Evaluation of MDA and ONT Sequencing for Developing a Point of Care Diagnostic for *Mycobacterium*

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

In

Biological Science

At Massey University, Manawatu

New Zealand

Bethany Emma Gannon

2023

To my Mum and Dad – Caroline and Patrick, for always encouraging me to go on every adventure life throws at me – especially this one.
And to my fiancé Matthew for supporting us through this chapter of our lives and for always cheering me on.
It is to them that this thesis is dedicated.

Abstract

Direct sequencing of pathogen DNA from clinical samples can significantly improve the speed of diagnosis, antimicrobial resistance prediction, and outbreak investigation. This approach is especially useful for slow-growing pathogens like *Mycobacterium tuberculosis* (MTB). However, since MTB may represent only 0.01% of the total DNA in a sputum sample, it is crucial to enrich MTB DNA by specific amplification and/or depletion of non-target DNA, including human and/or bacterial DNA. Here, we investigated the potential of selective multiple displacement amplification (MDA) using novel primers that might be used to characterise *Mycobacterium* DNA from young cultures or directly from sputum samples. In an earlier study, selective MDA primers were designed to preferentially bind to MTB DNA and not human DNA or the DNA from bacteria found to be commonly occurring in the upper respiratory tract of MTB patients. As suggested by analyses in this thesis, these primers will also amplify other closely related species of *Mycobacterium.* In the present study, MDA was coupled with Oxford Nanopore (ONT) MinION sequencing to help evaluate the potential MDA-nanopore sequencing as a point-of-care diagnostic. To assess the level of genome coverage of cultured *M. bovis* DNA with different selective MDA primers, both Illumina NovaSeq and Minion MK1C nanopore sequences were obtained. Selective primers that produced optimal coverage antimicrobial resistance genes in *M. bovis* were identified. It was found that mismatch priming played a significant role in the success of the MDA reactions. The implication of this for potential diagnostic applications has been discussed.

Acknowledgements

It is an absolute pleasure to thank everyone who helped make my master's study and this thesis possible.

First and foremost, I would like to acknowledge the support and wisdom of my supervisor, Professor Peter Lockhart, for his invaluable guidance and unwavering support throughout this thesis. His expertise, mentorship and belief in my abilities have influenced my academic journey. Thank you for being an expectational exceptional mentor. Secondly, I would like to acknowledge Trish McLenachan for assisting with countless hours of lab work, patiently guiding me through lab procedures, and sharing her knowledge and expertise. She has been instrumental in conducting the experiments that made this work possible.

Lastly, and most importantly, I would like to thank the following people for their support and generosity: I would like to give a massive thank you to my fiancé, Matthew Redmayne; my parents and brother, Caroline, Patrick, and Kieran Gannon; my loving and supportive friends Nathan and Rachel Muckley, Barbara Redmayne, and my (almost) mother, father, and sister-in-law Nives Botica Redmayne, John Redmayne and Ivanna Redmayne.

To Matthew, I would like to express my deepest appreciation and gratitude for your unwavering support throughout my master’s journey. Your support, patience, and love have guided me throughout this academic journey. Your belief in my potential and tireless encouragement pushed me forward, even during the most challenging moments. You were always there to celebrate my triumphs and offer a steady hand during times of stress. Thank you for supporting us while I completed my studies.
Furthermore, to my parents, Caroline and Patrick - thank you for cheering me on, even from thousands of miles away. Thank you all for your constant love, support, and guidance and for pushing me to pursue my loves and passions always. Words cannot begin to describe what that means to me.

I would also like to acknowledge Taylor Swift for solely providing the soundtrack (Taylor's Version, of course) that this thesis was written.

“Not a single one of us here today has done it alone. We are each a patchwork quilt of those who have loved us. Those who have believed in our future".” - Dr.h.c Taylor Swift.

Table of Contents

Abbreviations

Abbreviation	Meaning
°C	Degrees Celsius
ul	Microlitre(s)
bp	Base pair(s)
AMR	Antimicrobial Resistance
DNA	Deoxyribonucleic Acid
DST	Drug susceptibility testing
DTT	Dithiothreitol
IDT	Integrated DNA Technologies
IPP	Isopentenyl pyrophosphate
LAMP	Loop-mediated isothermal amplification
MAC	*Mycobacterium avium* complex
M. avium	*Mycobacterium avium*
M. bovis	*Mycobacterium bovis*
M. canettii	*Mycobacterium canettii*
MDA	Multiple Displacement Amplification
MIC	Minimum Inhibitory Concentration
M. intracellulare	*Mycobacterium intracellulare*
M. microti	*Mycobacterium microti*
MTBC	*Mycobacterium tuberculosis* complex
MDR-TB	Multi-drug resistant tuberculosis
M. tuberculosis	*Mycobacterium tuberculosis*
MTB	*Mycobacterium tuberculosis*
ng	Nanogram
ONT	Oxford Nanopore Technologies

pmol	Picomoles per litre
RAP	Receptor Associated Protein
sMDA	Selective Multiple Displacement Amplification
SWGA	Selective Whole Genome Amplification
TB	Tuberculosis
Tm	Melting Temperature
WGS	Whole Genome Sequencing
WHO	World Health Organization
XDR-TB	Extensively Drug-Resistant Tuberculosis

List of Figures

List of Tables

"Never be so kind, you forget to be clever. Never be so clever, you forget to be kind."
- 'Marjorie' by Taylor Swift from 'Evermore'

Chapter 1: Introduction

1.1 Mycobacterium

Mycobacteria is a genus comprising over 190 species within the phylum Actinomycetota, with its own family Mycobacteriaceae. Mycobacterial species are usually aerobic and non-motile, capable of growth with minimal nutrients. They are commonly found in varying environments, including soil, water, and animal and human hosts. Within this genus are pathogenic bacteria responsible for tuberculosis diseases. These include *Mycobacterium tuberculosis* (MTB: human tuberculosis) and *Mycobacterium bovis* (M. bovis: bovine tuberculosis). There are also *Mycobacteria* that cause non-tuberculosis disease. These include *Mycobacterium leprae* (leprosy) and *Mycobacterium avium* complex (MAC infection). A feature of the genus is that the genomes of *Mycobacteria* exhibit a high degree of plasticity that has accompanied the evolution of diverse lineages and subspecies that have distinct characteristics and potential pathogenicity (Niemann & Supply., 2014; Sanoussi *et al*., 2021).

1.2 The Problem of Antimicrobial Resistance in *Mycobacterium tuberculosis*

The incidence of both tuberculosis and non-tuberculosis disease is rising globally (World Health Organization, 2021; Ratnatunga *et al*., 2020), and the burden of disease is predicted to worsen. This is attributed to the development of increasing levels of resistance against conventional antibiotics in both tuberculosis and non-tuberculosis-causing *Mycobacteria*.

MTB is a significant global public health issue, particularly in low-income countries (Siddiqi *et al*., 2003; World Health Organization, 2021). Despite the early successes of treatment in controlling MTB, it remains one of the leading causes of death from a single infectious agent worldwide, second to SARS-CoV-2 (COVID-19) and above HIV/AIDS (World Health Organization, 2021). It is currently the 13th overall leading cause of death worldwide.

In 1993, the World Health Organization (WHO) classified TB as a global health emergency. At that time, there were approximately seven million cases and 1.3 million deaths per year worldwide. In 2020, 1.5 million people died from TB, with over 95% of TB-related deaths occurring in low-income countries. Despite the increased prevalence of TB, global TB programs estimate that three million cases are unaccounted for due to underdiagnosis and

underreporting of international TB programs (World Health Organization, 2021). While TB is treatable using current anti-TB drugs, the regimen prescribed to patients is complex and lengthy, which has led to poor adherence by patients to treatment regimens (Fox *et al*., 2023). This has promoted the selection of MTB strains resistant to one or more of the available first-line drugs, compounding the already challenging global control of TB.

The emergence of drug-resistant Mycobacterium strains can reverse progress to reduce morbidity and mortality due to these infectious diseases, especially tuberculosis. Of particular concern are multidrug-resistant strains resistant to isoniazid and rifampicin, the two most effective and commonly used anti-tuberculosis drugs. The emergence of extensively drug-resistant TB (XDR-TB), a form of multidrug-resistant TB with additional resistance to even more anti-TB medicines and less responsive to fewer medicines, has raised significant concern.

The WHO "End TB Strategy" aims to reduce the number of TB-related deaths to meet the sustainable development goal of ending TB by 2030 (World Health Organization, 2015). In 2020, global reports indicated an estimated 10 million new cases of TB. Of these, an estimated 6% were multi-drug resistant strains, with resistant strains primarily presenting in patients with previous TB infections (CDC, 2021). Since 2018, the WHO has reported a 10% increase in MDR-TB strains. The identification and treatment management of MDR-TB is associated with additional resistance-related problems (Jang & Chung, 2020). The advance of MDR-TB poses a significant threat to controlling and combating the WHO End TB Strategy goal of eliminating TB by 2030 (defined as less than one case per million worldwide).

The emergence of drug-resistant TB and non-tuberculosis *Mycobacteria* has the potential to reverse progress made in the last few decades to reduce the morbidity and mortality of these diseases. The emergence and spread of drug-resistant MTB is a complex process involving genetic, environmental, and social factors. It is critical to address these factors to prevent the further development and spread of drug-resistant TB strains and other infectious *Mycobacterium* species.

1.3 Point-of-care diagnostics are needed for rapid diagnosis and informed decision-making

Due to limited drug options, treating drug-resistant *Mycobacterium* cases is quite challenging. Existing diagnostics are inadequate due to slow turnaround time and a lack of drug susceptibility testing (DST). Prolonged treatment regimens using the same few drugs have resulted in poor patient compliance due to unpleasant side effects. As a result, *Mycobacterium* strains that are becoming increasingly resistant to the available first-line anti-TB drugs have emerged.

Managing drug-resistant tuberculosis and non-tuberculosis mycobacterial infections in clinics can be very difficult. The drugs available for treatment cause many side effects that patients find hard to tolerate, such as hepatitis, diarrhoea, vomiting, nephrotoxicity, ototoxicity, and psychological disturbances (Sturdy *et al*., 2011). By improving the capability to conduct and quickly report drug susceptibility testing, we can improve patient outcomes and prevent the misuse of antibiotics.

There is an urgent need for an affordable and durable diagnostic tool that can be used in laboratories with limited resources to effectively and efficiently identify MDR-TB and XDR-TB infections. The high costs of current sample processing methods make it difficult for these labs to identify MDR-TB accurately. Rapid diagnostic testing for MTB is necessary to diagnose the diseases and test for antibiotic resistance profiles quickly. This is crucial for prescribing appropriate treatments, preventing unnecessary use of antibiotics, and ultimately helping prevent further drug resistance. Whole genome sequencing (WGS) can be a practical tool in clinical settings for identifying genotypic mutations that predict antibiotic resistance profiles. A study by Witney *et al*., 2015 showed this and found that in several cases of TB, the speed of generation and the comprehensive nature of WGS data helped inform decisions in clinical settings when focused on a group of extensively drug-resistant TB (XDR-TB) patients. This finding was significant. Making informed decisions when prescribing antibiotics for TB and non-tuberculosis *Mycobacterium* infections is imperative, and WGS provides comprehensive information on the infecting agent. The rapid prediction and detection of drug resistance is important for detecting appropriate antibiotic treatments, which can significantly increase cure rates (Zhao *et al*., 2022). This information can guarantee the most efficacious treatment, diminishes the possibility of antibiotic resistance, and encourages

treatment compliance. Alternative methods of diagnosis have several shortcomings, as described below.

1.4 Limitations of Current Diagnostics

Correct diagnosis and antimicrobial detection are essential for appropriately treating tuberculosis and non-tuberculosis infections. Current methods are time-consuming, can increase the risk of poor clinical outcomes, and fail to identify and control transmission (Javaid *et al*., 2018; Ji *et al*., 2021). A widely used diagnostic test for TB is the microscopic examination of sputum, which looks for acid-fast bacilli. This takes no more than an hour to complete, but this method is costly, lacks both specificity and sensitivity and cannot test for antimicrobial resistance. A positive result produced by this test cannot differentiate between different *Mycobacterium* species (Dezemon *et al*., 2014). Another widely used test, and one that is considered the gold standard in suspected pulmonary TB cases, is the Lowenstein-Jensen culture. This is more sensitive than smear microscopy but is much more time-consuming, taking from 4 to 9 weeks in solid media culture. On top of its lengthy timeframe, this method also requires well-trained laboratory staff and adapted infrastructure. (N'Guessan *et al.,* 2017). This delay in diagnosis can impact effective medical interventions, and therefore, the need for new, rapid, and accurate diagnostic methods has emerged.

PCR tests for TB have a high sensitivity and specificity, giving results within 24 hours. PCR can be performed on many specimen types, such as sputum, bronchial fluid, and cerebrospinal fluid. PCR methods can also detect TB bacteria at low concentrations, as well as detect specific mutations associated with drug resistance. PCR-based methods for detecting drug resistance are typically used to complement traditional drug susceptibility testing (DST) methods, such as culture-based ones. PCR tests can provide rapid results and detect specific drug resistance mutations directly from clinical samples, but they may only cover part of the spectrum of drug resistance mutations. The sensitivity of PCR tests can vary between 77% and 95%, while their specificity is greater than 95% for smear-positive specimens. However, for patients with negative smear tests, the sensitivity of PCR tests has been reported to be lower than 90% (Kivihya-Ndgugga *et al*., 2004). The main disadvantages of this diagnostic method are cost and laboratory personnel training. PCR requires a costly investment in specialised equipment, and the protocols for preparing samples for analysis are

complex. The Xpert MTB/ RIF assay is a PCR test with an incredibly fast and highly efficient method for detecting tuberculosis. Within just two hours from patient sample collection, , it can detect the presence of microbes from the *Mycobacterium tuberculosis* complex and resistance to rifampin. This groundbreaking diagnostic tool received approval from the World Health Organization in 2010, and for good reason. It is incredibly user-friendly, requiring minimal technical skills due to its full integration, self-contained design, and automation. A meta-analysis by Zhang *et al*., 2020 found that for detecting rifampin resistance, Xpert has a sensitivity of 95% and specificity of 98%, making it a highly accurate diagnostic tool for detecting resistance to this first-line drug. The high cost of this test is a considerable hurdle for widespread implementation, especially for low and middle-income countries that bear most of the global burden of tuberculosis. This means these countries still rely on slow and outdated TB tests, such as sputum smear microscopy, which lack sensitivity and specificity. Early opportunities to intervene in the progress of the disease are therefore lost in these cases.

1.5 Minion Sequencing Protocols and Rapid Barcoding Protocol; Poor Efficiency Due to Low Copy Numbers and the Need to Make MinION Sequencing More Effective

Nanopore sequencing is a unique and scalable technology that utilizes real-time sequence analysis of long or short DNA and RNA fragments. This technology monitors electrical currents as nucleic acids pass through a protein nanopore located on a Flow cell or Flongle. The resulting signal is decoded to give the nucleotide sequence of DNA or RNA fragment. The Oxford Nanopore MinION MK1C device is a small standalone sequencing device that allows sequencing without the need for expensive laboratory infrastructure. The Rapid Barcoding transposase kits are also relatively simple to use. MinION data can be analysed on a laptop computer; even an electrical supply is not required. These features of the MinION present sequencing analysis opportunities in remote areas, such as in low-income countries, where a high initial investment in equipment is not affordable.

Adopting ONT sequencing for the diagnosis of MTB strains offers several other advantages for DNA sequencing, including increased read lengths for mapping and improved de novo assemblies. The sequencing accuracy of reads obtained from this technology is constantly improving, with raw read accuracy at 99.92% when using the most recent version 14 chemistry, R10.4.1 flow cells and MinKNOW, or 99.6% when using the latest flow cells and

Guppy software (Oxford Nanopore Technologies, 2023). However, the MK1C device does not have accurate base calling available. A Graphics Processing Unit (GPU) is required to run the most accurate, real-time base calling with the MK1C device.

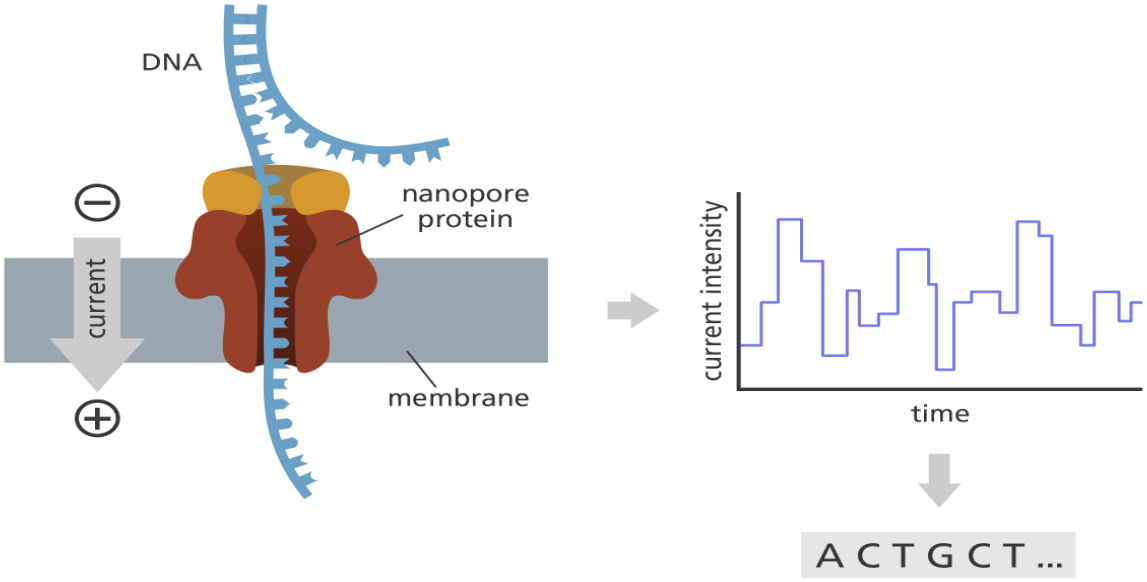


Figure 1.1: Mechanism of Nanopore Sequencing (Genome Research Limited, 2021)

The MinION has been demonstrated to be somewhat comparable to the Illumina MiniSeq process in terms of accuracy – Illumina has an accuracy of 99%, whereas the MinION has an accuracy of 95% (Stevens *et al*., 2023), which is increasing as new chemistries and flow cells are developed and released (Wang *et al*., 2021; Ni *et al*., 2023). An advantage of the MinION is that it is very portable and low-cost in comparison to the Illumina MiniSeq, and genome assembly is facilitated by having longer reads (Wasswa *et al*., 2022).

