.1
strain NO098	GCA_900186815.1
strain OSU THCO2-90	GCA_900130145.1

strain P15/8B-11	GCA_016803815.1
strain P30-2B/09	GCA_016804065.1
strain P7-7B/10	GCA_016803795.1
strain PG2	GCF_000971645.1
strain V1-20 strain phage resistant V1-20	GCA_003433315.1
strain V2-20 strain phage resistant V2-20	GCA_003433335.1
strain V4-28 strain phage resistant V4-28	GCA_003433355.1
strain VQ50	GCF_000971685.1
V3-5	GCA_000831225.1
V4-24	GCA_000831185.1
v4-33	GCA_000831205.1

Moritella viscosa

Strain name	Genome accession
strain 06/09/139	GCF_900120285.1
strain F57	GCF_900120015.1
strain K56	GCF_900120145.1
strain K58	GCF_900119925.1
strain LFI 5006	GCF_900120065.1
strain MT 2528	GCF_900120075.1
strain NCIMB13584	Sequenced in this project
strain NVI 3632	GCF_900120025.1
strain NVI 4917	GCF_900120115.1
strain NVI 5450	GCF_900120035.1
strain NVI 5482	GCF_900119985.1
strain Vvi-11	GCF_900120105.1
strain Vvi-7	GCF_900120305.1
	GCF_000953735.1

Renibacterium salmoninarum

Strain name	Genome accession
ATCC 33209	GCA_000018885.1
strain DJ2R	GCA_003343265.1
strain H2	GCA_003343285.1
strain NCIMB2235	Sequenced in this project

Tenacibaculum maritimum

Strain name	Genome accession
NBRC 15946	GCA_000509405.1
strain 902	GCA_902705365.1
strain Aq16-85	GCA_902705305.1
strain Aq16-88	GCA_902705275.1
strain Aq16-89	GCA_902705375.1
strain CVI1001048	GCA_902705265.1

strain DPIF 89/0239-1	GCA_902705355.1
strain DPIF 89/3001-6.2	GCA_902705315.1
strain FC	GCA_902705415.1
strain FS08(1)	GCA_902705395.1
strain JIP 10/97	GCA_902705285.1
strain JIP 32/91-4	GCA_902705385.1
strain JIP 46/00	GCA_902705435.1
strain NAC SLCC MFF	GCA_902705345.1
strain NCIMB 2154T	GCA_900119795.1
strain NCIMB 2158	GCA_902705425.1
strain P1-39	GCA_902705535.1
strain P2-27	GCA_902705465.1
strain P2-48	GCA_902705555.1
strain P4-45	GCA_902705495.1
strain TFA4	GCA_902705565.1
strain TM-KORJJ	GCA_004803875.1
strain UCD SB2	GCA_902705445.1
strain USC SE30.1	GCA_902705525.1
strain USC SP9.1	GCA_902705515.1

Yersinia ruckeri	
Strain name	Genome accession
ATCC 29473	GCA_000173755.1
ATCC 29473	GCF_000754815.1
CSF007-82	GCA_000824965.1
SC09	GCA_000775355.1
strain 00/0652-K3	GCF_001883275.1
strain 00/1445	GCF_001883325.1
strain 00/1793-k2	GCF_001883365.1
strain 00/2994	GCF_001883405.1
strain 01/0230-7	GCF_001883415.1
strain 01/0298	GCF_001883425.1
strain 02/0972-2br	GCF_001883435.1
strain 02/0981-4br	GCF_001883485.1
strain 04/1749	GCF_001883565.1
strain 04/1779	GCF_001883575.1
strain 04/2640-8k	GCF_001883605.1
strain 05/0285-1K	GCF_001883615.1
strain 05/0297-5eye	GCF_001883065.1
strain 07/3342-1k	GCF_001882895.1
strain 07/3642-5K-b	GCF_001883075.1
strain 07/3828-6E	GCF_001883105.1
strain 08/0188-3K	GCF_001882925.1
strain 09/0217-5k	GCF_001882945.1
strain 11/4175-3k	GCF_001882955.1
strain 11/4666-4k	GCF_001883115.1
strain 12/3871-3K	GCF_001883145.1

strain 14/0125-1k	GCF_001883155.1
strain 150	GCA_001750505.1
strain 37551	GCF_000737165.1
strain 87/3421-SP	GCF_001880465.1
strain 88/3837	GCF_001880365.1
strain 88/3873	GCF_001880355.1
strain 88/4281-4	GCF_001880425.1
strain 89/3717-10	GCF_001880475.1
strain 89/4243	GCF_001880435.1
strain 90/0961-C9	GCF_001883765.1
strain 90/4316	GCF_001883745.1
strain 91/4311 A1	GCF_001883755.1
strain 91/4316	GCF_001882675.1
strain 92/5354-1	GCF_001882625.1
strain 93/1038-1	GCF_001882715.1
strain 93/5839-1	GCF_001882465.1
strain 95/4881-1	GCF_001883335.1
strain 95/6654-1	GCF_001883245.1
strain 96/5134-k	GCF_001883255.1
strain 97/0226-1 vc	GCF_001883345.1
strain 97/1152-1	GCF_001883265.1
strain AHL1	GCF_001883185.1
strain AHL2	GCF_001883195.1
strain AHL3	GCF_001882975.1
strain AHL4	GCF_001883225.1
strain AHL5	GCF_001882995.1
strain AHL6	GCF_001883025.1
strain AHL7	GCF_001883035.1
strain Big Creek 74	GCF_000964565.1
strain Feb-00	GCF_001883535.1
strain Feb-68	GCF_001883495.1
strain FMV-22	GCA_008086925.1
strain IP27752	GCA_902173145.1
strain IP27754	GCA_902173085.1
strain KMM821	GCA_017498685.1
strain Mar-05	GCF_001883525.1
strain NCIMB2194	Sequenced in this project
strain NCTC10476	GCA_900460675.1
strain NCTC12986	GCA_900460715.1
strain NHV_3758	GCA_002442495.1
strain OMBL4	GCF_001172905.1
strain QMA0440	CP017236 (Genbank accession)
strain RS41	GCF_001166725.1
strain SCPM-O-B-8085	GCA_002738395.1
YRB	GCA_000834255.1

Appendix 2

See end of appendix for product manufacturers.

