Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

A Selective MDA - ONT diagnostic for *Vibrio parahaemolyticus*

School of Agriculture and Environment

College of Sciences

Massey University

Thomas Liebregts

2026

Abstract

Accurate detection of *Vibrio parahaemolyticus* is important for food safety, public health, and environmental monitoring. Although whole-genome sequencing provides high-resolution characterisation of bacterial pathogens, direct sequencing can be limited by low DNA abundance and uneven genome representation. This study evaluated whether selective multiple displacement amplification (sMDA) combined with Oxford Nanopore Technologies (ONT) sequencing could provide an effective strategy for enriching *V. parahaemolyticus* DNA prior to sequencing. Three complementary approaches were used. Existing short-primer sets were first evaluated experimentally. A new primer set was then designed using computational analysis of genome-wide primer-binding frequency and distribution across publicly available *V. parahaemolyticus* reference genomes. Finally, the selected primer set was validated experimentally using DNA extracted from cultured isolates, followed by ONT sequencing and genome coverage analysis. Computational analyses identified substantial differences in primer abundance and genomic distribution, allowing candidate primers with broad chromosome I representation to be selected. Experimental evaluation demonstrated that the optimised primer set generated sufficient DNA for ONT sequencing. Mapping of sequencing reads to the reference genome showed that 99.956% of chromosome I was covered at a minimum depth of 1×, while 98.611% was covered at a depth of at least 10×. Despite this extensive overall coverage, sequencing depth was heterogeneous, reflecting the distribution of primer-binding sites. Regions containing few suitable primer-binding sites exhibited reduced coverage, and several chromosome II-associated virulence loci showed substantially lower coverage than chromosome I targets. Comparison of computational predictions with experimental genome coverage demonstrated that primer-binding site distribution was a major determinant of sequencing performance. These findings support the computational primer design strategy while identifying primer distribution as a key limitation affecting genome-wide enrichment. Overall, this study demonstrates that selective MDA combined with ONT sequencing provides a feasible proof-of-concept approach for targeted enrichment and genomic characterisation of *V. parahaemolyticus*. The work establishes a rational framework for computational primer design and experimental validation while identifying optimisation of primer coverage, particularly across chromosome II, as a priority for developing a comprehensive genomic surveillance and diagnostic platform.

Acknowledgments

I would like to firstly thank my supervisor, Professor Peter Lockhart, one of the kindest and most dedicated people I have had the pleasure to meet and work with. I am profoundly grateful for his mentorship and for giving me a firm grounding in my endeavour for research. His depth and expertise of the field have been instrumental to the success of this project.

I would also like to thank my laboratory supervisor Trish McLenachan for her invaluable help during the practical part of my research. I am extremely grateful for her wisdom and patience as a newbie to the field, which has been critical in forming my skills and knowledge.

I would also like to express my special gratitude to Professor Paletasa Havea and Dr. Sesimani Havea. Their generosity and kindness have been integral to my post graduate education. I am extremely grateful to them for their continued support and guidance, and this work is merely a small example of the tremendous contribution they have made.

Last but not least I would like to thank my parents for their love, support and faith in me. The unwavering encouragement they provided has been a beacon of strength and motivation throughout my academic journey.

List of Figures

		Page
Figure 1.1	Structure of *Vibrio parahaemolyticus*	11
Figure 1.2	T3SS1 process based on induction signals	12
Figure 2.1	Loading of the MinION flow cell prior to sequencing	27
Figure 3.1	Genome-wide coverage of *Vibrio parahaemolyticus* isolate PFR30J09	47
Figure 3.2	Genome-wide coverage of *Vibrio parahaemolyticus* isolate PFR21C03	47
Figure 3.3	Sequencing coverage across the type III secretion system 1 (T3SS1) locus	52
Figure 3.4	Sequencing coverage across the type VI secretion system 1 (T6SS1) locus	53
Figure 3.5	Sequencing coverage across the multivalent adhesion molecule 7 (*mam7*) gene	53
Figure A3.1	Sequencing coverage across the thermolabile haemolysin (*tlh*) gene	77
Figure A3.2	Sequencing coverage across the ferric vibrioferrin receptor (*pvuA*) gene	78
Figure A3.3	Sequencing coverage across the flagellar motor protein (*motY*) gene	79
Figure A3.4	Sequencing coverage across the lateral flagellin (*lafA*) gene	80
Figure A3.5	Sequencing coverage across the flagellar motor switch protein (*fliM*) gene	81
Figure A3.6	Sequencing coverage across the flagellar hook protein (*flgE*) gene	83

List of Tables

		Page
Table 2.1	Primer sequences used for sMDA of *Vibrio parahaemolyticus*	24
Table 3.1	DNA concentrations following MDA using the VpCl primer set	29
Table 3.2	Coverage Results of VpCl Primers	30
Table 3.3	DNA concentrations following MDA using the cpmt primer set	31
Table 3.4	Coverage Results of Cpmt Primers	32
Table 3.5	Predicted genome-wide abundance of candidate primer recognition sites across three complete chromosome I genomes of *Vibrio parahaemolyticus*	34
Table 3.6	Predicted spatial distribution of primer recognition sites across chromosome I of three complete *Vibrio parahaemolyticus* genomes	36
Table 3.7	Correlation of predicted primer recognition-site abundance among three complete chromosome I genomes of *Vibrio parahaemolyticus*	38
Table 3.8	Predicted recognition-site abundance of the Vib1 primer set within selected virulence-associated loci of *Vibrio parahaemolyticus*	40
Table 3.9	DNA concentrations obtained from two *Vibrio parahaemolyticus* isolates following extraction using the Roche High Pure PCR Template Preparation Kit and the LysoFAST BOOST protocol	42
Table 3.10	DNA concentrations following selective multiple displacement amplification using the Vib1 primer set	43
Table 3.11	Summary of Oxford Nanopore sequencing quality following trimming and quality filtering	44
Table 3.12	Read length statistics following trimming and quality filtering	45
Table 3.13	Genome-wide coverage statistics for *Vibrio parahaemolyticus* isolate PFR30J09	46
Table 3.14	Recovery of representative virulence-associated loci following selective multiple displacement amplification and Oxford Nanopore sequencing	49
Table 3.15	Distribution of sequencing depth across representative virulence-associated loci	50

Contents

Abstract i

Acknowledgments ii

List of Figures iii

List of Tables iv

Chapter 1 Introduction 1

1.1 *Vibrio parahaemolyticus* as a pathogen. 1

1.1.1 Impact on Human Health 2

1.1.2 A Global and Local Problem 3

1.1.3 The Need for Timely Detection 4

1.2. The Biology and Pathogenesis of *Vibrio parahaemolyticus* 5

1.2.1 Genetic diversity of *Vibrio* parahaemolyticus 8

1.2.2 Key virulence factors 8

1.2.3 Mechanisms of infection and host-pathogen interactions. 10

1.2.4 Challenges in identifying pathogenic versus non-pathogenic strains. 13

1.3 Methods of identification 14

1.3.1 Culture based methods 14

1.3.2 Serology and biochemical tests 15

1.3.3 Molecular diagnostics 16

1.3.4 Whole genome sequencing 16

1.4 Selective MDA–ONT Workflow for *Vibrio parahaemolyticus* 17

1.4.1 Mechanism and Advantages of Selective MDA 17

1.4.2 Integration with ONT Sequencing 18

1.4.3 Applications and Implications 18

1.5 *Vibrio parahaemolyticus* Genomic Database 19

1.5.1 *Vibrio parahaemolyticus* Genomic Database 19

1.5.2 Integration with sMDA–ONT Workflow 20

1.6 Thesis Aims 22

Chapter 2 Materials and Methods 23

2.1 Primers 23

2.2 DNA source 24

2.3 Cultures of *Vibrio parahaemolyticus* 24

2.4 DNA Extraction: LysoFAST BOOST 24

2.5 DNA extraction using the High Pure PCR Template Preparation Kit 25

2.6 DNA Quantification: Qubit 25

2.7 DNA Quantification: Electrophoresis 25

2.8 Multiple Displacement Amplification 26

2.9 Column clean up 26

2.10 ONT Sequencing: Rapid barcoding 26

2.11 ONT Sequencing: loading a flowcell 27

2.12 Bioinformatic Analysis 28

Chapter 3 Results 29

3.1 Experimental evaluation of the VpCl primer set 29

3.2 Experimental evaluation of the cpmt primer set 31

3.3 Computational evaluation of candidate Vib1 primers 33

3.3.1 Predicted genome-wide abundance of candidate primer recognition sites 34

3.3.2 Predicted spatial distribution of primer recognition sites 35

3.3.3 Robustness of primer-binding site abundance estimates across reference genomes 37

3.3.4 Predicted distribution of *Vib*1 primer recognition sites within virulence-associated loci 39

3.4 Experimental validation of computational predictions 41

3.4.1 DNA extraction and template quality 41

3.4.2 Whole-genome amplification using the *Vib*1 primer set 42

3.5 Oxford Nanopore sequencing quality 44

3.6 Validation of computational predictions by genome mapping 46

3.6.1 Genome-wide coverage 46

3.6.2 Recovery of virulence-associated loci 48

3.6.3 Comparison of computational predictions with experimental genome recovery 54

Chapter 4 Discussion 57

4.1 Performance of Primer Sets in MDA 57

4.1.1 Comparative amplification efficiency of the VpCl and cpmt primer sets 57

4.1.2 Amplification bias and genomic motif distribution 58

4.1.3 Rationale and necessity of computationally optimised *Vibrio*-specific primer design 60

4.1.4 Implications for optimisation and diagnostic performance 61

4.2 Feasibility of MDA from Culture-Derived DNA 63

4.3 Genome recovery following selective multiple displacement amplification 65

4.4 Recovery of Virulence-Associated Loci 67

4.4.1 Recovery of adhesion, motility and iron acquisition genes 67

4.4.2 Implications for virulence profiling 67

4.5 Public Health Relevance in a Regional and Global Context 69

4.5.1 Public health significance of the workflow 69

4.5.2 Applications in outbreak investigation and surveillance 69

4.5.3 Implications for climate change and Pacific communities 70

4.5.4 Limitations and future public health applications 71

4.6 Limitations and Future Improvements 72

Chapter 5 Conclusion 74

Appendix 76

Coverage profiles of additional virulence-associated loci following sMDA 76

Ranking of 6mer primer-binding site frequencies 83

References 84

Chapter 1 Introduction

1.1 *Vibrio parahaemolyticus* as a pathogen.

With nearly 100 documented species, *Vibrio* bacteria significantly impact various aspects of human life, particularly public health and the international seafood industry (Haque et al., 2023). These impacts are most pronounced in cases of vibriosis, an infection commonly acquired through the consumption of raw or undercooked shellfish (Bienlien et al., 2022). The rising incidence of vibriosis outbreaks has heightened the importance of effective identification and prevention measures (Neil et al., 2023). Although most *Vibrio* species are non-pathogenic, a select few warrants monitoring due to their potential to cause disease. Understanding the genetic and ecological factors driving *Vibrio* pathogenicity is critical for developing effective mitigation strategies.

Among these pathogenic species, *Vibrio parahaemolyticus* is a leading cause of seafood-associated illnesses, responsible for symptoms including diarrhoea, stomach cramps, nausea, and vomiting (Bell & Bott, 2021). This Gram-negative, halophilic bacterium thrives in marine environments, particularly in warm, rainy seasons and regions near the equator (Fearnley et al., 2024). Its prevalence is strongly influenced by environmental factors, such as rising sea surface temperatures and changes in salinity, which enhance its growth and survival (Baker-Austin et al., 2024). These conditions are particularly prominent in estuarine and coastal ecosystems, where *V. parahaemolyticus* can proliferate and contaminate seafood. A key factor in its pathogenesis is its ability to produce haemolysins—toxins that lyse red blood cells, enabling the bacteria to acquire iron from haemoglobin and disrupt cellular integrity (Raghunath, 2015).

V. parahaemolyticus has continued to evolve since its discovery in 1950 following a large outbreak in Japan, which resulted in 272 cases with 20 deaths (Fujino et al., 1953). Outbreaks of *V. parahaemolyticus* within the 20th century were rare and were vastly overshadowed by its closely related species, *V.cholerae* (Hu et al., 2016). Today, *V. parahaemolyticus* is the leading cause of seafood infections globally with an estimated half a million cases in 2020 (Trinanes & Martinez-Urtaza, 2021). These infections have become a major public health concern in countries such as the United States, Japan, and, more recently, Australia (Neil et al., 2023; Ueno et al., 2016). This leaves queries on how it will impact countries which depend on seafood as a resource and as a source of income.

1.1.1 Impact on Human Health

Building on its status as a significant global foodborne pathogen, *Vibrio parahaemolyticus* has profound implications for human health as the leading cause of seafood related illnesses. Its ability to cause a wide spectrum of illnesses, ranging from mild gastroenteritis to severe systemic infections underscores its public health importance (Wang et al., 2015). The pathogen has been documented to cause wound infections such as cellulitis, fasciitis and rapid tissue spread (Guillod et al., 2019; W. Xu et al., 2024). Outbreaks of the bacterium are prone to occur during warmer months and at coastal locations where high seafood consumption is a necessity (Brumfield et al., 2023). Understanding the transmission pathways, clinical manifestations, and potential severity of infections caused by *V. parahaemolyticus* is critical for mitigating its impact on affected populations.

The primary route of *V. parahaemolyticus* transmission is through the consumption of contaminated raw or undercooked shellfish. Outbreaks have also been linked to improperly cooked seafood, including fish, crabs, and lobsters (Marudhupandi et al., 2017; Seelman et al., 2023). Inadequate refrigeration and cross-contamination during seafood handling further increase the risk of infection by promoting bacteria growth (Cabello-Olmo et al., 2020). These improper cooking and storage practices highlight the increasing risk of infection, emphasizing the need for proper handling and preparation of seafood.

Although rare, *V. parahaemolyticus* can be transmitted through physical contact, particularly in individuals with open wounds. The bacterium can invade these exposed areas, leading to soft tissue infections such as cellulitis and fasciitis (Tan et al., 2017). For instance, a recent case of a victim who contracted the bacterium through contact with contaminated sewage water, which ultimately proved fatal (W. Xu et al., 2024). This alludes to the transmission of the bacterium predominantly appearing in unsanitary environments.

Symptoms of the bacteria typically include diarrhoea, abdominal cramps, nausea and vomiting, which most commonly appear within 16-24 hours of exposure (Wu et al., 2023). These symptoms are nonspecific and overlap with many other gastrointestinal illnesses, making accurate diagnosis challenging (Letchumanan et al., 2014). Although stool culture remains the gold standard for confirming infection in patients, the method is time-consuming (Davis et al., 2017). The delays in diagnosis not only prolong patient suffering but may hinder timely outbreak control, highlighting the urgent need for rapid and accessible identification tools for

V. parahaemolyticus. Emerging molecular diagnostic techniques offer promising solutions to address these limitations (Letchumanan et al., 2014).

1.1.2 A Global and Local Problem

V. parahaemolyticus represents a growing global and local public health issue with significant economic and societal impacts. The pathogen disrupts international seafood trade, particularly in major seafood exporting countries such as China, Vietnam and Thailand, where shellfish are a key commodity (Nakaguchi, 2013; Vu et al., 2022; Zhong et al., 2022). Its presence has led to implementation of stricter food safety regulations in key importing nations, including the U.S, Japan and European countries (Hara-Kudo & Kumagai, 2014). These regulations often result in export restrictions and bans when the bacterium is detected (Hu et al., 2021; Ngasotter et al., 2022). Consequently, such measures result in significant economic losses across the supply chain, impacting exporters, distributors and processors.

Consumer confidence also suffers during outbreaks, resulting in reduced demand for seafood and destabilizing aquaculture-dependent economies. Consumers often change their dietary habits when a resource has been perceived as unsafe to consume (Rahmaniya & Sekharan N, 2018). This shift disproportionately affects small-scale producers, particularly in developing nations, who often lack the resources necessary to implement effective pathogen detection and prevention measures (Hoque et al., 2022). This is particularly evident in Pacific Island nations, where seafood is vital for both nutrition and income (Brewer et al., 2023; Georgeou et al., 2022).

Climate change exacerbates the situation, creating warmer marine environments that enable *Vibrio* species to thrive (Brumfield et al., 2023). This has led to an expanded geographical range, with temperate regions in Europe, North America, and the Gulf of Mexico now experiencing outbreaks (Hara-Kudo & Kumagai, 2014; León-Sicairos et al., 2022). The increased prevalence of *Vibrio* infections poses additional burdens on public health systems and the seafood industry, as longer outbreaks strain resources and affect seafood quality (Zhong et al., 2022). Altered pH levels and rising temperatures in coastal waters further elevate the risk of infections.

Although warmer seawater temperatures generally favour the growth of *Vibrio parahaemolyticus* and are associated with increased infection risk during summer, infections in New Zealand are not restricted to the warmer months. Surveillance data from 1998–2024

indicate that infections occurred throughout the year, with a higher proportion of cases recorded between January and March (King et al., 2026). Furthermore, outbreaks have been documented during winter, including events in May–June 2019 and June–July 2020 (Jefferies et al., 2026). These findings demonstrate that seasonal increases in risk should not be interpreted as an absence of risk during colder months.

In Pacific Island nations such as Fiji, Samoa, and Vanuatu, the reliance on seafood creates increased vulnerabilities to *Vibrio*-related illnesses (Singh et al., 2012). Limited access to advanced diagnostic tools and prevention strategies compounds the risks in these regions. Addressing these challenges requires collaborative global efforts, including climate adaptation strategies, enhanced food safety measures, and targeted support for vulnerable communities (Baker-Austin et al., 2024).

1.1.3 The Need for Timely Detection

Accurate and rapid diagnosis is crucial in controlling foodborne pathogens. Prompt diagnosis facilitates the timely implementation of identification, treatment and prevention measures (Phillips et al., 2024). Conventional culture-based methods can identify Vibrio parahaemolyticus but do not necessarily establish the pathogenic potential of an isolate. PCR assays targeting virulence-associated genes, including tdh and trh, can identify important pathogenicity markers, although these markers do not fully characterise the pathogenic potential of every strain (Eldin et al., 2019). Emerging technologies like CRISPR-based methods and Multiple Displacement Amplification (MDA) are revolutionizing pathogen detection, offering greater speed and accuracy (Ghouneimy et al., 2023).

The consequences of delayed diagnosis are significant, impacting public health, economic stability, and timely responses to outbreaks. Misdiagnosis or late identification can prolong infections, exacerbate symptoms, and increase disease severity, particularly in vulnerable populations such as the elderly and immunocompromised (Eldin et al., 2019). Rapid and precise diagnosis enables effective clinical management, ensuring timely treatment and reducing morbidity (Belard et al., 2017). Economically, delayed detection contributes to higher healthcare costs, extended hospitalizations, and economic losses in affected communities reliant on aquaculture (Suneja et al., 2022).

Furthermore, rapid detection stunts potential outbreaks by enabling public health authorities to issue timely health warnings and advisories about contaminated seafood. Public awareness

campaigns can further complement these efforts, educating consumers on safe handling and consumption practices (Hara-Kudo & Kumagai, 2014). This is particularly crucial as *Vibrio parahaemolyticus* exhibits growing antibiotic resistance (Tan et al., 2017).

Early identification of resistant strains informs appropriate treatment strategies, prevents the use of ineffective antibiotics, and aids in the development of containment measures to mitigate risks (Suneja et al., 2022). These efforts align with international public health goals, emphasizing the need for robust surveillance systems and global cooperation to manage the threat posed by this pathogen effectively (Liu et al., 2024).

1.2. The Biology and Pathogenesis of *Vibrio parahaemolyticus*

Bacterial Classification

Vibrio parahaemolyticus is a Gram-negative, rod-shaped bacterium belonging to the family *Vibrionaceae*. It is a halophilic microorganism, meaning it requires saline environments to survive and proliferate, making it particularly common in marine and estuarine ecosystems (Zhang & Orth, 2013). Taxonomically, it is closely related to other *Vibrio* species, such as *V. cholerae, V. alginolyticus* and *V. vulnificus*, which are also known to be capable of causing infections in humans (Montieri et al., 2010). This classification underpins its ecological niche and informs strategies for its identification and control in both environmental and clinical contexts.

Optimal Growth Conditions

Environmental parameters are crucial for the growth of *V. parahaemolyticus* and for its prevalence as a pathogen. The optimal growth conditions for *Vibrio* depend on the life stage either planktonic or biofilm formation. Within the planktonic stage, optimal temperatures for *V. parahaemolyticus* range between 36–37 °C, whereas in the biofilm stage optimally grows around *30 °C* (Velez et al., 2023). Non-pathogenic strains tend to have negative trends of growth when exposed to higher temperatures whereas positive trends are observed in pathogenic strains (Wang et al., 2020).

V. parahaemolyticus has been shown to grow in large temperature ranges as well. A recent study observed growth of *V. parahaemolyticus* within bivalve molluscan shellfish from New

Zealand at 5-30 °C (Fletcher et al., 2024). Another study on the complex relationship of growth of *Vibrio.spp.* and temperature found that *V. parahaemolyticus* are able to grow within 10–35 °C, indicating potential harm to humans, due to body temperature (Sheikh et al., 2022). The bacterium also has a high tolerance range of 5-45 ^{0}C in various setups such as broth system, inoculated in liquids and Minimal Marine Medium agar.

In the case of salinity levels, *V. parahaemolyticus* can grow within 1-8% salt concentrations, with optimal growth achieved at around 2-3% (20-25 ppt) (Pazhani et al., 2021) (Namadi & Deng, 2021). High NaCl levels can inhibit the survival of the bacterium, effectively by draining water out of the cells (Pang et al., 2020). Studies show that low salinity levels increased pathogenicity of *V. parahaemolyticus*, reduced gut microbiota species richness and distort microbial functions (Zhang et al., 2023) (Chang et al., 2024).

Additionally, extreme precipitation weathers can cause drops in salinity levels, leading to optimal growth of the bacterium and outbreaks (Gassen et al., 2024). There is also a distinct salinity difference near the equator and at both poles. At the equator, there are warm temperate oceans which is an optimal habitat for its rainy seasons and for its salinity levels (Röthig et al., 2023).

Furthermore, the pH level of an environment influences the growth of *V. parahaemolyticus* as well. Recent studies on the bacterium's optimum environmental conditions generally shows that growth can be seen throughout the pH range 5-9.5, with the optimum growth at pH level 7.96 ± 0.1(Pumipuntu & Indrawattana, 2017) (Namadi & Deng, 2023). Surface ocean water is generally slightly basic, which further solidifies as the optimum habitat for the bacterium.

Ocean acidification could be detrimental, however, adaptations seen in other *Vibrio* show great affinity with long exposure to slightly acidic environments (Asplund, 2013) (Pumipuntu & Indrawattana, 2017). *V. parahaemolyticus* has shown enhanced survival at lower pH levels, increasing protection against low salinity levels and temperature. In contrast to alkaline conditions, the adapted cells were found to be more resistant towards heat, crystal violet, deoxycholic acid and hydrogen peroxide (Velez et al., 2023).

V. parahaemolyticus, like other *Vibrio.spp.,* are facultative anaerobic. Its growth can be stunted by strict anaerobic levels but not limited-oxygen conditions (Tran et al., 2020). During host infection, the bacterium grows optimally in the presence of oxygen compared to anaerobic

environments (Letchumanan et al., 2014) . An *in* vivo environment study showed that low oxygen conditions are more beneficial than higher conditions (Davis et al., 2017).

On the other hand, studies have shown that *V. parahaemolyticus* has a much stronger ability to develop biofilms under anaerobic conditions in comparison with micro-aerobiosis and aerobiosis(Qian et al., 2020). The low oxygen level in the human gastrointestinal tract also challenges the survival and infection of *V. parahaemolyticus* as biofilms are essential for growth and reproduction (H. Xu et al., 2024). Other optimal attributes which provide growth are high concentrations of bile acids and amino acids, which have been shown to promote the production of pathogenic haemolysins (Lane et al., 2023).

Life Cycle

The life span and life cycle of *V. parahaemolyticus* are intricately tied to its environment. The bacterium has three life stages; i.) a free-living planktonic swimming cell, ii.) a sessile swarmer cell along with other swarmers attached to surfaces, and iii.) an infection-associated state (Freitas et al., 2019). The bacterium demonstrates remarkable persistence in its free-living form, attaching to particulate matter or biofilms, which aids in nutrient acquisition and survival during unfavourable conditions (Broberg et al., 2011).

In its swimming stage, *V. parahaemolyticus* is monotrichous (one flagella), however, during swarming and in response to surface contact, it differentiates into a filamentous flagellated cell (Letchumanan et al., 2014). This results in numerous flagella positioned laterally along the cell body. It's second stage allows for a symbiotic relationship with sea animals (eg. Shellfish) where it can attach to and be provided protection (Chimalapati et al., 2020).

When ingested by a host, *V. parahaemolyticus* can establish an infection in a susceptible host, exploiting its host's resources to replicate and spread (Letchumanan et al., 2014). Strains which are pathogenic to humans may not affect marine animals due to differences in host susceptibility and the specific virulence factors require to cause disease in different organisms (Ngasotter et al., 2022). Its ability to cycle between these states increases its ecological success and pathogenic potential, particularly in seafood and in host animals.

1.2.1 Genetic diversity of *Vibrio* parahaemolyticus

The widespread prevalence of *Vibrio parahaemolyticus* is partly attributed to its genetic diversity, which facilitates adaptation to diverse environmental conditions and may contribute to the emergence and spread of antimicrobial resistance (Thaotumpitak et al., 2024). This genetic diversity of *Vibrio parahaemolyticus* allows for high rates of survivability within multiple environments and diverse conditions. Thus, the bacteria is able to keep up with the ever changing world around (He et al., 2021). The genome consists of two chromosomes, with recent studies finding roughly consisting of ~4500-5500 genes (Ghouneimy et al., 2023; LaPorte et al., 2023; Vandeputte et al., 2024).