The ONT MinION device has the potential to revolutionize the diagnosis of TB and improve patient care (Liu *et al*., 2023). One of the significant advantages of MinION is its ability to rapidly sequence the entire genome of MTB in a matter of hours at the point of care from young cultures and sputum samples.

The Rapid Barcoding protocol that ONT offers drastically reduces the hands-on time needed to make a library when compared to other next-generation sequencing platforms (King *et al*., 2020). This kit also reduces the amount of labour required and consumable costs required to run the device (Freed *et al*., 2020; King *et al*., 2020).

These features make the MinION potentially suitable for use in clinical mycobacteriology and clinical microbiology laboratories, where this platform has already been used to identify bacterial MDR pathogens (van Belkum *et al*., 2020). Further implementation in low-resource settings would allow for a more accurate and timely diagnosis of TB and help to identify emerging drug-resistant strains. It could be used to track the spread of TB, monitor the effectiveness of treatment, and identify mutations that could lead to drug resistance. Faster diagnostic turnaround times have the potential to help clinicians make more informed decisions about patient care and improve outcomes for those affected by TB.

Utilising the real-time base calling feature on the ONT MinION device allows real-time analysis, which allows a faster turnaround of results. This is crucial in a clinical diagnostic setting. The Rapid Barcoding kit could be especially useful in the field or remote areas where access to laboratory infrastructure is limited.

Despite this potential, whole genome sequencing of pathogens directly from clinical samples is technically challenging, as samples can vary in volume, numbers of human and bacterial cells, and the concentration of the target organism within the sample. TB detection from patient samples also faces this challenge since *Mycobacterium* DNA can represent as little as 0.01% of the total DNA within a sputum sample (Votintseva *et al*., 2017). Sputum is a standard sample type used when diagnosing TB, but the presence of other organisms and host DNA in sputum can impact methods of MTB detection. Enriching the target DNA present in samples is a crucial consideration in clinical settings and when considering diagnostic applications of WGS.

The quality of the DNA template is another consideration for successful WGS, and MTB, in this respect, can cause technical challenges. TB bacilli are scarce in clinical samples, limiting the availability of any MTB genomic DNA. The MTB cell wall is also remarkably tough and lipid-rich, making it difficult to disrupt the TB bacterium and impacting the quality and yield of genomic DNA. The genome of MTB is also quite durable. It has an average guanine/cytosine content of 65% across its genome, with some regions exceeding 80% (Cole *et al*., 1998), making this genome more thermally stable. As a result, careful consideration should be given to DNA extraction methods, library preparation, and sequencing platforms when considering TB diagnosis through NGS entering clinical practice. Currently, WGS is

performed following the isolation and culture of the bacteria. However, as previously mentioned, this can take weeks, and it would be advantageous to sequence directly from a clinical sputum sample (Doyle *et al*., 2018).

Sputum is a complex substance that contains mucus, human cells, various bacteria and viruses, cell debris and sometimes blood and pus. The quality of the sequence data obtained – coverage and depths of reads – depends on the purity and integrity of the DNA submitted. In order to process the sputum into a solution that can be directly sequenced and still have high-quality DNA, several approaches are possible, such as heating or chemical treatments. All of these approaches risk damaging the bacterial cell wall and exposing genomic DNA to extreme conditions, as well as the release of genomic DNA, which could be lost in subsequent steps that involve washing.

1.6 Selective MDA: Novel Primers Developed at Massey to Amplify MDA and Select Against Commonly Occurring Respiratory Bacteria in the Sputum of MTB Patients

Multiple displacement amplification (MDA) is a method that rapidly amplifies minute amounts of DNA within samples to a reasonable quantity for genomic analysis. Researchers have typically used the Φ29 DNA polymerase for this whole genome amplification reaction with small amounts of DNA template, ranging from 0.3ng up to 300ng DNA template in some cases (Dean *et al*., 2002). This enzyme catalyzes both rolling circle and multiple displacement amplification naturally in several microbial species, including members of both Bacteria and Archaea (del Solar *et al*., 1998). Φ29 DNA polymerase produces an accurate template for DNA sequencing as it has a 3’ to 5’ exonuclease activity(Garmendia *et al*., 1992).

MDA is an exponential process, which has made it a popular technique for amplifying circular and linear DNA molecules (Dean *et al*., 2002), and its potential for selective amplification of the MTB genome has been investigated (Clarke *et al*., 2017). Long DNA fragments produced by MDA primers and (typically) Φ29 DNA polymerase facilitate coverage of the genome, and if priming sites are present in sufficient density across the genome, they can produce consistent and accurate locus representation (Clarke *et al*., 2017).

Figure 1.2: Process of Multiple Displacement Amplification (Long et al., 2020)

Enrichment of MTB DNA by selective amplification and depletion of non-target DNA, including human and upper respiratory tract microbial DNA, is essential for the success of *Mycobacterium* detection in sputum samples, as Mycobacteria DNA can represent as little as 0.01% of the total amount of DNA extracted from sputum samples (Votintseva *et al*., 2017). Selective MDA has been seen as one possible means for increasing the sensitivity of mycobacterial detection and identification from low-biomass samples (Clarke *et al*. 2017). Selective multiple displacement amplification (sMDA) involves using specific primers to selectively amplify genomes, which can be sequenced with either next-generation or 3rd generation sequencing technologies.

sMDA on MTB has been attempted by Clarke *et al*., (2017) with limited success. Here, a program called SWGA (Selective Whole Genome Amplification) was developed to identify primers that could detect the presence of MTB within blood samples. The binding evenness and selectivity for the target genome varied, and this was found to impair the efficacy of

whole genome amplification. These primers were developed for use with the enzyme Φ29 DNA polymerase, which has an optimal temperature of 30ºC. This temperature optimum reduces the length of primers that can be used for sMDA, and thus their selectivity.

Novel primers for MDA of *Mycobacterium* genomes have also been developed at Massey University (PCT WO2020/095252 Rapid identification of bacterial pathogens 08 November 2019) and designed to amplify MTB and select against commonly occurring bacteria present in the sputum of MTB patients. These primers are able to utilise the Thermo Scientific™ EquiPhi29™ DNA Polymerase, which has an optimal temperature of 37-42ºC. These primers were identified based on analyses of the genomes of bacteria found to be occurring in a microbiome study conducted by Htun *et al*. 2018. In this study, the researchers examined the microbiomes of the upper respiratory tract of patients diagnosed with TB, as well as HIV and TB in Myanmar – a country with a high burden of TB. The microbiomes were compared to healthy individuals from within the same population.

Mycobacterium tuberculosis is a very slow-growing organism in culture. The use of MDA with selective primers to amplify young cultures would aid in the characterisation of cultured multi-drug resistant tuberculosis strains. Such primers also have potential application for same-day diagnosis of *Mycobacterium tuberculosis* strains from DNA-extracted sputum samples. This successful application would contribute to finding a real-world solution to the problem of MTB since rapid and accurate diagnosis is essential for choosing the most appropriate antibiotics for treatment. Undiagnosed cases of multi-drug resistance in MTB are currently contributing to poor patient outcomes and the emergence of drug resistance (Aung *et al*., 2021).

1.7 Studies Using *Mycobacterium bovis* as a Proxy for *Mycobacterium tuberculosis*

Mycobacterium bovis is part of the *Mycobacterium tuberculosis* complex (MTBC) and as such, shares a considerable amount of genetic similarity with MTB. *M. bovis* and *M. tuberculosis* show 99.95% identity at the nucleotide level (Garnier *et al*., 2003), and *M. bovis* has the ability to cause TB in both humans and animal hosts, just as MTB can. All of this makes *M.bovis* a suitable proxy to use in experimental circumstances due to the genetic similarity of the two species.

Safety is also a consideration when using *M.bovis* as a proxy. MTB research requires high-level biosafety equipment due to its pathogenicity. While isolated *M. bovis* DNA was used here, using *M. bovis* as a proxy for MTB is much safer to work with in laboratory conditions as *M. bovis* is less pathogenic to humans, though appropriate precautions are still necessary.

1.8 Aims

The main aim of this thesis has been to evaluate the potential of the novel selective MDA primers developed at Massey University for MDA-Nanopore sequencing as a point-of-care diagnostic. This evaluation involved characterising the distribution of MDA priming sites on *Mycobacterium* genomes and sequencing *M. bovis* template amplified using different MDA primer sets. This was done for cultured *M. bovis,* and cultured M. *bovis* spiked into DNA from human blood and sputum samples. *M. bovis* was used, as it shows 99% similarity to MTB (Brosch *et al*., 2002) and is more readily and easily available in New Zealand. An overview of specific aims relevant for evaluating the potential of the novel MDA primers are described in section 1.8.1 – 1.8.3:

1.8.1 Amplification of Cultured *M. bovis* Template with "High Binder" Primers

The coverage of genome and antibiotic resistance loci in MTB produced by using the Massey "high binder" primers was assessed using both Illumina and ONT MinION (Rapid Barcoding SQK-RBK004) sequencing. The Rapid Barcoding protocol was used as it is perhaps the ONT protocol most suitable as a point-of-care diagnostic for a low-income resource setting.

1.8.2 Amplification of Cultured *M. bovis* Template with "High Binder" Primers and Additional Complementary Primer

Additional primers were included in the MDA reaction to test whether this improved the sequencing coverage of the *M. bovis* genome and antimicrobial resistance (AMR) loci. This analysis was done to identify an optimal primer set for MTB amplification. All primers used have previously been analysed for their selectivity of *Mycobacterium* (PCT WO2020/095252 Rapid identification of bacterial pathogens 08 November 2019).

1.8.3 Amplification of M.bovis Template When Spiked into Sputum DNA

Coverage of the *M. bovis* genome and antibiotic gene loci was examined using the optimal priming set when DNA from sputum was mixed with the *M. bovis* template. The ratio of *M. bovis* DNA to sputum DNA ranged from 1 (5ng):1 (5ng) to 0.1 (0.5ng): 1 (5ng).

Chapter 2: Materials and Methods

2.1 Samples for Analysis/ Primer Design

2.1.1 Primer Design for Selective MDA

In previous work, a set of 15 primers were designed at Massey using an in-house Python script. In filed Massey patent documents (PCT WO2020/095252 Rapid identification of bacterial pathogens 9 Nov 2019; Table 2.1-2.1, Figure 2.1 below), these are named P1-P15.

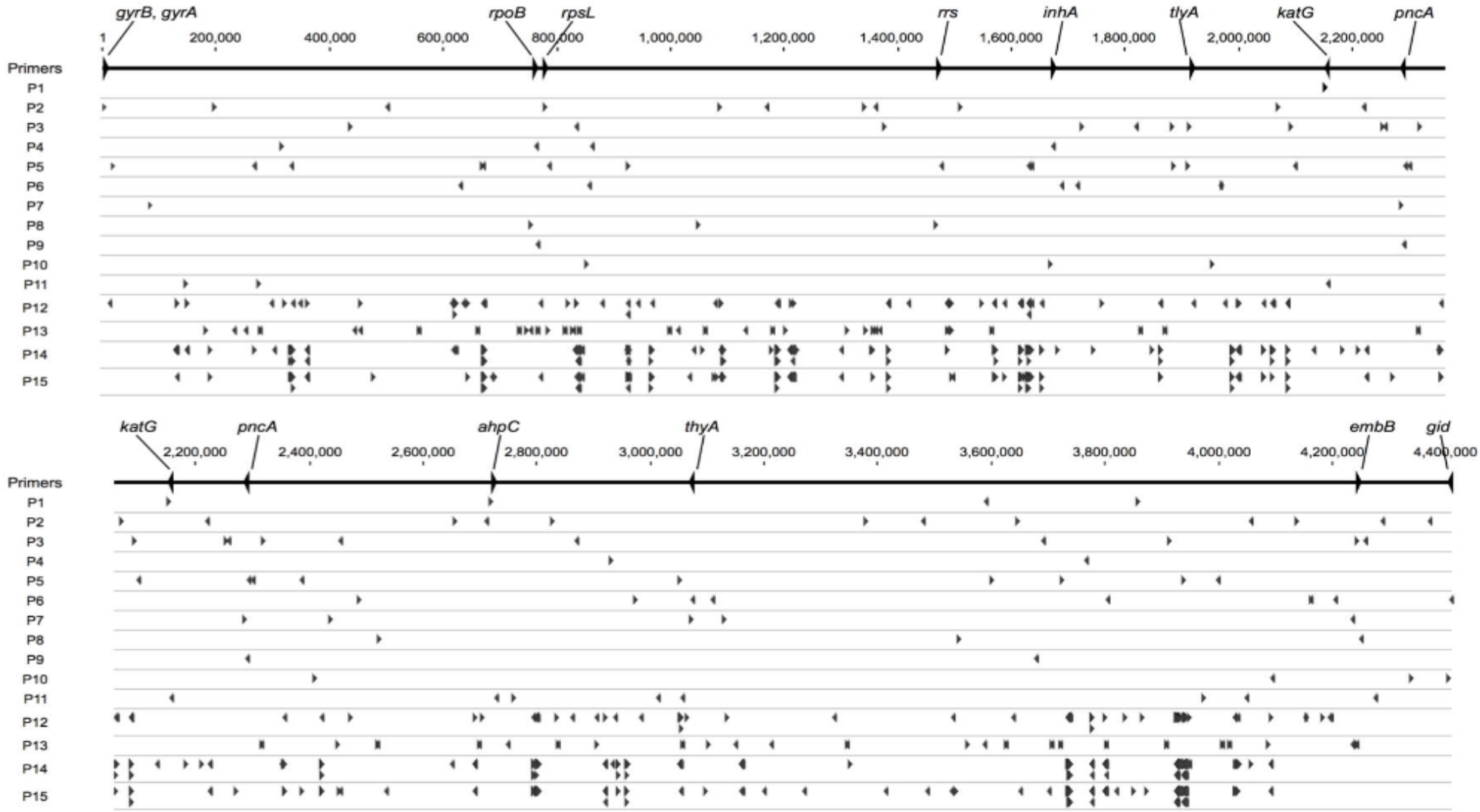
P1-P11 target specific Antimicrobial resistance (AMR) loci while primers P12, P13, P14 and P15 occur with high frequency across the MTB genome.

Table 2.1: Number of binding sites for the 15 MTB oligonucleotide primers on the MTB H37rv genome and on the genomes of other bacteria in the upper respiratory tract of humans

	Taxa[a]															
Primer name	*Mycobacterium tuberculosis*	*Haemophilus influenzae*	*Chlamydophila pneumoniae*	*Pseudomonas aeruginosa*	*Escherichia coli*	*Bordetella pertussis*	*Neisseria meningitidis*	*Listeria monocytogenes*	*Lactobacillus brevis*	*Leuconostoc mesenteroides*	*Clostridioides difficile*	*Porphyromonas gingivalis*	*Veillonella parvula*	*Moraxella catarrhalis*	*Enterobacter aerogenes*	*Staphylococcus aureus*
P1	4	0	0	0	1	2	0	0	1	0	0	0	1	0	3	0
P2	21	0	0	4	2	1	1	0	0	0	0	0	0	0	0	0
P3	17	0	0	5	0	1	1	0	1	0	0	1	0	0	1	0
P4	6	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0
P5	21	0	0	4	4	4	0	0	0	0	0	0	0	0	0	0
P6	15	0	0	3	0	0	5	0	0	0	0	0	0	0	2	0
P7	6	0	0	1	0	2	0	0	0	0	0	1	0	0	1	0
P8	6	0	0	1	0	4	0	0	0	0	0	0	0	0	2	0
P9	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P10	7	0	0	1	0	1	0	0	0	1	0	0	0	0	2	0
P11	10	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0
P12	179	0	0	47	2	30	12	0	1	0	0	0	0	0	24	0
P13	87	0	0	7	1	3	0	0	0	0	0	1	0	0	1	0
P14	567	1	0	42	27	16	48	1	4	0	0	0	0	3	28	0
P15	475	2	0	43	25	24	46	0	5	1	0	1	0	2	25	0

[a] GenBank accession numbers for genomes included in comparison. *Mycobacterium tuberculosis* (NC_000962), *Haemophilus influenzae* (NC_000907), *Chlamydophila pneumoniae* (NC_000922), *Pseudomonas aeruginosa* (NC_002516), *Escherichia coli* (NC_002695), *Bordetella pertussis* (NC_002929), *Neisseria meningitidis* (NC_003112), *Listeria monocytogenes* (NC_003210), *Lactobacillus brevis* (NC_008497), *Leuconostoc mesenteroides* (NC_008531), *Clostridioides difficile* (NC_009089), *Porphyromonas gingivalis* (NC_010729), Veillonellaparvula (NC_013520), *Moraxella catarrhalis* (NC_014147), *Enterobacter aerogenes* (NC_015663), *Staphylococcus aureus* (NZ_CP010295).