Medium 1 – Nutrient agar / Nutrient Broth “E” (NBE)

Valid for *Aeromonas salmonicida* and *Yersinia ruckeri*.
Ingredients per litre:

Ingredient	Quantity
Beef extract	1.0 g
Yeast extract	2.0 g
Peptone	5.0 g
NaCl	5.0 g
(Agar)	15.0 g

pH 7.2-7.6

Comments: normally used 13.0g of NBE powder, which contains the ingredients listed above, and added agar or not, depending on the desired medium type (broth or agar).

Medium 98 – Enriched cytophaga agar

Valid for *Flavobacterium* species.
Ingredients per litre:

Ingredient	Quantity
Tryptone	2.0 g
Beef extract	0.5 g
Yeast extract	0.5 g
Sodium acetate	0.2 g
(Agar)	15.0 g

pH 7.2-7.4

Comments: *Flavobacterium branchiophilum* would no longer be viable after being passaged through this medium a small number of times. Medium CSCA was therefore preferred to grow *Flavobacterium* species.

Medium 398 – Tryptone Soya Agar plus NaCL

Valid for *Moritella viscosa*.
Ingredients per litre:

Ingredient	Quantity
Tryptone	15.0 g
Soya peptone	5.0 g
NaCl	25.0 g
(Agar)	15.0 g

pH 7.2-7.4

Comments: Normally prepared directly with ready-made Tryptic Soya Agar; this contains 15.0 g tryptone, 5.0 g soya peptone, 5.0 g NaCl and 15.0 g agar. 20.0 g of NaCl would then be added to the medium.

Medium 223 – Marine cytophaga medium A

Valid for *Tenacibaculum maritimum*.
Ingredients per litre:

Ingredient	Quantity
Tryptone	2.0 g
Beef extract	0.5 g
Yeast extract	0.5 g
Sodium acetate	0.2 g
Artificial seawater salts	24.5 g
(Agar)	15.0 g

pH 7.2-7.4

Charcoal Salts Cytophaga Agar (CSCA)

Valid for *Flavobacterium* species.
Ingredients per litre:

Ingredient	Quantity
Tryptone	0.5 g
Beef extract	0.2 g
Yeast extract	0.5 g
Sodium acetate	0.2 g
Gelatin	2.0 g
100 mg mL^{-1} KCl	0.5 mL
100 mg mL^{-1} $MgCl_2$	0.5 mL
100 mg mL^{-1} $CaCl_2$	0.5 mL
Charcoal powder	1 g
(Agar)	15.0 g

pH 7.2-7.4

Comments: From Skulska, 2014.

Kidney Disease Medium with Charcoal (KDM-C)

Valid for *Renibacterium salmoninarum*.
Ingredients per litre:

Ingredient	Quantity
Peptone	10.0 g
Yeast extract	0.5 g

L-cysteine hydrochloride	1.0 g
Charcoal powder	1.0 g
(Agar)	15.0 g

pH 6.7-6.9

Comments: Recipe from Daly and Stevenson, 1985.

Tryptone Soya Broth + 2% NaCl

Used to grow *Moritella viscosa* in liquid cultures.

Ingredient	Quantity
Pancreatic digest of casein	17.0 g
Enzymatic digest of soya bean	3.0 g
NaCl	25.0 g
K_2HPO_4	2.5 g
Glucose	2.5 g

pH 7.1-7.5

Comments: Normally prepared directly with medium CM0129 (Oxoid); this contains 17.0 g pancreatic digest of casein, 3.0 g enzymatic digest of soya (soya peptone), 5.0 g NaCl, 2.5 g K_2HPO_4, and 2.5 g glucose. 20.0 g of NaCl would then be added to the broth.

General comments:

All media were prepared by mixing powders in a Duran bottle or flask, adding distilled water to the desired volume, measuring pH, and finally adding, where appropriate, agar, gelatin or charcoal powder. pH correction was carried out by adding small volumes of NaOH (1 M) and HCl (5% v/v) until pH was within desired range. Media were then sterilized at 115 °C for 15 minutes or 121 °C for 20 minutes. The latter approach was used for media KDM-C and CSCA. Agar plates were poured in a sterile laminar flow hood.

Manufacturers:

Agar Number 2 bacteriological: Neogen®, Heywood, UK.

CM0129: Oxoid, Basingstoke, UK

L-cysteine hydrochloride: Sigma®, Sigma-Aldrich

NBE (Nutrient Broth “E”): Neogen®, Heywood, UK.

Peptone (bacteriological): Oxoid, Basingstoke, UK.

Tryptic Soya Agar: Merck, Darmstadt, Germany

Tryptone: Duchefa Biochimie B.V., Haarlem, the Netherlands

Yeast extract: Oxoid, Basingstoke, UK.

Appendix 3

This appendix includes the detailed instructions for the protocols compared during Chapter 4 (sections 4.2.4 and 4.2.6.4). All the reagents were sterilised either by autoclaving at 115 °C for 15 minutes or filtering through a 0.22 µm-pore filter, with the exception of ethanol and acetone.