Vibrio parahaemolyticus exhibits considerable genetic diversity, and many environmental isolates lack the classical virulence-associated genes tdh and trh. These genes are more frequently associated with clinical isolates, although their presence or absence does not provide a complete assessment of pathogenic potential (Wang et al., 2020). All *V. parahaemolyticus* strains are often identified by their somatic antigen - O (lipopolysaccharide) and capsular antigen - K (capsular polysaccharide) (Chen et al., 2010). There are 13 O serotypes and 71 K serotypes which have been recorded. For instance, strain O3:K6 is known as a globally distributed pathogenic serotype, causing multiple outbreaks (Velazquez-Roman et al., 2014). Those with pathogenic strains are often characterized by the presence of specific virulence genes, such as *tdh*: thermostable direct haemolysin and *trh*: TDH-related haemolysin

1.2.2 Key virulence factors

The pathogenicity of *Vibrio parahaemolyticus* is driven by a suite of virulence factors that enable it to invade host tissues, evade immune defences, and cause disease. Central to its pathogenesis are haemolysins and secretion systems. These major virulence factors include:

Haemolysins

Thermostable Direct Haemolysin (TDH):

Thermostable direct haemolysin (TDH) is an important virulence-associated toxin produced by some pathogenic strains of V. parahaemolyticus. Encoded by the tdh gene, TDH contributes to haemolytic activity and is associated with gastrointestinal disease (Raghunath, 2015). Encoded by the *tdh* gene, it is primarily associated with clinical isolates, while being rare in

environmental strains (Paranjpye et al., 2012). Composed of over 150 amino acids, this toxin plays a critical role in the survival and fitness of the bacterium in pathogenic contexts (Yanagihara et al., 2010).

TDH-Related Haemolysin (TRH):

Similar in function to TDH, TRH is another significant virulence factor encoded by the *trh* gene. It differs from TDH by being thermolabile (Gutierrez West et al., 2013). It forms a tetramer and increases intracellular calcium, opening chloride channels, causing watery diarrhoea. TRH is divided into *trh*1 and *trh*2, differing in nucleotide composition. *trh*1 tends to be more prevalent in clinical strains whereas *trh*2 is more commonly associated within environmental isolates (Paria et al., 2019).

Thermolabile Haemolysin (TLH):

Encoded by the *tlh* gene, TLH is another haemolysin whose role is a lot less defined than TDH and TRH. The tlh gene is commonly used as a molecular marker for identifying V. parahaemolyticus, but its detection does not, by itself, establish that an isolate is pathogenic (Park & Zhang, 2024).

Secretion Systems

Type III Secretion Systems (T3SS):

These systems inject effector proteins into host cells, disrupting signalling pathways and promoting bacterial invasion and infection. The bacterium employs three T3SSs: T3SS1, T3SS2α, and T3SS2β (Plaza et al., 2021). The first system is linked to causing cytotoxicity (destruction on cells) whereas T3SS2α and T3SS2β are mainly associated with enterotoxicity (digestive system damage) leading to gastroenteritis symptoms (Li et al., 2019; Park et al., 2004). The common targets of T3SS effectors include the actin cytoskeleton, the immune signalling and autophagy (Letchumanan et al., 2014).

Type VI Secretion Systems (T6SS):

Type VI secretion systems (T6SSs) are contact-dependent secretion apparatuses that deliver effector proteins into neighbouring cells. Unlike type III secretion systems, which primarily

target eukaryotic host cells, T6SSs can target bacterial and eukaryotic cells. V. parahaemolyticus possesses two major T6SSs, T6SS1 and T6SS2, which contribute to interbacterial competition, environmental adaptation and host interactions (Ben-Yaakov & Salomon, 2019; Tchelet et al., 2023). Although the distribution of T6SS-associated genes varies among lineages, their presence alone is insufficient to establish that a strain is pathogenic or belongs to a pandemic lineage.

1.2.3 Mechanisms of infection and host-pathogen interactions.

Adhesion to Host Cells

Initial attachment of V. parahaemolyticus to host cells is facilitated by multivalent adhesion molecule 7 (MAM7), an outer membrane adhesin containing seven mammalian cell entry (MCE) domains. MAM7 interacts with fibronectin, an extracellular matrix glycoprotein, and phosphatidic acid, a phospholipid associated with host-cell membranes. These interactions facilitate bacterial attachment to host cells and contribute to the early stages of infection. Blocking these interactions can reduce bacterial adhesion (Liu & Chen, 2015; Isom et al., 2017; Li et al., 2019).

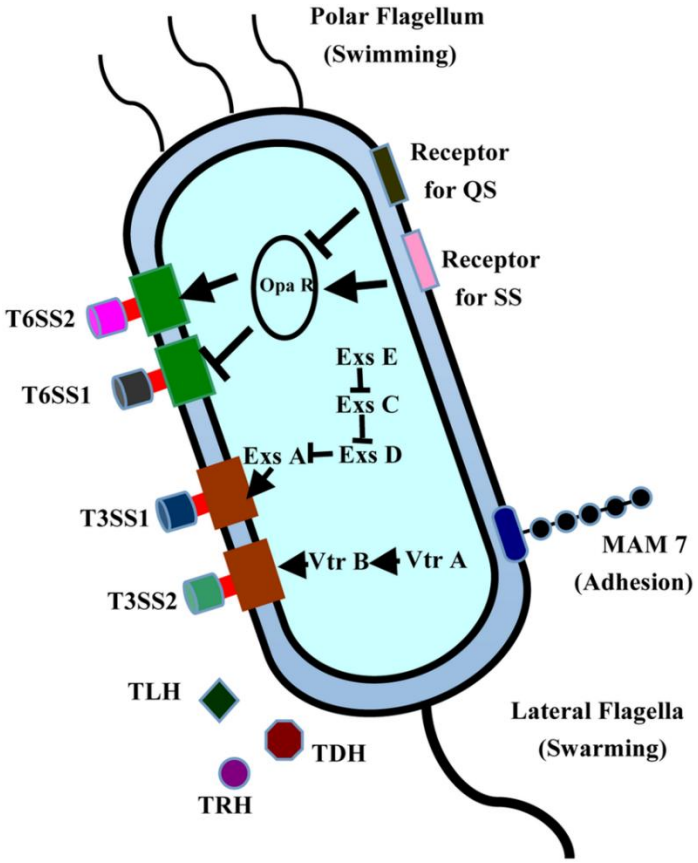


Figure 1.1: Structure of V. parahaemolyticus (Wang et al., 2015)

T3SS1

The T3SS1 system in *Vibrio parahaemolyticus* is regulated through the coordinated action of three regulatory proteins—ExsC, ExsD and ExsE—which control the master transcriptional activator ExsA (Li et al., 2019). Under non-inducing conditions, ExsD binds to ExsA and inhibits its activity, while ExsE binds to ExsC. Under inducing conditions, secretion of ExsE releases ExsC, which subsequently binds to ExsD, thereby releasing ExsA from inhibition and allowing transcription of T3SS1 genes (Gu et al., 2020; Letchumanan et al., 2014). Activation of T3SS1 enables the secretion of effector proteins that disrupt host-cell processes, contributing to autophagy, membrane blebbing, cell rounding and ultimately cell lysis (Nguyen et al., 2020).

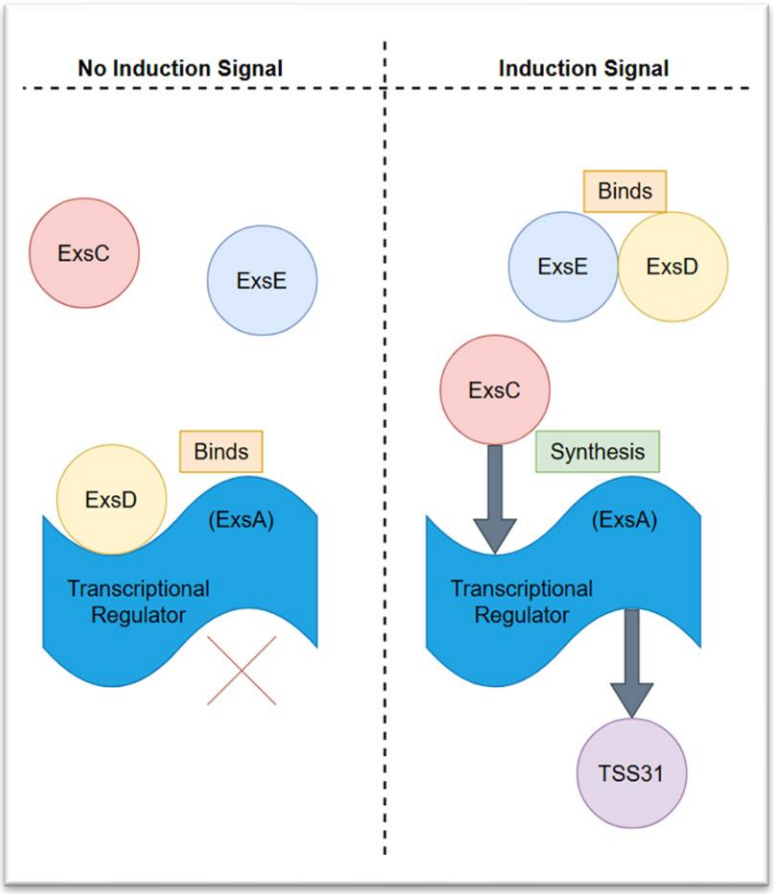


Figure 1.2: T3SS1 process based on induction signals

The pathogenesis is further driven by three main effector proteins: VopQ, VPA0450, and VopS.

1) **VopQ:** inhibits phagocytosis by inducing autophagy in host cells, forcing them to degrade their own components, thereby evading immune destruction (Nguyen et al., 2020).

2) **VPA0450:** destabilizes the host cell by detaching the plasma membrane from the actin cytoskeleton, leading to membrane blebbing and rapid cell lysis (Waddell et al., 2014).

3) **VopS:** Catalyses the AMPylation of Rho-family GTPases, thereby inhibiting their interactions with downstream signalling proteins. This disrupts actin cytoskeletal regulation and contributes to changes in host-cell morphology and function (Li et al., 2019).

This coordinated action of the T3SS1 system and its effectors demonstrates the sophisticated mechanisms *V. parahaemolyticus* employs to manipulate host cells, evade immune responses, and establish infection.

T3SS2

Unlike T3SS1, which primarily targets bacterial evasion of host immune responses, T3SS2 plays a significant role in the enterotoxicity and cytotoxicity of *V. parahaemolyticus (Plaza et al., 2024)*. There are in total 11 effector proteins which can be translocated to the eukaryotic cell. The main effectors of this secretion system are—VopC, VopT, VopA/P, and VopL—each target different aspects of host cell function to promote infection and evade host defences (Letchumanan et al., 2014). These effectors work synergistically to manipulate host cell signalling and structure, facilitating bacterial invasion and survival.

Together, these effectors enable *V. parahaemolyticus* to effectively manipulate the host cell's internal membrane. Collectively, these mechanisms illustrate the complexity of *V. parahaemolyticus* pathogenesis and its ability to circumvent host immune defences, contributing to its virulence and persistence within the host (Kaval et al., 2023).

1.2.4 Challenges in identifying pathogenic versus non-pathogenic strains.

Various difficulties can be brought up when differentiating between pathogenic and non-pathogenic strains. This can be alluded to its high genetic similarity (Guerrero et al., 2017). This similarity can complicate the diagnosis of pathogenic strains based solely on genetic markers (Vandeputte et al., 2024). For instance, a comparative genomic analysis of six *V. parahaemolyticus* strains sampled from different geographical regions found that 94% of the genetic differences were within a 141 kb region around the O- and K-antigen gene clusters. These findings indicate that the core genome is highly conserved among these strains (Chen et al., 2010)

In addition to genetic similarity, the phenotypic plasticity exhibited by *V. parahaemolyticus* poses another significant challenge. Phenotypic plasticity refers to the ability of an organism to modify its behaviour, physiology, or morphology in response to environmental conditions. While the species exhibits high genetic conservation (Chen et al., 2011), it also demonstrates diverse strain variation, with behaviours differentiating over time and across environments (Kumar et al., 2020).

For instance, pathogenic strains such as RIMD 2210633, may carry key virulence genes but fail to express them in poor conditions. In such environments, these strains exhibit little to no haemolytic activity, imitating the behaviour of non-pathogenic strains (Gutierrez West et al.,

2013). Furthermore, the absence of host-specific signals, such as bile salts, can suppress the expression of virulence genes, further reducing pathogenicity (Han et al., 2017). This adaptability often leads to the misclassification of pathogenic strains as non-pathogenic, further complicating efforts to develop a universal indicator.

1.3 Methods of identification

The identification of *Vibrio parahaemolyticus* is a critical first step in developing preventive measures to control the bacteria. Methods for identification are evaluated based on several factors, including accuracy, sensitivity, specificity, cost-effectiveness, ease of use, scalability, portability, and the ability to detect virulence markers and antibiotic resistance (Arbefeville et al., 2024). Furthermore, distinguishing pathogenic strains from non-pathogenic ones is essential for accurate identification.

Early identification methods relied on microscopy and Gram staining, which revealed *V. parahaemolyticus* as a Gram-negative, curved rod-shaped bacterium, characteristic of the *Vibrio* genus (Broberg et al., 2011). These tests led to the Kanagawa Phenomenon, used to identify pathogenic strains of *V. parahaemolyticus*. This study was used to observe the haemolytic factors produced by the bacteria (Chun et. al 1975). Subsequent testing on Thiosulfate-Citrate-Bile Salts-Sucrose (TCBS) agar allowed for successful culturing and differentiation of *V. parahaemolyticus* from other *Vibrio* species, such as *Vibrio cholerae (Yeung & Thorsen, 2016)*. These foundational methods laid the groundwork for the development of more advanced identification techniques, which are now integral to modern diagnostics.

1.3.1 Culture based methods

In early research of *V. parahaemolyticus*, agar-based media helped confirm its halophilic nature, distinguishing it from many non-halophilic foodborne pathogens (Bottone & Robin, 1978). The introduction of TCBS Agar (Thiosulfate-Citrate-Bile Salts-Sucrose) as a selective medium enabled the isolation of *Vibrio* species (Di Pinto et al., 2011). Its ease of use, affordability, and suitability for a variety of settings make it a preferred choice for isolating *V. parahaemolyticus*.

V. parahaemolyticus forms green colonies on TCBS agar due to its inability to ferment sucrose, which differentiates it from sucrose-fermenting *Vibrio* species, such as *V.cholerae* (Di Pinto et al., 2011*).* To date, TCBS Agar remains the most widely used medium for isolating *Vibrio parahaemolyticus.* The medium is selective for the *Vibrio* species, as it contains bile salts and thiosulfate that inhibit the growth of most non-*Vibrio* organisms. Other common mediums used are CHROM agar and *Vibrio* Selective Agar (Nigro & Steward, 2015).

However, TCBS agar lacks the ability to differentiate pathogenic from non-pathogenic strains. This agar is commonly used in combination with molecular or biochemical methods to confirm pathogenicity (Tagliavia et al., 2019). Additionally, the 12–48-hour incubation period can be a limitation in time-sensitive outbreak situations.

1.3.2 Serology and biochemical tests

Serological tests are widely used for detecting specific antigens, such as the O and K antigens, on the surface of *Vibrio parahaemolyticus*. These tests are typically performed using the agglutination method, where bacterial cells are mixed with specific antisera on a glass slide. If the antisera recognise the bacterial antigens, clumping occurs, indicating a positive agglutination reaction. O- and K-antigen serotyping is particularly useful for the epidemiological characterisation of *V. parahaemolyticus* isolates (Chen et al., 2010; U.S. Food and Drug Administration [FDA], 2004).

The agglutination test is quick, inexpensive and suitable for initial characterisation. However, serological tests alone are insufficient for definitive species identification because cross-reactivity can occur with other marine bacteria (FDA, 2004).

Biochemical tests such as indole production and salt tolerance tests can help differentiate *V. parahaemolyticus* from other species within the genus *Vibrio*. These tests assess characteristic physiological and metabolic properties, including salt requirements and the ability to metabolise tryptophan. However, biochemical characteristics can vary among isolates, and identification generally requires interpretation of several test results together (FDA, 2004).

While serological and biochemical tests provide valuable initial information for identifying *V. parahaemolyticus*, their limitations highlight the importance of integrating molecular techniques for more reliable and rapid pathogen detection.

1.3.3 Molecular diagnostics

Molecular diagnostic techniques, particularly polymerase chain reaction (PCR) and loop-mediated isothermal amplification (LAMP), are widely used for identifying *V. parahaemolyticus*. These methods can provide rapid and sensitive detection of specific DNA sequences. PCR amplifies selected regions of the *V. parahaemolyticus* genome using primers targeting species-associated genes or virulence-associated genes such as *tdh* and *trh* (Park & Zhang, 2024). LAMP provides an alternative amplification approach that operates at a constant temperature and has been successfully applied to the detection of *V. parahaemolyticus* in seafood samples (Nemoto et al., 2011). PCR generally involves extracting DNA from the sample, combining it with specific primers, nucleotides and DNA polymerase, and amplifying the target sequences before detection.

The main advantages of molecular diagnostics include their sensitivity, specificity and relatively rapid turnaround times, enabling detection of *V. parahaemolyticus* at low concentrations when assays have been appropriately validated (Park & Zhang, 2024; Nemoto et al., 2011). PCR can provide results within hours and multiplex assays can detect several genetic targets in a single reaction. However, these techniques require suitable equipment and trained personnel. Appropriate laboratory procedures are also necessary to minimise contamination, which can produce false-positive results, while reaction inhibition or inadequate DNA recovery can produce false-negative results.

1.3.4 Whole genome sequencing

Whole Genome Sequencing (WGS) is a molecular technique that analyses the complete genetic composition of an organism, providing high-resolution data on genome structure, virulence factors, antimicrobial resistance (AMR) genes, and evolutionary relationships. Compared to culture-based, serological, and standard molecular diagnostics, WGS offers unparalleled detail, detecting single nucleotide variants, structural variants, and mobile genetic elements across the entire genome. In *Vibrio parahaemolyticus*, WGS allows detection of unique genetic markers such as *tdh* and *trh*, enabling differentiation between pathogenic and non-pathogenic strains (Letchumanan et al., 2014). It also supports comparative genomics and strain typing through phylogenetic analysis.

Despite its advantages, WGS has notable limitations for identifying *Vibrio parahaemolyticus*. Firstly, the high genomic diversity of strains complicates diagnostics, as pathogenic traits are not always associated with classical markers such as *tdh* and *trh*, making functional predictions of pathogenicity unreliable (Raghunath, 2015; Santos et al., 2021). Secondly, environmental isolates often carry virulence-associated genomic islands and genes typically linked to pathogenic strains, which can mimic clinical strains and make differentiation between harmful and benign isolates difficult (Santos et al., 2021)

Third, even when full genome sequences are available, WGS may not provide clear, actionable results for rapid diagnostics, as analysis requires pure isolates and complex bioinformatics pipelines, increasing cost and turnaround time (Raghunath, 2015). Collectively, these challenges highlight the need for more targeted and efficient approach to accurately identify pathogenic *V. parahaemolyticus* such as MDA-ONT sequencing.

1.4 Selective MDA–ONT Workflow for *Vibrio parahaemolyticus*

1.4.1 Mechanism and Advantages of Selective MDA

Selective multiple displacement amplification (sMDA) is a targeted DNA amplification technique designed to preferentially amplify DNA from a selected organism. By employing short primers with recognition sites that are abundant in the target genome but less frequent in background DNA, selective amplification can enrich target genomic sequences from mixed samples (Leichty & Brisson, 2014). Unlike conventional MDA using random primers, selective MDA aims to favour amplification of the target genome, potentially improving the recovery of genomic sequence data for downstream analysis.

Selective whole-genome amplification has been successfully applied to other pathogens, including *Plasmodium falciparum* and *P. vivax*, where selective primer sets have enabled genome sequencing from clinical samples containing low concentrations of parasite DNA (Oyola et al., 2016; Cowell et al., 2017). These studies demonstrate the potential of selective amplification for recovering genomic information from challenging samples. In the context of *V. parahaemolyticus*, sMDA may improve the feasibility of genome sequencing from samples containing limited amounts of target DNA. However, its performance for direct detection in seafood and environmental samples requires further validation.

1.4.2 Integration with ONT Sequencing

Oxford Nanopore Technologies (ONT) sequencing complements sMDA by providing long-read sequencing of amplified DNA, with sequence data becoming available during the sequencing run. Long reads can facilitate genome assembly, resolution of repetitive genomic regions and identification of structural variants. These capabilities provide opportunities for detailed genomic characterisation of *V. parahaemolyticus*, including the investigation of virulence-associated genes and antimicrobial resistance determinants.

ONT sequencing platforms such as the MinION are portable and have been used for real-time pathogen genomic surveillance (Quick et al., 2016). Combining ONT sequencing with selective MDA therefore has potential applications in situations where the target organism is present at low abundance. However, successful genomic characterisation depends on amplification efficiency, sequencing performance and the completeness of genome recovery. Further validation would be required before the combined workflow could be used routinely for direct testing of seafood, environmental or clinical samples.

1.4.3 Applications and Implications

The selective MDA–ONT workflow has potential applications in public health, environmental monitoring and food safety. By combining targeted genomic amplification with long-read sequencing, this approach could facilitate genomic characterisation of *V. parahaemolyticus* from samples containing limited quantities of bacterial DNA. Selective amplification has previously been used to recover pathogen genomes from complex samples (Oyola et al., 2016; Cowell et al., 2017), while portable ONT sequencing has demonstrated its value for genomic surveillance (Quick et al., 2016). Together, these technologies could support future strain typing, outbreak investigations and environmental surveillance.

Collectively, this workflow represents a promising approach for targeted genomic enrichment and sequencing of *V. parahaemolyticus*. The present study provides a proof of concept using DNA extracted from cultured isolates. Further validation will be required to determine its performance for direct pathogen detection in seafood, environmental and clinical samples. The resulting genomic data could ultimately be integrated with curated reference databases to support identification and epidemiological investigations.

1.5 *Vibrio parahaemolyticus* Genomic Database

The establishment of a dedicated genomic database for *Vibrio parahaemolyticus* represents a critical tool for improving pathogen identification, surveillance, and outbreak response. By centralising genomic information from diverse strains, the database enables rapid comparison and precise identification of newly isolated strains. This resource reduces reliance on time-consuming conventional diagnostics and facilitates timely interventions in both environmental and clinical contexts. Furthermore, a curated database supports public health authorities and seafood safety regulators by providing a reliable reference for detecting pathogenic strains, monitoring antimicrobial resistance patterns, and tracing contamination sources in supply chains. The success of other bacterial genomic databases, such as BacDive, GlobDB, and PATRIC, demonstrates that centralised, well-structured genomic resources can enhance pathogen detection, support comparative genomics, and inform epidemiological studies (Liu et al., 2024; van der Graaf-van Bloois et al., 2023). Standardisation of data formats and quality metrics will ensure consistent, reliable, and reproducible analyses across studies, while scalability ensures the database can accommodate an increasing number of strains as more genomic data becomes available.

1.5.1 *Vibrio parahaemolyticus* Genomic Database

The establishment of a genomic database for *Vibrio parahaemolyticus* represents a critical step toward improving pathogen identification, surveillance, and outbreak response. Given the high genetic similarity between pathogenic and non-pathogenic strains, rapid and accurate classification requires access to well-characterised reference genomes and associated metadata. Centralised genomic resources enable comparative analyses that can support the identification of virulence-associated markers, track strain diversity, and improve the reliability of molecular diagnostics.

Existing bacterial genomic databases, such as BacDive, GlobDB, and PATRIC, demonstrate the value of structured and accessible genomic repositories in advancing pathogen research and public health monitoring. These platforms facilitate standardised data storage, enable reproducible analyses, and support large-scale comparative genomics. For *V. parahaemolyticus*, a dedicated genomic database would provide similar benefits, particularly in distinguishing clinically relevant strains and supporting food safety surveillance.

In the context of this study, the concept of a genomic database is closely linked to the selective MDA–ONT workflow. Selective amplification enables enrichment of target DNA, while ONT sequencing generates long-read data that can capture genomic variation, including virulence genes and structural features. Together, these approaches have the potential to generate sequence data that could be incorporated into a genomic reference framework for rapid comparison against known strains.

However, the development and implementation of a fully curated genomic database was beyond the scope of this thesis. This study instead focuses on evaluating the upstream components required for such a system, specifically the effectiveness of selective primer design and the feasibility of generating informative sequencing data using the sMDA–ONT approach. Publicly available *V. parahaemolyticus* genomes were used as reference datasets for primer evaluation and comparative analysis, rather than constructing a new database resource.

Despite this limitation, the results presented here highlight the importance of primer selection and genome coverage in determining the quality and utility of sequencing data for downstream applications. These findings provide a foundation for future work aimed at integrating sequencing outputs into a structured genomic database. Such a resource would enable rapid identification of strains, improved detection of virulence-associated genes, and enhanced capacity for epidemiological surveillance.

In summary, while a genomic database was not developed as part of this study, its importance remains central to the long-term application of the sMDA–ONT workflow. Future research should focus on the systematic collection, curation, and integration of genomic data into a dedicated database to support scalable, real-time pathogen detection and monitoring.

1.5.2 Integration with sMDA–ONT Workflow

Integration of a genomic database with the selective MDA–ONT sequencing workflow has the potential to enhance the speed and accuracy of *Vibrio parahaemolyticus* identification. Selective MDA enables enrichment of target DNA from complex samples, while ONT sequencing provides long-read genomic data that can capture structural variation, virulence-associated genes, and broader genomic context. In a fully developed system, these sequencing outputs could be incorporated into a genomic reference framework, enabling rapid comparison of newly sequenced isolates with previously characterised strains.