Figure 2.1: Simplified map of the MTB H37Rv reference genome with labels of thirteen gene loci commonly associated with antibiotic resistance



The sequence of the 15 Massey primers is shown in Table 2.2

Table 2.2: Sequence of Massey Primers

Primer Name	Nucleic Acid Sequence' (5' – 3')
P1	AATGGCCGTCGC
P2	GGTCGGTGCGGG
P3	TGGCCGGGGTGT
P4	GCAACACCGGGT
P5	GCGGGCACGGTG
P6	CGTCGGCTGCGG
P7	CCACCCGCGCAA
P8	GACGCGCCCACG
P9	TCGCTACCCACG
P10	ATGTTGGTGATC
P11	GGTGTCGACGAG
P12	CGGCGACGGCGG
P13	TGCGTCTGCTCG
P14	CCGCCGTTGCC
P15	CCGTTGCCGCC

As indicated in Table 2.1, the 15 primers shown in Table 2.2 are expected to exhibit greater specificity for the MTB genome over the genomes of commonly occurring bacteria in the upper respiratory tract of patients (Htun *et al*., 2018). MDA initiation sites had previously been identified in the *Mycobacterium tuberculosis* H37Rv genome, and these are shown in Figure 2.1. Potential MDA initiation sites in the genomes of other bacteria belonging to the MAC complex have been similarly calculated using Geneious Prime and these have been shown in Appendix I.

Figure 2.1 shows a simplified map of the MTB H37Rv reference genome with labels of thirteen gene loci commonly associated with antibiotic resistance. Binding sites for each of the 15 Massey MDA primers are indicated in the genome map.

To evaluate the Massey primers for an MDA nanopore diagnostic, empirical results of genome coverage were obtained for MDA reactions with different high binder primer combinations. Consideration was then given to the distribution of binding sites with less frequently occurring primers so that the primer sets evaluated would complement each other to achieve maximum coverage. The high binders (P12-P15) had a melting temperature (T_m) of 40-46^{o}C. Table 2.3 indicates individual T_m values, the total number of binding sites for each primer, and the combination thereof on the MTB H37Rv reference genome. Since the primers may bind somewhat non-specifically at other sites, "mismatch priming" may also contribute to the success of the MDA reactions. The number of binding sites where there is one mismatch has also been shown.

Table 2.3: Primer initiation sites and their estimated Tm

Primers	**No. of binding sites (no mismatches)**	**No. of binding sites (1 mismatch)**	**Tm**
P12	179	2020	46
P13	87	120	40
P14	567	1992	40
P15	475	1841	40
P12 + P13	266	2140	
P12 + P14	766	4012	
P12 + P15	656	3861	
P13 + P14	664	2112	
P13 + P15	562	1961	
P14 + P15	1042	3833	
P12 + P13 + P14	833	4132	
P12 + P13 + P15	741	3981	
P12 + P14 + P15	1241	5853	
P13 + P14 + P15	1149	3953	
P12 + P13 + P14 + P15	1308	5973	

Heterodimer formation between primer pairs was previously determined using the OligoAnalyser tool on the Integrated DNA Technologies (IDT) website

(https://sg.idtdna.com/pages/tools/oligoanalyzer). Table 2.4 shows the delta G calculated for each primer combination. A value less than -9 is considered to be problematic.

Table 2.4: Heterodimers of Massey Primers

	12	13	14	15
12	-3	-6.13	-17.84	-12.89
13		-3	-6.75	-6.75
14			-3	-3
15				-3

Primers were ordered from IDT with two phosphorothioated 3' bases. This modification is recommended for MDA primers (Skerra *et al*., 1992; Dean *et al*., 2002) and increases the stabilisation of the oligonucleotide backbone, making it more resistant to nuclease degradation.

2.1.2 Template DNA AgR4

DNA from *Mycobacterium bovis* was used as a template. *Mycobacterium bovis* is very similar to MTB (>99% similarity (Brosch *et al*., 2002)). DNA was kindly provided by Dr. Marian Price-Carter from the Vet School at Massey University. Sample AgR4 was used, and the concentration of this sample was 20ng/ul as determined on a Qubit 2® Fluorometer (Invitrogen). The DNA was also run on an agarose gel to check the size and integrity; it appeared as a tight, high molecular weight band, indicating that most fragments were >12kb in length and that there was minimal degradation. The DNA was diluted to 5ng/ul, and 2ul (10ng) was used as a template for MDA.

2.2 Selective MDA

For all MDA reactions, the enzyme used was the EquiPhi29^{TM} DNA polymerase from ThermoFisher Scientific. This polymerase is a mutant version of Φ29 with an increased incubation temperature of 42-45ºC (compared to 30-37 ºC) and a shorter incubation time of only 3 hours (compared to 16 hours). To enhance the reaction, inorganic pyrophosphatase (IPP) from New England BioLab was added to the mix as recommended by Dean et. al., 2002

and in the notes for the *Protocol for DNA amplification* provided by ThermoFisher with the enzyme. Dithiothreitol (DTT) solution, provided with the enzyme, was aliquoted into single use amounts (20ul) and frozen, to prevent repeated cycles of freezing and thawing.

To prepare MDA reactions, we followed the recommended protocol, which involved combining template DNA (10ng) and primers (62.5 pmol, final concentration in the MDA reaction is 3.125uM) with 1X reaction buffer in a final volume of 5ul. This mixture was denatured at 95ºC for 3 minutes and then immediately placed on ice for 3 minutes. Afterwards, DTT (1mM), dNTPs (1mM each dNTP), IPP (0.1U), and 10U of EquiPhi29™ in 1X Buffer were added to the solution. The reaction was incubated at 42ºC for 3 hours and then subjected to a heat kill at 65ºC for 10 minutes.

The concentration of DNA post-MDA reactions were determined using a Qubit 2 fluorometer from ThermoFisher Scientific and the DNA Broad Range Solution. Following that, primers, buffers and enzymes were removed from the reaction using a Genomic DNA Clean and Concentrator column from Zymo research. The DNA was eluted in 15ul of water, and the DNA concentration was measured again. The clean-up step was found to be important for achieving successful sequencing reactions.

The MDA template was found to be suitable for both Illumina and ONT sequencing, and results have been presented for both.

2.3 Illumina Sequencing by Custom Science

DNA extracts were submitted to Custom Science (https://shop.customscience.co.nz/) for sequencing on an Illumina HiSeq instrument using a 250bp paired-end read protocol.

The library preparations used for next-generation Illumina sequencing were made according to the instructions provided by the manufacturer (NEBNext Ultra™ DNA Library Prep Kit for Illumina). For each sample, 1 µg genomic DNA was randomly fragmented to <500 bp by sonication (Covaria S220). The fragments were treated with End Prep Enzyme Mix for end repairing, 5’ Phosphorylation and dA-tailing in one reaction, followed by a T-A ligation to add adaptors to both ends. Size selection of Adaptor-ligated DNA was then performed using

AxyPrep Mag PCR Clean-up (Axygen), and fragments of ~410 bp (approximate inset size of 350 bp) were recovered. Each sample was then amplified by PCR for eight cycles using P5 and P7 primers, with both primers carrying sequences which can anneal with flowcell to perform bridge PCR and P7 primer carrying a six-base index, allowing for multiplexing. The PCR products were cleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit2.0 Fluorometer (Invitrogen).

The libraries with various indexes were combined and loaded onto an Illumina HiSeq machine as per the manufacturer's instructions. (Illumina, USA). Sequencing was carried out using a 2x150 paired-end configuration. Image analysis and base calling were conducted by the HiSeq Control Software (Illumina) on the HiSeq instrument.

Demultiplexing was performed by bcl2fastq 2.17 and raw data was filtered by discarding paired-end reads with adaptor, discarding paired-end reads when the content of N bases was more than 10% in either reads, and paired-end reads were discarded when the ratio of bases of low quality (Q<20) was more than 0.5 in either read.

2.4 Oxford Nanopore Rapid Barcoding Protocol

The Oxford Nanopore Rapid Barcoding kit allows for a 10-minute library preparation before loading the flow cell, providing that there is enough DNA available (400 ng in 7.5 ul). The Rapid Barcoding kit was selected and used for the nanopore sequencing throughout this project as this is what is more likely to be used in the field or in remote areas that have limited access to laboratory infrastructure.

MDA templates were sequenced in the School of Natural Sciences at Massey University using the Rapid Barcoding Sequencing Kit (SQK-RBK004). Briefly, 200ng of template DNA was combined with fragmentation mix RB01 (-12) and water in a total volume of 5ul. The reaction was incubated at 30ºC for 1 minute, followed by 80ºC for 1 minute, and cooled on ice. The fragmentation mix contains a transposase, which cleaves the DNA template and adds the barcodes simultaneously. Barcoded samples were pooled following fragmentation, and a 5ul aliquot was carried forward for sequencing. Next, 0.5ul of Rapid Adapter (RAP) mix was added, and the reaction was incubated at room temperature for 5 minutes. The DNA library

was stored on ice before loading. RAP mix contains the sequencing adapter to carry the fragments through the nanopores on the flow cell. Both Flongles (FLO-001) and Flow Cells (FLO-MIN106D) were used for MinION sequencing in this project.

2.5 Bioinformatic Analyses

Matched quality-checked Illumina Paired-end reads provided by Custom Science were mapped to the *M.bovis* genome (NCBI Reference Sequence: NZ_CP039850.1) using BWA2 to create SAM files. These were then converted to BAM files and sorted BAM files using Samtools mPILEUP and a script created by Roger Moraga from TEABREAK Bioinformatics was then used to create tables of average read coverage in 50bp windows across the *M.bovis* genome.

Fast base-calling and barcode demultiplexing was implemented using MinKNOW software (Oxford Nanopore Technology, Oxford, UK) to produce read sets with a q > 9 quality score. Porechop (v0.2.4 https://github.com/rrwick/Porechop) was then used for adapter removal.

Seqtk (v 1.4 (r122) https://github.com/lh3/seqtk) was used to randomly subsample reads to produce reads sets of similar number. Fifty thousand reads were subsampled from each sample. Seqtk is a tool used to process sequencing data in FASTA or FASTQ format. Each sample read was mapped to an *M. bovis* reference genome (NCBI Reference Sequence: NZ_CP039850.1) using Minimap2 (v2.26 (r1175) https://github.com/lh3/minimap2).
The detection of chimeras was done through Yacrd (v1.0.0 https://github.com/natir/yacrd). Samtools (v1.18 https://github.com/samtools/samtools) was used to convert the resulting SAM files from Minimap2 and Yacrd into BAM files and then used again to convert these into sorted BAM files. The Samtools mpileup utility was used to create a table of the data that Circa (OMGenomics https://omgenomics.com/Circa/) could read. Circa was used to create Circa plots to visualize read coverage. Circa was also used to visualize primer binding site distributions.

Chapter 3: Genome Coverage of *M.bovis* with Selective Multiple Displacement Amplification

3.1 Genome and Antibiotic Resistance Loci Coverage of Cultured *M. bovis* Template with "High Binders" (Illumina sequencing)

The average read coverage of the *M.bovis* genome produced by selective MDA with the Massey high binders (P12-P15) was determined.

3.1.1 MDA Results

Table 3.1 shows the concentration of MDA template measured after the MDA reaction and again after clean-up .

Table 3.1: Concentration of MDA template post MDA and clean-up

Primers	**Qubit Post MDA ng/ul**	**Qubit Post-clean-up ng/ul**
P12	686	540
P13	Too low	-
P14	107	257
P15	810	272
P12 + P13	336	124
P12 + P14	626	279
P12 + P15	848	399
P13 + P14	624	410
P13 + P15	848	220
P14 + P15	624	368
P12 + P13 + P14	368	80.4
P12 + P13 + P15	138	69.8
P12 + P14 + P15	298	170
P13 + P14 + P15	320	170
P12 + P13+ P14 + P15	360	144

3.1.2 Discussion

The results show all reactions produced MDA template under the amplification conditions used, with the exception of P13. Surprisingly, the highest concentrations obtained were when a single primer was used for the MDA reaction: P12, P15. In all cases, the clean-up step reduced the estimated concentration of template.

3.1.3 Illumina Reads Obtained for MDA Product

The template was sent to Custom Science for paired-end Illumina sequencing to evaluate these results further.

3.1.3.1 Results

Assuming that the Illumina sequence provider sequenced similar volumes of the MDA template concentration sent, the number of raw and processed reads shown in Table 3.2 indicates that with the exception of P13, the ”high binders” were successful in amplifying the MTB genome template when only one primer was used in the amplification reaction.

Table 3.2 provides an overview of the number of paired reads obtained for each MDA reaction.

Table 3.2: Number of binding sites and paired-end reads of each primer and primer combinations

Primers	# Binding Sites	# paired end reads (raw)	# paired end reads (processed)
P12	179	4729690	468549
P13	87	120	-
P14	567	5828062	446704
P15	475	5215061	435760
P12 + P13	266	4021600	383569
P12 + P14	766	4871126	411201
P12 + P15	656	4411004	400759
P13 + P14	664	4958063	390237
P13 + P15	562	4861213	473322
P14 + P15	1042	5274464	461317
P12 + P13 + P14	833	4283858	192661
P12 + P13 + P15	741	5252865	268283
P12 + P14 + P15	1241	4255821	161819
P13 + P14 + P15	1149	4914304	460543
P12 + P13+ P14 + P15	1308	4941898	207990

Table 3.2 shows the number of paired-end raw and processed reads obtained from each MDA reaction. The number of reads decreases in all cases once processed. Primer P12 and P13 + P15 achieved the highest number of processed reads.

3.1.3.2 Discussion

As indicated in Figure 2.1 and Table 2.3, P13 has a relatively lower number of MDA priming sites than P14 and P15, and this may explain its poor performance in amplification reactions where one primer was used. Some combinations of primers produced lower amounts of MDA template than others, and it is possible that primer interactions may have contributed to these results. However, this could not be determined from interactions predicted in Table 2.4. For example, P12 is predicted to interact negatively with P14 and P15. In the combination P12+

P14 + P15, it did produce lower amounts of template, but when combined in the reactions P12+ P14 or P12+ P15 it produced higher amounts of MDA template. The results did not suggest that increasing the number of primers would necessarily increase the amount of MDA template produced.

3.1.4 Mapping of Illumina Reads to AMR Loci in *M. bovis*

The paired-end reads in each of the MDA products were mapped using BWA2 to each of the 37 AMR loci in the *M.bovis* genome. A normalised number of reads (50,000 reads per MDA reaction) are shown in Table 3.3.

3.1.5 Results

The average coverage for 37 AMR genes of mapped Illumina processed reads was determined as described in Chapter 2. Table 3.3 indicates that combining primers did not increase average AMR coverage across the *M. bovis* genome. The best coverage of AMR gene loci was obtained with single MDA primers (P12, P14 and P15). Similar and lower levels of coverage were seen for each AMR region across the genome in mixed primer samples.

Table 3.3: Average AMR Region Reads (Illumina)

AMR Gene	P 12	P 14	P 15	P12 + 13	P 12 + P14	P 12 + P15	P 13 + P14	P 13 + P15	P 14 + P15	P 12+13+14	P 12+13+15	P 13+14+15	P 12+14+15	P 12+14+14+15
gryB	14.7	16.6	14.3	24.4	11.2	7.0	10.0	21.42	3.6	2.0	6.8	24.5	2.0	0.8
ponA1	3.2	3.9	8.6	2.3	1.3	2.2	1.8	3.65	2.3	0.6	1.8	5.7	1.1	0.3
fgd1	13.3	13.1	2.4	8.5	6.3	5.6	10.8	8.82	10.0	1.9	3.0	8.1	3.0	0.4
rpoB	4.1	1.9	1.1	1.6	0.9	1.3	1.9	1.22	3.4	0.1	0.9	1.6	0.2	0.4
rv0678	2.4	2.3	2.9	0.6	1.4	1.2	1.1	2.23	4.4	0.4	0.2	1.5	0.4	0.4
rspL	4.5	0.5	2.3	0.8	1.9	3.4	0.6	2.13	2.7	0	3.5	1.8	0	0
rplC	4.7	3.5	4.1	2.0	1.4	2.8	5.0	3.68	4.4	0.4	1.1	6.1	0.6	1.2
atpE	16.1	8.8	20.3	6.1	4.6	6.5	6.9	18.13	12.6	3.2	1.1	16.3	1.2	2.3
rrs	1.9	2.9	1.8	8.1	3.1	2.1	2.1	3.18	9.6	1.4	2.4	9.0	1.6	1.4
Rrl	2.8	4.0	2.2	6.2	3.7	1.8	1.8	4.04	10.2	1.4	3.2	9.2	1.3	1.0
fabG1	5.1	6.6	3.6	1.1	2.3	3.7	3.5	4.48	4.2	0	1.8	1.9	1.9	0.7
inhA	10.5	4.9	2.6	2.2	3.4	4.8	5.0	5.03	6.3	0.8	1.4	5.1	1.0	0.5
rpsA	3.0	2.3	2.3	0.5	1.0	1.1	0.6	0.90	0.7	0	0.4	0.5	0.3	0.2
tlyA	6.0	1.9	2.4	0.2	1.5	2.0	0.5	1.94	0.8	0.3	0.1	0.1	0	0
cycA	3.5	0.6	3.6	2.1	2.1	2.8	1.9	1.00	4.7	0.42	1.3	0.6	0.4	0
KatG	9.8	26.4	33.9	12.5	22.3	8.7	9.2	77.16	44.5	3.28	5.6	72.8	6.7	3.4
pncA	1.6	1.0	0.8	0.3	0.9	1.4	2.2	0.76	1.7	0	0.1	0.7	0	0.4
eis	6.7	1.7	1.6	1.6	0.9	1.3	1.5	2.24	3.5	0.1	1.4	0.7	0.7	0.4
aphC	11.3	1.3	1.4	1.5	1.3	2.5	1.9	0.27	7.6	1.1	0	0.4	0.8	0
folC	2.1	0.3	0.4	3.1	1.4	7.7	9.8	0	0.8	0.6	23.1	0.3	7.5	0.8
pepQ	3.6	1.9	4.6	1.0	0.9	1.5	0.9	1.43	0.7	0.1	0.3	0.5	0.3	0
ribD	6.0	2.3	3.3	2.1	5.2	8.9	8.8	3.29	5.5	1.2	0	1.5	0	0.3
thyX	3.6	4.2	5.1	2.9	2.5	2.2	2.5	4.36	4.9	0.1	1.2	5.5	0.4	0.6
dfrA	5.9	4.7	5.5	2.2	3.1	2.0	3.5	1.998	4.7	0.5	1.8	4.2	1.0	1.2

thyA	6.6	6.2	7.6	1.0	4.6	3.5	6.4	3.4	2.9	0.1	0.7	4.1	0.2	0.9
ald	3.1	1.9	2.2	4.5	0.9	1.3	1.3	0.8	1.3	0.2	0	1.1	0.1	0
ddlA	3.4	1.8	3.0	1.1	1.7	1.1	1.1	1.4	2.2	0.4	0.4	3.1	0	0.4
gryA	4.3	2.4	3.8	1.0	0.7	0.6	0.7	3.2	2.6	0.1	1.0	2.2	0.3	0.2
alr	2.7	1.5	1.7	1.0	0.7	1.5	0.7	0.4	1.01	0	0.4	2.0	0	0.7
ddn	2.6	0.9	4.4	2.7	2.7	2.4	0.3	2.8	4.9	1.0	0	1.7	0	0
panD	1.3	0.2	2.1	0.2	0	0.8	0.8	0.3	1.5	0.3	0.6	0.5	0	0
embC	2.8	1.3	1.7	0.4	0.3	1.1	0.9	0.9	0.9	0.2	0.4	0.4	0.5	0.3
embB	2.8	1.3	1.2	1.0	1.2	1.0	1.2	1.2	2.8	0.1	0.5	3.2	0.3	0.2
ubiA	9.8	3.6	4.1	3.1	3.2	2.7	1.7	2.2	8.1	0.6	1.6	6.6	0.5	0.8
ethA	4.6	4.9	4.1	1.1	2.5	3.6	2.3	5.1	4.7	0.6	1.3	7.3	0.8	1.1
ethR	2.3	4.6	3.6	1.1	0.9	0.7	1.9	4.1	2.2	0.5	0.3	6.8	0.8	0.7
gidB	5.4	8.3	1.8	6.3	4.6	5.8	5.5	6.7	20.4	1.3	3.1	10.2	0.4	1.3

3.1.6 Discussion

The experiments and analyses described in section 3.1 aimed to assess whether the 'high binder' primers P12, P13, P14 and P15 would amplify the target genome and, more specifically, AMR gene loci. The results were somewhat surprising, as it was expected that combining multiple primers would result in a higher level of coverage across the genome, as theoretically, more binding sites would be available. However, this was found to not be the case, and single primers were found to produce better overall coverage. This might be explained by negative interactions between different primers when combined, making them less efficient. Perhaps important to note is that Primer P12 is longer and more GC rich than the other high binder primers. Thus, it has a slightly higher Tm than the other high-binder primers. This may prove advantageous in assay development, as there would be the potential to increase the temperature of the MDA reaction for selectivity if this primer was used.