Based on Boom *et al*. (1990)

Reagents:
- L6 lysis buffer -> Preparation as per instructions: 120 g guanidinium thiocyanate in 100 mL 0.1 M Tris-HCl (pH 6.4) (heat to ~60 °C and shake to dissolve), plus 22 mL 0.2 M EDTA (pH 8.0, adjusted with NaOH), plus 2.6 g Triton X-100. **Carried out in fume hood due to potential release of cyanide gas (lethal)**.
- Silica Coarse suspension -> 60 g of silica particles steeped in 500 mL of distilled water for 24 h, 430 mL removed, another 430 mL distilled water added, silica resuspended by shaking, steeped for 5 h, 440 mL water removed, add HCl to reduce pH to 2. Result is a slurry.
- L2 washing buffer -> Preparation as per instructions: 120 g of GuSCN in 100 mL of 0.1 M Tris-HCl (pH6.4) (heat to 60 °C and shake to dissolve). **Carried out in fume hood due to potential release of cyanide gas (lethal)**.
- TE buffer -> 10 mM Tris-HCl, 1 mM EDTA, pH 8.
- 70% ethanol
- Acetone

Equipment:
- Heating block (56 °C)
- Vortex
- Aliquot tubes (sterile 15 mL Falcon tubes should do)
- Centrifuge (12,000 x *g*)

Consumables:
- Preassembled reaction tubes -> 900 µL L6 buffer, 40 µL silica coarse suspension in a 1.5 mL tube. Vortex right before use to homogenise.
- 1.5 mL tubes
- P1000, P200 and wide-bore P200 tips

Method:

1. Add 900 µL of L6 solution to sample tube, then add 40 µL of silica coarse suspension with wide-bore P200 tips.
2. Vortex 5 s (approx.), incubate 10 min, vortex 5 s (approx.).
3. Centrifuge 12,000 x *g*, 15 s
4. Dispose of supernatant by suction
5. Wash silica-NA pellet with buffer L2 (twice), ethanol (twice) and acetone (once). Washing implies addition of 1 mL of each solution, vortex until complete pellet resuspension, and centrifugation at 12,000 x *g* for 15 s, followed by removal of supernatant.
6. After removal of acetone, dry in a heat block (56 °C) for 10 mins, lids open, to evaporate remaining solvents
7. Add 100 µL TE buffer (amount unspecified), vortex (“briefly”)
8. Incubate for 10 mins, 56 °C
9. Vortex (“briefly”)
10. Centrifuge 2 mins at 12,000 x *g*
11. Remove supernatant and store in 1.5 mL microcentrifuge tube

Adapted from Bordelon *et al.* (2013)

Reagents:
- Lysis buffer -> 50 mM Tris-HCl, 10 mM EDTA, 0.1 M NaCl, 5% Triton X-100
- Lysozyme solution -> 10 mg mL^{-1} lysozyme
- DNA precipitation buffer -> 80% ethanol, 5 mM K_2HPO_4, pH 8.5
- DNA wash buffer -> 70% ethanol
- DNA eluent -> Molecular grade water
- DNA silica adsorption buffer -> 4 M guanidine thiocyanate, 25 mM sodium citrate, pH 7.0, $6x10^8$ Dynabeads MyOne Silane magnetic beads

Equipment:
- Magnetic separating rack
- Pipettes (P1000, P200)
- Regular rack (for 1.5 – 2 mL tubes)

Consumables:
- Pipette tips (P1000, P200)
- 1.5 mL microcentrifuge tubes (3x)

Method:

1. Add 100 µL lysis solution + 25 µL lysozyme solution to tube containing sample, vortex, and digest at 25 °C with soft agitation for 20 mins. This can be done in the shaking incubator.
2. Centrifuge 5 mins, 10,000 x *g*
3. Remove supernatant to another tube, mix with 300 µL of DNA-silica adsorption buffer, add 20 µL of beads. Do not forget to shake and warm beads to room temperature before use. Put in shaking incubator again for 5 mins.
4. Transfer mix to a 1.5 mL microcentrifuge tube
5. Put tubes in magnetic rack, leave to rest 2 mins.
6. Remove buffer without touching the beads.
7. Add 300 µL DNA precipitation buffer, separate from magnets and mix with pipette.
8. Put tubes in magnetic rack, leave to rest for 2 minutes and remove precipitation buffer.
9. Add 300 µL DNA wash buffer (ethanol 70%), remove tubes from magnetic rack and mix with pipette.
10. Place tubes in magnetic rack, leave to rest 2 mins, remove the wash buffer.
11. Add DNA eluent (100 µL), remove tubes from magnetic rack, mix by pipetting up and down gently.
12. Put tubes back on the magnetic rack, leave to rest for 2 mins, and remove the eluate with a pipette to a new 1.5 mL tube.
13. DNA is ready for downstream applications

Adapted from Casas *et al*. (1995):

Reagents:
- Lysis buffer -> 4 M guanidinium thiocyanate, 0.5% N-laurolyl sarcosine, 1 mM dithiothreitol, 25 mM sodium citrate, 20 µg of glycogen per tube.
- Isopropyl alcohol (cold)
- 70% Ethanol
- Double distilled water
- 50 mg mL^{-1} lysozyme

Equipment:
- Cold centrifuge (14,000 x *g*)
- Rack

Consumables:
- 1.5 mL tubes (1x)
- Pipette tips

Method:

1. Add 100 µL of 50 mg mL^{-1} lysozyme to the sample tube, mix by pipetting or vortexing (10 s). Incubate at 37 °C for 30 minutes.
2. Add 400 µL lysis buffer to the tube containing the sample.
3. Vortex 10 s, incubate 10 minutes at room temperature.
4. Add 40 µg of glycogen and 500 µL of cold (-20 °C) isopropyl alcohol per tube. This precipitates nucleic acids.
5. Centrifuge samples at 4 °C and 14,000 x *g* for 10 minutes.
6. Remove isopropyl alcohol and wash pellet with 1 mL of 70% ethanol.
7. Centrifuge samples at 4 °C and 14,000 x *g* for 10 minutes.
8. Remove ethanol.
9. Dissolve pellet at bottom of tube in 100 µL water.
10. DNA is ready for downstream applications.