Such integration would support applications including strain identification, detection of virulence determinants, and monitoring of antimicrobial resistance. By linking sequencing outputs to a structured reference resource, the workflow could facilitate real-time analysis and improve the responsiveness of surveillance and outbreak investigations. This approach aligns with existing genomic surveillance strategies, where sequencing data are continuously incorporated into reference datasets to enhance diagnostic accuracy and epidemiological insight.

However, this integration was not implemented in the present study. Instead, sequencing data generated from the sMDA–ONT workflow were compared against publicly available *V. parahaemolyticus* reference genomes to assess genome coverage and similarity. These analyses provide a preliminary indication of how sequencing outputs from this workflow could be used within a genomic reference framework.

Overall, while the integration of sequencing workflows with a dedicated genomic database represents an important future direction, this study focuses on evaluating the upstream components required to support such a system, particularly primer performance and the feasibility of generating informative sequencing data.

1.6 Thesis Aims

The overarching aim of this thesis was to investigate whether selective multiple displacement amplification (sMDA) combined with Oxford Nanopore Technologies (ONT) sequencing could provide an effective strategy for targeted enrichment and genomic characterisation of *Vibrio parahaemolyticus*. Specifically, this thesis addressed the following objectives:

Aim 1

To identify and evaluate short primer sets capable of selectively enriching *Vibrio parahaemolyticus* DNA by combining experimental testing of existing primer sets with computational prediction of genome-wide primer binding frequency and distribution.

Aim 2

To evaluate the performance of a selective MDA–ONT sequencing workflow using DNA extracted from cultured *Vibrio parahaemolyticus* isolates, including assessment of DNA amplification, sequencing performance, and genome coverage.

Aim 3

To determine how computational predictions of primer binding correspond with experimentally observed genome coverage and to evaluate the suitability of the resulting sequencing data for comparison with reference genomes as a foundation for future genomic surveillance.

Chapter 2 Materials and Methods

2.1 Primers

Three sets of primers were evaluated in this study. The VpCl and cpmt primer sets had previously been developed within the laboratory, whereas the Vib1 primer set was designed specifically for this project using a computational primer-selection strategy. The complete primer sequences used throughout this study are listed in Table 2.1.

The expected abundance of each primer within the *Vibrio parahaemolyticus* genome was estimated by counting exact primer matches across complete chromosome I genome sequences. Initial analyses were undertaken using the reference strain RIMD 2210633 (*V. parahaemolyticus* O3:K6), which was also used as the mapping reference for the Oxford Nanopore sequencing analyses presented later in this thesis. To determine whether primer-binding site abundance depended on the choice of reference genome, the newly designed Vib1 primers were subsequently evaluated against two additional complete chromosome I sequences from environmental isolates (AM43962 and BM24A). These analyses demonstrated highly consistent primer rankings among genomes, indicating that primer-binding site abundance estimates were robust to the choice of reference sequence.

The VpCl and cpmt primer sets had previously been developed within the laboratory, whereas the Vib1 primer set was designed specifically for this study using a computational primer-selection strategy. Primer-binding site abundance was initially estimated by counting exact recognition-site matches within chromosome I of the reference strain *Vibrio parahaemolyticus* RIMD 2210633. The Vib1 primer set was subsequently evaluated against two additional complete chromosome I genomes (AM43962 and BM24A) to assess the robustness of primer-binding site abundance estimates across diverse isolates. Detailed computational analyses of primer-binding site abundance, genomic distribution and occurrence within virulence-associated loci are presented in Chapter 3.

All primers were synthesised by Integrated DNA Technologies (IDT, Coralville, IA, USA). Primers were ordered with two phosphorothioate linkages at the 3′ terminus to reduce exonuclease degradation and improve primer stability during multiple displacement amplification. Lyophilised primers were resuspended in sterile nuclease-free water to a stock concentration of 1 nmol μL^{-1} and diluted to working concentrations before use. The nucleotide

sequences of all primers used in this study are listed in Table 2.1. Their predicted genomic abundance and distribution were determined computationally and are presented in Chapter 3.

Table 2.1. Primer sequences used for sMDA of *Vibrio parahaemolyticus*

Primer set	Primer	Sequence (5′→3′)	Length (bp)
VpCl	VpCl9	TTCATCGCCT	10
	VpCl11	TCATCGCCTT	10
	VpCl17	CATCGCCTTC	10
	VpCl24	ACTTCATCGC	10
cpmt	cpmt3	ATTATT	6
	cpmt4	TTAATT	6
	cpmt7	ATTAAA	6
	cpmt8	TAATAA	6
Vib1	Vib1-1	CATCAA	6
	Vib1-2	TCTTCA	6
	Vib1-3	TTGATG	6
	Vib1-4	TTGCTG	6

2.2 DNA source

DNA isolates of *V. parahaemolyticus* generated during a previous laboratory study (Singh et al, 2012) were available for the initial primer evaluation experiments. In that study, bacterial isolates were cultured from fish collected from markets in Fiji. Species identity was confirmed by comparison of 16S rRNA gene sequences with publicly available reference databases.

2.3 Cultures of *Vibrio parahaemolyticus*

Two cultured isolates of *V. parahaemolyticus* (PFR21C03 and PFR30J09) were provided by Professor Steve Flint (School of Food Technology and Natural Sciences, Massey University).

2.4 DNA Extraction: LysoFAST BOOST

LysoFAST BOOST is a product from the company DENATURE, designed for rapid DNA extraction from bacterial cultures and other samples such as yeast, tissue, insects, blood and plants. Briefly, a good scoop of culture was taken with a loop and mixed with 80ul of LysoFAST

reagent in a 1.7ml microcentrifuge tube. The mixture was vortexed well to disperse any lumps. The tubes were sat on the bench for 5 min then spun at top speed in a benchtop microcentrifuge (13,000 rpm) for 5 min. The supernatant was collected in a clean microcentrifuge tube.

2.5 DNA extraction using the High Pure PCR Template Preparation Kit

Culture was suspended in 200 µL phosphate-buffered saline (PBS) in a microcentrifuge tube. Lysozyme (5 µL of 10 mg/mL in 10 mM Tris-HCl, pH 8) was added, and the mixture was incubated at 37°C for 15 min. DNA was then extracted following the manufacturer's protocol (Roche Life Science) for the isolation of nucleic acids from bacteria and yeast.

2.6 DNA Quantification: Qubit

Following DNA extraction, DNA was quantified using the Qubit 4.0 Fluorometer (ThermoFisher Scientific). Briefly, 10ul of two standards provided with the kit were combined (separately) with 190ul of working solution in two Qubit assay tubes. The standards were read in the Qubit machine and a standard curve was generated. Then 2ul of sample was mixed with 198ul of working solution, and the DNA concentration determined using the standard curve. There are different working solutions and standards for different requirements; we used both the Qubit 1X dsDNA HS Assay Kit (High Sensitivity) which measures between 0.1 and 120ng and the Qubit 1X dsDNA BR Assay Kit (Broad Range) which measures between 0.2 and 4,000ng.

2.7 DNA Quantification: Electrophoresis

Gel electrophoresis was used to check the size and integrity of the DNA. After determining the concentration of DNA, 30-50ng was loaded into a well on a 1% (w:v) agarose gel in Tris-acetate- EDTA (TAE) Buffer (40 mM Tris-acetate, 1 mM EDTA, pH ~8). An aliquot of the 1kb+ Ladder (ThermoFisher Scientific) was loaded into an adjacent well and provided a ladder of markers from 100bp to 20kb. The gel was run at 100V for approximately 1hr.

2.8 Multiple Displacement Amplification

Multiple displacement amplification (MDA) is a technique that uses the DNA polymerase Phi29 and primers (random hexamers or selective primers) to amplify whole genomes from very small amounts of DNA. In this thesis we investigated different primer sets and different DNA extraction methods on the amplification of the genome of *V. parahaemolyticus*. This information forms a basis for being able to test for the presence of *V. parahaemolyticus* in a mixed sample eg. from a fish swab or a water metagenomic sample.

The polymerase used was EquiPhi29 DNA Polymerase from ThermoFisher Scientific. This enzyme has been modified from Phi29 polymerase to work at a higher temperature (42 °C instead of 30–37 °C) and to work faster (3-4hrs rather than overnight).

We followed the manufacturer's protocol: briefly, the primers, DNA, dNTPs, DTT, Buffer and water to 18 ul were mixed and denatured at 95 °C for 5 min. Following a brief sit on ice, the EquiPhi polymerase and IPP (inorganic pyrophosphate) were added to the mix and it was incubated at 37 – 42 °C for 3-4 hours, followed by a heat kill step of 65 °C for 10 min to deactivate the enzyme. We used concentrations recommended in the protocol. The amount of template DNA varied between 2-5ng; 125 pmol of primer was added to the reaction and where there were multiple primers, the amount of each primer added was 125 pmol.

2.9 Column clean up

We passed the amplified DNA through a Genomic DNA Clean and Concentrator-25 column (Zygem) to remove excess primers and enzymatic components. Briefly, a binding buffer is added to the amplified DNA and the solution is passed through a column by centrifugation, washed twice with wash buffer and eluted in molecular grade water.

2.10 ONT Sequencing: Rapid barcoding

The Oxford Nanopore Technologies (ONT) Rapid Barcoding Kit 24 V14 was used to sequence our genomic and amplified DNA samples according to the latest protocol from the web. Briefly, 200ng of sample was simultaneously nicked and tagged with a barcode. Samples were pooled and cleaned with beads. The library was quantified using the Qubit fluorometer.

2.11 ONT Sequencing: loading a flowcell

Prior to sequencing, each flow cell was assessed on the MinION MK1C device to confirm an adequate number of active pores. Flow cells were then primed by flushing with ONT flush buffer containing tether to prepare the nanopores for sequencing. Following the initial flush, the flow cell was incubated for approximately 5 minutes while the sequencing library was prepared.

Library preparation for loading involved combining an appropriate volume of DNA library—adjusted according to fragment size and concentration—with sequencing buffer and loading beads, following the Oxford Nanopore Technologies protocol. A second flush was performed immediately prior to loading to ensure optimal pore availability. The prepared library was then carefully loaded onto the flow cell, and the sequencing run was initiated using the MinION MK1C device.

Flongle flow cells, which contain approximately 126 nanopores compared to ~2000 in standard MinION flow cells, were also used in this study. The loading procedure for Flongle devices followed the same general protocol (priming, library preparation, and loading), but with reduced reagent volumes and modified loading ports specific to the Flongle configuration.

All flow cells and Flongle devices used in this study were compatible with Oxford Nanopore Technologies V14 chemistry.



Figure 2.1. Loading of the MinION flow cell prior to sequencing (ONT supplied image)

Data was saved on the device and transferred to a hard drive after the conclusion of the run. A detailed run report was provided for each run. Analysis of the sequence data was performed on a desktop computer.

2.12 Bioinformatic Analysis

Raw sequencing data generated from the MinION MK1C device were base called using Dorado (Oxford Nanopore Technologies) with the super-accuracy (SUP) model to produce FASTQ files. The version used was the release available at the time of analysis. Sequence reads were aligned to publicly available *Vibrio parahaemolyticus* reference genomes, including strain RIMD 2210633, using minimap2 https://github.com/lh3/minimap2 with parameters optimised for Oxford Nanopore long reads (-x map-ont). Alignment files were processed using samtools https://www.htslib.org/ to generate sorted BAM files. Genome coverage was assessed as the proportion of reference genome positions (chromosome I) covered by at least one sequencing read (coverage breadth). Coverage depth across specific genomic regions was also examined, and the presence of reads mapping to specific virulence-associated genes was used to evaluate gene-level representation. The computational workflow comprised quality assessment, genome mapping, primer-binding site abundance estimation and genome coverage analyses. Statistical summaries were generated using custom Python scripts, and graphical outputs were prepared for subsequent interpretation.

Chapter 3 Results

3.1 Experimental evaluation of the VpCl primer set

The initial stage of this study evaluated the performance of the VpCl primer set, which had previously been developed within the laboratory for selective multiple displacement amplification (MDA) of *Vibrio parahaemolyticus*. The objective of this experiment was to determine whether individual primers, or a combination of all four primers, were capable of generating sufficient amplification to recover substantial regions of the bacterial genome following Oxford Nanopore sequencing.

DNA isolate RS24, obtained from the Fiji collection, was selected as the template for these preliminary experiments. Prior to MDA, the genomic DNA had a concentration of 51 ng μL^{-1} and consisted predominantly of high molecular weight DNA exceeding 20 kb, with minimal evidence of degradation when examined by agarose gel electrophoresis. The DNA was diluted to a working concentration of 5 ng μL^{-1} before amplification.

Four individual MDA reactions were performed using primers VpCl9, VpCl11, VpCl17 and VpCl24, together with a fifth reaction containing all four primers. Amplification reactions were incubated at 37 °C for three hours, a temperature selected to accommodate the relatively low melting temperatures (30–38 °C) of the primers. DNA concentrations were measured immediately following MDA and again after column purification to assess both amplification efficiency and DNA recovery. Table 3.1 summarises DNA concentrations obtained following amplification and purification.

Table 3.1. DNA concentrations following MDA using the VpCl primer set

Reaction	Primer(s)	Template DNA	MDA	Qubit post cleanup	BARCODE
1	VpCL 9	5ng	128 ng μL^{-1}	21 ng μL^{-1}	16
2	VpCL 11	5ng	108 ng μL^{-1}	8 ng μL^{-1}	17
3	VpCL 17	5ng	125 ng μL^{-1}	17 ng μL^{-1}	18
4	VpCL 24	5ng	184 ng μL^{-1}	23.9 ng μL^{-1}	19
5	all 4 primers	5ng	149ng μL^{-1}	24 ng μL^{-1}	20

The VpCl primers successfully generated amplified DNA in all reactions. Post-MDA DNA concentrations ranged from 108 to 184 ng μL^{-1}, representing substantial amplification from the original 5 ng template. Following column purification, DNA concentrations decreased to

between 8 and 24 ng μL^{-1}, reflecting expected losses during clean-up while still providing sufficient DNA for Oxford Nanopore library preparation. Although all four individual primers produced amplified DNA, differences in yield were observed, with VpCl24 producing the greatest post-MDA concentration and VpCl11 the lowest.

Approximately 200 ng of purified DNA from each reaction was barcoded using the Oxford Nanopore Rapid Barcoding Kit (SQK-RBK114.24), pooled, bead purified and sequenced on a MinION Mk1C instrument for 20 hours. Sequencing reads were mapped against chromosome I of *Vibrio parahaemolyticus* reference strain RIMD 2210633 to determine the proportion of the genome recovered by each primer.

Genome coverage obtained for each primer is summarised in Table 3.2.

Table 3.2 – Coverage Results of VpCl Primers

	Total bases	**Covered bases**	**Mean depth**	**Max depth**	**%coverage**
VpCL9	3584100	114050	0.095566096	4.22	3.18210987
VpCL11	3584050	44600	0.047762681	3.295	1.24440228
VpCL17	3550900	6550	0.01192679	1.82	0.18446028
VpCL24	3583800	58150	0.05533006	4.01	1.62257939

Although all primers generated measurable amplification products, sequencing demonstrated that only a small proportion of chromosome I was recovered. Genome coverage ranged from only 0.18% for VpCl17 to 3.18% for VpCl9. Mean sequencing depth remained below 0.10× for all primers, indicating that amplification products were concentrated within relatively small genomic regions rather than being distributed uniformly across the chromosome. Maximum sequencing depths of between 1.82× and 4.22× further suggest that amplification occurred preferentially from a limited number of initiation sites.

The disparity between the substantial quantities of amplified DNA produced during MDA and the very limited genome coverage obtained following sequencing indicates that amplification efficiency alone was not a reliable indicator of genome recovery. Instead, the VpCl primers appeared to amplify relatively small regions of the chromosome repeatedly, producing high DNA yields but poor representation of the complete genome.

These findings demonstrated that, while the VpCl primer set was capable of supporting MDA, it was unsuitable for comprehensive recovery of *V. parahaemolyticus* chromosome I. Consequently, alternative primer designs were investigated to determine whether broader genome representation could be achieved.

3.2 Experimental evaluation of the cpmt primer set

Following the limited genome coverage obtained using the VpCl primer set, a second series of experiments was undertaken to evaluate the performance of the cpmt primers. These primers had previously been developed within the laboratory and were expected to provide improved amplification because of their shorter six-nucleotide sequences, which were anticipated to occur more frequently throughout the *Vibrio parahaemolyticus* genome. The objective of this experiment was to determine whether the cpmt primers could improve genome representation relative to the VpCl primers while maintaining efficient multiple displacement amplification.

As in the previous experiment, genomic DNA from isolate RS24 was diluted to 5 ng μL^{-1} and used as template for MDA. Individual amplification reactions were performed using primers cpmt3, cpmt4, cpmt7 and cpmt8 together with an additional reaction containing all four primers. Amplification reactions were incubated at 37 °C for three hours before DNA concentrations were measured both immediately after MDA and following column purification. DNA yields obtained for each amplification reaction are summarised in Table 3.3.

Table 3.3. DNA concentrations following MDA using the cpmt primer set

***Pr*imer(s)**	**Template DNA**	Qubit **post MDA**	Qubit **post cleanup**	**BARCODE**
cpmt3	5ng	111ng μL^{-1}	28.3 ng μL^{-1}	1
cpmt4	5ng	55.9 ng μL^{-1}	33.8 ng μL^{-1}	2
cpmt7	5ng	64.6 ng μL^{-1}	50.9 ng μL^{-1}	3
cpmt8	5ng	42.5 ng μL^{-1}	48.1 ng μL^{-1}	4
all 4 primers	5ng	54.8 ng μL^{-1}	44.1 ng μL^{-1}	5

All four cpmt primers successfully generated amplified DNA. Post-MDA DNA concentrations ranged from 42.5 to 111 ng μL^{-1}, while purified DNA concentrations ranged between 28.3 and 50.9 ng μL^{-1}. Unlike the VpCl primer set, differences among individual primers were relatively small following purification, with all reactions producing sufficient DNA for Oxford Nanopore

library preparation. The combined-primer reaction also generated a comparable quantity of amplified DNA, indicating that combining multiple primers did not substantially reduce amplification efficiency.

Approximately 200 ng of purified DNA from each reaction was barcoded, pooled and prepared for sequencing using the Oxford Nanopore Rapid Barcoding protocol. The final sequencing library had a concentration of 46.3 ng μL^{-1}, and 463 ng of DNA was loaded onto a MinION flow cell. Sequencing reads were mapped against chromosome I of *Vibrio parahaemolyticus* RIMD 2210633 to determine genome recovery. Genome coverage statistics are presented in Table 3.4.

Table 3.4 – Coverage Results of Cpmt Primers

	Total bases	**Covered bases**	**Mean depth**	**Max depth**	**% coverage**
cpmt3	3311850	3700	0.007425925	1.49	0.11172004
cpmt4	3584100	5200	0.015478489	2.855	0.14508524
cpmt7	3555900	32100	0.040535405	3.01	0.90272505
cpmt8	3583150	58850	0.064537135	4.315	1.64240961
cpmt_all	3555900	118750	0.100589253	6.375	3.33952023

The cpmt primers generally produced broader genome coverage than the original VpCl primers, although substantial differences were observed among individual primers. Genome coverage ranged from only 0.11% for cpmt3 to 1.64% for cpmt8, while the reaction containing all four primers achieved the greatest chromosome coverage (3.34%). Mean sequencing depth also increased in the combined-primer reaction, reaching 0.10× compared with values below 0.07× for the individual primers.

The improved performance of the combined cpmt primer set suggests that increasing the number of independent priming sites can increase the proportion of the genome recovered during MDA. Nevertheless, even the best-performing reaction recovered only approximately 3.3% of chromosome I, indicating that genome representation remained poor. Although this represented a modest improvement over the VpCl primers, the extent of chromosome recovery was still insufficient for reliable whole-genome amplification.

Comparison of the VpCl and cpmt experiments revealed an important limitation of selective MDA. Successful amplification, as measured by DNA yield, did not necessarily result in

extensive genome recovery. Instead, the sequencing data indicate that primer-binding characteristics, including the abundance and distribution of primer-binding sites throughout the chromosome, are likely to have a much greater influence on genome representation than amplification efficiency alone.

These observations motivated the development of a new set of candidate primers specifically designed to maximise the number and distribution of primer-binding sites across the *V. parahaemolyticus* genome. Before experimental testing, these candidate primers were evaluated computationally to determine their expected amplification characteristics.

3.3 Computational evaluation of candidate Vib1 primers

The limited genome coverage obtained using both the VpCl and cpmt primer sets suggested that primer-binding characteristics, rather than amplification efficiency alone, were likely to determine the success of selective multiple displacement amplification. Although both primer sets generated substantial quantities of amplified DNA, sequencing demonstrated that only a small proportion of chromosome I was recovered. These observations indicated that primers producing a large number of well-distributed initiation sites across the genome would be expected to provide more uniform amplification and improved genome representation.

A computational approach was therefore adopted to identify candidate primers with more favourable genomic characteristics. Candidate hexamer primers were evaluated using complete chromosome I genome sequences of *Vibrio parahaemolyticus*. Initial analyses were performed using two oyster-derived isolates (AM43962 and BM24A), which were representative of environmental strains available within this project. To determine whether primer-binding site abundance estimates depended upon the reference genome selected, analyses were subsequently extended to include the complete chromosome I sequence of the reference strain RIMD 2210633.

Four complementary computational analyses were undertaken. First, the abundance of primer-binding sites throughout chromosome I was determined for each candidate primer. Second, the spatial distribution of primer-binding sites was evaluated by measuring the distances between successive priming sites around the chromosome. Third, primer-binding site abundance estimates were compared among the three independently assembled genomes to determine the robustness of primer selection. Finally, the occurrence of primer-binding sites within major virulence-associated loci was examined to assess whether selective MDA would be expected

to recover genes associated with pathogenicity. Together, these analyses provided a quantitative framework for evaluating candidate primers before laboratory testing.

3.3.1 Predicted genome-wide abundance of candidate primer recognition sites

The first stage of the computational analysis quantified the number of exact primer recognition sites present throughout chromosome I. Because multiple displacement amplification is initiated independently from each primer-binding site, primers possessing a greater number of recognition sites were predicted to generate more extensive genome amplification than primers occurring only rarely. Recognition-site abundance was determined by searching both DNA strands of three complete chromosome I genomes for exact matches to each primer sequence. Forward and reverse-complement matches were combined to obtain the total number of predicted amplification initiation sites. Mean values and coefficients of variation were subsequently calculated across the three independently assembled genomes.

Table 3.5. Predicted genome-wide abundance of candidate primer recognition sites across three complete chromosome I genomes of *Vibrio parahaemolyticus*

Primer set	Primer	Sequence (5′–3′)	Length (bp)	Mean binding sites	Density (sites Mb^{-1})	CV (%)
VpCl	VpCl9	TTCATCGCCT	10	39.3	11.89	3.88
	VpCl11	TCATCGCCTT	10	41.7	12.60	6.04
	VpCl17	CATCGCCTTC	10	40.0	12.09	5.00
	VpCl24	ACTTCATCGC	10	40.0	12.09	2.50
Cpmt	cpmt3	ATTATT	6	1984.0	599.88	1.21
	cpmt4	TTAATT	6	2044.3	618.09	2.27
	cpmt7	ATTAAA	6	2509.0	758.58	2.02
	cpmt8	TAATAA	6	1681.0	508.23	2.37
Vib1	Vib1-1	CATCAA	6	5134.3	1552.50	0.42
	Vib1-2	TCTTCA	6	4047.7	1223.94	0.38
	Vib1-3	TTGATG	6	5134.3	1552.50	0.42
	Vib1-4	TTGCTG	6	4289.3	1296.99	0.51

Substantial differences in predicted primer-binding site abundance were observed among the three primer sets (Table 3.5). The original VpCl primers, which were 10 nucleotides in length, possessed only approximately 40 recognition sites throughout chromosome I. In contrast, the six-nucleotide cpmt primers exhibited approximately 1,700–2,500 recognition sites, reflecting the greater frequency expected for shorter primer sequences. The newly designed Vib1 primers displayed the greatest abundance, with CATCAA (Vib1-1) and TTGATG (Vib1-3) each possessing more than 5,100 recognition sites per chromosome, while TCTTCA (Vib1-2) and TTGCTG (Vib1-4) each possessed more than 4,000 recognition sites.

Recognition-site densities showed the same pattern, increasing from approximately 12 sites Mb^{-1} for the VpCl primers to between 508 and 759 sites Mb^{-1} for the cpmt primers and more than 1,200 sites Mb^{-1} for all Vib1 primers. These results predict that the Vib1 primer set would initiate substantially more independent DNA synthesis events during multiple displacement amplification than either of the previously available primer sets.

Primer-binding site abundance estimates were highly reproducible among the three independently assembled genomes, with coefficients of variation below 6.1% for all primers and below 0.6% for every Vib1 primer. The low level of variation indicates that primer-binding site abundance is largely determined by the conserved nucleotide composition of the *V. parahaemolyticus* genome rather than isolate-specific sequence variation. Consequently, primer-binding site abundance represents a robust characteristic for computational primer evaluation and supports the selection of the Vib1 primer set for subsequent experimental validation.

3.3.2 Predicted spatial distribution of primer recognition sites

The abundance of primer recognition sites alone does not necessarily ensure uniform genome amplification because recognition sites may be clustered within particular genomic regions, leaving large primer-free intervals elsewhere on the chromosome. To evaluate the predicted uniformity of amplification, the spatial distribution of primer recognition sites was examined by calculating the distances between adjacent recognition sites throughout chromosome I. Chromosomes were treated as circular molecules so that the final and initial recognition sites were considered contiguous. For each primer, the mean and median inter-site distances, maximum primer-free interval, 95th percentile gap and Uniformity Index were calculated and averaged across the three complete chromosome I genomes.