3.2 Genome and Antibiotic Resistance Loci Coverage of Cultured *M. bovis* template with "High Binders": P12, P14 and P15 (MinION Sequencing)

Independent MDA reactions were also carried out with *M.bovis* DNA, using each of the primers P12, P14 and P15 to determine whether similar coverage would be obtained with MinION nanopore sequencing. MDA product was produced as described in section 2.2, using 10ng each of DNA. The yields obtained are reported in Table 3.4. These experiments compared the relative level of antibiotic resistance loci coverage with Massey primers P12, P14 and P15. A visual comparison of coverage was also made between the Illumina sequencing results and ONT MinION sequencing results for MDA amplification with primers P12, P14 and P15.

3.2.1 *MDA Results*

Table 3.4 shows the concentrations of MDA template produced in duplicate MDA reactions with each of the three high binder primers.

Table 3.4: Concentration of duplicate primer samples post MDA and clean-up

Primer	Qubit Post MDA (ng/ul)	Qubit Post Clean-up (ng/ul)
12	540	271
12	520	247
14	148	81.7
14	188	90.4
15	146	67.9
15	116	51

The results show that under the amplification conditions used, all MDA reactions produced MDA template with each of the primers. Primer P12 had the highest template yield, followed by primer P14 and then P15. The clean-up step reduced the expected concentration of templates in all sample types.

3.2.2 Discussion

High binder P13 was not investigated in this experiment since it failed to produce MDA template in earlier investigations. This may be because it binds less frequently than primers P12, P14 and P15 (Table 2.3). The amount of MDA product produced using P12 was similar to the amount obtained in the MDA reaction with P12 used for Illumina sequencing. There was some variation in the concentration of duplicates for P14 and P15, and the amount of MDA template produced differed more with these two primers compared to the amount of MDA template produced in earlier reactions with the same primers (Table 3.1).
This is interesting as P15 is reported to have 567 binding sites and P12 179 (Table 3.2). Higher binding sites generally lead to higher coverage when amplifying a specific DNA target. Primer P12 may have higher specificity to the target genome, meaning it anneals to the exact sequence of interest and is more optimised so that it can maximise specificity and efficacy. This can also result in amplification bias, however, where some regions are preferentially amplified over others, which can distort the representation of the target genome in the data, and lead to uneven coverage.

3.2.3 ONT Sequencing and Genome Coverage

Samples 1-6 were barcoded as per section 2.4.1, using 400ng each of DNA and Barcodes 1-6. The reads were processed following the bioinformatics pipeline described in Chapter 2. Samples were prepared for sequencing using the Rapid Barcoding kit and run on a flow cell as described in 2.4.1.

3.2.4 Results

CIRCA plots were produced to compare the coverage profiles of Illumina sequencing and MinION sequencing for three independent MDA reactions for each of the primers P12, P14 and P15. These profiles are shown in Figures 3.1-3.3. Replicates for each of the primers produced similar profiles to each other, and the profiles were observed with Illumina sequencing. P12, P14 and P15 profiles differed in terms of their coverage of antimicrobial resistance loci. The average coverage for 37 AMR genes is shown in greater detail in Table 3.5 (MinION sequencing) and 3.6 (Illumina sequencing).

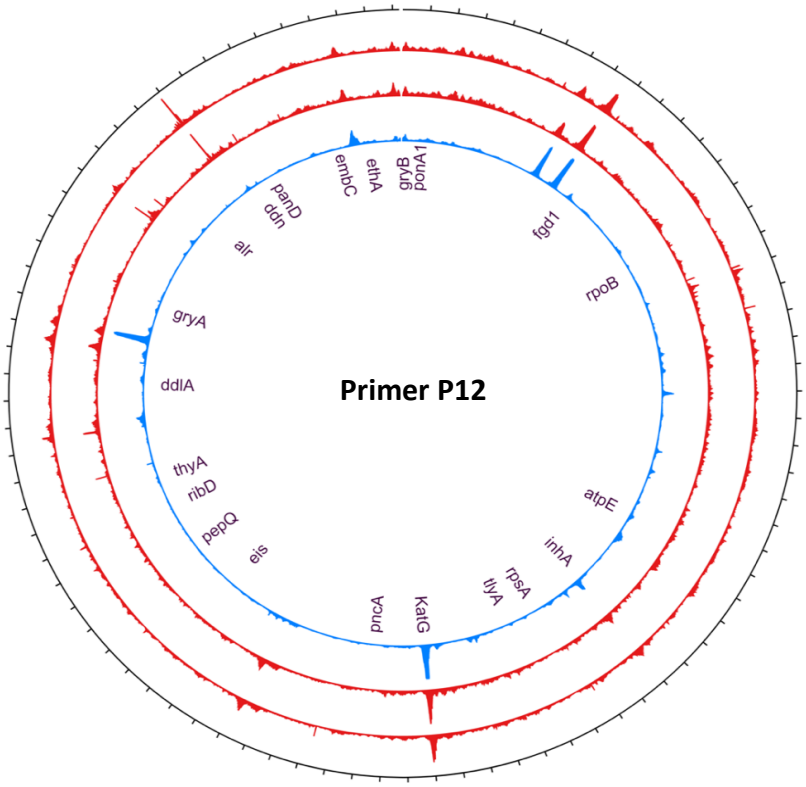


Figure 3.1: Circa Plot of primer P12 MinION duplicates (red) and Illumina data (blue)

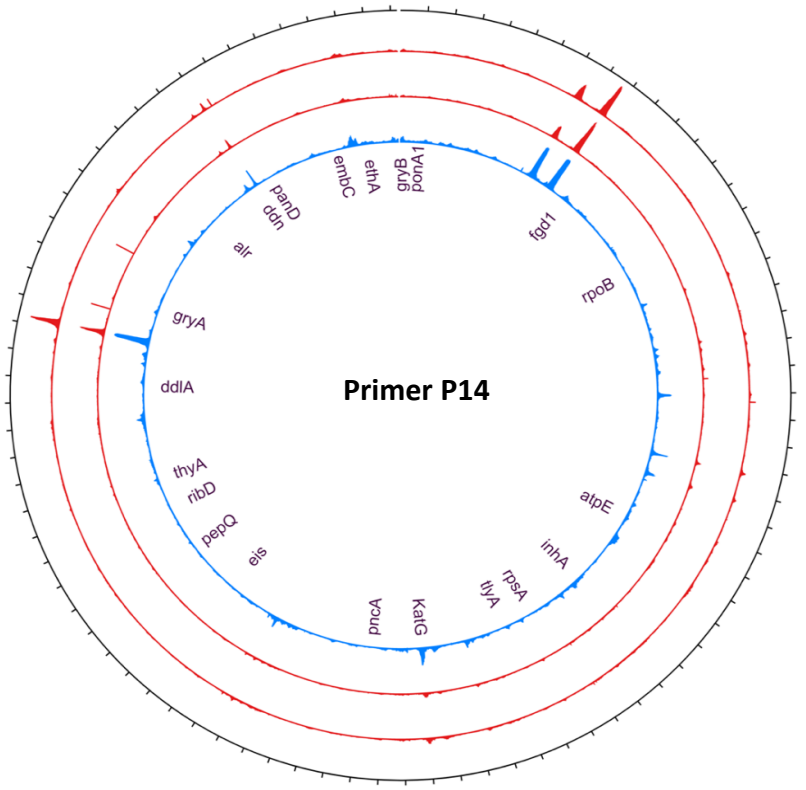


Figure 3.2: Circa Plot of primer P14 MinION duplicates (red) and Illumina data (blue)

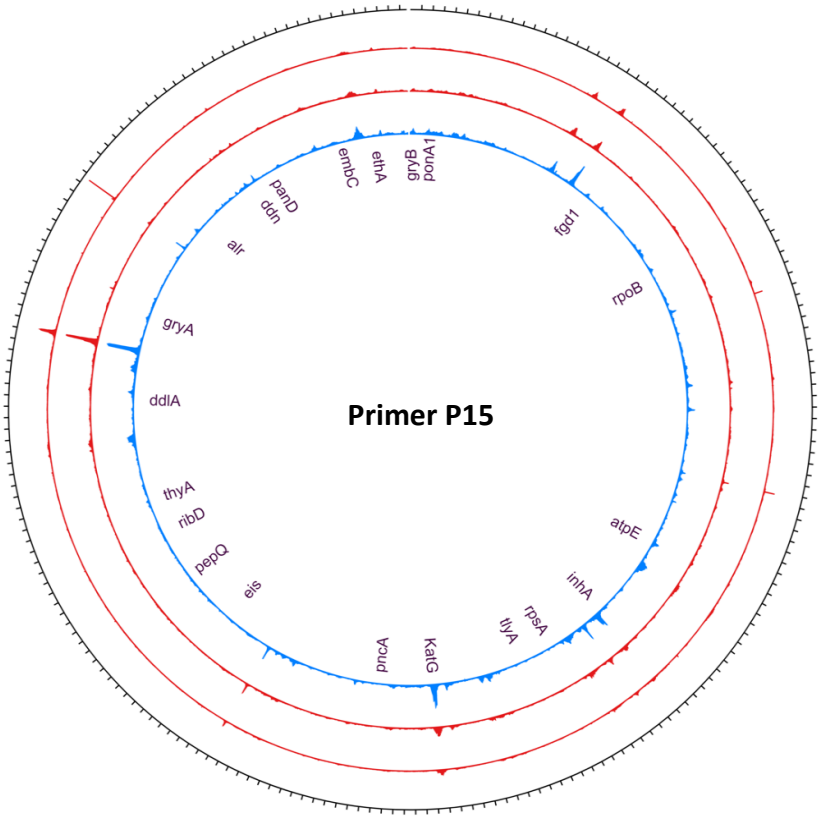


Figure 3.3: Circa Plot of primer P15 MinION duplicates (red) and Illumina data (blue)

Table 3.5: Average coverage of AMR regions - MinION duplicate data

AMR Gene	Primer P12	Primer P12	Primer P14	Primer P14	Primer P15	Primer P15
gryB	113.72	102.38	86.45	62.18	60.81	64.73
ponA1	12.08	22.73	21.02	15.50	30.36	19.41
fgd1	53.54	47.94	36.94	24.74	9.22	13.38
rpoB	20.15	12.25	12.63	6.06	8.05	10.87
rv0678	16.86	13.08	8.75	4.20	7.76	4.30
rspL	26.34	20.44	14.36	6.81	11.06	4.50
rplC	17.41	22.87	24.62	15.03	19.28	24.48
atpE	39.68	43.70	49.12	40.78	60.53	49.14
rrs	10.91	9.77	7.77	3.63	7.09	5.03
Rrl	12.54	8.01	14.49	6.92	6.26	4.01
fabG1	51.36	28.06	30.07	12.97	19.13	19.64
inhA	49.88	19.93	33.74	16.78	19.64	16.48
rpsA	17.23	5.02	9.58	11.92	11.65	4.13
tlyA	10.16	12.21	11.24	6.34	12.07	14.18
cycA	21.93	19.18	2.45	3.70	6.30	12.23
KatG	87.32	66.80	104.58	83.27	123.72	134.53
pncA	4.78	12.93	3.76	4.17	4.85	2.146
eis	16.82	15.03	11.43	9.79	9.71	11.67
aphC	32.83	27.08	5.42	4.38	7.92	6.09
folC	1.68	4.76	1.59	3.80	4.47	3.79
pepQ	9.73	14.52	10.10	9.60	19.03	11.83
ribD	19.17	14.93	25.08	19.52	13.13	17.00
thyX	10.28	12.31	15.97	9.45	10.21	17.16
dfrA	12.06	14.22	23.25	25.15	9.37	18.72
thyA	15.82	16.43	39.61	37.52	18.59	26.65
ald	10.52	6.51	9.69	5.16	7.4	5.52
ddlA	16.82	13.65	9.82	10.81	8.93	14.58
gryA	10.41	14.51	11.34	12.06	21.01	16.35
alr	9.65	2.15	15.44	7.96	5.60	19.30
ddn	26.55	15.47	19.82	20.07	17.15	7.41
panD	7.38	1.675	0.3	1.41	2.07	6.15
embC	11.67	8.45	8.72	7.61	9.40	12.92
embB	18.99	17.97	7.20	3.28	6.26	12.31
ubiA	22.91	18.26	15.64	13.68	8.81	19.84
ethA	15.87	14.74	23.21	20.04	20.65	27.16
ethR	11.79	11.90	27.69	17.44	22.32	22.31
gidB	43.49	40.92	33.15	13.72	8.00	8.29

Table 3.6: Average coverage of AMR regions - Illumina data

AMR Gene	Primer P12	Primer P14	Primer P15
gryB	14.66	16.59	14.32
ponA1	3.26	3.94	8.62
fgd1	13.28	13.14	2.42
rpoB	4.05	1.91	1.11
rv0678	2.39	2.31	2.85
rspL	4.46	0.51	2.29
rplC	4.68	3.45	4.12
atpE	16.11	8.79	20.35
rrs	1.94	2.93	1.83
Rrl	2.81	4.00	2.18
fabG1	5.07	6.58	3.60
inhA	10.45	4.94	2.62
rpsA	2.99	2.33	2.27
tlyA	5.93	1.89	2.44
cycA	3.52	0.62	3.58
KatG	9.75	26.38	33.93
pncA	1.6	1.02	0.82
eis	6.67	1.73	1.63
aphC	11.28	1.28	1.36
folC	2.13	0.32	0.37
pepQ	3.58	1.38	4.58
ribD	6.03	2.33	3.30
thyX	3.57	4.24	5.08
dfrA	5.93	4.73	5.52
thyA	6.60	6.15	7.62
ald	3.07	1.89	2.23
ddlA	3.43	1.79	3.01
gryA	4.27	2.41	3.83
alr	2.66	1.45	1.65
ddn	2.60	0.92	4.38
panD	1.33	0.15	2.12
embC	2.82	1.30	1.73
embB	2.80	1.29	1.21
ubiA	9.84	3.57	4.08
ethA	4.57	4.94	4.13
ethR	2.32	4.57	3.62
gidB	5.40	8.28	1.78

3.2.5 Discussion

This experiment aimed to deduce the relative level of antibiotic resistance loci coverage with the P12, P14 and P15 primers using *M.bovis* as a reference. CIRCA plots were drawn to compare the coverage and similarity of profiles obtained using Illumina and ONT sequencing. Read counts were normalized to 50,000 reads for both Illumina and ONT sequences. This simplified the analyses but meant the absolute levels of coverage were not directly comparable between the Illumina and ONT sequencing since the average read length of ONT reads was much greater than that of Illumina reads. As expected, the Illumina sequencing data produced fewer reads over each of the AMR regions when compared to the number of reads produced by the MinION over the same regions.
In the ONT profiles, Primer P12 had the fewest sites where there was little to no coverage and Primer P12 consistently produced the greatest level of overall coverage of AMR loci across the genome. Primers P14 and P15 exhibited levels of coverage that varied from one AMR gene to another.

The coverage profiles for Illumina sequencing and ONT MinION sequencing of MDA template for primers P12, P14 and P15 revealed similar results. This can be visualized in the CIRCA plots and tables shown above. which indicate that the MDA reactions were reproducible and produced similar patterns of genome coverage irrespective of whether Illumina or ONT sequencing was used to sequence the MDA template. This finding indicates the potential for using ONT sequencing to predict drug susceptibility based on the sequencing of MDA product. These CIRCA plots show even coverage, though a few peaks within the same regions can be seen. These peaks may appear due to the primers preferentially binding to certain regions of the genome and not others, which is something to consider when interpreting the data. Even coverage is considered highly desirable here, as increasingly more and more gene loci have been implicated in antibiotic resistance of MTB. There is also the possibility for interactions between the products of different gene loci. The peaks seen on the CIRCA plots may also be due to the high GC content of the Mycobacteria genome (65.6% for *M. tuberculosis*) (Cole *et al*., 1998). The primers themselves are GC-rich, which may cause them to preferentially bind to these regions.