Adapted from Griffiths *et al.* (2000)

Reagents:
- CTAB buffer -> 10% w/v CTAB (hexadecyltrimethylammonium bromide) 0.7 M NaCl 240 mM potassium phosphate buffer, (pH 8.0)
- Phenol-chloroform-isoamyl alcohol (25:24:1)
- Chloroform-isoamyl alcohol (24:1)
- Polyethylene glycol (PEG) solution -> 30% w/v Polyethylene glycol 6000, 1.6M NaCl
- 70% ethanol
- TE buffer -> 1mM Tris (pH 8.0), 1 mM EDTA

Equipment:
- Refrigerated microcentrifuge (for 2 mL and 1.5 mL tubes)
- Vortex
- Bead-beating mill
- P1000 pipette

Consumables:
- 2 mL Screw-cap tubes (2x)
- Glass beads
- P1000 pipette tips
- 1.5 mL microcentrifuge tubes (2x)

1. Defrost samples if frozen.
2. Under the flow hood, add 1 mL of CTAB extraction buffer to filters.
3. Vortex 10 s and incubate 15 mins. Vortex again, then transfer 500 µL to another 2 mL screwcap tube.
4. Pour in approximately 7 small glass beads.
5. In the fume hood, add 500 µl phenol-chloroform-isoamyl alcohol (250 µl Chloroform + 250 µl Trizol reagent).
6. Vortex samples for 10 s.
7. Make sure caps are firmly screwed on tubes and then bead beat for 30 s on highest setting of bead-beating mill.
8. Centrifuge at 13,000 x *g* for 10 mins.
9. Remove aqueous phase (approx. 500 µl) and transfer into 1.5 ml microcentrifuge tube.
10. In the fume hood, add 500 µl chloroform-isoamyl alcohol to the aqueous phase extracted. Mix up and down with pipette.
11. Vortex for 10 s.
12. Centrifuge at 13,000 x *g* for 10 mins.
13. Remove supernatant and add to 1.5 ml microcentrifuge tubes containing 1 mL PEG (polyethelene glycol 6000) solution.
14. Leave to precipitate at room temp. for 1.5 – 2 h.
15. Centrifuge at 15,000 x *g* at 4 °C For 10 mins.
16. Using a 1 ml pipette remove all solution and add 1 ml 70% ethanol.
17. Centrifuge at 15,000 x *g* at 4 °C for 10 mins.

18. Using a 1 ml pipette, extract as much ethanol solution as possible without disturbing the DNA pellet, which should have formed at the base of the tube.
19. Add 1 ml 70% ethanol and repeat (Steps 16 - 18).
20. After the second wash, remove as much ethanol as possible from the tube by pipetting. Air dry the pellet in a heating block at 37 °C. Check liquid volume regularly and remove from heating block as soon as tube is dry.
21. Resuspend the pellet in 50 µl 1X TE buffer. Ensure the pellet is fully suspended by raking tubes along the rack and vortexing well.
22. Use in downstream applications.

Adapted from Martin-Platero *et al*. (2007)

Reagents:
- TES - L buffer -> 10% sucrose, 25mM Tris-HCl, 10 mM EDTA, pH8
- Lysozyme -> 10 mg mL^{-1} in TES buffer
- Lysis buffer -> 100 mM Tris-HCl, 100 mM EDTA, 10 mM NaCl, 1% SDS
- Proteinase K solution -> 10 mg mL^{-1} proteinase K
- NaAc (Sodium acetate) solution -> 3M NaAc, pH 5.2
- TE buffer -> 10 mM Tris-HCl, 1mM EDTA, pH8
- Isopropanol
- 70% ethanol

Equipment:
- Centrifuge (microfuge, 17,000 x *g*)
- P200, P1000 pipette
- Heating block (80 °C, 37 °C)
- Vortex

Consumables:
- 1.5 mL tubes (2x)

Methods:

1. Add 100 µL of TES - L buffer to the 2 mL tube with the sample. Vortex 5s or pipette up and down to mix.
2. Add 10 mg mL^{-1} of lysozyme to each tube (1 mg, as each tube has 0.1 mL)
3. 30 minute incubation at 37 °C. This “protoplasts” cells
4. Add 600 µL of lysis buffer, gently invert tube several times to mix. Incubate 10-15 mins at room temperature
5. Optional: digest proteins in the lysate with 10 µL of proteinase K for 15 mins at 37 °C
6. Heat to 80 °C for 5 mins, then cool to room temperature (inactivates most enzymes)
7. Add 200 µL sodium acetate and vortex-mix for 10-15 s
8. Chill on ice for 10-15 mins
9. Centrifuge at 17,000 x *g* for 10 mins to precipitate proteins
10. Transfer supernatant to new 1.5 mL tube (should be ~600 µL)
11. Wash with 1 volume of isopropanol; gently invert several times to mix
12. Centrifuge at 17,000 x *g* for 5 mins to pellet DNA
13. Wash with 1 volume of 70% ethanol
14. Resuspend in 100 µL TE buffer
15. Use in downstream applications

Adapted from Orsini and Romano-Spica (2001)

Reagents:
-Washing solution -> 50 mM Tris-HCl, pH 7.7, 25 mM EDTA, 0.1% SDS, 0.1% PVP
-Lysis buffer -> 50 mM Tris-HCl, 25 mM EDTA, 3% SDS, 1.2% PVP
-Extraction solution -> 10 mM Tris HCl, pH 8, 1 mM EDTA, 0.3 M NaOAc (Sodium acetate), 1.2% PVP
-Phenol-chloroform -> 50:50 phenol:chloroform
-Isopropyl alcohol -> unspecified volume (usually is 1 vol. in other protocols)
-Ethanol
-TE buffer -> (pH 8.0)

Equipment:
-Centrifuge (12,000 x *g*)
-Microwave (700 W)

Consumables:
-1.5 mL microcentrifuge tubes
-Pipette tips (P200, P1000)

Method:

1. Add 1 mL washing solution to tube containing sample. Vortex 10 s and incubate 15 mins.
2. Centrifugation at 6,000 x *g* for 1 min.
3. Discard supernatant and resuspend in 35 µL of lysis buffer.
4. Microwaved at 700 W for 45 s. Tube caps may be slightly unscrewed to relieve pressure.
5. 400 µL of pre-warmed extraction solution (~65 °C) added to tube.
6. DNA extraction with phenol-chloroform
   Add 435 µL of phenol-chloroform (50:50), vortex to mix and centrifuge (12,000 x *g*). Remove aqueous phase to 1.5 mL microcentrifuge tube.
7. Isopropyl alcohol precipitation (use 1 volume, ie. 435 µL approx.).
8. 70% ethanol wash.
9. Resuspension of precipitated DNA in 100 µL TE buffer.
10. Ready for use in downstream applications.