Table 3.6. Predicted spatial distribution of primer recognition sites across chromosome I of three complete *Vibrio parahaemolyticus* genomes

Primer set	Primer	Mean inter-site distance (bp)	Median inter-site distance (bp)	Maximum gap (bp)	95th percentile gap (bp)	Uniformity index
VpCl	VpCl9	84,153	35,117	511,100	338,558	6.07
	VpCl11	79,554	46,924	333,120	248,118	4.20
	VpCl17	82,804	47,781	440,333	299,262	5.34
	VpCl24	82,708	59,145	366,879	248,317	4.44
cpmt	cpmt3	1,667	988	15,049	5,624	9.03
	cpmt4	1,618	1,008	13,136	5,512	8.12
	cpmt7	1,318	821	10,234	4,205	7.76
	cpmt8	1,968	1,175	23,139	6,427	11.76
Vib1	Vib1-1	644	450	6,627	1,966	10.29
	Vib1-2	817	521	7,808	2,590	9.55
	Vib1-3	644	450	6,627	1,966	10.29
	Vib1-4	771	520	13,656	2,274	17.70

The spatial distribution of primer recognition sites closely mirrored the abundance patterns described in Section 3.3.1. The Vib1 primers exhibited substantially shorter inter-site distances than either the VpCl or cpmt primer sets (Table 3.6), with mean inter-site distances ranging from 644 to 817 bp, compared with 1,318 to 1,968 bp for the cpmt primers and approximately 80 kb for the VpCl primers. Consequently, the Vib1 primers were predicted to initiate DNA synthesis at much more frequent intervals throughout chromosome I, providing a substantially denser network of potential amplification initiation sites. The distribution of larger primer-free intervals further distinguished the three primer sets. Maximum primer-free regions ranged from 6.6 to 13.7 kb for the Vib1 primers, compared with 10.2 to 23.1 kb for the cpmt primers and 333 to 511 kb for the VpCl primers. Although maximum gaps provide an indication of the most extreme primer-free regions, they are influenced by single outlying intervals. The 95th percentile gap therefore provides a more representative measure of genome-wide spacing by describing the upper limit of inter-site distances across the majority of the chromosome. For

the Vib1 primers, the 95th percentile gap ranged from 1.97 to 2.59 kb, approximately one-half of that observed for the cpmt primers and more than two orders of magnitude smaller than that observed for the VpCl primers. These results indicate that primer recognition sites were distributed much more uniformly throughout the chromosome, substantially reducing the likelihood of extensive genomic regions lacking nearby amplification initiation sites. The Uniformity Index provided a complementary summary of spacing regularity and generally supported the conclusions drawn from the inter-site distance analyses. Although individual primers differed modestly in their Uniformity Index values, all Vib1 primers combined high recognition-site abundance with consistently short inter-site distances and relatively small primer-free regions. Collectively, these characteristics predict that the Vib1 primer set should produce more uniform whole-genome amplification than either the VpCl or cpmt primer sets by increasing the number of independent amplification initiation events while minimising the occurrence of large under-amplified genomic regions.

The analyses presented in Sections 3.3.1 and 3.3.2 demonstrate that the Vib1 primers combine both a high abundance of recognition sites and a favourable genomic distribution. The final stage of the computational evaluation therefore examined whether these favourable characteristics remained consistent across independently assembled reference genomes before proceeding to experimental validation.

3.3.3 Robustness of primer-binding site abundance estimates across reference genomes

A potential limitation of computational primer design is that predicted primer-binding site abundance may depend upon the particular reference genome selected for analysis. To determine whether primer-binding site abundance estimates were sensitive to genome choice, the predicted numbers of primer recognition sites were compared among the three complete chromosome I genomes using both Pearson product-moment correlation coefficients and Spearman rank correlation coefficients. Pearson correlations assessed agreement in the absolute numbers of recognition sites, whereas Spearman correlations evaluated the consistency of primer rankings independently of absolute abundance.

Table 3.7. Correlation of predicted primer recognition-site abundance among three complete chromosome I genomes of *Vibrio parahaemolyticus*

Genome comparison	Pearson correlation coefficient	Spearman rank correlation coefficient	*P*-value
AM43962 vs BM24A	0.99984	0.97707	7.59×10^{-19}
AM43962 vs RIMD2210633	0.99993	0.99118	1.17×10^{-20}
BM24A vs RIMD2210633	0.99978	0.99472	4.12×10^{-18}

Predicted primer-binding site abundance was remarkably consistent among the three independently assembled chromosome I genomes (Table 3.7). Pearson correlation coefficients exceeded 0.999 for all pairwise genome comparisons, demonstrating that the absolute numbers of predicted primer recognition sites were almost identical irrespective of the reference genome analysed. Similarly, Spearman rank correlation coefficients ranged from 0.977 to 0.995, indicating that the relative ranking of candidate primers was also highly conserved across genomes.

The strongest agreement was observed between the AM43962 and RIMD2210633 genomes (Pearson $r = 0.99993$; Spearman $\rho = 0.99118$), although all three pairwise comparisons produced extremely high correlation coefficients and highly significant probabilities ($P < 10^{-18}$). These findings demonstrate that both the absolute abundance and relative ranking of candidate primers are largely independent of the reference genome selected for analysis.

Collectively, the results presented in Sections 3.3.1–3.3.3 demonstrate that the computational primer-selection approach consistently identified the Vib1 primers as possessing the greatest abundance of recognition sites together with the most favourable genome-wide distribution across three independently assembled *V. parahaemolyticus* genomes. The high level of agreement among reference genomes indicates that primer selection was robust to genome choice and therefore likely to be broadly applicable across diverse *V. parahaemolyticus* isolates. These computational predictions provided the rationale for the subsequent experimental evaluation of the Vib1 primer set using selective multiple displacement amplification.

3.3.4 Predicted distribution of *Vib*1 primer recognition sites within virulence-associated loci

Following evaluation of genome-wide primer-binding site abundance, spatial distribution and robustness across reference genomes, the final stage of the computational analysis examined the occurrence of Vib1 primer recognition sites within major virulence-associated loci. Although whole-genome amplification is desirable, the successful representation of genes associated with pathogenicity is particularly important for downstream genomic characterisation and comparative analyses. Consequently, primer recognition sites were quantified within representative genes involved in host adhesion, secretion systems, motility, iron acquisition and haemolysin production.

Table 3.8. Predicted recognition-site abundance of the Vib1 primer set within selected virulence-associated loci of *Vibrio parahaemolyticus*

Gene	**Function**	**Vib1-1**	**Vib1-2**	**Vib1-3**	**Vib1-4**
MAM7	Adhesion	7	3	7	2
T3SS1	Type III secretion system 1	75	65	75	31
T3SS2	Type III secretion system 2	0	1	0	0
T6SS1	Type VI secretion system 1	88	69	88	60
tdh	Thermostable direct haemolysin	0	0	0	0
tlh	Thermolabile haemolysin	3	2	3	0
trh	TDH-related haemolysin	2	5	2	1
flgE	Flagellar hook protein	8	8	8	10
fliM	Flagellar motor switch protein	76	51	76	68
lafA	Lateral flagellin	18	18	18	10
motY	Flagellar motor protein	2	2	2	5
pvuA	Ferric vibrioferrin receptor	3	1	3	5

The predicted distribution of Vib1 primer recognition sites varied among virulence-associated loci (Table 3.8), reflecting differences in nucleotide composition among individual genes. Recognition-site abundance was highest within the Type III secretion system 1 (T3SS1), Type VI secretion system 1 (T6SS1) and the flagellar motor gene fliM, where between 51 and 88 recognition sites were identified depending on the primer. These loci therefore contain numerous potential amplification initiation sites and are predicted to be well represented following selective multiple displacement amplification.

In contrast, relatively few recognition sites were identified within MAM7, motY, pvuA, tlh and *trh*, while no recognition sites were detected for three of the four Vib1 primers within the thermostable direct haemolysin gene (*tdh*) or within T3SS2. These observations indicate that primer representation is influenced by local sequence composition and that some virulence-associated loci may be intrinsically less accessible to selective amplification than others.

The complementary primers CATCAA (Vib1-1) and TTGATG (Vib1-3) exhibited identical recognition-site abundances for every locus examined, consistent with their complementary sequences and identical genome-wide occurrence demonstrated in Sections 3.3.1–3.3.3. The remaining Vib1 primers showed similar but not identical distributions, providing additional recognition sites within several loci while maintaining extensive coverage across the majority of genes examined.

Collectively, the computational analyses presented in Sections 3.3.1–3.3.4 demonstrate that the Vib1 primer set combines a high abundance of genome-wide recognition sites with favourable spatial distribution, robust performance across independently assembled reference genomes and extensive representation within many major virulence-associated loci. These characteristics predict that the Vib1 primers should provide substantially more effective selective multiple displacement amplification than the previously available VpCl and cpmt primer sets. The following sections evaluate these predictions experimentally using genomic DNA extracted from cultured *V. parahaemolyticus* isolates.

3.4 Experimental validation of computational predictions

The computational analyses presented in Section 3.3 identified the Vib1 primer set as possessing substantially more favourable characteristics for selective multiple displacement amplification than the previously developed VpCl and cpmt primer sets. In particular, the Vib1 primers exhibited a high abundance of predicted recognition sites, short inter-site distances, reduced primer-free genomic regions and consistent performance across three independently assembled *Vibrio parahaemolyticus* genomes. The following experiments were undertaken to determine whether these computational predictions translated into improved whole-genome amplification under laboratory conditions.

3.4.1 DNA extraction and template quality

Experimental validation was performed using two cultured *Vibrio parahaemolyticus* isolates, PFR30J09 and PFR21C03, obtained from the collection described by Wang et al. (2023). These isolates had previously been characterised as strong and weak biofilm-forming strains, respectively, providing representative environmental isolates for evaluating selective multiple displacement amplification.

Genomic DNA was extracted independently using the Roche High Pure PCR Template Preparation Kit and the LysoFAST BOOST extraction protocol. Agarose gel electrophoresis indicated that both methods produced predominantly high molecular weight DNA, although varying degrees of fragmentation were evident in all preparations. DNA extracted using the Roche protocol consistently exhibited less fragmentation than DNA extracted using the LysoFAST BOOST protocol, suggesting superior preservation of template integrity. DNA concentrations obtained using the two extraction methods are summarised in Table 3.9.

Table 3.9. DNA concentrations obtained from two *Vibrio parahaemolyticus* isolates following extraction using the Roche High Pure PCR Template Preparation Kit and the LysoFAST BOOST protocol

DNA extraction method	PFR30J09 (ng μL^{-1})	PFR21C03 (ng μL^{-1})
LysoFAST BOOST	37.9	21.8
Roche High Pure	204.0	76.0

The Roche extraction protocol consistently produced substantially greater DNA yields than the LysoFAST BOOST method (Table 3.9). DNA concentrations obtained using the Roche protocol were approximately 5.4-fold greater for isolate PFR30J09 and 3.5-fold greater for isolate PFR21C03. Together with the reduced DNA fragmentation observed following agarose gel electrophoresis, these results indicate that the Roche protocol produced DNA of both higher quantity and higher integrity than the LysoFAST BOOST method.

Because the objective of the subsequent experiments was to evaluate amplification performance rather than biological differences between isolates, all remaining analyses were performed using isolate PFR21C03. This isolate yielded sufficient DNA using both extraction methods to permit direct comparison of selective multiple displacement amplification from equivalent template concentrations.

3.4.2 Whole-genome amplification using the *Vib*1 primer set

To ensure that amplification efficiency reflected primer performance rather than differences in template concentration, DNA preparations from both extraction methods were diluted to a common concentration of 1 ng μL^{-1} before amplification using the complete Vib1 primer set. DNA concentrations measured immediately following selective multiple displacement amplification and after purification are summarised in Table 3.10.

Table 3.10. DNA concentrations following selective multiple displacement amplification using the *Vib*1 primer set

DNA extraction method	Template DNA (ng μL^{-1})	Post-MDA DNA (ng μL^{-1})	Post-cleanup DNA (ng μL^{-1})	Nanopore barcode
LysoFAST BOOST	1.0	162.0	29.7	BC22
Roche High Pure	1.0	216.0	61.3	BC21

Selective multiple displacement amplification successfully increased DNA concentration for both extraction methods (Table 3.10). Starting from identical template concentrations of 1 ng μL^{-1}, amplification produced 162 ng μL^{-1} and 216 ng μL^{-1} DNA from the LysoFAST BOOST and Roche DNA preparations, respectively. Following purification, sufficient DNA remained for Oxford Nanopore library preparation in both cases.

Amplification of Roche-derived DNA yielded approximately twice as much purified DNA as amplification of LysoFAST-derived DNA, indicating that template integrity continued to influence overall amplification efficiency despite identical primer composition and reaction conditions. Nevertheless, successful amplification was achieved from DNA prepared using both extraction methods.

These findings support the computational predictions presented in Section 3.3 by demonstrating that the Vib1 primer set efficiently amplified *V. parahaemolyticus* genomic DNA under laboratory conditions. At the same time, the results indicate that primer performance and template quality contribute independently to successful multiple displacement amplification. Consequently, both amplified and non-amplified DNA preparations were retained for Oxford Nanopore sequencing to determine whether the computationally predicted improvements in primer performance resulted in improved genome recovery.

3.5 Oxford Nanopore sequencing quality

The objective of the sequencing analysis was to determine whether DNA amplified using the Vib1 primer set generated Oxford Nanopore sequencing data of sufficient quality for reliable genome mapping and comparison of genome coverage. Because the computational analyses presented in Section 3.3 predicted improved genome representation rather than increased sequencing read length, sequencing quality was evaluated to establish whether the resulting reads were suitable for downstream mapping analyses.

Table 3.11. Summary of Oxford Nanopore sequencing quality following trimming and quality filtering

Metric	Value
Total reads analysed	158,522
Total bases sequenced	42,949,616
Bases with Q ≥ 8	76.63%
Bases with Q ≥ 10	66.14%
Mean Phred quality score	17.28

Sequencing generated 158,522 reads comprising approximately 42.9 Mb of sequence data following trimming and quality filtering (Table 3.11). More than three-quarters (76.6%) of all sequenced bases exceeded the minimum quality threshold of Q8, while 66.1% exceeded Q10. The mean Phred quality score of 17.3 corresponds to an average base-calling accuracy greater than 98%, indicating that the sequencing reads were of sufficient quality for reliable genome alignment and subsequent genome coverage analyses. Although read quality was consistently high, sequencing quality alone does not indicate whether long genomic fragments were retained throughout the amplification and library preparation workflow. Consequently, the distribution of sequencing read lengths was examined separately.

Table 3.12. Read length statistics following trimming and quality filtering

Metric	Value
Mean read length (bp)	2,687
Median read length (bp)	1,199
Read length N50 (bp)	6,119
Minimum read length (bp)	1
Maximum read length (bp)	642,146
GC content (%)	42.68

The filtered sequencing dataset exhibited a broad distribution of read lengths (Table 3.12). Although the longest individual read exceeded 640 kb, indicating that very long DNA molecules were present within the sequencing library, the majority of sequencing reads were considerably shorter, with a median read length of only 1,199 bp and a mean read length of 2,687 bp. These values indicate that the sequencing output was dominated by relatively short DNA fragments despite the presence of occasional ultra-long reads.

The cultured *Vibrio parahaemolyticus* isolates used in this study were selected because they produced high molecular weight genomic DNA suitable for long-read sequencing, and the Roche extraction protocol further yielded DNA of higher quality and quantity than the alternative extraction method (Section 3.4). Consequently, the predominance of relatively short sequencing reads is unlikely to be explained solely by degraded starting DNA. Instead, the results suggest that the amplification and library preparation workflow did not fully preserve long DNA fragments, even though sufficient DNA was generated for successful Oxford Nanopore sequencing.

Importantly, the primary objective of the Vib1 primer set was not to increase sequencing read length but to improve the number and genomic distribution of amplification initiation sites, thereby increasing the uniformity of genome representation following multiple displacement amplification. The sequencing data obtained here were therefore considered suitable for evaluating this hypothesis because the reads possessed sufficient quality for reliable genome mapping, even though the amplified DNA remained relatively fragmented. The following section examines whether the computationally predicted improvements in primer-binding site

abundance and genomic distribution translated into improved genome coverage following read mapping.

3.6 Validation of computational predictions by genome mapping

The computational analyses presented in Section 3.3 predicted that the Vib1 primer set would generate a substantially greater number of amplification initiation sites that were distributed more uniformly throughout the *Vibrio parahaemolyticus* genome than either the VpCl or cpmt primer sets. Following selective multiple displacement amplification and Oxford Nanopore sequencing, genome mapping was undertaken to determine whether these predictions translated into extensive genome recovery.

3.6.1 Genome-wide coverage

Genome-wide sequencing coverage was evaluated using isolate PFR30J09. Sequencing reads mapped across almost the entire reference genome (Figure 3.1), demonstrating that selective multiple displacement amplification generated broad genome representation. Although sequencing depth varied considerably along the chromosome, only a very small proportion of the genome lacked mapped reads.

Table 3.13. Genome-wide coverage statistics for *Vibrio parahaemolyticus* isolate PFR30J09

Metric	**Value**
Genome length (bp)	5.336×10^6
Bases ≥1× coverage	99.956%
Bases ≥10× coverage	98.611%
Bases ≥30× coverage	73.104%
Bases ≥50× coverage	39.780%
Mean sequencing depth	68.3×
Median sequencing depth	43×
Maximum sequencing depth	50,493×

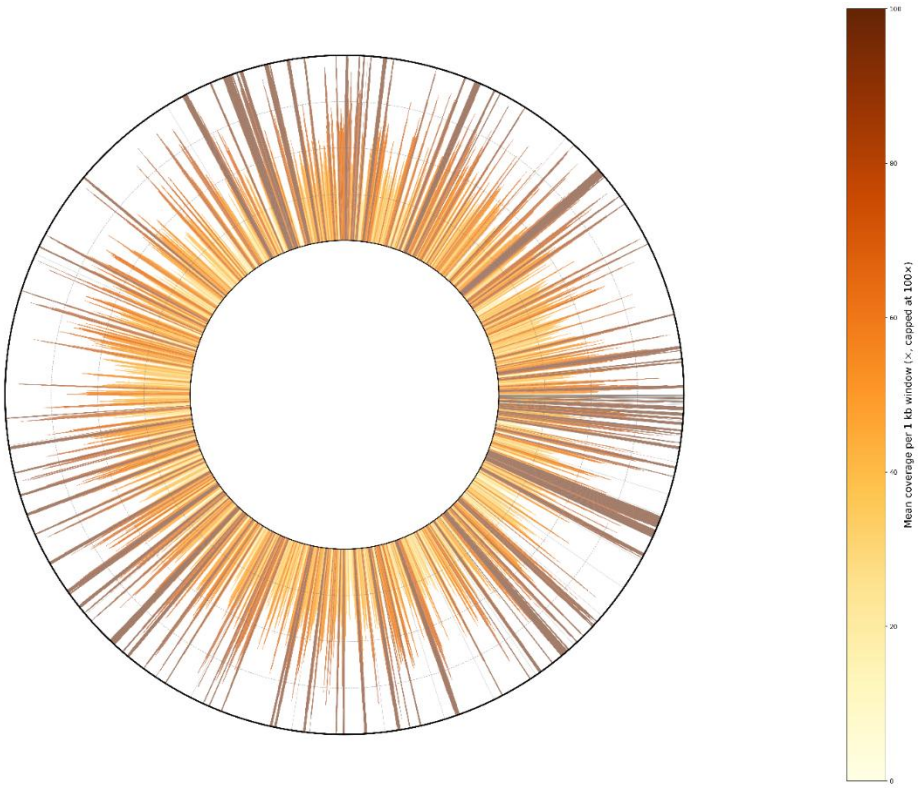


Figure 3.1. Genome-wide coverage of *Vibrio parahaemolyticus* isolate PFR30J09

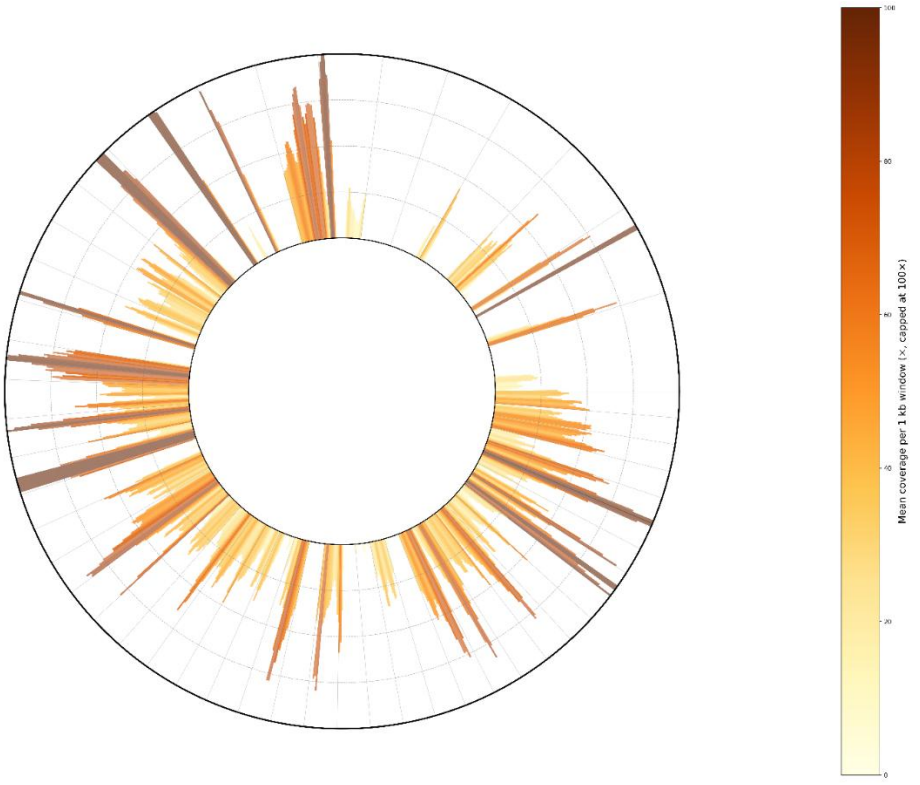


Figure 3.2. Genome-wide coverage of *Vibrio parahaemolyticus* isolate PFR21C03

The sequencing data provided almost complete genome recovery, with 99.96% of the reference genome represented by at least one sequencing read (Table 3.13). More than 98% of genomic positions achieved a minimum depth of 10×, while approximately 73% exceeded 30× coverage. These results indicate that the Vib1 primer set successfully generated genome-wide amplification suitable for downstream sequence analysis.

Although genome recovery was almost complete, sequencing depth varied markedly across the chromosome, with a small number of genomic regions exhibiting exceptionally high coverage (>19,000×), indicative of substantial amplification bias. Coverage of isolate PFR21C03 also exhibited a patchy distribution of sequencing depth across the chromosome (Figure 3.2), with substantial variation in local coverage. The presence of heterogeneous coverage in both independently prepared isolates suggests that uneven genome representation is an inherent characteristic of the selective multiple displacement amplification process rather than a consequence of isolate-specific genomic differences.

3.6.2 Recovery of virulence-associated loci

Although genome-wide coverage provides an overall measure of amplification success, the principal objective of selective multiple displacement amplification is to recover biologically important genomic regions with sufficient completeness to permit downstream genomic analyses. Coverage was therefore examined across representative virulence-associated loci involved in adhesion, secretion systems, motility, iron acquisition and haemolysin production. Gene recovery was assessed by determining the number of matching reference loci, mean sequencing depth and the percentage completeness of each locus following genome mapping.

Table 3.14. Recovery of representative virulence-associated loci following selective multiple displacement amplification and Oxford Nanopore sequencing

Gene / locus	**Number of BLAST hits**	**Mean coverage (×)**	**Completeness (%)**
MAM7	1	34.02	100.00
T3SS1	2	65.44	53.24
T3SS2	0	0.00	0.00
T6SS1	57	59.53	60.61
tdh	0	0.00	0.00
trh	0	0.00	0.00
tlh	1	20.69	100.00
lafA	3	26.79	82.88
motY	1	29.12	99.94
pvuA	1	38.19	100.00
fliM	9	41.82	93.87
flgE	1	28.71	100.00

Recovery varied among individual virulence-associated loci (Table 3.14). Four genes (MAM7, tlh, pvuA and flgE) were recovered completely, each achieving 100% sequence completeness. Similarly, the motility-associated gene motY was recovered almost completely (99.9%), while fliM also exhibited high overall recovery (93.9%). These results indicate that selective multiple displacement amplification successfully recovered a broad range of functionally important loci distributed throughout the chromosome.

In contrast, recovery of several larger or more complex virulence loci was less complete. The T3SS1 pathogenicity island achieved only 53.2% completeness, while T6SS1 reached 60.6% completeness despite exhibiting relatively high mean sequencing depths. The lateral flagellin gene lafA showed intermediate recovery (82.9%), suggesting that local variation in amplification efficiency remained evident despite the extensive genome-wide coverage observed in Section 3.6.1.

No sequences corresponding to T3SS2, *tdh* or *trh* were detected in the mapped dataset. These virulence determinants are known to vary among *Vibrio parahaemolyticus* isolates, and their

absence from the mapped reads may therefore reflect genuine strain-specific differences in gene content, incomplete recovery during selective multiple displacement amplification, or a combination of both factors. The available data do not permit these alternative explanations to be distinguished.

Coverage profiles for each recovered virulence-associated locus are presented in Figures 3.3–3.11. These plots illustrate the distribution of sequencing depth across individual genes and demonstrate that variation in locus completeness generally reflected local fluctuations in amplification rather than complete loss of individual genomic regions. Genes recovered with 100% completeness typically exhibited continuous coverage across their entire coding sequences, whereas partially recovered loci contained extended regions of reduced sequencing depth or isolated gaps in coverage.