3.2.6 Genome and Antibiotic Resistance Loci Coverage of Cultured *M. bovis* Template with "High Binders" and Additional Primers

To determine if coverage could be improved for P12, MDA reactions with locus specific primers P5 and P6 were investigated. As shown in Figures 3.4 and 3.5, the MDA priming sites of P5 and P6 occur in regions not well covered by P12. ONT MinION Rapid Barcoding (SQK-RBK114.24) was used to sequence the MDA product, as this sequencing protocol is the simplest protocol that could be used in the field.

3.2.6.1 Results

Table 3.7: MDA Results of P12 + P5 + P6

Primer	Qubit Post MDA (ng/ul)	Qubit Post Clean-up (ng/ul)
5 + 6	460	348
12 + 5	700	388
12 + 6	498	342
12 + 5 + 6	600	448

All reactions produced MDA template under the amplification conditions. As observed previously, in all cases the clean-up step reduced the amount of template. The primer combination of P12 + P5 + P6 produced the greatest yield in cleaned MDA template. Samples were prepared for sequencing using the Rapid Barcoding kit and run on a flow cell as described in 2.4.

Figure 3.4: Circa Plot of Primers with no mismatches - P12 (outermost/ dark purple), P5 (middle/ light purple), and P6 (innermost/ blue)

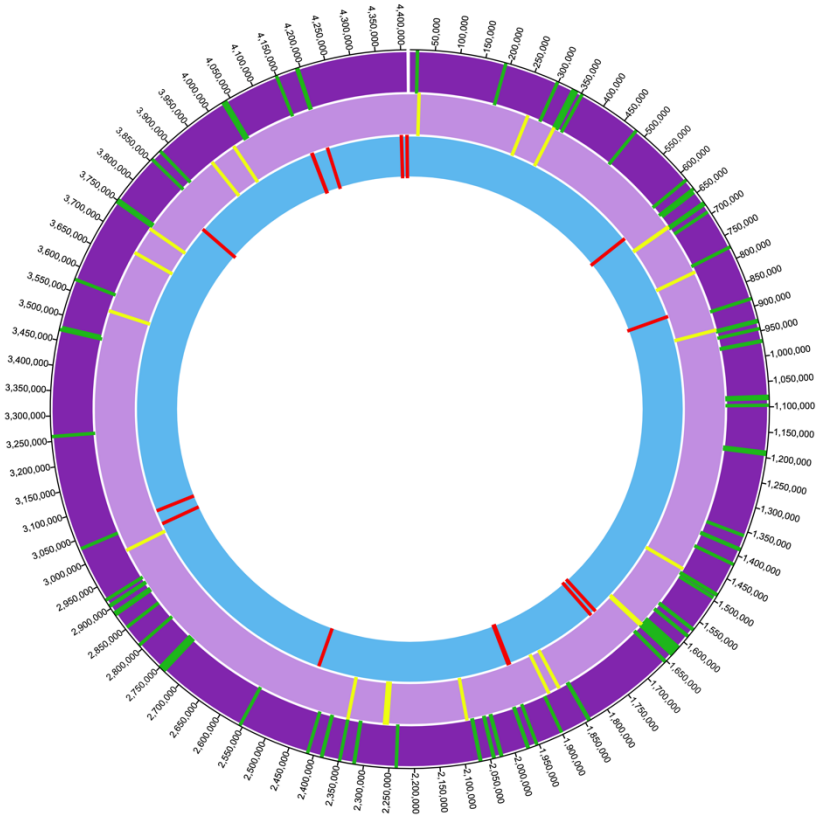
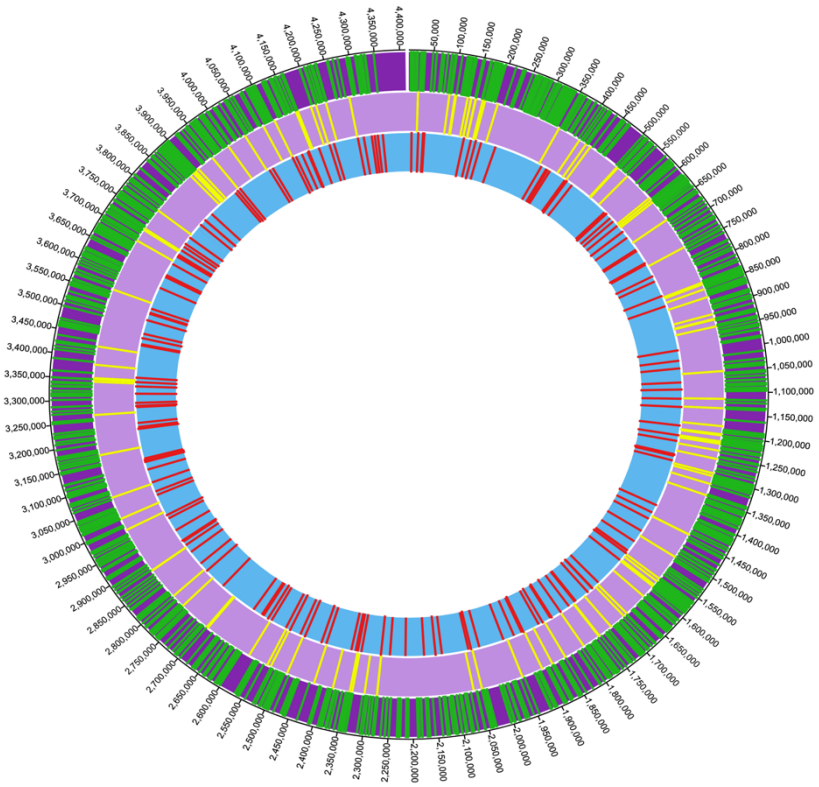


Figure 3.5: Circa Plot of Primers with one mismatch - P12 (outermost/ dark purple), P5 (middle/ light purple), and P6 (innermost/ blue)



3.2.6.2 Discussion

The Qubit results were encouraging since the MDA reactions with additional locus specific primers appeared to increase the yield of MDA template. Based on the Qubit results, primers P5 and P6 also produced MDA product when used together without P12. This was a surprising result as it was predicted from the exact matching profile of P5 and P6 that there were relatively few priming sites in the *M. bovis* genome. The result suggests that mismatch priming likely contributes to the success of the MDA reaction. This was investigated further by sequencing the MDA products.

3.2.7 ONT Sequencing and Genome Coverage

3.2.7.1 Results

Table 3.8 shows the AMR gene loci coverage obtained with the different primer combinations. Figure 3.6 shows the CIRCA plot and overview of the coverage of the *M. bovis*

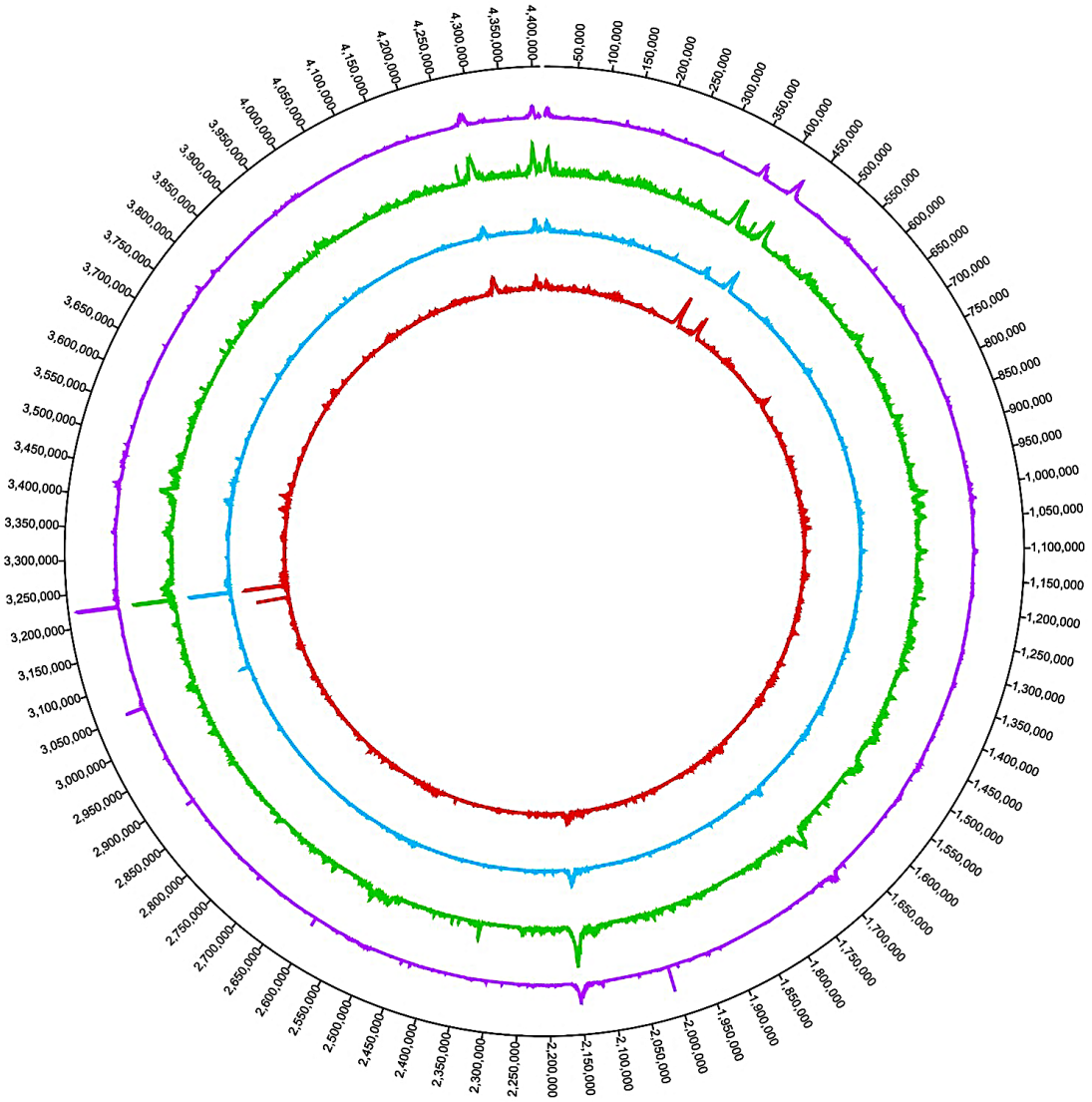
genome for different primer combinations. An interesting observation is that while the depth of coverage varies across the genome, some regions of the *M. bovis* genome exhibit similar spikes in coverage for different primer combinations (e.g. in the gyrB and KatG gene regions)

Table 3.8: Average AMR Coverage for P5, P6 and P12 Primer Combinations

AMR Gene	P12	P12	P5 + P6	P12 + P5	P12 + P6	P12 + P5 + P6
gryB	113.72	102.38	215.27	262.98	279.69	236.83
ponA1	12.08	22.73	13.42	15.62	40.14	16.37
fgd1	53.54	47.94	52.33	39.4	48.49	32.48
rpoB	20.15	12.25	22.66	28.27	19.47	16.46
Rv0678	20.15	12.25	38.11	18.61	11.64	21.21
rspL	26.34	20.44	37.47	21.23	18.85	18.75
rplC	17.41	22.87	22.66	18.31	20.92	16.05
atpE	39.68	43.70	50.11	50.53	72.54	48.22
rrs	10.91	9.77	51.25	52.35	48.45	59.79
Rrl	12.54	8.01	46.76	61.91	34.08	56.31
fabG1	51.36	28.06	40.04	30.67	28.33	19.51
inhA	49.88	19.93	45.74	33.9	39.92	31.67
rpsA	17.23	5.02	16.39	6.89	6.64	5.58
tlyA	10.16	12.21	15.92	8.76	10.49	10.71
cycA	21.93	19.18	15.66	12.24	26.23	14.14
KatG	87.32	66.80	103.46	179.54	175.69	212.28
pncA	4.78	12.93	13.17	5.51	2.68	2.15
eis	16.82	15.03	44.14	6.96	10	6.53
aphC	32.83	27.08	18.42	17.71	14.92	14.97
folC	1.68	4.76	0.75	0	0.83	0
pepQ	9.73	14.52	11.33	10.43	7.93	7.34
ribD	19.17	14.93	9.03	6.07	17.23	9.74
thyX	10.28	12.31	28.83	20.47	17.52	22.35
dfrA	12.06	14.22	18.24	15.78	10.8	10.25
thyA	15.82	16.43	23.28	12.82	14.82	13.03
ald	10.52	6.51	5.54	3.26	11.86	9.53
ddlA	16.82	13.65	11.99	10.36	17.06	8.88
gryA	10.41	14.51	19.76	16.53	21.03	12.77
alr	9.65	2.15	5.22	7.91	6.2	9.11
ddn	26.55	15.47	33.46	23.06	29.78	20.91
panD	7.38	1.675	8.02	1.88	2.35	4.39
embC	11.67	8.45	10.18	3.86	5.64	5.17
embB	18.99	17.97	13.13	8.46	6.91	11.17
ubiA	22.91	18.26	45.91	25.92	42.17	36.9
ethA	15.87	14.74	33.35	24.15	24.2	22.08
ethR	11.79	11.90	32.95	22.24	21.96	19.93
gid	43.49	40.92	95.48	110.25	98.36	80.76

Figure 3.6: Circa Plot of P12, P5 and P6 Coverage

Order from innermost ring: P5+P6 (red), P12+P5 (blue), P12+P6 (green), P12+P5+P6 (purple)



3.2.7.2 Discussion

With some genes, the inclusion of primers P5 or P6 with P12, P5+P6+P12 resulted in higher levels of AMR region coverage than was obtained with just P12 (Table 3.7, 3.8). However, this was not always the case (e.g. PncA, folC).

Primers P5 and P6 yielded a high level of coverage when combined without P12 present. This was unexpected, given the number of exact matching P5 and P6 priming sites in the *M. bovis* genome (Figure 3.4). This higher than expected coverage is likely explained by mismatch priming since there is a high density of such sites across the *M. bovis* genome (Figure 3.5). Figure 3.6 shows that similar regions of the genome produced spikes in coverage, even when different primers were used for amplification. Although further analyses could be undertaken, preliminary analyses (not shown) suggest that these regions do not have elevated GC% contents compared to other regions of the genome, nor do they appear to have a higher density of mismatch priming sites for P5, P6 or P12. It is possible that secondary structure considerations contribute to such spikes (Fan *et al*., 2019).

3.3 MDA Genome and Antibiotic Resistance Loci Coverage of Cultured *M.bovis* Template Spiked into Sputum

To assess the performance of the P12 + P5 +P6 combination to amplify *M. bovis* DNA in the presence of DNA from sputum and human blood, MDA reactions were set up for a range of *M. bovis* DNA concentrations with either a) sputum DNA, b) human DNA and c) no background DNA. MDA reactions were also performed on a) sputum DNA and b) human DNA in the absence of *M. bovis* DNA. The MDA amplification results are shown in Table 3.10.

3.3.1 MDA Results

MDA amplified product was produced in all reactions, but the reaction containing only sputum DNA or Human DNA yielded the lowest amount of product. In *M. bovis* spiked reactions (human DNA + *M. bovis* DNA; sputum DNA + *M. bovis* DNA), greater yields of amplified DNA were obtained in reactions with higher concentrations of *M. bovis* DNA.

Table 3.9: Qubit Results of reaction containing sputum DNA, M. bovis DNA and sputum DNA/or Human DNA

Reaction Number	Reaction Contents	QUBIT post MDA (ng/ul)	QUBIT post clean-up (ng/ul)
1	5ng sputum+5ng *M.bovis*	349	182
2	5ng sputum+0.5ng *M.bovis*	236	156
3	5ng sputum+0.05ng *M. bovis*	203	152
4	5ng human DNA+5ng *M. bovis*	590	289
5	5ng human DNA+0.5ng *M. bovis*	418	151
6	5ng human DNA+0.05ng *M. bovis*	224	104
7	10ng sputum	240	80
8	10ng Human DNA	106	58
9	0.5ng *M.bovis*	560	266
10	0.05ng *M. bovis*	370	195

3.3.2 *Discussion*

The lower yield in reactions without *M. bovis* DNA and higher amplification yields in the presence of *M. bovis* suggested that while the MDA primers did amplify something in human DNA and sputum backgrounds, the *M. bovis* DNA may be preferentially amplified with primers P12+P5+P6. To determine if this was the case, MDA product was sequenced for reaction products. The AMR gene loci coverage for 50,000 reads for each reaction is shown in Table 3.10.