Pitcher and colleagues (1989) method:

Reagents:
- Lysozyme -> Lysozyme 50 mg mL^{-1}
- TE buffer -> 10 mM Tris-HCl, 1 mM EDTA, pH 8
- GES reagent -> 5 M guanidinium thiocyanate, 100 mM EDTA, 0.5% v/v sarkosyl.
Prepare as follows: Mix 60 g guanidinium thiocyanate into 20 mL 0.5 M EDTA (pH8) and 20 mL deionised water while heating at 65 °C until complete dissolution. Once cooled, add 5 mL of 10% v/v sarkosyl solution. Make up to 100 mL with deionised water, filter through a 0.45 µm pore size filter.
- Ammonium acetate -> 2.5 M – Use cold (4 °C)
- Chloroform 2-pentanol
- Propan-2-ol
- 70% ethanol
- Pure sterile water

Equipment:
- Vortex
- Centrifuge (17,000g)
- P200 and P1000 pipettes
- Heating block (37 °C)

Consumables:
- 1.5 mL microcentrifuge tubes (2x)
- P200, P1000 pipette tips
- Box full of ice

Method:

1. Add 1 mL of TE buffer to the tube containing the sample.
2. Vortex 10 s, incubate 15 mins, vortex 10 s.
3. Transfer 100 µL from this tube to a 1.5 mL microcentrifuge tube, then add 50 mg mL^{-1} Lysozyme (5 mg). Incubate at 37 °C for 30 minutes.
4. Add 500 µL of GES reagent (Guanidinium thiocyanate solution).
5. Vortex cell solution briefly (5-10 mins).
6. Cool on ice.
7. Add cold ammonium acetate (250 µL) and mix.
8. Hold on ice, 10 mins.
9. In fume hood, add chloroform 2-pentanol (24:1) mixture (500 µL).
10. Thorough mixing of phases.
11. Centrifuge at 17,000 x *g*, 10 mins.
12. Supernatant transferred to 1.5 mL microcentrifuge tubes and 0.54 volumes of cold 2-propanol were added.
13. Tube inversion for 1 minute to mix solutions.
14. Centrifuge at 6500 x *g* for 20 s.
15. Wash in 70% ethanol (five times).
16. Vacuum dry.
17. Dissolved DNA overnight at 4 °C in 100 µL sterile water.
18. DNA is ready for downstream applications.

Adapted PowerSoil protocol

Reagents:
- Buffers and solutions C1 – C6, supplied with kit

Equipment:
- Centrifuge (10,000 x *g*)
- Vortex (including tube-holding adapter)
- P1000 pipette

Consumables:
- Box of ice
- 2 mL tubes, supplied with kit (6x)
- Spin columns with waste-collection tube
- P1000 pipette tips

Method:

1. Empty contents of PowerBead tube into sample tube.
2. Add 60 µL solution C1, vortex 5 s.
3. Fasten tubes onto vortex adapter and vortex for 10 min at maximum speed.
4. Remove tubes to centrifuge, spin at 10,000 x *g* for 30 s.
5. Siphon supernatant off with P1000 pipette and place in new 2 mL tube
6. Add 250 µL of solution C2, vortex for 5 s.
7. Incubate on ice for 5 mins.
8. Centrifuge tubes at 10,000 x *g* for 1 min.
9. Transfer 600 µL (or as much as possible) of supernatant to new 2 mL tube, avoiding pellet.
10. Add 200 µL of solution C3, vortex 5 s.
11. Incubate on ice for 5 mins.
12. Centrifuge tubes at 10,000 x *g* for 1 min.
13. Avoiding pellet transfer 750 µL (or as much as possible) of supernatant to new 2 mL tube
14. Add 1200 µL solution C4 (well mixed) to the tube (2 volumes of 600 µL). Vortex for 5 s.
15. Transfer 675 µL to a spin column and centrifuge at 10,000 x *g* for 1 min.
16. Discard flow-through and repeat previous step until all the solution has been processed (2 more times).
17. Add 500 µL solution C5 to column, centrifuge at 10,000 x *g* for 30 s.
18. Discard flow-through and centrifuge again for 1 min.
19. Move spin column to 2 mL collection tube and add 100 µL of solution C6.
20. Centrifuge at 10,000 x *g* for 30 s.
21. DNA in tube is ready for downstream applications.

Appendix 4

Accesions and subspecies of 16S sequences used in the phylogenetic analysis of the *Tenacibaculum maritimum* strain obtained from the NCIMB.