To further characterise amplification across individual loci, the proportion of bases falling within different sequencing-depth categories was calculated (Table 3.15).

Table 3.15. Distribution of sequencing depth across representative virulence-associated loci

Gene	**0–1×**	**1–5×**	**5–10×**	**10–25×**	**25–50×**	**50–100×**	**>100×**
flgE	0.00	0.01	0.15	36.31	63.53	0.00	0.00
fliM	0.00	0.00	0.24	10.61	63.98	23.54	1.62
lafA	0.07	0.50	0.69	41.60	45.64	11.48	0.00
MAM7	0.00	0.00	0.00	5.21	94.79	0.00	0.00
motY	0.00	0.03	0.03	9.47	90.47	0.00	0.00
pvuA	0.00	0.00	0.00	6.45	82.89	10.66	0.00
T3SS1	0.00	0.00	0.01	16.08	39.55	42.48	1.87
T6SS1	0.02	0.04	0.04	10.01	48.37	31.30	10.21
tlh	0.00	0.00	0.64	99.36	0.00	0.00	0.00
tdh	100	0	0	0	0	0	0
trh	100	0	0	0	0	0	0
T3SS2	100	0	0	0	0	0	0

Analysis of sequencing-depth distributions demonstrated that the majority of recovered loci were covered by at least 25× sequencing depth across most of their lengths, although substantial local variation remained evident. Genes such as MAM7, motY, pvuA and flgE displayed relatively uniform coverage profiles, whereas T3SS1 and T6SS1 exhibited considerably greater heterogeneity, with both low- and high-depth regions occurring within the same locus. These findings are consistent with the genome-wide coverage analyses presented in Section 3.6.1 and indicate that, although the Vib1 primer set achieved extensive genome recovery, amplification bias was not eliminated.

Collectively, these results demonstrate that selective multiple displacement amplification using the Vib1 primer set successfully recovered the majority of biologically important virulence-associated loci while maintaining near-complete genome representation. However, variation in sequencing depth among individual genes indicates that local amplification efficiency remained heterogeneous, providing the basis for the final comparison between the computational predictions and the experimentally observed genome recovery presented in the following section.

Representative coverage profiles illustrating complete, intermediate and incomplete locus recovery are presented in Figures 3.3 and 3.4. Coverage profiles for the remaining virulence-associated loci are provided in Appendix A (Figures A3.1–A3.6). Although most regions were recovered, substantial variation in sequencing depth resulted in incomplete overall recovery (60.6% completeness).

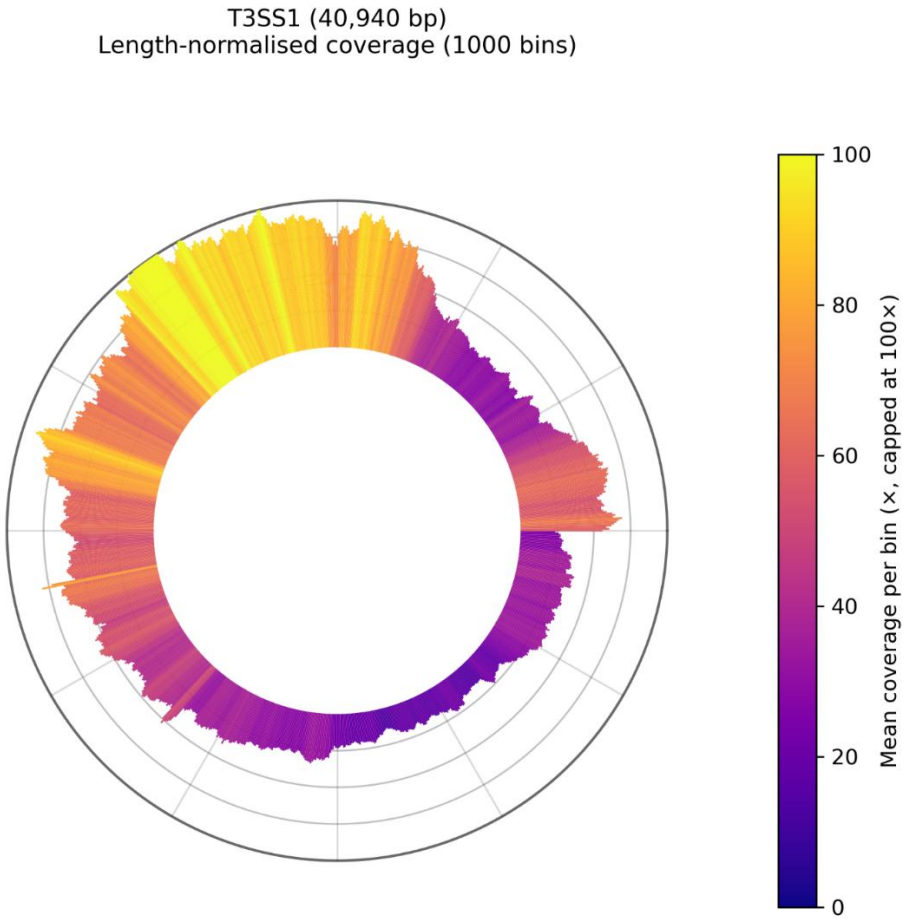


Figure 3.3. Sequencing coverage across the T3SS1 pathogenicity island following selective multiple displacement amplification using the *Vib*1 primer set**.** Coverage depth is capped at 100× to improve visualisation. Extended regions of reduced sequencing depth contributed to partial recovery of the locus (53.2% completeness).

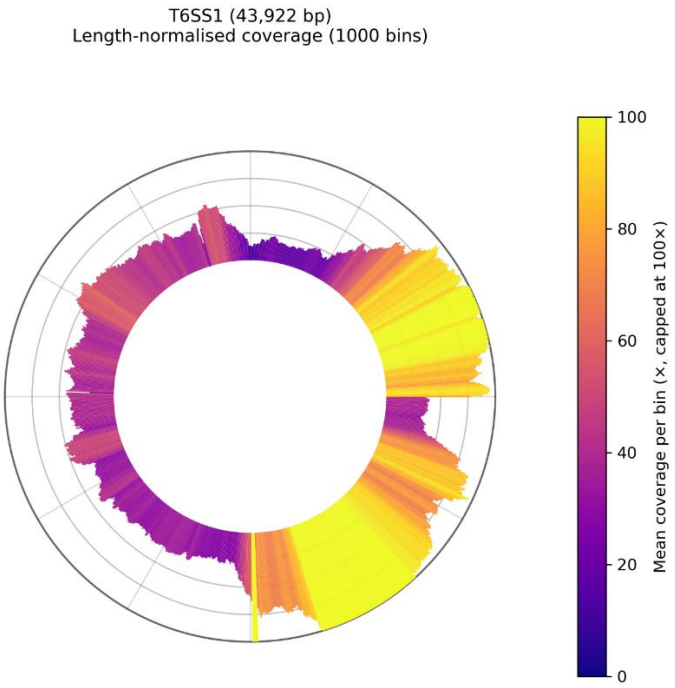


Figure 3.4. Sequencing coverage across the T6SS1 locus following selective multiple displacement amplification using the Vib1 primer set. Coverage depth is capped at 100×

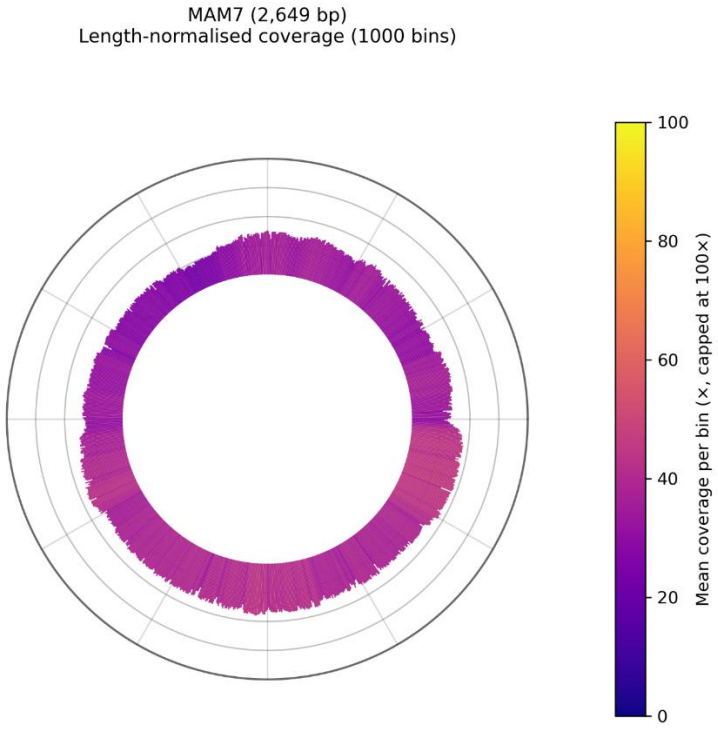


Figure 3.5. Sequencing coverage across the multivalent adhesion molecule 7 (*mam7*) gene following selective multiple displacement amplification using the Vib1 primer set. Coverage depth is capped at 100×. Continuous sequencing coverage across the entire locus resulted in complete gene recovery

3.6.3 Comparison of computational predictions with experimental genome recovery

The computational analyses presented in Section 3.3 predicted that the Vib1 primer set would provide substantially improved whole-genome amplification compared with the previously available VpCl and cpmt primer sets. These predictions were based on four independent characteristics of the candidate primers, including genome-wide abundance of recognition sites, spatial distribution across the chromosome, consistency among independently assembled reference genomes and representation within major virulence-associated loci. The Oxford Nanopore sequencing results provide an opportunity to evaluate the extent to which these computational predictions were supported experimentally.

The most strongly supported prediction concerned overall genome representation. Computational analyses demonstrated that the Vib1 primers possessed approximately 100-fold more recognition sites than the original VpCl primers and approximately two-fold more recognition sites than the cpmt primers, while maintaining substantially shorter distances between adjacent primer-binding sites. Experimentally, these characteristics translated into almost complete genome recovery, with 99.96% of the reference genome represented by at least one sequencing read and more than 98% of genomic positions achieving at least 10× sequencing coverage. These findings indicate that the increased abundance and improved genomic distribution of Vib1 recognition sites successfully promoted amplification across almost the entire chromosome.

The computational analyses also predicted that improved spacing between primer-binding sites would reduce the occurrence of large unamplified genomic regions. This prediction was likewise supported experimentally. Genome mapping demonstrated extensive chromosome-wide coverage with very few regions lacking sequencing reads, indicating that the dense distribution of Vib1 recognition sites was sufficient to initiate amplification throughout most of the genome. Consequently, the principal computational objective of achieving near-complete genome representation was successfully achieved.

However, the computational analyses did not predict that amplification would occur uniformly across all genomic regions. Genome coverage remained highly heterogeneous, with sequencing depth varying substantially throughout the chromosome and a small number of loci exhibiting exceptionally high read depths. Similar variation was observed among virulence-associated loci, where several genes were recovered completely while others exhibited only partial

recovery despite relatively high average sequencing depth. These observations indicate that increasing primer-binding site abundance and improving primer distribution substantially enhanced genome representation but did not eliminate preferential amplification of particular genomic regions.

Recovery of individual virulence-associated loci broadly reflected the computational analyses of primer recognition-site distribution. Genes predicted to contain numerous Vib1 recognition sites, including MAM7, fliM, flgE, motY and pvuA, were recovered either completely or almost completely following sequencing. Conversely, recovery of the larger secretion system loci (T3SS1 and T6SS1) remained incomplete, while T3SS2, *tdh* and *trh* were not detected in the mapped dataset. Because these virulence determinants vary naturally among *Vibrio parahaemolyticus* isolates, their absence cannot be attributed solely to amplification efficiency, and the present data do not distinguish between genuine biological absence and incomplete recovery.

The sequencing results also demonstrated that the principal limitation of the experimental workflow differed from the computational limitations addressed during primer design. The Vib1 primers were specifically designed to increase the number and genomic distribution of amplification initiation sites rather than to maximise sequencing read length. Although the Oxford Nanopore libraries were dominated by relatively short sequencing reads, these reads nevertheless provided sufficient quality and genome-wide representation to evaluate amplification performance. The principal objective of the computational framework was therefore achieved, even though additional optimisation of the selective multiple displacement amplification protocol will likely be required to improve the recovery of long DNA fragments and reduce local amplification bias.

Overall, the experimental results provide strong support for the computational primer-design strategy developed in this study. The Vib1 primer set consistently generated near-complete genome representation while recovering the majority of biologically important virulence-associated loci. Although amplification remained quantitatively heterogeneous, the computational predictions of increased primer-binding site abundance, improved genome-wide distribution and enhanced genome recovery were largely confirmed experimentally. Collectively, these findings provide strong evidence that computational optimisation of primer recognition-site abundance represents an effective strategy for improving selective multiple displacement amplification.

This chapter demonstrates that computational evaluation of primer recognition-site abundance provides a practical framework for rational primer design in selective multiple displacement amplification. Primer sets containing only modest numbers of genomic recognition sites generated substantial quantities of amplified DNA but recovered only a small proportion of the chromosome following sequencing. In contrast, computational optimisation of primer-binding site abundance and genomic distribution produced the Vib1 primer set, which achieved near-complete genome recovery while substantially improving representation of biologically important loci. Although amplification bias remained evident, the combined computational and experimental analyses establish primer-binding site abundance as a major determinant of genome recovery and provide a foundation for further optimisation of selective whole-genome amplification.

Chapter 4 Discussion

4.1 Performance of Primer Sets in MDA

This study demonstrated that computational optimisation of primer recognition-site abundance substantially improved genome representation following selective multiple displacement amplification of *Vibrio parahaemolyticus*. Comparison of the previously developed VpCl and cpmt primer sets with the newly designed Vib1 primers showed that primer-binding site abundance and genomic distribution were major determinants of amplification performance. Together, the computational and experimental analyses indicate that rational primer design provides a more effective strategy for improving selective MDA than empirical primer selection alone.

An important outcome of this comparison is that DNA yield alone proved to be a poor indicator of primer performance. Both the VpCl and cpmt primer sets generated substantial quantities of amplified DNA, yet sequencing demonstrated that genome representation remained extremely limited. Consequently, optimisation of primer design should focus on maximising genome coverage rather than simply increasing the total quantity of amplified DNA.

4.1.1 Comparative amplification efficiency of the VpCl and cpmt primer sets

The experimental evaluation of the VpCl and cpmt primer sets demonstrated that successful DNA amplification alone does not necessarily result in effective whole-genome recovery. Both primer sets generated substantial quantities of amplified DNA under identical multiple displacement amplification conditions (37°C for 3 h), confirming that Phi29-mediated strand displacement and DNA synthesis proceeded efficiently in all reactions. However, subsequent Oxford Nanopore sequencing revealed that high DNA yield did not correspond to broad genome representation, indicating that DNA quantity alone is an inadequate measure of primer performance.

The VpCl primer set, comprising four 10-mer primers with approximately 38–42 predicted recognition sites per chromosome, consistently produced high post-amplification DNA concentrations despite their relatively low genomic abundance. Nevertheless, genome mapping demonstrated that amplification was confined to only a small proportion of chromosome I, with genome coverage ranging from 0.18% to 3.18% and mean sequencing depths remaining below 0.1×. These findings indicate that relatively few amplification initiation sites were sufficient to

generate substantial quantities of DNA but failed to provide representative amplification across the chromosome.

The cpmt primer set was developed to increase the number of potential amplification initiation sites by using shorter 6-mer primers. Computational analyses predicted approximately 1,700–2,500 recognition sites per chromosome for the individual cpmt primers, representing an approximately 50-fold increase compared with the VpCl primer set. Although this increase in primer-binding site abundance again generated substantial quantities of amplified DNA, genome coverage remained limited, ranging from 0.11% to 1.64% for individual primers. These observations indicate that increasing primer-binding site abundance alone was insufficient to achieve proportional improvements in whole-genome representation.

Pooling all four cpmt primers produced a modest improvement in genome recovery, increasing coverage to 3.34%, marginally exceeding the best-performing VpCl primer (3.18%). The improvement obtained by combining multiple primers suggests that increasing primer diversity can increase the probability of initiating amplification across additional genomic regions. Nevertheless, the relatively small improvement observed indicates that empirical optimisation through increased primer number alone is insufficient to overcome the intrinsic amplification bias associated with selective multiple displacement amplification.

The comparison of the VpCl and cpmt primer sets provided an important foundation for the computational primer design strategy developed in this study. Collectively, these experiments demonstrated that neither high DNA yield nor increased primer-binding site abundance alone reliably predicted genome-wide amplification performance. Instead, they indicated that successful primer design requires consideration of both the abundance and genomic distribution of primer-binding sites. These observations motivated the development of the Vib1 primer set using a computational framework that explicitly optimised both characteristics before experimental validation.

4.1.2 Amplification bias and genomic motif distribution

One of the principal findings of this study was that primer-binding site abundance and genomic distribution strongly influenced the extent of genome recovery achieved following selective multiple displacement amplification. Computational analyses demonstrated that the Vib1 primer set contained substantially more recognition sites than the previously evaluated VpCl and cpmt primer sets, while the binding sites were also distributed more evenly throughout

chromosome I. Experimental genome mapping subsequently showed that these characteristics were associated with near-complete genome representation, with more than 99.9% of the reference chromosome recovered following sequencing. These findings indicate that computational optimisation of primer-binding site abundance represents an effective strategy for improving whole-genome recovery.

Although genome representation was greatly improved, sequencing depth remained heterogeneous across the chromosome, with some genomic regions exhibiting exceptionally high read depths while others were represented at considerably lower coverage. Similar patterns were observed among virulence-associated loci, where genes such as *tlh*, *pvuA*, *flgE* and MAM7 were recovered completely, whereas larger loci including T3SS1 and T6SS1 exhibited only partial recovery. These observations demonstrate that increasing primer-binding site abundance alone does not eliminate amplification bias.

The heterogeneous coverage observed in this study is likely to reflect multiple factors acting during selective multiple displacement amplification. The computational analyses showed that the Vib1 primers possessed shorter average distances between adjacent recognition sites and substantially smaller primer-free regions than either the VpCl or cpmt primer sets. These characteristics would be expected to increase the probability that DNA synthesis is initiated throughout the chromosome, thereby reducing large regions that fail to amplify. Nevertheless, local sequence composition, secondary DNA structure, accessibility of primer-binding sites and competitive amplification dynamics are also likely to influence amplification efficiency following primer annealing. Consequently, primer-binding site abundance provides a plausible mechanistic explanation for the substantial improvement in genome recovery, although additional biological and biochemical factors almost certainly contribute to the residual variation in sequencing depth.

Previous studies have similarly reported that multiple displacement amplification can produce uneven genome coverage despite generating large quantities of amplified DNA (Lasken & Stockwell, 2007; Hosono et al., 2003). These amplification biases arise because early amplification events are preferentially amplified during the highly processive rolling-circle synthesis catalysed by Phi29 DNA polymerase, resulting in stochastic over-representation of some genomic regions. The present study demonstrates that rational primer selection can substantially reduce, but not completely eliminate, these intrinsic biases by increasing the number and genomic distribution of amplification initiation sites.

An important outcome of this work is the distinction between genome representation and coverage uniformity. The computationally designed Vib1 primers successfully achieved near-complete genome recovery, confirming the value of optimising primer-binding site abundance and genomic distribution prior to experimental testing. However, the persistence of local variation in sequencing depth indicates that future improvements in selective multiple displacement amplification will require optimisation of additional aspects of the amplification workflow, including reaction conditions, primer concentration, amplification kinetics and library preparation. Consequently, computational primer design should be regarded as one component of a broader strategy for improving selective whole-genome amplification rather than a complete solution to amplification bias.

4.1.3 Rationale and necessity of computationally optimised *Vibrio*-specific primer design

The experimental evaluation of the VpCl and cpmt primer sets demonstrated that empirical primer selection alone was insufficient to achieve broad and representative genome amplification. These findings provided the rationale for developing the Vib1 primer set using a computational primer design strategy based on complete *Vibrio parahaemolyticus* reference genomes. Rather than selecting primer sequences solely on the basis of empirical testing, candidate primers were first evaluated according to their abundance, genomic distribution, conservation across independently assembled genomes and occurrence within biologically important virulence-associated loci. This computational framework enabled primer selection to be guided by predicted genome-wide performance before laboratory validation.

The Vib1 primers exhibited approximately 4,000–5,100 recognition sites per chromosome, substantially exceeding both the VpCl and cpmt primer sets. More importantly, these recognition sites were distributed relatively evenly throughout chromosome I, resulting in shorter average distances between adjacent primer-binding sites and substantially fewer large primer-free regions. Computational analyses further demonstrated that primer-binding site abundance and ranking were highly conserved across three independently assembled *V. parahaemolyticus* genomes, indicating that the selected motifs represented stable genomic characteristics rather than strain-specific sequence features.

Experimental validation supported these computational predictions. The Vib1 primer set generated near-complete genome recovery following Oxford Nanopore sequencing, with more than 99.9% of chromosome I represented by at least one sequencing read. Recovery of

biologically important loci was also substantially improved, with complete or near-complete recovery of several virulence-associated genes. Although amplification bias remained evident in some genomic regions, these results demonstrate that computational optimisation of primer-binding site abundance and genomic distribution can substantially improve genome representation compared with empirically selected primer sets.

Species-specific primer design has particular importance for diagnostic applications involving complex biological samples, where pathogen DNA frequently represents only a small proportion of the total nucleic acid present. Conventional random hexamer MDA amplifies all available DNA indiscriminately, whereas computationally selected primers increase the probability that amplification is initiated preferentially from pathogen-derived DNA.

Importantly, the Vib1 primer set was intentionally optimised using chromosome I because this chromosome provided an appropriate starting point for evaluating the computational primer-selection strategy and contains numerous conserved loci suitable for diagnostic development. Consequently, incomplete recovery of chromosome II and several toxin-associated loci was not unexpected and should not be interpreted as a failure of the computational approach. Rather, the present study demonstrates that computational optimisation can substantially improve recovery of the genomic regions for which primers were specifically designed. Extension of the same design strategy across both chromosomes represents the logical next stage in developing a comprehensive diagnostic assay.

4.1.4 Implications for optimisation and diagnostic performance

The combined computational and experimental analyses demonstrate that successful selective multiple displacement amplification depends on more than simply increasing primer-binding site abundance. The comparative evaluation of the VpCl, cpmt and Vib1 primer sets showed that primer performance is determined by a combination of primer-binding site abundance, genomic distribution and the representation of biologically important genomic regions. Optimising only one of these characteristics is unlikely to achieve uniform genome amplification. Instead, effective primer design requires consideration of both the number of available priming sites and their distribution throughout the target genome.

An important outcome of this study is the demonstration that computational analysis can substantially improve primer selection before laboratory validation. By evaluating candidate primers against complete reference genomes, it was possible to identify motifs that combined high genomic abundance with relatively uniform chromosomal distribution and consistent

occurrence across multiple *Vibrio parahaemolyticus* isolates. The subsequent experimental validation showed that this strategy produced near-complete genome recovery, providing strong support for the use of computational optimisation as the foundation for selective MDA primer design.

Despite these improvements, amplification bias was not completely eliminated. Although the Vib1 primer set achieved extensive genome representation, sequencing depth remained heterogeneous across the chromosome and several virulence-associated loci were recovered only partially. These observations indicate that primer optimisation alone cannot fully overcome the intrinsic biases associated with multiple displacement amplification. Additional factors, including local sequence composition, DNA secondary structure, amplification kinetics and library preparation, are also likely to influence the final distribution of sequencing coverage.

From a diagnostic perspective, these findings have important implications. Reliable pathogen detection requires not only recovery of the genome as a whole but also consistent representation of genomic regions containing clinically relevant virulence determinants. Future optimisation strategies should therefore combine genome-wide analyses of primer-binding site abundance and distribution with targeted evaluation of primer-binding site occurrence within diagnostically important genes. Such an approach would minimise the risk of systematic under-representation of key virulence loci while maintaining broad genome coverage.

Collectively, these findings establish a rational framework for the development of selective multiple displacement amplification systems. Rather than relying on empirical trial-and-error optimisation, computational evaluation of complete reference genomes can be used to guide primer selection prior to laboratory testing. Although further refinement of amplification chemistry and sequencing workflows will be required to reduce the remaining coverage bias, the approach developed in this study provides a robust foundation for improving targeted whole-genome amplification of bacterial pathogens.

4.2 Feasibility of MDA from Culture-Derived DNA

The experimental evaluation demonstrated that selective multiple displacement amplification could be successfully applied to cultured *Vibrio parahaemolyticus* DNA obtained using two independent extraction methods. Both the Roche High Pure Kit and the LysoFAST BOOST protocol produced DNA suitable for amplification and subsequent Oxford Nanopore sequencing, indicating that the Vib1 primer set was capable of amplifying DNA across a range of template qualities. These findings demonstrate that selective MDA is sufficiently robust for routine laboratory applications, even when DNA quality is not optimal.

The two extraction methods differed substantially in both DNA yield and apparent DNA integrity. DNA extracted using the Roche High Pure Kit exhibited higher concentrations and reduced smearing during agarose gel electrophoresis compared with DNA extracted using the LysoFAST BOOST protocol, indicating that the Roche method produced higher-quality template DNA. These differences were reflected in the subsequent amplification reactions, where Roche-derived DNA consistently generated greater post-MDA DNA concentrations than LysoFAST-derived DNA. Although both extraction methods ultimately produced sufficient DNA for Oxford Nanopore library preparation, the results suggest that template quality remained an important determinant of amplification efficiency.