3.3.3 *MinION Sequencing Results*

Coverage of AMR gene resistance loci was greatest for *M. bovis* DNA when no background human or sputum DNA was present. Relatively high levels of coverage were found with *M. bovis* DNA in the presence of human DNA. The lowest coverage of AMR resistance loci was observed in the presence of sputum DNA. As the *M. bovis* DNA concentration was reduced, the coverage of AMR gene resistance loci decreased in all ca

Table 3.10: Average coverage of AMR Regions from Mixed Samples

AMR Gene	5ng sputum+5ng *M.bovis*	5ng sputum+0.5ng *M.bovis*	5ng Sputum+0.05ng *M. bovis*	5ng Human DNA+5ng *M. bovis*	5ng Human DNA+0.5ng *M. bovis*	5ng Human DNA+0.05ng *M. bovis*	10ng Sputum DNA	10ng Human DNA	0.5ng *M.bovis DNA*	0.05ng *M.bovis* DNA
gryB	29.73	0.58	1.72	253.07	120.74	117.67	0	0	170.11	280.11
ponA1	2.77	0	0	8.66	11.92	4.88	0	0	16.73	10.62
fgd1	7.69	0.02	0	23.25	18.74	19.83	0	0	36.10	14.84
rpoB	3.24	0.63	1.11	6.08	6.26	2.38	0.35	0	10.00	7.78
rv0678	2.19	0.12	0	14.74	20.11	5.83	0	0	14.01	26.96
rspL	2.13	0.62	0	15.80	22.36	6.26	0	0	17.79	25.81
rplC	3.78	0.74	0	10.29	9.56	8.68	0	0	17.36	11.17
atpE	6.81	1.43	0	30.15	36.60	46.78	0	0	38.64	45.10
rrs	35.15	30.46	22.99	65.87	54.28	50.73	14.67	0	50.79	92.11
rrl	44.48	35.19	33.34	56.10	53.48	29.57	16.36	0	45.28	77.39
fabG1	2.55	0.68	0	12.33	15.22	9.35	0	0	15.13	12.69
inhA	2.11	0.67	0	18.87	16.473	13.36	0	0	17.55	16.54
rpsA	0	0	0	3.64	4.201	5.90	0	0	5.93	1.43
tlyA	2.22	0	0	6.24	9.62	4.78	0	0	7.59	8.08
cycA	2.91	0.12	0	5.96	15.32	6.00	0	0	9.89	8.59
KatG	22.46	5.41	0.45	167.57	230.22	160.73	0	0	278.03	403.06
pncA	1.18	0	0	3.34	1.16	1.45	0	0	3.47	1.24
eis	0.13	0	0	7.51	5.45	5.63	0	0	6.93	11.81
aphC	1.05	0	0	5.35	5.19	0.41	0	0	5.78	6.42
folC	1.09	0	0	0	0	0	0	0	0.06	0
pepQ	3.23	0	0	3.75	3.75	3.60	0	0	5.56	3.84
ribD	3.44	3.59	0.49	5.98	5.69	2.96	0	0	22.42	3.40
thyX	3.93	0	0	15.62	10.47	8.31	0	0	16.02	8.07
dfrA	1.57	0.10	0	4.30	9.44	9.60	0	0	16.65	7.75
thyA	1.43	0.02	0	4.912	9.95	9.34	0	0	16.67	11.33

ald	1.42	0	0	4.23	5.19	2.65	0	0	6.65	2.07
ddlA	2.71	0	0	7.38	16.93	8.97	0	0	5.63	9.76
gryA	0.90	0	0	7.21	13.97	3.20	0	0	9.46	12.26
alr	2.80	0.63	0	4.47	9.64	17.01	0	0	8.09	7.91
ddn	1.09	0.77	0	6.23	18.09	7.98	0	0	18.345	13.55
panD	0	0	0	2.86	2.08	2.43	0	0	4.06	4.18
embC	2.12	0	0	4.59	5.28	1.23	0	0	5.08	3.09
embB	0.85	0	0	5.12	2.71	2.03	0	0	6.99	1.67
ubiA	9.27	3.35	0	41.13	57.35	20.88	0	0	62.74	66.53
ethA	4.81	1.07	0	21.14	13.08	5.91	0	0	22.03	15.19
ethR	7.23	0	0	18.16	14.54	4.55	0	0	27.35	15.77
gidB	43.88	0	0	62.67	57.68	59.29	0	0	84.32	171.64

3.3.4 Discussion

Table 3.10 shows the average level of reads mapping to AMR gene loci. In the reactions that contained only *M. bovis* DNA, low levels of coverage for most loci could be obtained with as little as 50,000 reads when the concentration was as low as 0.05ng/ul. The presence of DNA from sputum clearly reduced the AMR loci coverage. To examine if this was due to MDA amplification of other bacteria in the sputum, a preliminary analysis was conducted by mapping the ONT reads from the three sputum treatments (sputum+5ng/0.5ng/0.05ng *M.bovis*) to a silva 16S rRNA gene database. Minimap2 and Python scripts from the Spaghetti pipeline were used for this purpose (https://github.com/adlape95/Spaghetti). The most abundant matches were to upper respiratory tract species of bacteria: *Streptococcus agalactiae, Rothia mucilaginosa, Megasphaera micronuciformis, Candidatus arsenophonus* and *Neisseria zalophi*.

Interestingly, with the exception of *Rothia*, none are particularly GC% rich and as discussed further in the section on future work, this may have implications for developing an MDA-ONT *Mycobacterium* diagnostic. Of note in Table 3.1.3 is that some reads mapped to rpoB (gene for RNA polymerase B) and rrS and rrL genes (genes for ribosomal RNAs). These are highly conserved genes in bacteria, and it is likely the matching reads are to bacteria present in the sputum.

Chapter 4: Evaluation of Selective MDA ONT Sequencing

4.1 Diagnosis of TB and the Potential of ONT Sequencing

Tuberculosis is a highly infectious and pathogenic disease that has a global burden with the greatest impact in low economy countries lacking adequate TB detection and monitoring measures (Bagcchi, 2022). The TB diagnostic methods currently in use have significant drawbacks, including low sensitivity, high costs, and complex operational procedures, which make them unsuitable for clinical settings (Kontsevaya *et al*., 2023). Some rapid detection methods, including PCR, are being used in MTB testing in limited settings with certain advantages. However, PCR tests are locus-specific, so only detect a subset of AMR bacteria. For effective TB diagnostics in low-resource settings, the ideal diagnostic tool must be fast, sensitive, affordable, specific, comprehensive and user-friendly.

Smear microscopy and culture techniques remain the most commonly used methods for the diagnosis of TB (Ben Selma *et al*., 2009), especially in resource-limited settings (Das *et al*., 2019). However, the time-consuming nature of these methods is not sufficient for informing treatments that could be helpful for interrupting the infection cycle (Pandey *et al*., 2019). This is why improved rapid diagnosis is important for the WHO's goal of eradicating TB (Bagcchi, 2022). The success of ONT sequencing for informed decision-making during the COVID-19 pandemic makes it an obvious choice for rapid diagnosis (Yakovleva *et al*., 2021; Barbé *et al*., 2022). Using ONT for TB diagnosis has come a long way. In July 2023, the WHO announced that a tool for MTB diagnosis was being developed by Oxford Nanopore Technologies and that this could lead to ONT devices being used in clinical decision-making for the treatment of drug-resistant TB (World Health Organization, 2023). Dippenaar *et al*. (2022) have reviewed recent publications using ONT sequencing for the detection and diagnosis of *Mycobacterium*.

4.2 Advantages and Challenges in Using ONT Sequencing for Clinical Applications and TB Research

As Nanopore sequencing is a rapidly evolving, the standardisation of protocols, quality control measures and validations against established TB diagnostics are essential to ensure the accuracy and reliability of results within a clinical setting. Important progress has been made in developing ONT sequencing for the diagnosis of *Mycobacterium* disease, but there also remains some challenges.

While nanopore sequencing is capable of producing reads of up to 2 Mb long (Payne *et al*., 2019), a significant drawback of this sequencing technology is its lower throughput of sequence data and suboptimal accuracy in older flow cell versions when compared to other sequencing technologies, such as Illumina. Sequencing accuracy can be measured by either a single read or a consensus accuracy. However, in 2015, the single-read accuracy (this refers to the percentage of similarity between a given sequence and its reference sequence when compared) of ONT sequencing data was only 60%. This low accuracy was due to random read errors. The raw read accuracy has since increased from <60% to >99% for the latest R10.4 flow cells after several updates in flow cell chemistry, the development of post-sequencing correction tools and improvements in base-calling software. This level of accuracy is comparable to that of Illumina and other next-generation sequencing methods. Consensus accuracy, dependent on systematic errors, assesses the identification of a consensus sequence created from numerous overlapping reads from the same genomic region. For *de novo* sequencing, MTB has an approximate consensus accuracy of 99.3% at 130x coverage and 99.92% at 238x coverage (Bainomigsa *et al*., 2018). Recent sequencing studies of SARS-Cov2 genomes have amply demonstrated that the sequencing accuracy of ONT technology is sufficient to inform understanding of emerging pathogen variants of concern (Barbé *et al*., 2022).

One of the most significant advantages for Mycobacterium diagnosis using ONT sequencing is that it allows for real-time data analysis while the sequencing process is still ongoing. This is a unique feature that is not available on any other sequencing platforms that are currently available, which has been a drawback for infectious disease diagnostics While the ONT Promethion and GridION platforms provide a much higher throughput of samples and output of data, the portability and real-time sequencing advantages of the MinION make it more

suitable for sequencing in remote areas or in the field where there are few lab resources, and where it can provide results within 48 hours of samples being received (Xing *et al*., 2022).

However, while the MinION's flow cell can handle multiple patient samples simultaneously and can be reused multiple times, direct sequencing of *Mycobacterium* present in sputum DNA requires a large number of reads, given the low frequency of MTB DNA in sputum DNA, and this limits the number of samples that can be analysed on any given flow cell (Dippenaar *et al*., 2022). This requirement adds an expense to diagnosis and means that ONT Flongles cannot be used for sputum samples without some means of selecting for MTB DNA or selecting against other DNA in the samples. In some situations, Flongles would be preferable if they could be used because they are disposable, and contamination between samples would be easier to control. The cost of Flongles is also approximately $90 USD, which is significantly less expensive than the MinION R9.4.1 flow cells, which cost $USD 900 each. In principle, it is possible to generate approximately 1GB of data on a Flongle in just 24 hours. Recent experiences in our lab using R9.4.1 Flongles show that they are less reliable than the R9.4.1 flow cells and cannot be reused. However, with version 14 chemistry and the newer R10 Flongles, it will be interesting to see if their performance has improved.

4.3 Selective Multiple Displacement Amplification for *Mycobacterium* Diagnosis (when combined with the MinION)

Selective multiple displacement amplification is one method that could potentially be used to selectively amplify *Mycobacterium* DNA in sputum samples. MDA typically uses Phi29 or the more recently improved mutant version of this DNA polymerase - EquiPhi29 (Thermo Fischer Scientific; Blanco *et al*., 1989). Due to polymerisation activity and excellent enzymatic fidelity, this DNA polymerase is highly valued for its ability to amplify genomic DNA in an adequate amount and quality (Mai *et al*., 2004).

Previous efforts have been made to design selective MDA primers for *Mycobacterium*, but these have produced limited success (Clarke *et al*. 2017). This earlier study used shorter primers than those investigated in our present study and used the Phi29 enzyme, which has a lower optimal temperature than Equiphi29. This enzyme also requires a much shorter time for amplification than does Phi29, and this makes it potentially more useful in a rapid diagnostic.

An advantage of MDA primer sets that selectively amplify AMR genes in the *Mycobacteria* genome is that this template would enable comprehensive analysis of MTB genome variants and also variants in *Mycobacterium* species closely related to MTB.

When coupled with ONT sequencing, such MDA primer sets have the potential to aid the early and rapid detection of *Mycobacterium* infections. Fast and efficient diagnosis is key for timely treatment and effective disease management, especially in areas with a high TB burden. Being able to identify DNA sequence variants in AMR genes is thus essential for appropriate treatment guidance strategies and in helping to prevent the further spread and development of drug-resistant TB. Potential applications include the characterisation of young cultures and or sputum samples. The present study has demonstrated the potential of the methodology for sequencing DNA from young cultures, and proposed future work is aimed at further evaluating their potential for direct characterisation of DNA from sputum samples.

4.4 Future Work

While not possible within the time constraints of this project, there are some opportunities to expand the research that has been undertaken. Future work includes:

4.4.1 Adaptive Sampling

Mycobacteria can represent as little as 0.01% of the total DNA extracted from sputum samples (Votintseva *et al*., 2017), which poses a challenge for the diagnosis and treatment of TB. Oxford Nanopore recently introduced an Adaptive Sampling method that has the potential to increase the coverage of specific targets in metagenomic samples (Martin *et al*., 2022; Weilguny *et al*., 2023; Su *et al*., 2023). In principle, it enables real-time enrichment or depletion of molecules of interest from a mixed sample and enables targeted enrichment or depletion of unwanted DNA within samples directly during sequencing. Adaptive sampling compares the raw data in real-time with reference data to decide if a DNA molecule should continue to be sequenced (accepted) or removed from the pore (rejected). Each pore can reverse its voltage to reject DNA molecules and sequence another DNA molecule, increasing the sequencing capacity for molecules of interest. In studies of mock bacterial communities, enrichment is up to 13.87-fold. However, the protocol also damages the nanopores, and this reduces the sequencing output and enrichment (Martin *et al*., 2022). In practice, enrichment

of MTB has been less than 4-fold (Su *et al*., 2023). For this reason, adaptive sampling alone may be insufficient to significantly increase the coverage of MTB sequence in DNA present in sputum samples. It may have greater potential to improve the sequencing coverage of selective MDA *Mycobacterium* template since the amount of target DNAs is enriched by the selective MDA step. DNA for upper respiratory tract bacteria were found to be amplified using primers 12+5+6 in the present study. However, bacterial, and human DNA sequences were a relatively minor component of the MDA amplified template. An important future experiment will be resequencing MDA product made with the Massey primers with an adaptive sampling protocol to determine the impact of this on coverage of spiked *Mycobacterium* DNA. The protocol could either select for *Mycobacterium* DNA or select against certain bacteria. Because mismatch priming could result in off-target DNA being amplified, selecting for *Mycobacterium* DNA might be a more successful strategy. By implementing adaptive sampling in MinION sequencing for MDA product made from DNA extracted from sputum samples, there is the potential to enhance the efficiency of detecting AMR variants in a protocol suitable for rapid point of care.

4.4.2 MDA and Other Isothermal Methods for Amplification

Another important direction for future work will be evaluating MDA-ONT adaptive sampling protocols on their accuracy to correctly call different MTB AMR bacteria that are also sequenced using Illumina sequencing technology. *M. bovis* was used as a proxy of MTB in the current project. Large-scale clinical trials will be essential to validate its efficacy. A point-of-care MDA-ONT-adaptive sampling diagnostic for AMR *Mycobacterium* bacteria will need to pass such a test of proof of principle.

Further work could also be made in respect of MDA amplification conditions and the choice of MDA primers. There was limited evaluation of the P1-P15 primer set made in the current study, and given the importance of mismatch priming, further empirical analyses could be conducted. P12 could alone be investigated for its performance at higher amplification temperatures, given its higher Tm. Higher temperatures for the MDA reaction could produce greater specificity.

Other isothermal amplification methods could be considered and assessed in the future. The Loop-mediated isothermal amplification (LAMP) method has also been assessed with TB

samples. This technique can be faster and more sensitive than PCR and, like MDA, is conducted under isothermal conditions without requiring laboratory infrastructure (Ağel *et al*., 2020). This means it could be implemented in a triage fashion to initially screen larger sample numbers prior to selecting specific samples for MDA-ONT sequencing. The study by Ağel *et al*., 2020 found that the LAMP-LFD and colourimetric LAMP protocols, when optimised using sputum samples, could be used as a reliable, sensitive, rapid and specific assay in the diagnosis of TB and is a good method to consider for diagnosis in the field. However, if paired with ONT MinION sequencing, the improved reliability of detecting positives would provide for rapid field detection and diagnosis of TB with increased reliability.

Chapter 5: Conclusions

It is a necessity across the world to be able to go from the study and research of sequence data to making clinical diagnoses of patients (Goldfeder *et al*., 2016). While the ONT MinION sequencing technology is only at an early stage of implementation in limited resource settings worldwide, the potential for MinION sequencing to revolutionise research and clinical testing in these contexts is already becoming clear.

Improved speed, sensitivity and specificity: Our analysis indicates that the utilisation of selective MDA primers in conjunction with ONT MinION sequencing could substantially enhance the speed, sensitivity and specificity of *Mycobacterium* detection and diagnosis. Our results suggest that it could already be used for the characterization of young MGIT cultures (Votintseva *et al*., 2015) and may have potential application for analyses of sputum samples. These applications could revolutionise the speed at which doctors and healthcare workers identify and initiate treatment for patients with TB. This could potentially reduce transmission rates and help stop the development of further drug resistance in this bacterium.

With further developments and optimisation, the approach of using selective MDA as well as MinION sequencing could be adapted and developed for point-of-care diagnostics. This could bring TB testing closer to patients and could make a difference in regions with a high burden of TB, especially in remote areas. However, it is also important to acknowledge that while this approach shows promise, it is not without challenges, especially in resource-limited settings. Isothermal DNA amplification (MDA) and ONT rapid barcoding used in the present study are some of the simplest protocols for DNA amplification and DNA sequencing on the ONT MinION sequencing platform. Nevertheless, challenges and limitations in future implementations as a diagnostic include the need for training personnel, the potential for technical errors, and the upfront cost of consumables.

There are several promising avenues for future research, and these have been discussed. These directions are important for realising the potential of MDA-ONT sequencing. However, the combination of novel MDA primers with Oxford Nanopore MinION sequencing has the potential to revolutionise TB diagnosis. Implemented together, they offer rapid, cost-effective, and accurate results. While challenges exist, the overall promise of this approach is substantial, and it represents a significant step forward in the fight against TB and

MDR-TB. Further collaborations between researchers, healthcare workers, and policymakers will be critical for reaching the full potential of this technology in clinical practice.