Species	Strain	Accession
Flavobacterium aquatile	DSM 1132	AM230485.1
Flexibacter maritimus		D14023.1
Polaribacter filamentus		U73726.1
Tenacibaculum adriaticum	B390	NR_042579.1
Tenacibaculum aestuarii	SMK-4	NR_043713.1
Tenacibaculum aestuariivivum	JDTF-79	NR_159126.1
Tenacibaculum agarivorans	HZ1	NR_159214.1
Tenacibaculum aiptasiae	a4	NR_044202.1
Tenacibaculum amylolyticum	NBRC 16310	NR_113842.1
Tenacibaculum amylolyticum	MBIC4355	NR_024737.1
Tenacibaculum ascidiaceicola	RSS1-6	NR_148772.1
Tenacibaculum caenipelagi	HJ-26M	NR_125675.1
Tenacibaculum crassostreae	JO-1	NR_044498.1
Tenacibaculum dicentrarchi	35/09	NR_108475.1
Tenacibaculum discolor	LL04 11.1.1	NR_042576.1
Tenacibaculum gallaicum	A37.1	NR_042631.1
Tenacibaculum gallaicum	A37.1T	AM746477.1
Tenacibaculum geojense	YCS-6	NR_117983.1
Tenacibaculum haliotis	RA3-2	NR_158003.1
Tenacibaculum holothuriorum	S2-2	NR_145845.1
Tenacibaculum insulae	JDTF-31	NR_159277.1
Tenacibaculum jejuense	CNURIC013	NR_116704.1
Tenacibaculum litopenaei	B-I	NR_043967.1
Tenacibaculum litoreum	CL-TF13	NR_043302.1
Tenacibaculum lutimaris	TF-26	NR_043080.1
Tenacibaculum maritimum	NBRC 15946	NR_113825.1
Tenacibaculum mesophilum	NBRC 16307	NR_113841.1
Tenacibaculum mesophilum	MBIC1140	NR_024736.1
Tenacibaculum ovolyticum	IFO 15947	NR_040912.1
Tenacibaculum ovolyticum	NBRC 15947	NR_113826.1
Tenacibaculum sediminilitoris	YKTF-3	NR_149768.1
Tenacibaculum skagerrakense	D30	NR_025229.1
Tenacibaculum soleae	LL04 12.1.7	NR_042630.1
Tenacibaculum soleae	LL04 12.1.7	AM746476.1
Tenacibaculum xiamenense	WJ-1	NR_109729.1

Appendix 5

List of NCBI Assembly accessions used during the design of LAMP primers during Chapter 5. The first table shows the set of target genomes, that is, the genomes that the primers designed were expected to target. These all belong to different strains of the same target species.
The second table shows the set of non-target, "background" genomes, which served to avoid primers binding to undesired close relatives of the target species.

Table 1:

Species	Accessions of species genomes
Aeromonas salmonicida	GCF_000204115.1, GCF_000388115.1, GCF_000687355.2, GCF_000708125.1, GCF_000764645.1, GCF_000764655.1, GCF_000764665.1, GCF_000819725.1, GCF_000819805.1, GCF_000819865.1, GCF_000819885.1, GCF_000820005.1, GCF_000820025.1, GCF_000820045.1, GCF_000820085.1, GCF_000820125.1, GCF_000820145.1, GCF_000820165.1, GCF_000820185.1, GCF_001306015.1, GCF_001481395.1, GCF_002795305.1, GCF_002812985.1, GCF_002906925.1, GCF_003265465.1, GCF_003265495.1, GCF_003491245.1, GCF_006246425.1, GCF_010974825.1, GCF_900476005.1, GCF_900637545.1
Flavobacterium branchiophilum	GCF_002217205.1, GCF_002530755.1, GCF_006716585.1, GCF_007990815.1

Flavobacterium psychrophilum	GCF_000064305.2, GCF_000739395.1, GCF_000754365.1, GCF_000767095.1, GCF_000801615.1, GCF_000801695.1, GCF_000801715.1, GCF_000831185.1, GCF_000831205.1, GCF_000831225.1, GCF_000971575.2, GCF_000971605.2, GCF_000971645.2, GCF_000971685.2, GCF_000971735.2, GCF_000971785.1, GCF_001485515.1, GCF_001510795.1, GCF_001510855.1, GCF_001510875.1, GCF_001887135.1, GCF_001887145.1, GCF_002150145.1, GCF_002150215.1, GCF_002150245.1, GCF_002150255.1, GCF_002150295.1, GCF_002207815.1, GCF_002217405.1, GCF_003433315.1, GCF_003433335.1, GCF_003433355.1, GCF_003732575.1, GCF_003788665.1, GCF_006777745.1, GCF_006777825.1, GCF_006777945.1, GCF_006778065.1, GCF_006778185.1, GCF_006778265.1, GCF_006778345.1, GCF_006778465.1, GCF_006778585.1, GCF_009730675.1, GCF_013343175.1, GCF_013343195.1, GCF_900101925.1, GCF_900130145.1, GCF_900186345.1, GCF_900186365.1, GCF_900186375.1, GCF_900186385.1, GCF_900186395.1, GCF_900186405.1, GCF_900186415.1, GCF_900186425.1, GCF_900186435.1, GCF_900186445.1, GCF_900186455.1, GCF_900186475.1, GCF_900186485.1, GCF_900186495.1, GCF_900186515.1, GCF_900186525.1, GCF_900186545.1, GCF_900186555.1, GCF_900186565.1, GCF_900186575.1, GCF_900186595.1, GCF_900186605.1, GCF_900186645.1, GCF_900186665.1, GCF_900186685.1, GCF_900186755.1, GCF_900186785.1, GCF_900186805.1, GCF_900186815.1, GCF_900186825.1, GCF_900609965.1, GCF_900609975.1, GCF_900609985.1, GCF_900609995.1, GCF_900610005.1, GCF_900610015.1, GCF_900610025.1, GCF_900610035.1, GCF_900610045.1, GCF_900610055.1, GCF_900610065.1, GCF_900610075.1, GCF_900610085.1, GCF_900610095.1, GCF_900610105.1, GCF_900610115.1, GCF_900610125.1, GCF_900610135.1, GCF_900610145.1, GCF_900610155.1, GCF_900610165.1, GCF_900610175.1, GCF_900610185.1, GCF_900610195.1, GCF_900610205.1, GCF_900610215.1, GCF_900610225.1, GCF_900610235.1, GCF_900610245.1, GCF_900610255.1, GCF_900610265.1, GCF_900610275.1, GCF_900610285.1, GCF_900610295.1, GCF_900610305.1, GCF_900610315.1, GCF_900610325.1, GCF_900610335.1, GCF_900610345.1, GCF_900610355.1, GCF_900610385.1
Moritella viscosa	GCF_900119925.1, GCF_900119985.1, GCF_900120015.1, GCF_900120025.1, GCF_900120035.1, GCF_900120065.1, GCF_900120075.1, GCF_900120105.1, GCF_900120115.1, GCF_900120145.1, GCF_900120285.1, GCF_900120305.1
Renibacterium salmoninarum	GCF_003343265.1, GCF_003343285.1
Tenacibaculum maritimum	GCF_004803875.1, GCF_902705265.1, GCF_902705275.1, GCF_902705285.1, GCF_902705305.1, GCF_902705315.1, GCF_902705345.1, GCF_902705355.1, GCF_902705365.1, GCF_902705375.1, GCF_902705385.1, GCF_902705395.1, GCF_902705415.1, GCF_902705425.1, GCF_902705435.1, GCF_902705445.1, GCF_902705465.1, GCF_902705495.1, GCF_902705515.1, GCF_902705525.1, GCF_902705535.1, GCF_902705555.1, GCF_902705565.1