An important observation from this study is that successful amplification should not be equated with successful genome recovery. Both extraction methods generated large quantities of amplified DNA following selective MDA, and all amplified samples produced sequencing libraries suitable for Oxford Nanopore sequencing. However, the sequencing data demonstrated that high DNA yield alone provides little information regarding the size distribution of amplified fragments or the uniformity of genome representation. This distinction reinforces one of the principal conclusions of the preceding chapter: DNA quantity is an inadequate surrogate for amplification performance, and evaluation of selective MDA should instead focus on genome coverage and representation.

Although high molecular weight genomic DNA was selected for these experiments, the resulting sequencing libraries remained dominated by relatively short reads, with a mean read length of approximately 2.7 kb and a median read length of only 1.2 kb. While occasional ultra-long reads were observed, these represented rare events rather than the predominant amplification products. The predominance of relatively short sequencing reads suggests that

the current selective MDA workflow did not consistently preserve long DNA fragments, despite producing sufficient DNA for successful genome mapping. Consequently, factors occurring during amplification, clean-up or library preparation may have contributed to fragmentation of amplified products.

This observation highlights an important limitation of the current workflow. The computational optimisation of primer design substantially improved genome representation, yet optimisation of primer-binding sites alone was insufficient to preserve long sequencing fragments. Future refinement of the selective MDA protocol should therefore focus not only on primer design but also on reaction conditions, amplification kinetics and library preparation procedures that minimise DNA fragmentation while maintaining uniform genome coverage.

Despite these limitations, the experimental results demonstrate that selective MDA is technically feasible using culture-derived *V. parahaemolyticus* DNA. Amplification was successful across both extraction methods, sufficient DNA was consistently obtained for Oxford Nanopore sequencing, and subsequent genome mapping demonstrated extensive genome recovery. These findings indicate that the principal remaining challenge is no longer obtaining sufficient amplified DNA, but improving the length and uniformity of amplified fragments while preserving the excellent genome representation achieved with the Vib1 primer set.

4.3 Genome recovery following selective multiple displacement amplification

One of the principal outcomes of this study was the demonstration that computational optimisation of primer design substantially improved genome recovery following selective multiple displacement amplification. Genome mapping of sequencing reads to the *Vibrio parahaemolyticus* reference genome showed that 99.956% of chromosome I was represented by at least one sequencing read, while 98.611% achieved a minimum sequencing depth of 10×. Compared with the previously evaluated VpCl and cpmt primer sets, which recovered only a small proportion of the genome, the Vib1 primer set achieved near-complete recovery of the chromosome for which it had been specifically designed. These findings provide strong experimental validation of the computational primer-selection strategy developed in this study.

Although genome recovery was almost complete, sequencing depth remained uneven across the chromosome. The mean sequencing depth (68×) exceeded the median depth (43×), indicating that a relatively small number of genomic regions were amplified to exceptionally high depths while others were represented at lower coverage. Such positively skewed coverage profiles are characteristic of Phi29-mediated multiple displacement amplification and reflect stochastic amplification of early priming events rather than widespread failure of genome recovery.

Importantly, amplification bias affected sequencing depth rather than overall genome representation. Only a very small proportion of chromosome I lacked sequencing coverage, indicating that complete amplification failure occurred rarely. Instead, local variation in coverage probably reflected differences in primer-binding site distribution, chromosomal architecture and amplification dynamics. These findings demonstrate an important distinction between genome representation and coverage uniformity: computational optimisation substantially improved genome recovery but did not eliminate the intrinsic amplification bias associated with multiple displacement amplification.

It is also important to interpret these results within the context of the primer design strategy. The Vib1 primers were intentionally designed using chromosome I because this chromosome contains many conserved loci suitable for developing a first-generation diagnostic assay. Consequently, incomplete recovery of chromosome II and several toxin-associated loci was not unexpected and should not be interpreted as a failure of the computational approach. Rather,

the results demonstrate that genome-informed primer optimisation can substantially improve recovery of the genomic regions for which primers are specifically designed. Extension of this strategy across both chromosomes represents the logical next step in developing a comprehensive diagnostic assay.

Overall, the present study demonstrates that computational optimisation of primer-binding site abundance and genomic distribution provides an effective framework for improving selective whole-genome amplification. Future optimisation should focus on improving coverage uniformity while extending primer selection to encompass clinically important loci located on chromosome II.

4.4 Recovery of Virulence-Associated Loci

Although genome-wide coverage provides an overall measure of amplification success, the principal objective of selective multiple displacement amplification is to recover biologically important genomic regions that support species identification and virulence assessment. Consequently, recovery of representative virulence-associated loci was examined to determine whether the improved genome representation achieved by the Vib1 primer set also translated into recovery of diagnostically relevant genes.

Recovery varied substantially among different classes of virulence loci, reflecting differences in primer-binding site distribution, chromosomal location and genomic organisation. Conserved chromosomal virulence genes were generally recovered with high completeness, whereas accessory virulence loci exhibited more variable recovery. These findings indicate that computational optimisation substantially improved recovery of chromosome I but that amplification efficiency remained influenced by local genomic characteristics.

4.4.1 Recovery of adhesion, motility and iron acquisition genes

Several genes associated with adhesion, motility and iron acquisition were recovered with high completeness. The adhesion gene *MAM7*, the iron acquisition gene *pvuA*, and the motility-associated genes *motY*, *fliM* and *flgE* all exhibited complete or near-complete recovery, demonstrating that the Vib1 primer set consistently amplified conserved chromosomal virulence determinants. Similarly, the species marker *tlh* was recovered completely, supporting its continued value for species identification.

The successful recovery of these loci provides strong experimental support for the computational primer optimisation strategy. Because these genes are distributed across chromosome I rather than being confined to a single genomic region, their consistent recovery indicates that optimisation of primer-binding site abundance and genomic distribution substantially improved amplification across the chromosome rather than within only isolated loci.

4.4.2 Implications for virulence profiling

Collectively, these findings demonstrate that computationally optimised selective MDA can reliably recover species markers together with many conserved virulence-associated loci while

substantially improving genome representation compared with the previously evaluated primer sets. However, recovery remained incomplete for toxin-associated determinants and several accessory genomic regions. Importantly, these loci were not primary targets during primer optimisation, as the Vib1 primer set was intentionally designed using chromosome I as the first stage in developing a selective diagnostic assay.

Consequently, the present workflow should be regarded as an effective first-generation genomic enrichment strategy rather than a complete diagnostic assay. Reliable recovery of chromosome I provides extensive information for species identification and structural virulence profiling, while future optimisation of primer-binding site distribution across chromosome II should improve recovery of toxin-associated genes and further enhance diagnostic capability.

4.5 Public Health Relevance in a Regional and Global Context

4.5.1 Public health significance of the workflow

The selective MDA–ONT workflow developed in this study demonstrated that *Vibrio parahaemolyticus* could be recovered and characterised from very small quantities of input DNA. Computationally optimised primer design enabled near-complete recovery of Chromosome I, reliable detection of the species marker *tlh*, and extensive recovery of several structural virulence determinants, including components of the T3SS1 and T6SS1 secretion systems. These findings demonstrate that selective MDA can substantially improve genome recovery while providing biologically meaningful information on the virulence potential of environmental isolates.

However, recovery of virulence determinants was not uniform. Although structural virulence systems were recovered reliably, the major toxin genes *tdh* and *trh* were not detected. Consequently, the present workflow is well suited to species confirmation and broad virulence profiling but does not yet provide complete pathogenicity assessment.

4.5.2 Applications in outbreak investigation and surveillance

Rapid identification of foodborne pathogens is essential during outbreak investigations. Conventional diagnostic polymerase chain reaction (PCR) provides a rapid, sensitive, and relatively inexpensive method for detecting *Vibrio parahaemolyticus* and selected virulence-associated genes. However, targeted PCR assays are restricted to predefined genetic markers and provide limited information about genomic characteristics beyond these targets. The selective multiple displacement amplification–Oxford Nanopore Technologies (sMDA–ONT) workflow investigated in this study offers a potentially complementary approach by enriching bacterial DNA for sequencing, with the potential to recover information from multiple genomic regions and virulence-associated loci. The recovery of 99.956% of chromosome I at a minimum sequencing depth of 1× demonstrates extensive chromosome representation, supporting the feasibility of this approach for genomic characterisation. However, the extent to which this broader genomic information could improve outbreak investigations remains to be established through further validation.

Despite these potential advantages, the current workflow has important limitations. Although extensive chromosome I coverage was achieved, sequencing coverage was heterogeneous, and

the *tdh* and *trh* virulence-associated genes were not recovered. These genes are commonly used as molecular markers of pathogenic potential in *V. parahaemolyticus*, and their absence from the sequencing data limits the ability to characterise these specific virulence determinants. However, because the analysis focused on chromosome I, their non-recovery does not establish their absence from the isolates. Targeted PCR could therefore complement the current workflow by providing specific confirmation of *tdh* and *trh*, while improvements to primer distribution and chromosome II coverage would be necessary to enhance the completeness of genomic characterisation. Furthermore, the present study did not directly compare sMDA–ONT with conventional PCR in terms of diagnostic sensitivity, specificity, cost, or turnaround time. Consequently, the workflow cannot yet be considered superior to established PCR-based methods and should instead be regarded as a proof-of-concept approach that may complement existing diagnostic strategies.

The portability and real-time sequencing capabilities of ONT, combined with the potential of sMDA to amplify low-input DNA, offer opportunities for future applications in decentralised genomic surveillance. If further optimised and validated, this approach could support the genomic characterisation of *V. parahaemolyticus* from seafood and environmental samples while providing information beyond that obtained through targeted PCR alone. However, its practical suitability for routine surveillance will depend on achieving more uniform genomic coverage, recovering relevant virulence-associated loci across both chromosomes, and establishing its diagnostic performance using diverse strains and representative sample types. Future studies should also compare the workflow directly with conventional PCR to determine whether the additional genomic information justifies its costs, processing requirements, and sequencing time. These improvements would help establish the role of sMDA–ONT as a complementary tool for food safety monitoring and public health investigations.

4.5.3 Implications for climate change and Pacific communities

The geographical distribution of pathogenic *Vibrio* species is expanding as sea surface temperatures increase and marine environments continue to change. Consequently, rapid genomic methods capable of identifying emerging pathogens from low quantities of DNA are likely to become increasingly valuable for environmental monitoring and public health surveillance.

The seasonal epidemiology of *V. parahaemolyticus* in New Zealand also has implications for food safety surveillance. Although warmer conditions can increase environmental bacterial concentrations and the risk of infection, documented winter outbreaks demonstrate that surveillance should not be restricted to summer (King et al., 2026; Jefferies et al., 2026). For New Zealand and the wider Pacific region, this supports the potential value of monitoring approaches capable of detecting *V. parahaemolyticus* across different seasons, particularly in seafood and coastal environments. The sMDA–ONT workflow investigated in this study may contribute to future surveillance strategies, although its suitability for routine seasonal monitoring requires further validation.

These considerations are particularly relevant for coastal Pacific communities, where seafood is culturally and economically important and access to centralised diagnostic laboratories may be limited. The combination of selective MDA with portable Oxford Nanopore sequencing offers the potential for regional genomic surveillance using relatively simple laboratory infrastructure. Reliable species identification together with recovery of multiple structural virulence determinants could improve local diagnostic capacity and support more timely public health responses.

4.5.4 Limitations and future public health applications

Although the present workflow demonstrated substantial improvements in genome recovery, incomplete representation of several clinically important virulence loci limits its immediate application as a comprehensive diagnostic method. Future optimisation should therefore focus on improving recovery of toxin-associated genes while maintaining the excellent genome representation achieved by the Vib1 primer set.

Overall, this study demonstrates that computationally optimised selective MDA provides a promising framework for rapid pathogen enrichment and genomic characterisation from low-input DNA. With further refinement of primer design and amplification chemistry, the approach has considerable potential to support future surveillance, outbreak investigation and environmental monitoring of *Vibrio parahaemolyticus* and other bacterial pathogens.

4.6 Limitations and Future Improvements

Although the present study demonstrated substantial improvements in selective whole-genome amplification of *Vibrio parahaemolyticus*, several limitations should be considered when interpreting the results.

The principal limitation was the persistence of amplification bias. Computational optimisation of primer-binding site abundance and genomic distribution substantially improved genome recovery; however, sequencing depth remained heterogeneous across the chromosome and several clinically important virulence loci, including *tdh*, *trh* and T3SS2, were not recovered. These findings indicate that primer-binding site distribution is only one determinant of amplification success and that additional factors, including local sequence composition, chromosomal organisation and amplification kinetics, also influence locus representation.

A second limitation is that primer design was based on exact sequence matches identified within a small number of complete reference genomes. Although this conservative approach ensured reproducible primer selection, it did not consider near-matches that may still support primer annealing during multiple displacement amplification. Similarly, the computational analyses did not explicitly incorporate local GC content, primer thermodynamics or DNA secondary structure, all of which may influence amplification efficiency.

The study was also limited to cultured *V. parahaemolyticus* isolates. Consequently, the performance of the selective MDA workflow was not evaluated using complex biological samples such as seafood, environmental water or clinical specimens, where host DNA, microbial background and inhibitory compounds may influence amplification success. Furthermore, although amplification was successfully achieved from 5 ng of input DNA, the analytical sensitivity and detection limits of the method were not determined, and no direct comparison was made with established PCR-based diagnostic assays.

Future development of the workflow should focus primarily on further optimisation of primer design. Expanding primer selection across a larger and more diverse collection of *V. parahaemolyticus* genomes, while explicitly considering primer distribution across both chromosomes and within clinically important virulence loci, should improve recovery of toxin-associated genes and reduce systematic amplification bias. Incorporating primer thermodynamics, local sequence composition and secondary structure into the computational design process may further improve primer performance.

Methodological optimisation also remains important. Evaluation of alternative amplification conditions, including primer concentrations, reaction temperature and incubation time, may improve coverage uniformity, while assessment of technical reproducibility across independent MDA reactions would provide greater confidence in the robustness of the workflow.

Finally, validation using environmental, seafood and clinical samples will be required to determine the practical diagnostic value of the approach. Comparative evaluation against conventional PCR, random hexamer MDA and untargeted whole-genome sequencing would establish the relative advantages of computationally optimised selective MDA and define the situations in which the method provides the greatest benefit.

Overall, this study demonstrates that computational optimisation of primer design provides a rational framework for improving selective multiple displacement amplification. Although further refinement is required before routine diagnostic implementation, the substantial improvements in genome recovery achieved here provide a strong foundation for the future development of targeted genomic enrichment strategies for *Vibrio parahaemolyticus* and other bacterial pathogens.

Chapter 5 Conclusion

The primary objective of this thesis was to determine whether computational optimisation of primer design could improve selective multiple displacement amplification of *Vibrio parahaemolyticus* for Oxford Nanopore sequencing. By combining genome-wide computational analyses with experimental validation, this study demonstrated that rational primer design can substantially improve genome recovery from low-input DNA and provides a practical framework for targeted whole-genome amplification.

Comparison of the previously developed VpCl and cpmt primer sets demonstrated that primer-binding site abundance alone is not a reliable predictor of amplification performance. Although both primer sets generated substantial quantities of amplified DNA, genome recovery remained limited, indicating that DNA yield is a poor surrogate for successful whole-genome amplification. These findings motivated the development of the Vib1 primer set using a computational strategy that explicitly optimised primer-binding site abundance, genomic distribution and conservation across multiple reference genomes.

Experimental validation confirmed the computational predictions. The Vib1 primer set achieved near-complete recovery of chromosome I following Oxford Nanopore sequencing, demonstrating that computational optimisation substantially improved genome representation compared with the previously evaluated primer sets. Recovery of several biologically important virulence-associated loci, including adhesion, motility, secretion system and iron acquisition genes, further demonstrated the value of genome-informed primer design for targeted genomic enrichment.

Despite these improvements, amplification bias remained evident. Importantly, however, the Vib1 primers were intentionally designed using chromosome I as the optimisation target because this represented the first stage in developing a selective diagnostic assay. Consequently, incomplete recovery of chromosome II and several toxin-associated loci was expected rather than representing a failure of the computational design strategy. Future optimisation should therefore extend primer selection across both chromosomes while refining amplification chemistry and library preparation.

The computational framework developed in this study provides a rational alternative to empirical primer selection. Rather than relying on iterative laboratory testing, candidate

primers can first be evaluated using complete genome sequences to identify motifs with favourable abundance, genomic distribution and representation within biologically important loci. Experimental validation demonstrated that this strategy substantially improved recovery of chromosome I, providing a proof of concept for genome-informed primer optimisation. Although further refinement will be required before routine diagnostic implementation, particularly through extension of primer optimisation across both chromosomes, the approach developed here establishes a practical foundation for targeted genomic enrichment using Oxford Nanopore sequencing and should be broadly applicable to other bacterial pathogens.

Appendix

Coverage profiles of additional virulence-associated loci following sMDA

Representative coverage profiles illustrating complete, intermediate and incomplete locus recovery are presented in the main text (Figures 3.3–3.5). The figures presented in this appendix provide the corresponding coverage profiles for the remaining virulence-associated loci examined in this study. All profiles were generated by mapping Oxford Nanopore sequencing reads obtained following selective multiple displacement amplification with the Vib1 primer set to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× in all figures to improve visualisation of local variation in sequencing depth.

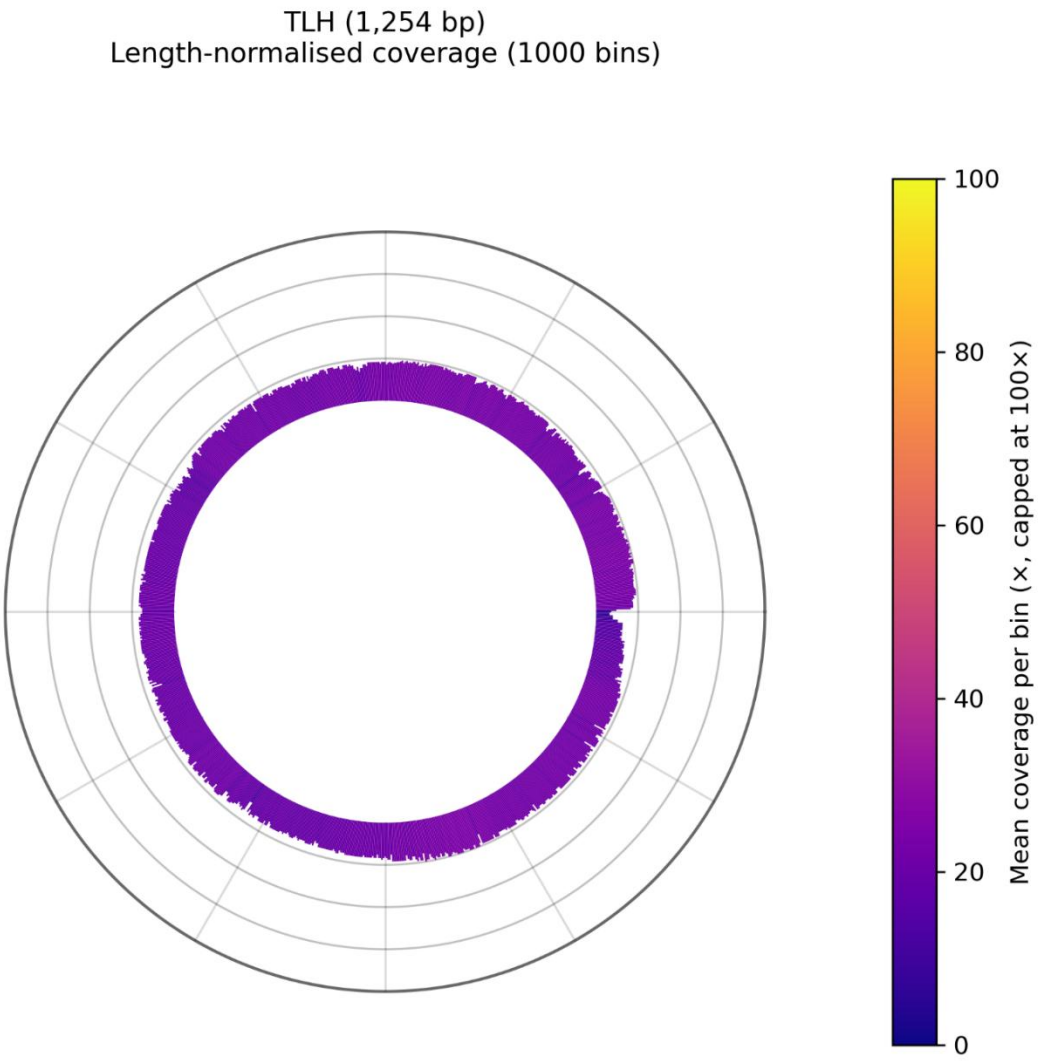


Figure A3.1. Sequencing coverage across the thermolabile haemolysin (tlh) gene following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× to improve visualisation of local variation in sequencing depth. Continuous coverage across the entire coding sequence resulted in complete recovery of the tlh locus.

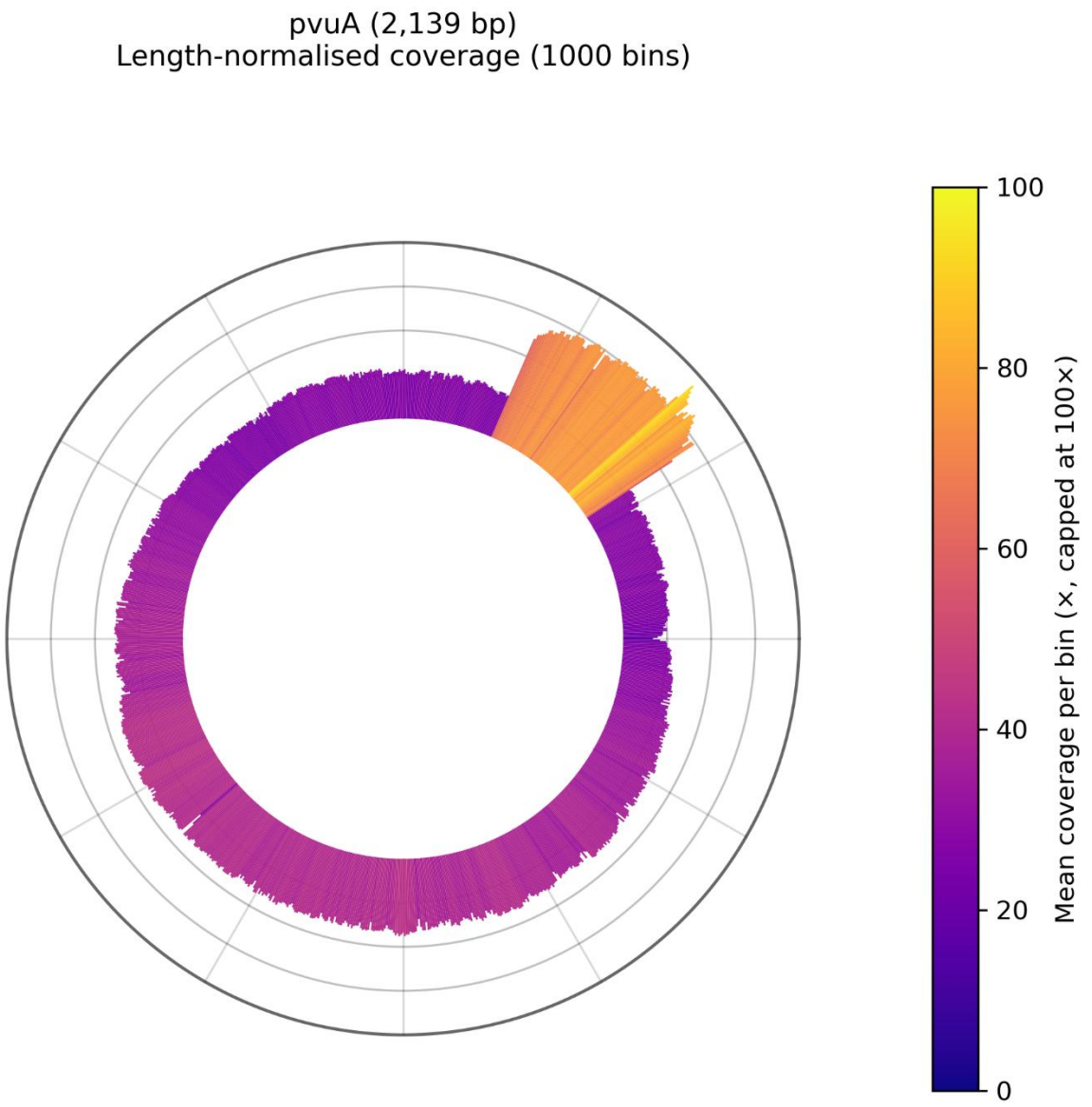


Figure A3.2. Sequencing coverage across the ferric vibrioferrin receptor (pvuA) locus following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× to improve visualisation of local variation in sequencing depth. Coverage was continuous across the gene despite moderate variation in sequencing depth, resulting in complete recovery of the locus.