References

Ağel, H. E., Sağcan, H., Ceyhan, I., & Durmaz, R. (2020). Optimization of isothermal amplification method for *Mycobacterium tuberculosis* detection and visualization method for fieldwork. *Turk J Med Sci*, *50*(4), 1069-1075. https://doi.org/10.3906/sag-1910-6

Alvarez, A. H., Estrada-Chávez, C., & Flores-Valdez, M. A. (2009). Molecular findings and approaches spotlighting *Mycobacterium bovis* persistence in cattle. *Vet Res*, *40*(3), 22. https://doi.org/10.1051/vetres/2009005

Amirkhani, A., Humayun, M., Ye, W., Worku, Y., & Yang, Z. (2021). Patient characteristics associated with different types of prison TB: an epidemiological analysis of 921 TB cases diagnosed at an Ethiopian prison. *BMC Pulm Med*, *21*(1), 334. https://doi.org/10.1186/s12890-021-01699-w

Aung, H. L., Nyunt, W. W., Fong, Y., Biggs, P. J., Winkworth, R. C., Lockhart, P. J., Yeo, T. W., Hill, P. C., Cook, G. M., & Aung, S. T. (2021). Genomic Profiling of *Mycobacterium tuberculosis* Strains, Myanmar. *Emerg Infect Dis*, *27*(11), 2847-2855. https://doi.org/10.3201/eid2711.210726

United Nations. (2016). United Nations meeting on antimicrobial resistance. *Bulletin of the World Health Organization, World Health Organization.* , *94 (9)*, 638 - 639. http://dx.doi.org/10.2471/BLT.16.020916

Bagcchi, S. (2023). WHO's Global Tuberculosis Report 2022. *Lancet Microbe*, *4*(1), e20. https://doi.org/10.1016/s2666-5247(22)00359-7

Bainomugisa, A., Duarte, T., Lavu, E., Pandey, S., Coulter, C., Marais, B. J., & Coin, L. M. (2018). A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. *Microb Genom*, *4*(7). https://doi.org/10.1099/mgen.0.000188

Barbé, L., Schaeffer, J., Besnard, A., Jousse, S., Wurtzer, S., Moulin, L., Le Guyader, F. S., & Desdouits, M. (2022). SARS-CoV-2 Whole-Genome Sequencing Using Oxford Nanopore Technology for Variant Monitoring in Wastewaters. *Front Microbiol*, *13*, 889811. https://doi.org/10.3389/fmicb.2022.889811

Ben-Selma, W., Ben-Kahla, I., Marzouk, M., Ferjeni, A., Ghezal, S., Ben-Said, M., & Boukadida, J. (2009). Rapid detection of *Mycobacterium tuberculosis* in sputum by Patho-TB kit in comparison with direct microscopy and culture. *Diagn Microbiol Infect Dis*, *65*(3), 232-235. https://doi.org/10.1016/j.diagmicrobio.2009.07.021

Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C., & Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *The Journal of Biological Chemistry*, *264*(15), 8935-8940. https://doi.org/10.1016/S0021-9258(18)81883-X. PMID 2498321.

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

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D., & Cole, S. T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*, *99*(6), 3684-3689. https://doi.org/10.1073/pnas.052548299
Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L. M., Pym, A. S., Samper, S., van Soolingen, D., & Cole, S. T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*, *99*(6), 3684-3689. https://doi.org/10.1073/pnas.052548299
Brown, A. C., Bryant, J. M., Einer-Jensen, K., Holdstock, J., Houniet, D. T., Chan, J. Z. M., Depledge, D. P., Nikolayevskyy, V., Broda, A., Stone, M. J., Christiansen, M. T., Williams, R., McAndrew, M. B., Tutill, H., Brown, J., Melzer, M., Rosmarin, C., McHugh, T. D., Shorten, R. J., . . . Breuer, J. (2015). Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples. *Journal of Clinical Microbiology*, *53*(7), 2230-2237. https://doi.org/doi:10.1128/JCM.00486-15

Brunker, K., Jaswant, G., Thumbi, S. M., Lushasi, K., Lugelo, A., Czupryna, A. M., Ade, F., Wambura, G., Chuchu, V., Steenson, R., Ngeleja, C., Bautista, C., Manalo, D. L., Gomez, M. R. R., Chu, M., Miranda, M. E., Kamat, M., Rysava, K., Espineda, J., . . . Hampson, K. (2020). Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes. *Wellcome Open Res*, *5*, 3. https://doi.org/10.12688/wellcomeopenres.15518.2

Cabibbe, A. M., Spitaleri, A., Battaglia, S., Colman, R. E., Suresh, A., Uplekar, S., Rodwell, T. C., & Cirillo, D. M. (2020). Application of Targeted Next-Generation Sequencing Assay on a Portable Sequencing Platform for Culture-Free Detection of Drug-Resistant Tuberculosis from Clinical Samples. *Journal of Clinical Microbiology*, *58*(10), e00632-00620. https://doi.org/doi:10.1128/JCM.00632-20

Clarke, E. L., Sundararaman, S. A., Seifert, S. N., Bushman, F. D., Hahn, B. H., & Brisson, D. (2017). swga: a primer design toolkit for selective whole genome amplification. *Bioinformatics*, *33*(14), 2071-2077. https://doi.org/10.1093/bioinformatics/btx118

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., . . . Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *396*(6707), 190-190. https://doi.org/10.1038/24206

Colman, R. E., Anderson, J., Lemmer, D., Lehmkuhl, E., Georghiou, S. B., Heaton, H., Wiggins, K., Gillece, J. D., Schupp, J. M., Catanzaro, D. G., Crudu, V., Cohen, T., Rodwell, T. C., & Engelthaler, D. M. (2016). Rapid Drug Susceptibility Testing of Drug-Resistant *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples by Use of Amplicon Sequencing: a Proof-of-Concept Study. *Journal of Clinical Microbiology*, *54*(8), 2058-2067. https://doi.org/doi:10.1128/JCM.00535-16

Coscolla, M., & Gagneux, S. (2014). Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Semin Immunol*, *26*(6), 431-444. https://doi.org/10.1016/j.smim.2014.09.012

Das, P. K., Ganguly, S. B., & Mandal, B. (2019). Sputum smear microscopy in tuberculosis: It is still relevant in the era of molecular diagnosis when seen from the public health perspective. *Biomedical and Biotechnology Research Journal (BBRJ)*, *3*(2), 77-79.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S. F., Egholm, M., & Lasken, R. S. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A*, *99*(8), 5261-5266. https://doi.org/10.1073/pnas.082089499

del Solar, G., Giraldo, R., Ruiz-Echevarria, M. J., Espinosa, M., & Diaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev*, *62*(2), 434-464. https://doi.org/10.1128/MMBR.62.2.434-464.1998

Dezemon, Z., Muvunyi, C. M., & Jacob, O. (2014). Staining techniques for detection of acid fast bacilli: what hope does fluorescein-diacetate (FDA) vitality staining technique represent for the monitoring of tuberculosis treatment in resource limited settings. *Trends in Bacteriology*, *1*(1). https://doi.org/10.7243/2057-4711-1-1

Dippenaar, A., Goossens, S. N., Grobbelaar, M., Oostvogels, S., Cuypers, B., Laukens, K., Meehan, C. J., Warren, R. M., & Rie, A. v. (2022). Nanopore Sequencing for *Mycobacterium tuberculosis*: a Critical Review of the Literature, New Developments, and Future Opportunities. *Journal of Clinical Microbiology*, *60*(1), e00646-00621. https://doi.org/doi:10.1128/JCM.00646-21

Dookie, N., Naidoo, K., & Padayatchi, N. (2018). Whole-Genome Sequencing To Guide the Selection of Treatment for Drug-Resistant Tuberculosis. *Antimicrob Agents Chemother*, *62*(8). https://doi.org/10.1128/aac.00574-18

Doyle, R. M., Burgess, C., Williams, R., Gorton, R., Booth, H., Brown, J., Bryant, J. M., Chan, J., Creer, D., Holdstock, J., Kunst, H., Lozewicz, S., Platt, G., Romero, E. Y., Speight, G., Tiberi, S., Abubakar, I., Lipman, M., McHugh, T. D., & Breuer, J. (2018). Direct Whole-Genome Sequencing of Sputum Accurately Identifies Drug-Resistant *Mycobacterium tuberculosis* Faster than MGIT Culture Sequencing. *J Clin Microbiol*, *56*(8). https://doi.org/10.1128/jcm.00666-18

Duffy, S. C., Srinivasan, S., Schilling, M. A., Stuber, T., Danchuk, S. N., Michael, J. S., Venkatesan, M., Bansal, N., Maan, S., Jindal, N., Chaudhary, D., Dandapat, P., Katani, R., Chothe, S., Veerasami, M., Robbe-Austerman, S., Juleff, N., Kapur, V., & Behr, M. A. (2020). Reconsidering *Mycobacterium bovis* as a proxy for zoonotic tuberculosis: a molecular epidemiological surveillance study. *Lancet Microbe*, *1*(2), e66-e73. https://doi.org/10.1016/S2666-5247(20)30038-0

Fan, H., Wang, J., Komiyama, M., & Liang, X. (2019). Effects of secondary structures of DNA templates on the quantification of qPCR. *J Biomol Struct Dyn*, *37*(11), 2867-2874. https://doi.org/10.1080/07391102.2018.1498804

Fox, W. S., Strydom, N., Imperial, M. Z., Jarlsberg, L., & Savic, R. M. (2023). Examining nonadherence in the treatment of tuberculosis: The patterns that lead to failure. *Br J Clin Pharmacol*, *89*(7), 1965-1977. https://doi.org/10.1111/bcp.15515

Freed, N. E., Vlková, M., Faisal, M. B., & Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding . *Biol Methods Protoc*, *5*(1), bpaa014. https://doi.org/10.1093/biomethods/bpaa014

Garmendia, C., Bernad, A., Esteban, J. A., Blanco, L., & Salas, M. (1992). The bacteriophage phi 29 DNA polymerase, a proofreading enzyme. *J Biol Chem*, *267*(4), 2594-2599.

Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrell, B. G., . . . Hewinson, R. G. (2003). The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A*, *100*(13), 7877-7882. https://doi.org/10.1073/pnas.1130426100

George, S., Xu, Y., Sanderson, N., Hubbard, A. T., Griffiths, D. T., Morgan, M., Pankhurst, L., Hoosdally, S. J., Foster, D., Thulborn, S., Robinson, E., Grace Smith, E., Rathod, P., Sarah Walker, A., Peto, T. E. A., Crook, D. W., & Dingle, K. E. (2018). MinION Nanopore Sequencing of Multiple Displacement Amplified Mycobacteria DNA Direct from Sputum. *bioRxiv*, 490417. https://doi.org/10.1101/490417

George, S., Xu, Y., Sanderson, N., Hubbard, A. T., Griffiths, D. T., Morgan, M., Pankhurst, L., Hoosdally, S. J., Foster, D., Thulborn, S., Robinson, E., Grace Smith, E., Rathod, P., Sarah Walker, A., Peto, T. E. A., Crook, D. W., & Dingle, K. E. (2018). MinION Nanopore Sequencing of Multiple Displacement Amplified Mycobacteria DNA Direct from Sputum. *bioRxiv*, 490417. https://doi.org/10.1101/490417

Gkika, E., Psaroulaki, A., Tselentis, Y., Angelakis, E., & Kouikoglou, V. S. (2019). Can point-of-care testing shorten hospitalization length of stay? An exploratory investigation of infectious agents using regression modelling. *Health Informatics J*, *25*(4), 1606-1617. https://doi.org/10.1177/1460458218796612

Health, M. o. (2019). *Guidelines for Tuberculosis Control in New Zealand*. Wellington: Ministry of Health.

Health, M. o. (2022). *Communicable Disease Control Manual: Tuberculosis*. Ministry of Health. Retrieved 23 August from https://www.health.govt.nz/our-work/diseases-and-conditions/communicable-disease-control-manual/tuberculosis

Hillemann, D., Hoffner, S., Cirillo, D., Drobniewski, F., Richter, E., Rusch-Gerdes, S., Baltic-Nordic, T. B. L. N., Tb, P.-N., & Networks, E. E.-T. (2013). First evaluation after implementation of a quality control system for the second line drug susceptibility testing of *Mycobacterium tuberculosis* joint efforts in low and high incidence countries. *PLoS One*, *8*(10), e76765. https://doi.org/10.1371/journal.pone.0076765

Htun, K. S., Fong, Y., Kyaw, A. A., Aung, S. T., Oo, K. Z., Zaw, T., Lockhart, P. J., Russell, B., Cook, G. M., Aung, H. L., & Hlaing, T. M. (2018). Microbiome dataset from the upper respiratory tract of patients living with HIV, HIV/TB and TB from Myanmar. *Data Brief*, *21*, 354-357. https://doi.org/10.1016/j.dib.2018.10.003

Huang, L., Ma, F., Chapman, A., Lu, S., & Xie, X. S. (2015). Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annu Rev Genomics Hum Genet*, *16*, 79-102. https://doi.org/10.1146/annurev-genom-090413-025352

Ince, D., & Hooper, D. C. (2003). Quinolone resistance due to reduced target enzyme expression. *J Bacteriol*, *185*(23), 6883-6892. https://doi.org/10.1128/jb.185.23.6883-6892.2003

Jang, J. G., & Chung, J. H. (2020). Diagnosis and treatment of multidrug-resistant tuberculosis. *Yeungnam Univ J Med*, *37*(4), 277-285. https://doi.org/10.12701/yujm.2020.00626

Javaid, A., Ullah, I., Masud, H., Basit, A., Ahmad, W., Butt, Z. A., & Qasim, M. (2018). Predictors of poor treatment outcomes in multidrug-resistant tuberculosis patients: a retrospective cohort study. *Clin Microbiol Infect*, *24*(6), 612-617. https://doi.org/10.1016/j.cmi.2017.09.012

Ji, H., Xu, J., Wu, R., Chen, X., Lv, X., Liu, H., Duan, Y., Sun, M., Pan, Y., Chen, Y., Lu, X., & Zhou, L. (2021). Cut-off Points of Treatment Delay to Predict Poor Outcomes Among New Pulmonary Tuberculosis Cases in Dalian, China: A Cohort Study. *Infect Drug Resist*, *14*, 5521-5530. https://doi.org/10.2147/idr.S346375

Kayigire, X. A., Friedrich, S. O., van der Merwe, L., & Diacon, A. H. (2017). Acquisition of Rifampin Resistance in Pulmonary Tuberculosis. *Antimicrob Agents Chemother*, *61*(4). https://doi.org/10.1128/aac.02220-16

Kivihya-Ndugga, L., van Cleeff, M., Juma, E., Kimwomi, J., Githui, W., Oskam, L., Schuitema, A., van Soolingen, D., Nganga, L., Kibuga, D., Odhiambo, J., & Klatser, P. (2004). Comparison of PCR with the routine procedure for diagnosis of tuberculosis in a population with high prevalences of tuberculosis and human immunodeficiency virus. *J Clin Microbiol*, *42*(3), 1012-1015. https://doi.org/10.1128/jcm.42.3.1012-1015.2004

Kontsevaya, I., Cabibbe, A. M., Cirillo, D. M., DiNardo, A. R., Frahm, N., Gillespie, S. H., Holtzman, D., Meiwes, L., Petruccioli, E., Reimann, M., Ruhwald, M., Sabiiti, W., Saluzzo, F., Tagliani, E., & Goletti, D. (2023). Update on the diagnosis of tuberculosis. *Clin Microbiol Infect*. https://doi.org/10.1016/j.cmi.2023.07.014

Liu, Z., Yang, Y., Wang, Q., Wang, L., Nie, W., & Chu, N. (2023). Diagnostic value of a nanopore sequencing assay of bronchoalveolar lavage fluid in pulmonary tuberculosis. *BMC Pulmonary Medicine*, *23*(1), 77. https://doi.org/10.1186/s12890-023-02337-3

Long, N., Qiao, Y., Xu, Z., Tu, J., & Lu, Z. (2020). Recent advances and application in whole-genome multiple displacement amplification. In *Quantitative Biology* (Vol. 8, pp. 279-294).

Loose, M., Malla, S., & Stout, M. (2016). Real-time selective sequencing using nanopore technology. *Nat Methods*, *13*(9), 751-754. https://doi.org/10.1038/nmeth.3930

Mai, M., Hoyer, J. D., & McClure, R. F. (2004). Use of multiple displacement amplification to amplify genomic DNA before sequencing of the alpha and beta haemoglobin genes. *J Clin Pathol*, *57*(6), 637-640. https://doi.org/10.1136/jcp.2003.014704

Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., & Leggett, R. M. (2021). Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *bioRxiv*, 2021.2005.2007.443191. https://doi.org/10.1101/2021.05.07.443191

Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., & Leggett, R. M. (2022). Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol*, *23*(1), 11. https://doi.org/10.1186/s13059-021-02582-x

Mashalla, Y., Setlhare, V., Massele, A., Sepako, E., Tiroyakgosi, C., Kgatlwane, J., Chuma, M., & Godman, B. (2017). Assessment of prescribing practices at the primary healthcare facilities in Botswana with an emphasis on antibiotics: Findings and implications. *Int J Clin Pract*, *71*(12). https://doi.org/10.1111/ijcp.13042

Mathew, P., Kuo, Y. H., Vazirani, B., Eng, R. H., & Weinstein, M. P. (2002). Are three sputum acid-fast bacillus smears necessary for discontinuing tuberculosis isolation? *J Clin Microbiol*, *40*(9), 3482-3484. https://doi.org/10.1128/jcm.40.9.3482-3484.2002

Minero, G. A. S., Bagnasco, M., Fock, J., Tian, B., Garbarino, F., & Hansen, M. F. (2020). Automated on-chip analysis of tuberculosis drug-resistance mutation with integrated DNA

ligation and amplification. *Anal Bioanal Chem*, *412*(12), 2705-2710. https://doi.org/10.1007/s00216-020-02568-x

Mongan, A. E., Tuda, J. S. B., & Runtuwene, L. R. (2020). Portable sequencer in the fight against infectious disease. *J Hum Genet*, *65*(1), 35-40. https://doi.org/10.1038/s10038-019-0675-4

Ni, Y., Liu, X., Simeneh, Z. M., Yang, M., & Li, R. (2023). Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput Struct Biotechnol J*, *21*, 2352-2364. https://doi.org/10.1016/j.csbj.2023.03.038

Niemann, S., & Supply, P. (2014). Diversity and evolution of *Mycobacterium tuberculosis*: moving to whole-genome-based approaches. *Cold Spring Harb Perspect Med*, *4*(12), a021188. https://doi.org/10.1101/cshperspect.a021188

Nikaido, H. (2009). Multidrug resistance in bacteria. *Annu Rev Biochem*, *78*, 119-146. https://doi.org/10.1146/annurev.biochem.78.082907.145923

Pandey, V., Singh, P., Singh, S., Arora, N., Quadir, N., Singh, S., Das, A., Dudeja, M., Kapur, P., Ehtesham, N. Z., Elangovan, R., & Hasnain, S. E. (2019). SeeTB: A novel alternative to sputum smear microscopy to diagnose tuberculosis in high burden countries. *Scientific Reports*, *9*(1), 16371. https://doi.org/10.1038/s41598-019-52739-9