Yersinia ruckeri	GCF_000173755.1, GCF_000737165.1, GCF_000775355.2, GCF_000824965.1, GCF_000834255.1, GCF_001166725.1, GCF_001172905.1, GCF_001750505.1, GCF_001880355.1, GCF_001880365.1, GCF_001880425.1, GCF_001880435.1, GCF_001880465.1, GCF_001880475.1, GCF_001882465.1, GCF_001882625.1, GCF_001882675.1, GCF_001882715.1, GCF_001882895.1, GCF_001882925.1, GCF_001882945.1, GCF_001882955.1, GCF_001882975.1, GCF_001882995.1, GCF_001883025.1, GCF_001883035.1, GCF_001883065.1, GCF_001883075.1, GCF_001883105.1, GCF_001883115.1, GCF_001883145.1, GCF_001883155.1, GCF_001883185.1, GCF_001883195.1, GCF_001883225.1, GCF_001883245.1, GCF_001883255.1, GCF_001883265.1, GCF_001883275.1, GCF_001883325.1, GCF_001883335.1, GCF_001883345.1, GCF_001883365.1, GCF_001883405.1, GCF_001883415.1, GCF_001883425.1, GCF_001883435.1, GCF_001883485.1, GCF_001883495.1, GCF_001883525.1, GCF_001883535.1, GCF_001883565.1, GCF_001883575.1, GCF_001883605.1, GCF_001883615.1, GCF_001883745.1, GCF_001883755.1, GCF_001883765.1, GCF_002192595.1, GCF_002442495.2, GCF_002738395.1, GCF_008086925.1, GCF_900460675.1, GCF_900460715.1, GCF_902173085.1, GCF_902173145.1

Table 2

Species	Accessions of genus representatives
Aeromonas salmonicida	GCF_000204115.1, GCF_000388115.1, GCF_000687355.2, GCF_000708125.1, GCF_000764645.1, GCF_000764655.1, GCF_000764665.1, GCF_000819725.1, GCF_000819805.1, GCF_000819865.1, GCF_000819885.1, GCF_000820005.1, GCF_000820025.1, GCF_000820045.1, GCF_000820085.1, GCF_000820125.1, GCF_000820145.1, GCF_000820165.1, GCF_000820185.1, GCF_001306015.1, GCF_001481395.1, GCF_002795305.1, GCF_002812985.1, GCF_002906925.1, GCF_003265465.1, GCF_003265495.1, GCF_003491245.1, GCF_006246425.1, GCF_010974825.1, GCF_900476005.1, GCF_900637545.1
Flavobacterium branchiophilum	GCF_000016645.1, GCF_000240075.2, GCF_000378485.1, GCF_000419685.1, GCF_000422685.1, GCF_000422705.1,

Flavobacterium psychrophilum	GCF_000422725.1, GCF_000425425.1, GCF_000425445.1, GCF_000425465.1, GCF_000425485.1, GCF_000425505.1, GCF_000430025.1, GCF_000455605.1, GCF_000498475.1, GCF_000498495.1, GCF_000498515.1, GCF_000498535.1, GCF_000611675.1, GCF_000686885.1, GCF_000695795.1, GCF_000735715.2, GCF_000737685.1, GCF_000737695.1, GCF_000757385.1, GCF_000769915.1, GCF_000832125.1, GCF_001278115.1, GCF_001404985.1, GCF_001602525.1, GCF_001637185.1, GCF_001686925.1, GCF_001761465.1, GCF_001831475.1, GCF_001857965.1, GCF_002217285.1, GCF_002217355.1, GCF_002217395.1, GCF_002217445.1, GCF_002217475.1, GCF_002222055.1, GCF_002251775.1, GCF_002251835.1, GCF_002846575.1, GCF_002917885.1, GCF_002920895.1, GCF_003055625.1, GCF_003076455.1, GCF_003076475.1, GCF_003096035.1, GCF_003097535.1, GCF_003122385.1, GCF_003148385.1, GCF_003254545.1, GCF_003254585.1, GCF_003254625.1, GCF_003254745.1, GCF_003259835.1, GCF_003268815.1, GCF_003268855.1, GCF_003293845.1, GCF_003312425.1, GCF_003344925.1, GCF_003350545.1, GCF_003385115.1, GCF_003385895.1, GCF_003590565.1, GCF_003634455.1, GCF_003634755.1, GCF_003688495.1, GCF_003858535.1, GCF_003865365.1, GCF_003865405.1, GCF_003996965.1, GCF_004329815.1, GCF_004345565.1, GCF_004352915.1, GCF_004355225.1, GCF_004362665.1, GCF_004363695.1, GCF_004634245.1, GCF_004745595.1, GCF_004797125.1, GCF_006385255.1, GCF_006491595.2, GCF_007341385.1, GCF_007830355.1, GCF_008369745.1, GCF_008806775.1, GCF_009363055.1, GCF_009741375.1, GCF_010645065.1, GCF_010645075.1, GCF_900099915.1, GCF_900100165.1, GCF_900100375.1, GCF_900101895.1, GCF_900106645.1, GCF_900107635.1, GCF_900108015.1, GCF_900108395.1, GCF_900108955.1, GCF_900110375.1, GCF_900110615.1, GCF_900111075.1, GCF_900111965.1, GCF_900112575.1, GCF_900112975.1, GCF_900115115.1, GCF_900129405.1, GCF_900129545.1, GCF_900129555.1, GCF_900129575.1, GCF_900129585.1, GCF_900129705.1, GCF_900142035.1, GCF_900142695.1, GCF_900142715.1, GCF_900142735.1, GCF_900142775.1, GCF_900142885.1, GCF_900148835.1, GCF_900182645.1 (GCF_000253275.1, GCF_000754405.1)
Moritella viscosa	GCF_000276805.1, GCF_008931805.1, GCF_900465055.1
Renibacterium salmoninarum	None
Tenacibaculum maritimum	GCF_000430545.1, GCF_001483385.1, GCF_001693415.1, GCF_001889045.1, GCF_001936575.1, GCF_002120225.1, GCF_003387615.1, GCF_003610735.1, GCF_003664185.1, GCF_003937815.1, GCF_004345825.1, GCF_004363005.1, GCF_008124875.1, GCF_008806755.1, GCF_900129475.1, GCF_900198195.1, GCF_900239485.1