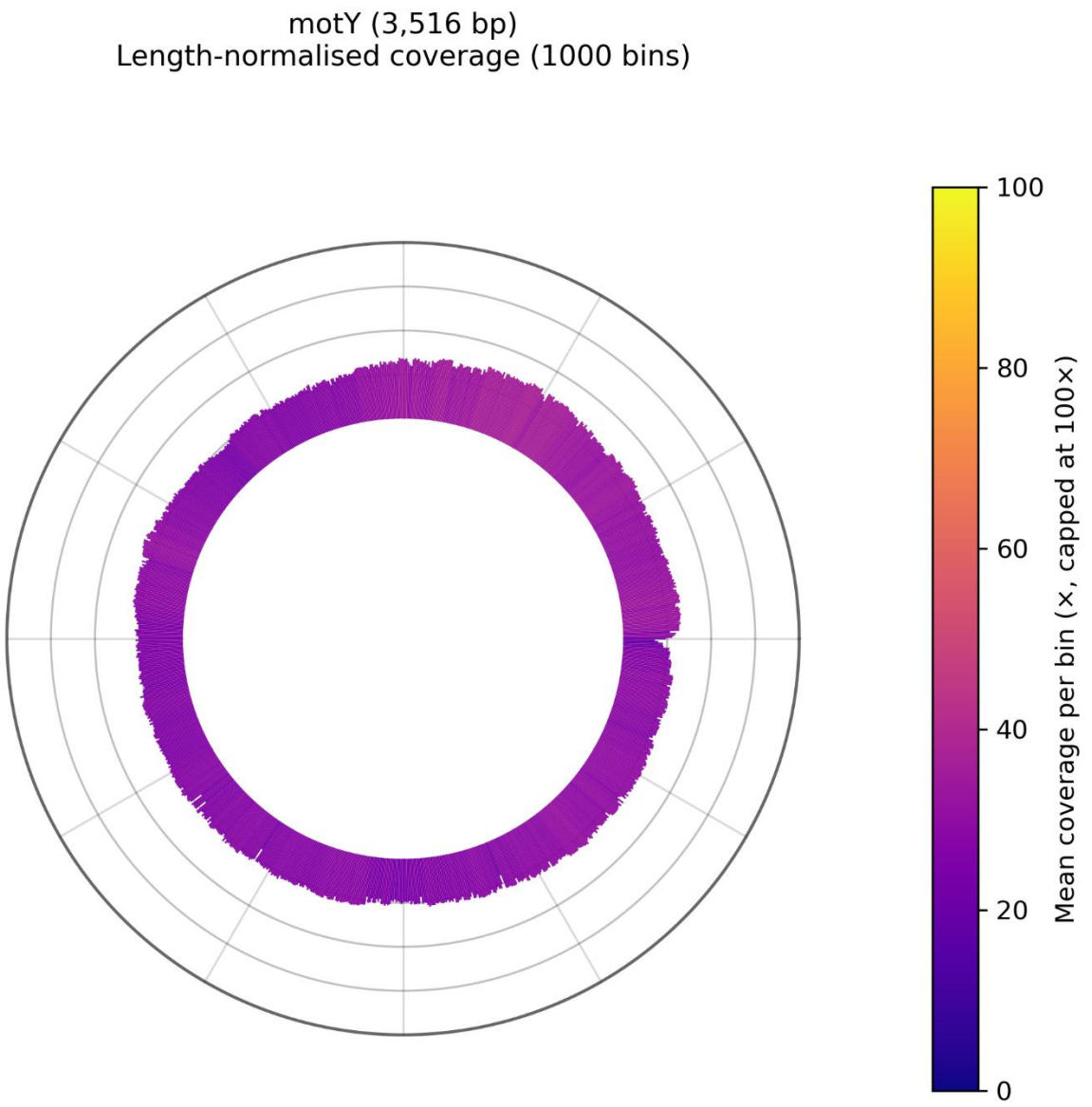


Figure A3.3. Sequencing coverage across the flagellar motor protein (motY) gene following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× to improve visualisation of local variation in sequencing depth. Sequencing coverage was continuous across almost the entire coding sequence, producing near-complete recovery of the gene.

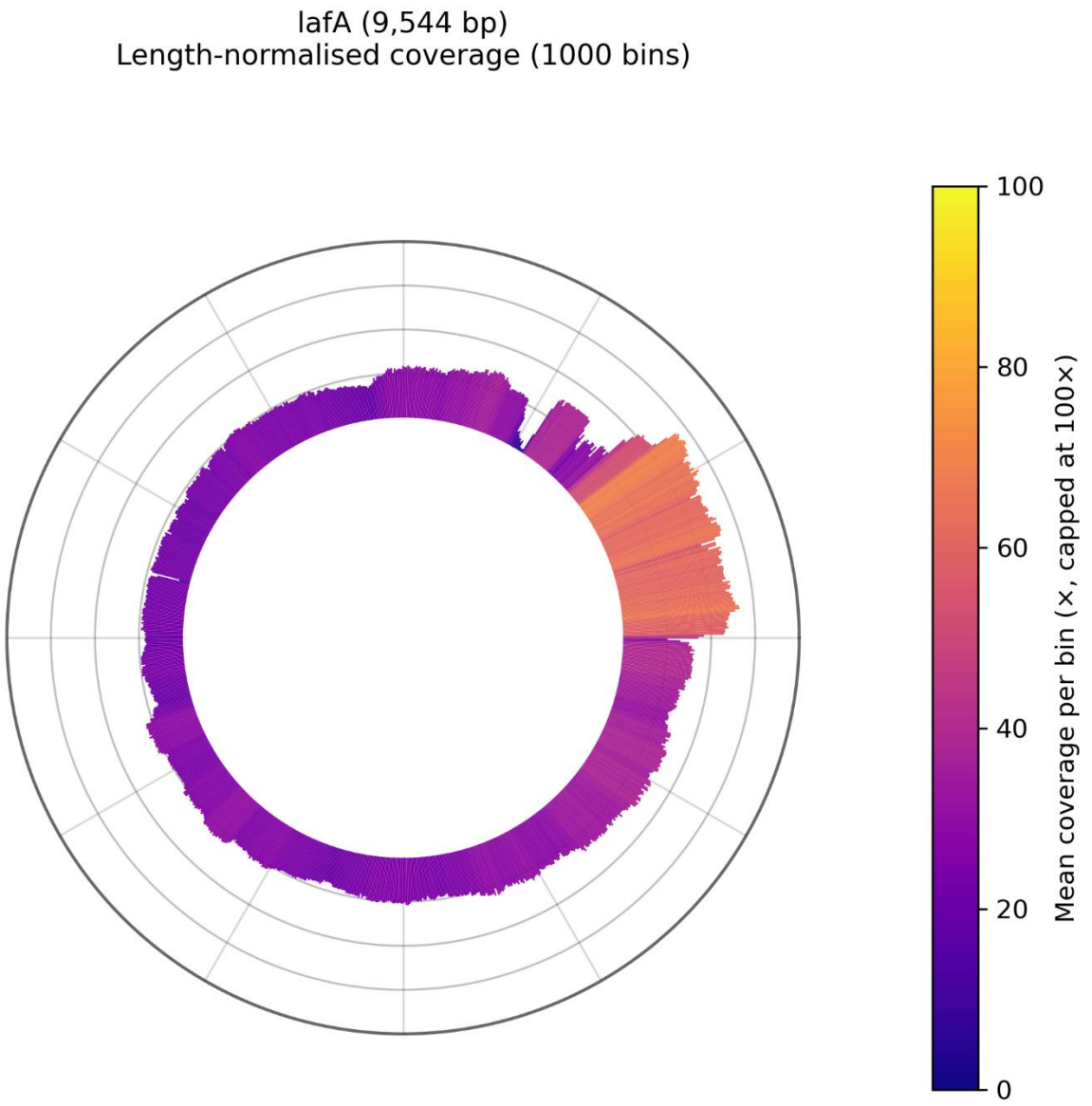


Figure A3.4. Sequencing coverage across the lateral flagellin (lafA) gene following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× to improve visualisation of local variation in sequencing depth. Localised reductions in sequencing depth contributed to incomplete recovery of the locus, illustrating moderate amplification bias.

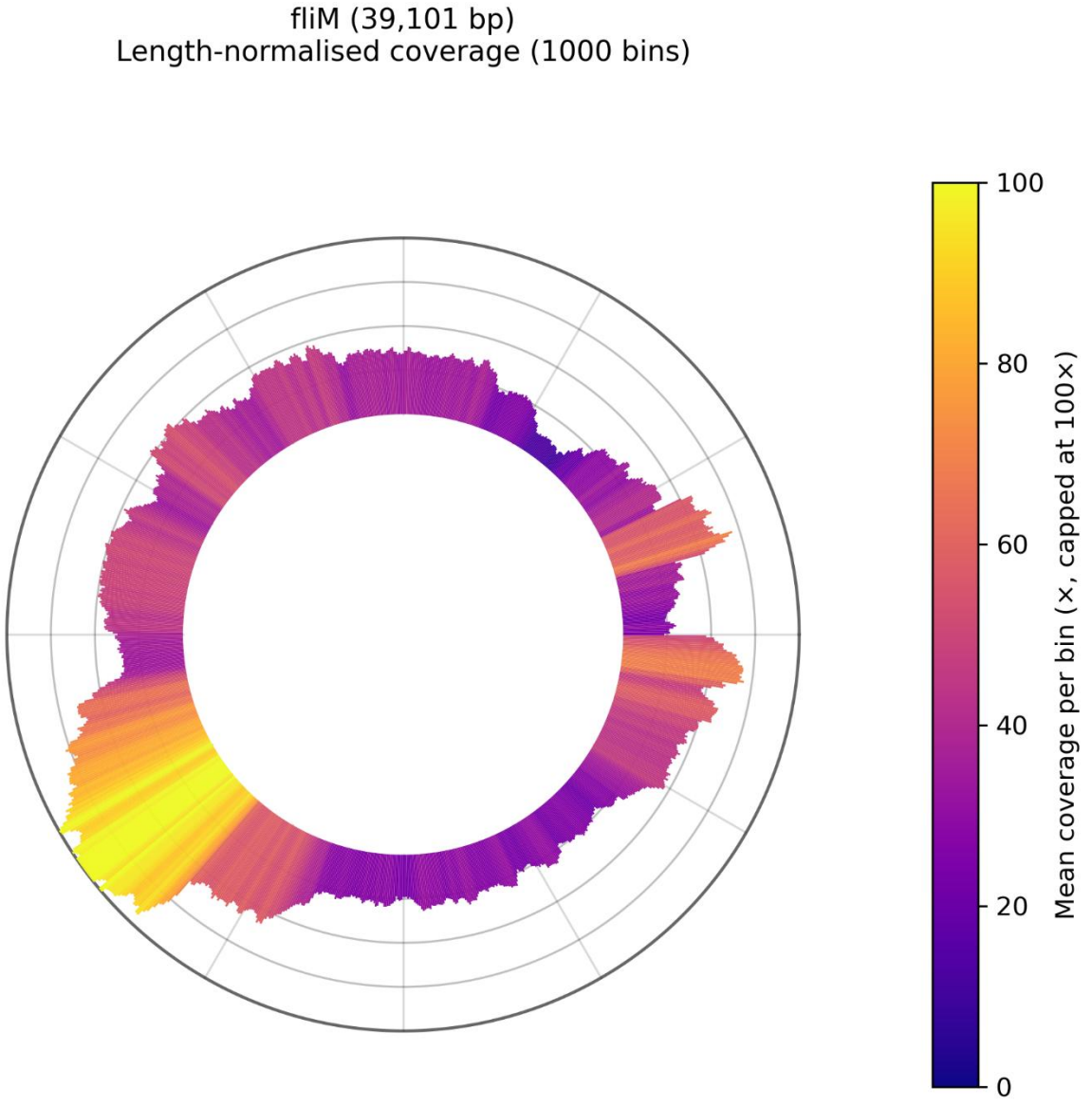


Figure A3.5. Sequencing coverage across the flagellar motor switch protein (fliM) gene following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at **100**× to improve visualisation of local variation in sequencing depth. Although several localised reductions in coverage were observed, the majority of the coding sequence was recovered, resulting in high overall completeness.

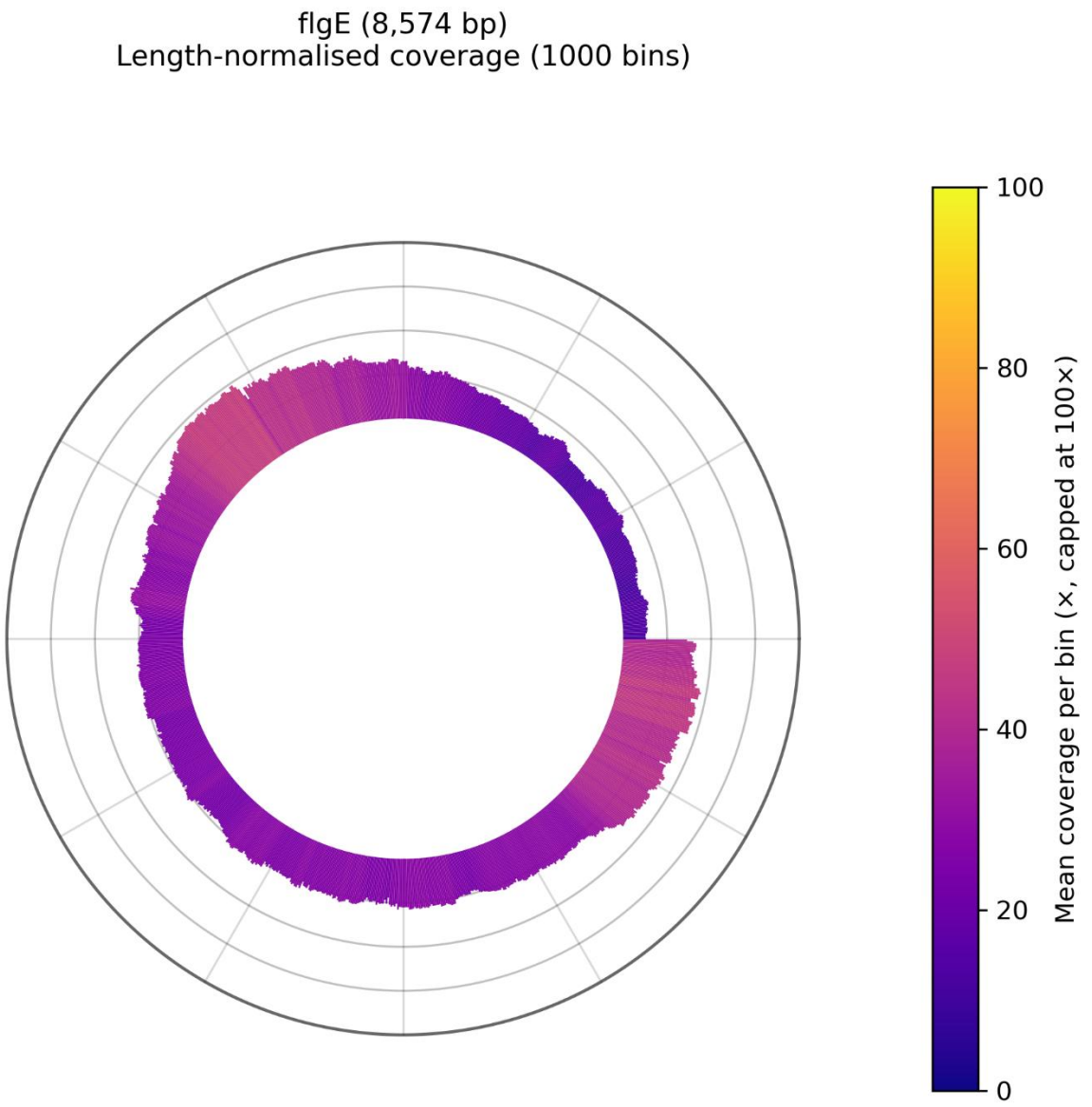


Figure A3.6. Sequencing coverage across the flagellar hook protein (flgE) gene following selective multiple displacement amplification using the *Vib*1 primer set.

Oxford Nanopore sequencing reads were mapped to the *Vibrio parahaemolyticus* reference genome. Coverage depth is capped at 100× to improve visualisation of local variation in sequencing depth. Sequencing coverage extended continuously across the coding sequence with only minor fluctuations in read depth, resulting in complete recovery of the locus.

Ranking of 6mer primer-binding site frequencies

Table A1: Occurrences of Suitable 6-mers in *V. parahaemolyticus* Chromosome I

Primer	BM24A chromosome I		AM43962 chromosome I	
	Counts - Occurrences	Rank Most-Common	Counts - Occurrences	Rank Most-Common
CATCAA	2579	1st	2563	1st
TTGATG	2557	2nd	2549	2nd
ATCAAA	2332	10th	2357	4th
TGATGA	2169	31st	2171	24th
CTTCAA	2117	39th	2105	42nd
TTGCTG	2165	32nd	2117	40th
CAGCAA	2116	40th	2197	20th
TCATCA	2084	47th	2060	46th
TCTTCA	2020	51st	2017	51st
TGAAGA	2010	55th	2037	48th
GATGAA	1974	60th	1970	56th

References

Arbefeville, S. S., Timbrook, T., & Garner, C. D. (2024). Evolving strategies in microbe identification: A comprehensive review of biochemical, MALDI-TOF MS and molecular testing methods. *J Antimicrob Chemother, 79*(12 Suppl. 2), i2–i8. https://doi.org/10.1093/jac/dkae275

Asplund, M. E. (2013). *Ecological aspects of marine Vibrio bacteria: Exploring relationships to other organisms and a changing environment.* https://news-oceanacidification-icc.org/category/science/

Baker-Austin, C., Lake, I., Archer, E., Hartnell, R., Trinanes, J., & Martinez-Urtaza, J. (2024). Stemming the rising tide of *Vibrio* disease. *The Lancet Planetary Health, 8*(7), e515–e520. https://doi.org/10.1016/S2542-5196(24)00124-4

Belard, A., Buchman, T., Forsberg, J., Potter, B. K., Dente, C. J., Kirk, A., & Elster, E. (2017). Precision diagnosis: A view of the clinical decision support systems (CDSS) landscape through the lens of critical care. *Journal of Clinical Monitoring and Computing, 31*(2), 261–271. https://doi.org/10.1007/s10877-016-9849-1

Bell, A., & Bott, M. (2021). Vibriosis: What you and your patients need to know. *2639-6378 Dela J Public Health, 7*(1), 14–21. https://doi.org/10.32481/djph.2021.001.005

Ben-Yaakov, R., & Salomon, D. (2019). The regulatory network of *Vibrio parahaemolyticus* type VI secretion system 1. *Environ Microbiol, 21*(7), 2248–2260. https://doi.org/10.1111/1462-2920.14594

Bienlien, L. M., Audemard, C., Reece, K. S., & Carnegie, R. B. (2022). Impact of parasitism on levels of human-pathogenic *Vibrio* species in eastern oysters. *Journal of Applied Microbiology, 132*(2), 760–771. https://doi.org/10.1111/jam.15287

Bottone, E. J., & Robin, T. (1978). *Vibrio parahaemolyticus*: Suspicion of presence based on aberrant biochemical and morphological features. *Journal of Clinical Microbiology, 8*(6), 760–763. https://doi.org/10.1128/jcm.8.6.760-763.1978

Brewer, T. D., Andrew, N. L., Abbott, D., Detenamo, R., Faaola, E. N., Gounder, P. V., Lal, N., Lui, K., Ravuvu, A., Sapalojang, D., Sharp, M. K., Sulu, R. J., Suvulo, S., Tamate, J. M. M. M., Thow, A. M., & Wells, A. T. (2023). The role of trade in Pacific food security and nutrition. *Global Food Security, 36*, 100670. https://doi.org/10.1016/j.gfs.2022.100670

Broberg, C. A., Calder Tj, & Orth, K. (2011). *Vibrio parahaemolyticus* cell biology and pathogenicity determinants. *Microbes Infect, 13*(12–13), 992–1001. https://doi.org/10.1016/j.micinf.2011.06.013

Brumfield, K. D., Chen, A. J., Gangwar, M., Usmani, M., Hasan, N. A., Jutla, A. S., Huq, A., & Colwell, R. R. (2023). Environmental factors influencing occurrence of *Vibrio parahaemolyticus* and *Vibrio vulnificus*. *Applied and Environmental Microbiology, 89*(6), e00307-00323. https://doi.org/10.1128/aem.00307-23

Cabello-Olmo, M., Oneca, M., Torre, P., Díaz, J., Encio, I., Barajas, M., & Araña, M. (2020). Influence of storage temperature and packaging on bacteria and yeast viability in a plant-based fermented food. *Foods,* 9(3):302. https://doi.org/10.3390/foods9030302

Chang, Y.-T., Huang, W.-T., Wu, P.-L., Kumar, R., Wang, H.-C., & Lu, H.-P. (2024). Low salinity stress increases the risk of *Vibrio parahaemolyticus* infection and gut microbiota dysbiosis in Pacific white shrimp. *BMC Microbiology, 24*. https://doi.org/10.1186/s12866-024-03407-0

Chen, Y., Dai, J. G., Jr., Morris, J. G., Jr., & Johnson, J. A. (2010). Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol* 2:10:274. https://doi.org/10.1186/1471-2180-10-274

Chen, Y., Stine, O. C., Badger, J. H., Gil, A. I., Nair, G. B., Nishibuchi, M., & Fouts, D. E. (2010). Comparative genomic analysis of *Vibrio parahaemolyticus*: Serotype conversion and virulence. *BMC Microbiology, 10*, 274. https://doi.org/10.1186/1471-2180-10-274

Chimalapati, S., Lafrance, A. E., Chen, L., & Orth, K. (2020). *Vibrio parahaemolyticus*: Basic techniques for growth, genetic manipulation, and analysis of virulence factors. *Curr Protoc Microbiol, 59*(1), e131. https://doi.org/10.1002/cpmc.131

Chun, D., Chung, J. K., Tak, R., & Seol, S. Y. (1975). Nature of the Kanagawa phenomenon of *Vibrio parahaemolyticus*. *Infect Immun, 12*(1), 81–87. https://doi.org/10.1128/iai.12.1.81-87.1975

Cowell, A. N., Loy, D. E., Sundararaman, S. A., Valdivia, H., Fisch, K., Lescano, A. G., Baldeviano, G. C., Durand, S., Gerbasi, V., Sutherland, C. J., Nolder, D., Vinetz, J. M., Hahn, B. H., & Winzeler, E. A. (2017). Selective whole-genome amplification is a robust method that enables scalable whole-genome sequencing of Plasmodium vivax from unprocessed clinical samples. mBio, 8(1), e02257-16. https://doi.org/10.1128/mBio.02257-16

Davis, B. J. K., Jacobs, J. M., Davis, M. F., Schwab, K. J., DePaola, A., & Curriero, F. C. (2017). Environmental determinants of *Vibrio parahaemolyticus* in the Chesapeake Bay. *Applied and Environmental Microbiology, 83*(21), e01147-01117. https://doi.org/10.1128/AEM.01147-17

Di Pinto, A., Terio, V., Novello, L., & Tantillo, G. (2011). Comparison between thiosulphate-citrate-bile salt sucrose (TCBS) agar and CHROMagar *Vibrio* for isolating *Vibrio*

parahaemolyticus. *Food Control, 22*(1), 124–127. https://doi.org/10.1016/j.foodcont.2010.06.013

Eldin, C., Parola, P., & Raoult, D. (2019). Limitations of diagnostic tests for bacterial infections. *Médecine et Maladies Infectieuses, 49*(2), 98–101. https://doi.org/10.1016/j.medmal.2018.12.004

Fearnley, E., Leong, L. E. X., Centofanti, A., Dowsett, P., Combs, B. G., Draper, A. D. K., Hocking, H., Howden, B., Horan, K., Wilmot, M., Levy, A., Cooley, L. A., Kennedy, K. J., Wang, Q., Arnott, A., Graham, R. M. A., Sinchenko, V., Jennison, A. V., Kane, S., & Wright, R. (2024). *Vibrio parahaemolyticus* foodborne illness associated with oysters, Australia, 2021–2022. *Emerging Infectious Diseases, 11*, 2271–2278. https://doi.org/10.3201/eid3011.240172

Fletcher, G. C., Cruz, C. D., & Hedderley, D. I. (2024). *Vibrio parahaemolyticus*: Predicting effects of storage temperature on growth in *Crassostrea gigas* harvested in New Zealand. *Aquaculture, 579*, 740128. https://doi.org/10.1016/j.aquaculture.2023.740128

Freitas, C., Glatter, T., & Ringgaard, S. (2019). The release of a distinct cell type from swarm colonies facilitates dissemination of *Vibrio parahaemolyticus* in the environment. *The ISME Journal, 14*(1), 230–244. https://doi.org/10.1038/s41396-019-0521-x

Fujino, T., Okuno, Y., Nakada, D., Aoyama, A., Fukai, K., Mukai, T., & Ueho, T. (1953). *On the bacteriological examination of shirasu-food poisoning.* https://www.cabidigitallibrary.org/doi/full/10.5555/19542702324

Gassen, L., Esters, L., Ribas-Ribas, M., & Wurl, O. (2024). The impact of rainfall on the sea surface salinity: A mesocosm study. *Scientific Reports, 14*. https://doi.org/10.1038/s41598-024-56915-4

Georgeou, N., Hawksley, C., Wali, N., Lountain, S., Rowe, E., West, C., & Barratt, L. (2022). Food security and small holder farming in Pacific Island countries and territories: A scoping review. *PLOS Sustain Transform, 1*(4), e0000009. https://doi.org/10.1371/journal.pstr.0000009

Ghouneimy, A., Mahas, A., Marsic, T., Aman, R., & Mahfouz, M. (2023). CRISPR-based diagnostics: Challenges and potential solutions toward point-of-care applications. *ACS Synth Biol, 12*(1), 1–16. https://doi.org/10.1021/acssynbio.2c00496

Gu, D., Zhang, Y., Wang, Q., & Zhou, X. (2020). S-nitrosylation-mediated activation of a histidine kinase represses the type 3 secretion system and promotes virulence of an enteric pathogen. *Nature Communications, 11*(1), 5777. https://doi.org/10.1038/s41467-020-19506-1

Guerrero, A., Lizárraga-Partida, M., Gómez Gil Rodríguez, B., Licea-Navarro, A. F., Revilla-Castellanos, V. J., Wong-Chang, I., & González-Sánchez, R. (2017). Genetic analysis of *Vibrio parahaemolyticus* O3:K6 strains that have been isolated in Mexico since 1998. *PLoS One, 12*(1), e0169722. https://doi.org/10.1371/journal.pone.0169722