Yersinia ruckeri	GCF_000222975.1, GCF_000582515.1, GCF_000754805.1, GCF_000834455.1, GCF_000834865.1, GCF_001047675.1, GCF_001112925.1, GCF_001152565.1, GCF_001182085.1, GCF_001319545.1, GCF_001656035.1, GCF_002188995.1, GCF_004124235.1, GCF_009831415.1, GCF_011765335.1, GCF_013282725.1, GCF_900637475.1, GCF_902168495.1

Appendix 6

Digital appendix with scripts showing the way GLAPD was used to design LAMP primers. These scripts and explanations are available at:
https://github.com/algaraber/Digital_appendix/tree/main/Appendix6

Password for compressed and encrypted data:
AaxE6Lj4kFCo

Appendix 7

Digital appendix with scripts used to plot distance between potentially amplifying primer pairs on non-target genomes. The results of this process, shown in section 5.2.3, were used to decide the LAMP primers that were tested in Chapter 5.

Accessible at:
https://github.com/algaraber/Digital_appendix/tree/main/Appendix7

Password for compressed and encrypted data:
AaxE6Lj4kFCo

Appendix 8

Reference sequence for primer design	Target species	Primer	Sequence (5' - 3')
Vibrio parahaemolyticus operon 1	*Aeromonas hydrophila* *Aeromonas salmonicida* *Aliivibrio salmonicida* *Edwardsiella ictaluri* *Edwardsiella tarda* *Moritella viscosa* *Photobacterium damselae* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Vibrio alginolyticus* *Vibrio anguillarum* *Vibrio parahaemolyticus* *Yersinia ruckeri*	F3	GGCAGATAGGGACCGAACT
		FIP	GTCGGCTCATCACATCCTGGGCAG CTCGCGTACCACTTT
		BIP	GCCAAACACCGCCGTCGATAGGTA CTCCGGGGATAACAGG
		B3	GAAGGGCCATCGCTCAAC
Pseudomonas fluorescens operon 1	*Aeromonas hydrophila* *Aeromonas salmonicida* *Aliivibrio salmonicida* *Edwardsiella ictaluri* *Edwardsiella tarda* *Moritella viscosa* *Photobacterium damselae* *Pseudomonas aeruginosa* *Pseudomonas fluorescens* *Vibrio alginolyticus* *Vibrio anguillarum*	F3	GAAGGGCCATCGCTCAAC
		FIP	GCCAAACACCGCCGTCGATAGGTA CTCCGGGGATAACAGG
		BIP	ACCTCGATGTCGGCTCATCACAGC GTACCACTTTAAATGGCG
		B3	GAACTGTCTCACGACGTTCT

	Vibrio parahaemolyticus *Yersinia ruckeri*		
Flavobacterium psychrophilum operon 1	*Flavobacterium branchiophilum* *Flavobacterium columnare* *Flavobacterium psychrophilum* *Tenacibaculum maritimum*	F3	TGCAATGGCATAAGGGAGC
		FIP	ATGGCCCTTCCATGCGGAACACAG GTCGATCAGGTACGAA
		BIP	AACAGGCTGATCTCCCCCAAGATG ACGAGCCGACATCGAG
		B3	TTCTCCAGCCCCAGGATG
Streptococcus iniae operon 1	*Lactococcus garvieae* *Streptococcus iniae*	F3	GAAAGCGTGGGGAGCAAA
		FIP	GCGGAGTGCTTAATGCGTTAGCTA GTCCACGCCGTAAACGA
		BIP	CTGGGGAGTACGACCGCAAGCAT GCTCCACCGCTTGTG
		B3	TCTTCGCGTTGCTTCGAATT
Renibacterium salmoninarum operon 1	*Nocardia seriolae* *Renibacterium salmoninarum*	F3	CGCGGCCTATCAGCTTG
		FIP	CGTGTCTCAGTCCCAGTGTGGACC AAGGCGACGACGG
		BIP	CTCCTACGGGAGGCAGCAGTCGTC GCTGCATCAGGC
		B3	AGGCCGTCATCCCTCAC

Appendix 9

Digital appendix that contains the BRL_classifier.sh script, which assigns DNA amplified though the broad-range LAMP to its bacterial species of origin. It also contains one sample dataset, derived from the one obtained in section 6.2.3.1.1 of the thesis (3500 reads for each of the samples sequenced).
Available at:
https://github.com/algaraber/Digital_appendix/tree/main/Appendix9

Password for compressed and encrypted data:
AaxE6Lj4kFCo