Guillod, C., Ghitti, F., & Mainetti, C. (2019). *Vibrio parahaemolyticus*-induced cellulitis and septic shock after a sea beach holiday in a patient with leg ulcers. *Case Reports in Dermatology, 11*(1), 94–100. https://doi.org/10.1159/000499478

Gutierrez West, C. K., Klein, S. L., & Lovell, C. R. (2013). High frequency of virulence factor genes *tdh*, *trh*, and *tlh* in *Vibrio parahaemolyticus* strains isolated from a pristine estuary. *Appl Environ Microbiol, 79*(7), 2247–2252. https://doi.org/10.1128/AEM.03792-12

Han, F., Gu, R.-R., Shen, X.-S., Chen, Y.-G., Tian, L.-L., Zhou, W.-F., & Cai, Y.-Q. (2017). Detection of total and pathogenic *Vibrio parahaemolyticus* in shellfish growing along the South Yellow Sea and the East China Sea. *Journal of Food Protection, 80*(11), 1882–1889. https://doi.org/10.4315/0362-028X.JFP-17-080

Haque, E., Hussain, J., Shankar, S., Haldar, S., & Chatterjee, S. (2023). Chapter 15 – Implication of *Vibrio* biofilms in human and seafood sector. In S. Das & N. A. Kungwani (Eds.), *Understanding microbial biofilms* (pp. 247–260). Academic Press. https://doi.org/10.1016/B978-0-323-99977-9.00038-7

Hara-Kudo, Y., & Kumagai, S. (2014). Impact of seafood regulations for *Vibrio parahaemolyticus* infection and verification by analyses of seafood contamination and infection. *Epidemiol Infect, 142*(11), 2237–2247. https://doi.org/10.1017/S0950268814001897

He, M., Lei, T., Jiang, F., Zhang, J., Zeng, H., Wang, J., Chen, M., Xue, L., Wu, S., Ye, Q., Pang, R., Ding, Y., & Wu, Q. (2021). Genetic diversity and population structure of *Vibrio parahaemolyticus* isolated from clinical and food sources. *Frontiers in Microbiology, 12*. https://doi.org/10.3389/fmicb.2021.708795

Hoque, M., Akhter, N., & Chowdhury, M. S. R. (2022). Consumers' preferences for the traceability information of seafood safety. *Foods* **2022**, *11*, 1675. https://doi.org/10.3390/foods11121675

Hu, D., Liu, B., Feng, L., Ding, P., Guo, X., Wang, M., Cao, B., Reeves, P. R., & Wang, L. (2016). Origins of the current seventh cholera pandemic. *Proceedings of the National Academy of Sciences, 113*(48), E7730–E7739. https://doi.org/10.1073/pnas.1608732113

Hu, Y. Q., Huang, X. H., Guo, L. Q., Shen, Z. C., Lv, L. X., Li, F. X., Zhou, Z. H., & Zhang, D. F. (2021). Rapid and visual detection of *Vibrio parahaemolyticus* in aquatic foods using *bla*(CARB-17) gene-based loop-mediated isothermal amplification with lateral flow dipstick

(LAMP-LFD). *J Microbiol Biotechnol, 31*(12), 1672–1683. https://doi.org/10.4014/jmb.2107.07022

Isom, G. L., Davies, N. J., Chong, Z.-S., Bryant, J. A., Jamshad, M., Sharif, M., Cunningham, A. F., Knowles, T. J., Chng, S.-S., Cole, J. A., & Henderson, I. R. (2017). MCE domain proteins: Conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. *Scientific Reports, 7*(1), 8608. https://doi.org/10.1038/s41598-017-09111-6

Jefferies, S., Paine, S., Wang, J., Ren, X., Winter, D., Fletcher, G. C., Perchec-Merien, A.-M., & Wright, J. (2026). Emergent Vibrio parahaemolyticus gastroenteritis outbreaks, New Zealand, 2019–2022. Emerging Infectious Diseases, 32(8). https://wwwnc.cdc.gov/eid/article/32/8/26-0097_article

Kaval, K. G., Chimalapati, S., Siegel, S. D., Garcia, N., Jaishankar, J., Dalia, A. B., & Orth, K. (2023). Membrane-localized expression, production and assembly of *Vibrio parahaemolyticus* T3SS2 provides evidence for transertion. *Nature Communications, 14*(1), 1178. https://doi.org/10.1038/s41467-023-36762-z

King, N., Addidle, M., Wright, J., et al. (2026). Vibrio infections in Aotearoa New Zealand (1998–2024) and the effect of current disease reporting structures. New Zealand Medical Journal, 139(1636), 14–24. https://doi.org/10.26635/6965.7257

Kumar, V., Roy, S., Baruah, K., Van Haver, D., Impens, F., & Bossier, P. (2020). Environmental conditions steer phenotypic switching in acute hepatopancreatic necrosis disease-causing, affecting PirAVP/PirBVP toxins production. *Environmental Microbiology, 22*(10), 4212–4230. https://doi.org/10.1111/1462-2920.14903

Lane, J. M., Brosschot, T. P., Gatti, D. M., Gauthier, C. M., Lawrence, K. M., Pluzhnikova, V., & Reynolds, L. A. (2023). Chronic small intestinal helminth infection perturbs bile acid homeostasis and disrupts bile acid signaling in the murine small intestine. *Frontiers in Parasitology, 2*. https://doi.org/10.3389/fpara.2023.1214136

LaPorte, J. P., Spinard, E. J., Cavanagh, D., Gomez-Chiarri, M., Rowley, D. C., Mekalanos, J. J., Mittraparp-Arthorn, P., & Nelson, D. (2023). Draft genome sequence of *Vibrio parahaemolyticus* PSU5579, isolated during an outbreak of acute hepatopancreatic necrosis disease in Thailand. *Microbiol Resour Announc, 12*(2), e0087322. https://doi.org/10.1128/mra.00873-22

Leichty, A. R., & Brisson, D. (2014). Selective whole genome amplification for resequencing target microbial species from complex natural samples. Genetics, 198(2), 473–481. https://doi.org/10.1534/genetics.114.165498

León-Sicairos, N., Zatarain-Lopez, R., Angulo-Zamudio, U., Velazquez-Roman, J., Flores-Villaseñor, H., Martinez-Garcia, J. J., Moreno-Pérez, M., Buelna-Romero, A., Hernández-Monroy, I., Lopez-Martinez, I., Cuen-Diaz, H. M., Diaz-Quiñonez, J., & Canizalez-Roman,

A. (2022). *Vibrio parahaemolyticus* is associated with diarrhea cases in Mexico, with a dominance of pandemic O3:K6 clones. *Int. J. Environ. Res. Public Health* 2022, *19*, 10318. https://doi.org/10.3390/ijerph191610318

Letchumanan, V., Chan, K. G., & Lee, L. H. (2014). *Vibrio parahaemolyticus*: A review on the pathogenesis, prevalence, and advance molecular identification techniques. *Front Microbiol, 5*, 705. https://doi.org/10.3389/fmicb.2014.00705

Li, L., Meng, H., Gu, D., Li, Y., & Jia, M. (2019). Molecular mechanisms of *Vibrio parahaemolyticus* pathogenesis. *Microbiological Research, 222*, 43–51. https://doi.org/10.1016/j.micres.2019.03.003

Liu, M., & Chen, S. (2015). A novel adhesive factor contributing to the virulence of *Vibrio parahaemolyticus*. *Scientific Reports, 5*, Article 14449. https://doi.org/10.1038/srep14449

Liu, S., Sun, M., Zhang, N., Sun, Z., Tian, X., Li, L., & Wang, Y. (2024). Shaping global health promotion: A comprehensive analysis of the 10 Global Conferences on Health Promotion Conferences (1986–2021). *Global Health Journal, 8*(2), 91–96. https://doi.org/10.1016/j.glohj.2024.05.002

Marudhupandi, T., Kumar, T. T. A., Prakash, S., Balamurugan, J., & Dhayanithi, N. B. (2017). *Vibrio parahaemolyticus*, a causative bacterium for tail rot disease in ornamental fish, *Amphiprion sebae*. *Aquaculture Reports, 8*, 39–44. https://doi.org/10.1016/j.aqrep.2017.09.004

Meparambu Prabhakaran, D., Ramamurthy, T., & Thomas, S. (2020). Genetic and virulence characterisation of *Vibrio parahaemolyticus* isolated from the Indian coast. *BMC Microbiology, 20*(1), 62. https://doi.org/10.1186/s12866-020-01746-2

Montieri, S., Ciccozzi, M., & Croci, L. (2010). Phylogenetic and evolutionary analysis of *Vibrio parahaemolyticus* and *Vibrio alginolyticus* isolates based on *toxR* gene sequence. *New Microbiol, 33*(4), 359–372. PMID: 21213595.

Nakaguchi, Y. (2013). Contamination by *Vibrio parahaemolyticus* and its virulent strains in seafood marketed in Thailand, Vietnam, Malaysia, and Indonesia. *Trop Med Health, 41*(3), 95–102. https://doi.org/10.2149/tmh.2011-06

Namadi, P., & Deng, Z. (2021). Modeling and forecasting *Vibrio parahaemolyticus* concentrations in oysters. *Water Research, 189*, 116638. https://doi.org/10.1016/j.watres.2020.116638

Namadi, P., & Deng, Z. (2023). Optimum environmental conditions controlling prevalence of *Vibrio parahaemolyticus* in the marine environment. *Marine Environmental Research, 183*, 105828. https://doi.org/10.1016/j.marenvres.2022.105828

Neil, W. A., Hard, C., Bowers, J. C., & Jones, J. L. (2023). Levels of *Vibrio parahaemolyticus* in Pacific oysters (*Crassostrea gigas*) from Washington State following ambient exposure and chilling. *Journal of Food Protection, 86*(6), 100092. https://doi.org/10.1016/j.jfp.2023.100092

Nemoto, J., Ikedo, M., Kojima, T., Momoda, T., Konuma, H., & Hara-Kudo, Y. (2011). Development and evaluation of a loop-mediated isothermal amplification assay for rapid and sensitive detection of Vibrio parahaemolyticus. Journal of Food Protection, 74(9), 1462–1467. https://doi.org/10.4315/0362-028X.JFP-10-519

Ngasotter, S., Mukherjee, S., Singh, S. K., Bharti, D., Haque, R., Varshney, S., Nanda, C., Waikhom, D., Devi, M. S., & Singh, A. S. (2022). Prevalence, virulence, and antibiotic resistance of *Vibrio parahaemolyticus* from seafood and its environment: An updated review. *Mediterranean Journal of Infection, Microbes and Antimicrobials, 11*(1), 1. https://doi.org/10.4274/mjima.galenos.2021.2021.1

Nguyen, A. Q., Shimohata, T., Hatayama, S., Tentaku, A., Kido, J., Bui, T. M. H., Uebanso, T., Mawatari, K., & Takahashi, A. (2020). Type III secretion effector VopQ of *Vibrio parahaemolyticus* modulates central carbon metabolism in epithelial cells. https://doi.org/10.1128/mSphere .00960-19

Nigro, O. D., & Steward, G. F. (2015). Differential specificity of selective culture media for enumeration of pathogenic vibrios: Advantages and limitations of multi-plating methods. *Journal of Microbiological Methods, 111*, 24–30. https://doi.org/10.1016/j.mimet.2015.01.014

Oyola, S. O., Ariani, C. V., Hamilton, W. L., Kekre, M., Amenga-Etego, L. N., Ghansah, A., Rutledge, G. G., Redmond, S., Manske, M., Jyothi, D., Jacob, C. G., Otto, T. D., Rockett, K., Newbold, C. I., Berriman, M., & Kwiatkowski, D. P. (2016). Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. Malaria Journal, 15, Article 597. https://doi.org/10.1186/s12936-016-1641-7

Pang, H., Xin, X., He, J., Cui, B., Guo, D., Liu, S., Yan, Z., Liu, C., Wang, X., & Nan, J. (2020). Effect of NaCl concentration on microbiological properties in NaCl-assisted anaerobic fermentation: Hydrolase activity and microbial community distribution. *Frontiers in Microbiology, 11*. https://doi.org/10.3389/fmicb.2020.589222

Paranjpye, R., Hamel, O. S., Stojanovski, A., & Liermann, M. (2012). Genetic diversity of clinical and environmental *Vibrio parahaemolyticus* strains from the Pacific Northwest. *Appl Environ Microbiol, 78*(24), 8631–8638. https://doi.org/10.1128/AEM.01531-12

Paria, P., Chakraborty, H. J., Behera, B. K., Das Mohapatra, P. K., & Das, B. K. (2019). Computational characterization and molecular dynamics simulation of the thermostable direct haemolysin-related haemolysin (TRH) amplified from *Vibrio parahaemolyticus*. *Microbial Pathogenesis, 127*, 172–182. https://doi.org/10.1016/j.micpath.2018.11.037

Park, K. S., Ono, T., Rokuda, M., Jang, M.-H., Okada, K., Iida, T., & Honda, T. (2004). Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. *Infection and Immunity, 72*, 6659–6665. https://doi.org/10.1128/IAI.72.11.6659-6665.2004

Park, S. B., & Zhang, Y. (2024). Innovative multiplex PCR assay for detection of *tlh*, *trh*, and *tdh* genes in *Vibrio parahaemolyticus* with reference to the U.S. FDA's Bacteriological Analytical Manual (BAM). *Pathogens, 13(9):*774. https://doi.org/10.3390/pathogens13090774

Pazhani, G. P., Chowhury, G., & Ramamurthy, T. (2021). Adaptations of *Vibrio parahaemolyticus* to stress during environmental survival, host colonization, and infection. *Frontiers in Microbiology, 12*, 737299. https://doi.org/10.3389/fmicb.2021.737299

Phillips, C., Parkinson, A., Namsrai, T., Chalmers, A., Dews, C., Gregory, D., Kelly, E., Lowe, C., & Desborough, J. (2024). Time to diagnosis for a rare disease: Managing medical uncertainty. A qualitative study. *Orphanet J Rare Dis, 19*(1), 297. https://doi.org/10.1186/s13023-024-03319-2

Plaza, N., Pérez-Reytor, D., Corsini, G., García, K., & Urrutia, Í. M. (2024). Contribution of the type III secretion system (T3SS2) of *Vibrio parahaemolyticus* in mitochondrial stress in human intestinal cells. *Microorganisms, 12*(4), 813. https://doi.org/10.3390/microorganisms12040813

Plaza, N., Urrutia, I. M., Garcia, K., Waldor, M. K., & Blondel, C. J. (2021). Identification of a family of *Vibrio* type III secretion system effectors that contain a conserved serine/threonine kinase domain. *mSphere, 6*(4), e0059921. https://doi.org/10.1128/mSphere.00599-21

Pumipuntu, N., & Indrawattana, N. (2017). *Vibrio parahaemolyticus*: A seafood-borne pathogen. *J Trop Med Parasitol, 40*, 50–62.

Qian, H., Li, W., Guo, L., Tan, L., Liu, H., Wang, J., Pan, Y., & Zhao, Y. (2020). Stress response of *Vibrio parahaemolyticus* and *Listeria monocytogenes* biofilms to different modified atmospheres. *Frontiers in Microbiology, 11*. https://doi.org/10.3389/fmicb.2020.00023

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

Raghunath, P. (2015). Roles of thermostable direct haemolysin (TDH) and TDH-related haemolysin (TRH) in *Vibrio parahaemolyticus*. *Front Microbiol, 5*, 805. https://doi.org/10.3389/fmicb.2014.00805

Rahmaniya, N., & Sekharan N, M. (2018). Consumer behaviour towards seafood and seafood safety. *International Journal of Current Advanced Research, 7*. https://doi.org/10.24327/ijcar.2018.8736.1417

Röthig, T., M., S., Trevathan-Tackett, V., Voolstra, C. R., Ross, C., Chaffron, S., Durack, P. J., Warmuth, L. M., & Sweet, M. (2023). Human-induced salinity changes impact marine organisms and ecosystems. *Global Change Biology, 29*(17), 4731–4749. https://doi.org/10.1111/gcb.16859

Santos, L. de O., de Lanna, C. A., Arcanjo, A. C. da C., Bisch, P. M., & von Krüger, W. M. A. (2021). Genotypic diversity and pathogenic potential of clinical and environmental Vibrio parahaemolyticus isolates from Brazil. Frontiers in Microbiology, 12, Article 602653. https://doi.org/10.3389/fmicb.2021.602653

Seelman, S., Whitney, B., Stokes, E. K., Elliot, E. L., Griswold, T., Patel, K., Bloodgood, S., Jones, J. L., Cripe, J., Cornell, J., Luo, Y., Williams, D. L., Boyle, M. M., Cahoon, J., Brennan, C., Wildey, L. M., Grover, V. M., Simonson, S., Crosby, A. J., ... Viazis, S. (2023). An outbreak investigation of *Vibrio parahaemolyticus* infections in the United States linked to crabmeat imported from Venezuela: 2018. *Foodborne Pathogens and Disease, 20*(4), 123–131. https://doi.org/10.1089/fpd.2022.0078

Sheikh, H. I., Najiah, M., Fadhlina, A., Laith, A. A., Nor, M. M., Jalal, K. C. A., & Kasan, N. A. (2022). Temperature upshift mostly but not always enhances the growth of *Vibrio* species: A systematic review. *Frontiers in Marine Science, 9*. https://doi.org/10.3389/fmars.2022.959830

Singh, R., Narayan, V., McLenachan, P., Winkworth, R. C., Mitra, S., Lockhart, P. J., Berry, L., Hatha, A. M., Aalbersberg, W., & Rao, D. (2012). Detection and diversity of pathogenic *Vibrio* from Fiji. *Environ Microbiol Rep, 4*(4), 403–411. https://doi.org/10.1111/j.1758-2229.2012.00344.x

Suneja, M., Beekmann, S. E., Dhaliwal, G., Miller, A. C., & Polgreen, P. M. (2022). Diagnostic delays in infectious diseases. *Diagnosis (Berl), 9*(3), 332–339. https://doi.org/10.1515/dx-2021-0092

Tagliavia, M., Salamone, M., Bennici, C., Quatrini, P., & Cuttitta, A. (2019). A modified culture medium for improved isolation of marine vibrios. *MicrobiologyOpen, 8*(9), e00835. https://doi.org/10.1002/mbo3.835

Tan, C. W., Malcolm, T. T. H., Kuan, C. H., Thung, T. Y., Chang, W. S., Loo, Y. Y., Premarathne, J. M. K. J. K., Ramzi, O. B., Norshafawatie, M. F. S., Yusralimuna, N., Rukayadi, Y., Nakaguchi, Y., Nishibuchi, M., & Radu, S. (2017). Prevalence and antimicrobial susceptibility of *Vibrio parahaemolyticus* isolated from short mackerels (*Rastrelliger brachysoma*) in Malaysia. *Frontiers in Microbiology, 8*. https://doi.org/10.3389/fmicb.2017.01087

Tchelet, D., Keppel, K., Bosis, E., & Salomon, D. (2023). *Vibrio parahaemolyticus* T6SS2 effector repertoires. *Gut Microbes, 15*(1), 2178795. https://doi.org/10.1080/19490976.2023.2178795

Thaotumpitak, V., Odoi, J. O., Anuntawirun, S., & Jeamsripong, S. (2024). Meta-analysis and systematic review of phenotypic and genotypic antimicrobial resistance and virulence factors in *Vibrio parahaemolyticus* isolated from shrimp. Antibiotics (Basel) 13(4):370. https://doi.org/10.3390/antibiotics13040370

Tran, N., Nguyen, P., & Nguyen, H. (2020). Effects of pH, temperature and oxygen-limited condition on the virulence of *Vibrio parahaemolyticus*. *Science and Technology Development Journal, 23*, First. https://doi.org/10.32508/stdj.v23i3.2057

Trinanes, J., & Martinez-Urtaza, J. (2021). Future scenarios of risk of *Vibrio* infections in a warming planet: A global mapping study. *The Lancet Planetary Health, 5*(7), e426–e435. https://doi.org/10.1016/S2542-5196(21)00169-8

Ueno, H., K., T., Kikuchi, K., Kobori, S., & Miyazaki, M. (2016). The first report of *Vibrio parahaemolyticus* strain O10:K60 in Japan, a new combination of O and K serotypes isolated from a patient with gastroenteritis. *Jpn. J. Infect. Dis., 69*, 28–32.

U.S. Food and Drug Administration. (2004). Bacteriological analytical manual: Chapter 9. Vibrio. FDA publication.

Vandeputte, M., Coppens, S., Bossier, P., Vereecke, N., & Vanrompay, D. (2024). Genomic mining of *Vibrio parahaemolyticus* highlights prevalence of antimicrobial resistance genes and new genetic markers associated with AHPND and *tdh*+/ *trh*+ genotypes. *BMC Genomics, 25*(1), 178. https://doi.org/10.1186/s12864-024-10093-9

Velazquez-Roman, J., León-Sicairos, N., de Jesus Hernández-Díaz, L., & Canizalez-Roman, A. (2014). Pandemic *Vibrio parahaemolyticus* O3:K6 on the American continent. *Front Cell Infect Microbiol, 3*, 110. https://doi.org/10.3389/fcimb.2013.00110

Velez, K. E. C., Leighton, R. E., Decho, A. W., Pinckney, J. L., & Norman, R. S. (2023). Modeling pH and temperature effects as climatic hazards in *Vibrio vulnificus* and *Vibrio parahaemolyticus* planktonic growth and biofilm formation. *GeoHealth, 7*(4), e2022GH000769. https://doi.org/10.1029/2022GH000769

Vu, T. T. T., Hoang, T. T. H., Fleischmann, S., Pham, H. N., Lai, T. L. H., Cam, T. T. H., Truong, L. O., Le, V. P., & Alter, T. (2022). Quantification and antimicrobial resistance of *Vibrio parahaemolyticus* in retail seafood in Hanoi, Vietnam. *Journal of Food Protection, 85*(5), 786–791. https://doi.org/10.4315/JFP-21-444

Waddell, B., Southward, C. M., McKenna, N., & DeVinney, R. (2014). Identification of VPA0451 as the specific chaperone for the *Vibrio parahaemolyticus* chromosome I type III-secreted effector VPA0450. *FEMS Microbiology Letters, 353*(2), 141–150. https://doi.org/10.1111/1574-6968.12416

Wang, R., Zhong, Y., Gu, X., Yuan, J., Saeed, A. F., & Wang, S. (2015). The pathogenesis, detection, and prevention of *Vibrio parahaemolyticus*. *Frontiers in Microbiology, 6*. https://doi.org/10.3389/fmicb.2015.00144

Wang, Y., Zhao, Y., Pan, Y., & Liu, H. (2020). Comparison of the growth variability of *Vibrio parahaemolyticus* coupled with strain source and genotype analyses in simulated gastric digestion fluids. *Frontiers in Microbiology, 11*, 212. https://doi.org/10.3389/fmicb.2020.00212

Wang, D., Palmer, J. S., Fletcher, G. C., On, S. L. W., Gagic, D., & Flint, S. H. (2023). Efficacy of commercial peroxyacetic acid on *Vibrio parahaemolyticus* planktonic cells and biofilms on stainless steel and Greenshell™ mussel (*Perna canaliculus*) surfaces. *International Journal of Food Microbiology, 405*, Article 110372. https://doi.org/10.1016/j.ijfoodmicro.2023.110372

Wu, X., Zhu, Y., Yan, W., Zhang, P., & Chen, L. (2023). Pathogenic characteristics of *Vibrio parahaemolyticus* that caused a gastroenteritis outbreak event in Huzhou. *FEMS Microbiology Letters*, Volume 371, 2024, fnad130, https://doi.org/10.1093/femsle/fnad130

Xu, H., Zhang, B., Yu, P., Sun, M., Xie, L., & Chen, L. (2024). Comparative genomic and transcriptomic analyses reveal distinct response strategies to hypoxia by *Vibrio parahaemolyticus* isolates of clinical and aquatic animal origins. *Annals of Microbiology, 74*(1), 20. https://doi.org/10.1186/s13213-024-01769-4

Xu, W., Chen, M., Chen, X., Su, Y., Tang, L., & Zhang, Y. (2024). *Vibrio parahaemolyticus* infection caused by market sewage: A case report and literature review. *Heliyon, 10*(1). https://doi.org/10.1016/j.heliyon.2023.e23461

Yanagihara, I., Nakahira, K., Yamane, T., Kaieda, S., Mayanagi, K., Hamada, D., Fukui, T., Ohnishi, K., Kajiyama, S., Shimizu, T., Sato, M., Ikegami, T., Ikeguchi, M., Honda, T., & Hashimoto, H. (2010). Structure and functional characterization of *Vibrio parahaemolyticus* thermostable direct haemolysin. *Journal of Biological Chemistry, 285*(21), 16267–16274. https://doi.org/10.1074/jbc.M109.074526

Yeung, M., & Thorsen, T. (2016). Development of a more sensitive and specific chromogenic agar medium for the detection of *Vibrio parahaemolyticus* and other *Vibrio* species. J Vis Exp (117):54493, https://doi.org/10.3791/54493

Zhang, L., & Orth, K. (2013). Virulence determinants for *Vibrio parahaemolyticus* infection. *Current Opinion in Microbiology, 16*(1), 70–77. https://doi.org/10.1016/j.mib.2013.02.002

Zhang, Y., Tan, X., Li, M., Liu, P., Jiao, X., & Gu, D. (2023). Transcriptome analysis reveals the effect of low NaCl concentration on osmotic stress and type III secretion system in *Vibrio parahaemolyticus*. *International Journal of Molecular Sciences, 24*, 2621. https://doi.org/10.3390/ijms24032621

Zhong, X., Pan, Z., Mu, Y., Zhu, Y., Zhang, Y., Ma, J., Yang, M., & Yao, H. (2022). Characterization and epidemiological analysis of *Vibrio parahaemolyticus* isolated from different marine products in East China. *International Journal of Food Microbiology, 380*, 109867. https://doi.org/10.1016/j.ijfoodmicro.2022.109867