Payne, A., Holmes, N., Rakyan, V., & Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, *35*(13), 2193-2198. https://doi.org/10.1093/bioinformatics/bty841

Pendleton, K. M., Erb-Downward, J. R., Bao, Y., Branton, W. R., Falkowski, N. R., Newton, D. W., Huffnagle, G. B., & Dickson, R. P. (2017). Rapid Pathogen Identification in Bacterial Pneumonia Using Real-Time Metagenomics. *Am J Respir Crit Care Med*, *196*(12), 1610-1612. https://doi.org/10.1164/rccm.201703-0537LE

Peters, J. S., Ismail, N., Dippenaar, A., Ma, S., Sherman, D. R., Warren, R. M., & Kana, B. D. (2020). Genetic Diversity in *Mycobacterium tuberculosis* Clinical Isolates and Resulting Outcomes of Tuberculosis Infection and Disease. *Annu Rev Genet*, *54*, 511-537. https://doi.org/10.1146/annurev-genet-022820-085940

Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., & Tsongalis, G. J. (2019). Third-Generation Sequencing in the Clinical Laboratory: Exploring the Advantages and Challenges of Nanopore Sequencing. *Journal of Clinical Microbiology*, *58*(1), 10.1128/jcm.01315-01319. https://doi.org/doi:10.1128/jcm.01315-19

Centres for Disease Control and Prevention. (2021). *Be Antibiotics Aware: Smart Use, Best Care*. https://www.cdc.gov/patientsafety/features/be-antibiotics-aware.html

Ratnatunga, C. N., Lutzky, V. P., Kupz, A., Doolan, D. L., Reid, D. W., Field, M., Bell, S. C., Thomson, R. M., & Miles, J. J. (2020). The Rise of Non-Tuberculosis Mycobacterial Lung Disease. *Front Immunol*, *11*, 303. https://doi.org/10.3389/fimmu.2020.00303

Raymond, N. G., Adegbele, J., Coulibaly, I., Kouame-N'takpé, N., Seck-Angu, H., Guei, A., Kouakou, J., & Dosso, M. (2017). Clinical Performances of Pure TB-Lamp Kit for M. tuberculosis Complex Detection in Sputum Samples. *Journal of Tuberculosis Research*, *5*, 129-138. https://doi.org/10.4236/jtr.2017.52014

Robicsek, A., Strahilevitz, J., Jacoby, G. A., Macielag, M., Abbanat, D., Park, C. H., Bush, K., & Hooper, D. C. (2006). Fluoroquinolone-modifying enzyme: a new adaptation of a common aminoglycoside acetyltransferase. *Nat Med*, *12*(1), 83-88. https://doi.org/10.1038/nm1347

Sandegren, L., & Andersson, D. I. (2009). Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol*, *7*(8), 578-588. https://doi.org/10.1038/nrmicro2174

Sanoussi, C. N., Coscolla, M., Ofori-Anyinam, B., Otchere, I. D., Antonio, M., Niemann, S., Parkhill, J., Harris, S., Yeboah-Manu, D., Gagneux, S., Rigouts, L., Affolabi, D., de Jong, B. C., & Meehan, C. J. (2021). *Mycobacterium tuberculosis* complex lineage 5 exhibits high levels of within-lineage genomic diversity and differing gene content compared to the type strain H37Rv. *Microb Genom*, *7*(7). https://doi.org/10.1099/mgen.0.000437

ThermoFischer Scientific. (2023). *EquiPhi29™ DNA Polymerase*. Retrieved 10 August from https://www.thermofisher.com/order/catalog/product/A39390

National Health Service. (2019). *Diagnosis: Tuberculosis (TB)*. Retrieved September 25 from https://www.nhs.uk/conditions/tuberculosis-tb/diagnosis/#:~:text=The%20interferon%20gamma%20release%20assay,be%20reliable%20in%20these%20cases

Sharma, S. K., Kohli, M., Yadav, R. N., Chaubey, J., Bhasin, D., Sreenivas, V., Sharma, R., & Singh, B. K. (2015). Evaluating the Diagnostic Accuracy of Xpert MTB/RIF Assay in Pulmonary Tuberculosis. *PLoS One*, *10*(10), e0141011. https://doi.org/10.1371/journal.pone.0141011

Siddiqi, K., Lambert, M. L., & Walley, J. (2003). Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: the current evidence. *Lancet Infect Dis*, *3*(5), 288-296. https://doi.org/10.1016/s1473-3099(03)00609-1

Skerra, A. (1992). Phosphorothioate primers improve the amplification of DNA sequences by DNA polymerases with proofreading activity. *Nucleic Acids Res*, *20*(14), 3551-3554. https://doi.org/10.1093/nar/20.14.3551

Stevens, B. M., Creed, T. B., Reardon, C. L., & Manter, D. K. (2023). Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Scientific Reports*, *13*(1), 9323. https://doi.org/10.1038/s41598-023-36101-8

Sturdy, A., Goodman, A., Jose, R. J., Loyse, A., O'Donoghue, M., Kon, O. M., Dedicoat, M. J., Harrison, T. S., John, L., Lipman, M., & Cooke, G. S. (2011). Multidrug-resistant tuberculosis (MDR-TB) treatment in the UK: a study of injectable use and toxicity in practice. *J Antimicrob Chemother*, *66*(8), 1815-1820. https://doi.org/10.1093/jac/dkr221

Su, J., Lui, W. W., Lee, Y., Zheng, Z., Siu, G. K., Ng, T. T., Zhang, T., Lam, T. T., Lao, H. Y., Yam, W. C., Tam, K. K., Leung, K. S., Lam, T. W., Leung, A. W., & Luo, R. (2023). Evaluation of *Mycobacterium tuberculosis* enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Sci Rep*, *13*(1), 5237. https://doi.org/10.1038/s41598-023-32378-x

Oxford Nanopore Technologies. (2023). *Nanopore Sequencing Accuracy* Retrieved 28 March from https://nanoporetech.com/accuracy

Tyler, A. D., Christianson, S., Knox, N. C., Mabon, P., Wolfe, J., Van Domselaar, G., Graham, M. R., & Sharma, M. K. (2016). Comparison of Sample Preparation Methods Used for the Next-Generation Sequencing of *Mycobacterium tuberculosis*. *PLoS One*, *11*(2), e0148676. https://doi.org/10.1371/journal.pone.0148676

van Belkum, A., Burnham, C. D., Rossen, J. W. A., Mallard, F., Rochas, O., & Dunne, W. M., Jr. (2020). Innovative and rapid antimicrobial susceptibility testing systems. *Nat Rev Microbiol*, *18*(5), 299-311. https://doi.org/10.1038/s41579-020-0327-x

Votintseva, A. A., Bradley, P., Pankhurst, L., Del Ojo Elias, C., Loose, M., Nilgiriwala, K., Chatterjee, A., Smith, E. G., Sanderson, N., Walker, T. M., Morgan, M. R., Wyllie, D. H., Walker, A. S., Peto, T. E. A., Crook, D. W., & Iqbal, Z. (2017). Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. *J Clin Microbiol*, *55*(5), 1285-1298. https://doi.org/10.1128/JCM.02483-16

Votintseva, A. A., Pankhurst, L. J., Anson, L. W., Morgan, M. R., Gascoyne-Binzi, D., Walker, T. M., Quan, T. P., Wyllie, D. H., Elias, C. D. O., Wilcox, M., Walker, A. S., Peto, T. E. A., & Crook, D. W. (2015). Mycobacterial DNA Extraction for Whole-Genome Sequencing from Early Positive Liquid (MGIT) Cultures. *Journal of Clinical Microbiology*, *53*(4), 1137-1143. https://doi.org/doi:10.1128/jcm.03073-14

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, *39*(11), 1348-1365. https://doi.org/10.1038/s41587-021-01108-x

Wasswa, F. B., Kassaza, K., Nielsen, K., & Bazira, J. (2022). MinION Whole-Genome Sequencing in Resource-Limited Settings: Challenges and Opportunities. *Current Clinical Microbiology Reports*, *9*(4), 52-59. https://doi.org/10.1007/s40588-022-00183-1

Weilguny, L., Maio, N. D., Munro, R., Manser, C., Birney, E., Loose, M., & Goldman, N. (2022). Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *bioRxiv*, 2020.2002.2007.938670. https://doi.org/10.1101/2020.02.07.938670

Witney, A. A., Cosgrove, C. A., Arnold, A., Hinds, J., Stoker, N. G., & Butcher, P. D. (2016). Clinical use of whole genome sequencing for *Mycobacterium tuberculosis*. *BMC Med*, *14*, 46. https://doi.org/10.1186/s12916-016-0598-2

Witney, A. A., Gould, K. A., Arnold, A., Coleman, D., Delgado, R., Dhillon, J., Pond, M. J., Pope, C. F., Planche, T. D., Stoker, N. G., Cosgrove, C. A., Butcher, P. D., Harrison, T. S., & Hinds, J. (2015). Clinical application of whole-genome sequencing to inform treatment for multidrug-resistant tuberculosis cases. *J Clin Microbiol*, *53*(5), 1473-1483. https://doi.org/10.1128/JCM.02993-14

World Health Organization. (2021). *Tuberculosis Fact Sheet* Retrieved 25 August from https://www.who.int/news-room/fact-sheets/detail/tuberculosis

World Health Organization. (2021). *Global tuberculosis report 2021*. Geneva. https://www.who.int/publications/i/item/9789240037021

World Health Organization. (2023). Use of targeted next-generation sequencing to detect drug-resistant tuberculosis. Rapid Communication, July 2023c

Yakovleva, A., Kovalenko, G., Redlinger, M., Liulchuk, M. G., Bortz, E., Zadorozhna, V. I., Scherbinska, A. M., Wertheim, J. O., Goodfellow, I., Meredith, L., & Vasylyeva, T. I. (2021). Tracking SARS-COV-2 Variants Using Nanopore Sequencing in Ukraine in Summer 2021. *Res Sq*. https://doi.org/10.21203/rs.3.rs-1044446/v1

Yang, Y., Che, Y., Liu, L., Wang, C., Yin, X., Deng, Y., Yang, C., & Zhang, T. (2022). Rapid absolute quantification of pathogens and ARGs by nanopore sequencing. *Sci Total Environ*, *809*, 152190. https://doi.org/10.1016/j.scitotenv.2021.152190

Zhang, Y., & Yew, W. W. (2009). Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis*, *13*(11), 1320-1330. https://www.ncbi.nlm.nih.gov/pubmed/19861002

Zhao, K., Tu, C., Chen, W., Liang, H., Zhang, W., Wang, Y., Jin, Y., Hu, J., Sun, Y., Xu, J., & Yu, Y. (2022). Rapid Identification of Drug-Resistant Tuberculosis Genes Using Direct PCR Amplification and Oxford Nanopore Technology Sequencing. *Can J Infect Dis Med Microbiol*, *2022*, 7588033. https://doi.org/10.1155/2022/7588033

Appendix I

MDA Binding Site Distribution Maps For Disease Causing Mycobacteria

The *Mycobacterium tuberculosis* complex consists of a group of organisms that are of chief clinical importance in causing TB in both humans and animals and have previously been thought to be highly conserved. As a result, any variation would not be considered clinically significant (Coscolla & Gagneux., 2014). This group includes *M. tuberculosis, M. bovis, M. caprae, M.canettii, and M microti.* MTB, the leading cause of tuberculosis in humans, is comprised of considerable diversity within its genetic lineages. Despite their close relationship genetically, these mycobacteria differ pathogenetically, epidemiologically, and geographically in host preference and severity of human disease. In genetic similarity, all the members of the MAC complex are highly conserved, with a similarity of 99.9% at the nucleotide level and identical 16S rRNA sequences (Peters *et al*., 2020).

Mycobacterium avium-intracellular complex (MAC) is a common environmental pathogen whose pathogenicity ranges from colonisation to disease, especially in immunocompromised individuals. The species that make up the MAC complex (*M.intracellulare* and *M.avium*) are a group of closely related mycobacterial species to MTB that share similar characteristics and cause similar infections.

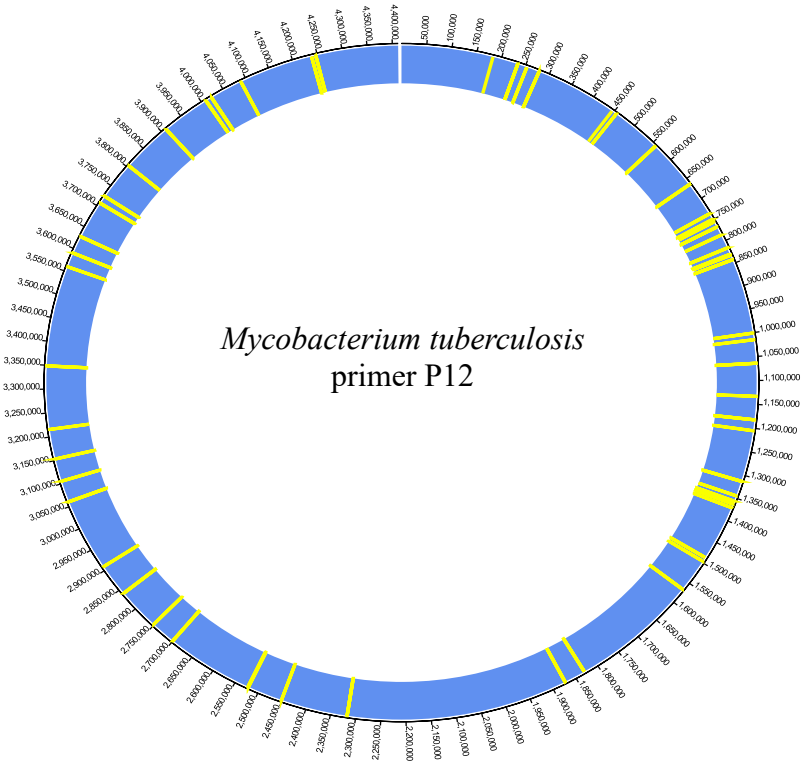
Other *Mycobacterium* species that could be suitable for further exploration of the Massey MDA primers are those that are closely related and genetically similar to MTB. This ensures that the amplified DNA accurately represents the target species, as the closer species are genetically, the more likely the MDA will produce reliable and reproducible results.

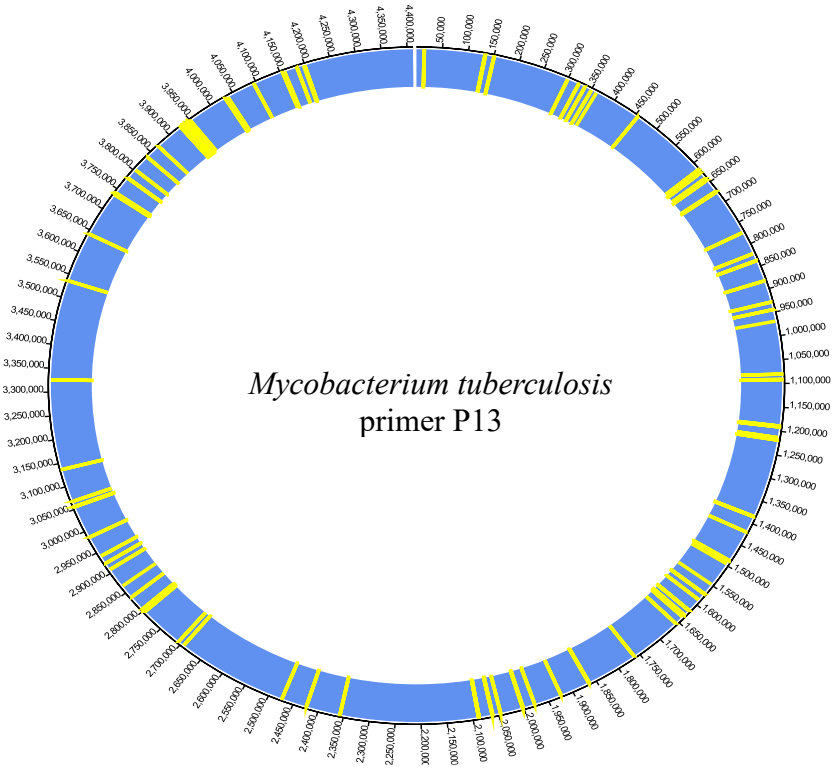
It is important recognise the close evolutionary relationship and shared ancestry of *Mycobacterium bovis* and *Mycobacterium tuberculosis.* Both MTB and *M.bovis* belong to the *Mycobacterium tuberculosis* complex, and they also share a common ancestor. The CIRCA plots of both species highlights their genetic similarities, with a very similar level of primer binding sites seen in each, implying a significant degree of genetic similarity. This makes *M.bovis* a good proxy for genomic studies on MTB. As they belong to the same lineage and share a recent common ancestor, any insights gained from studying *M.bovis* are likely to be applicable to MTB, too.

It is crucial, however, to note that even though they are closely related, there are still notable genetic and phenotypic differences between *M.bovis* and MTB, differences that could have implications for certain aspects of genomic studies (such as drug susceptibility profiles). Unlike MTB, which only causes disease in humans, *M. bovis* is a zoonotic pathogen that can cause disease in humans and animals.

Mycobacterium strains belonging to *the M. tuberculosis* MAC complex include *M. tuberculosis, M. bovis, M. microti*, *M. canettii, M. intracellulare and M. avium.* They are all organisms that are capable of causing disease in humans. The MDA binding site distribution maps for these species were mapped to examine the specificity and diagnostic potential of the Massey primers. Maps for *M. tuberculosis* and *M. bovis* were constructed for forward and reverse high binder primer sequences: P12, P13, P14 and P15. P12 + P5 + P6 were mapped onto the genomes of *M. tuberculosis, M. bovis, M. microti*, *M. canettii, M. intracellulare and M. avium* using a Python script written for this purpose that produced a table that could be loaded into CIRCA (Appendix II).

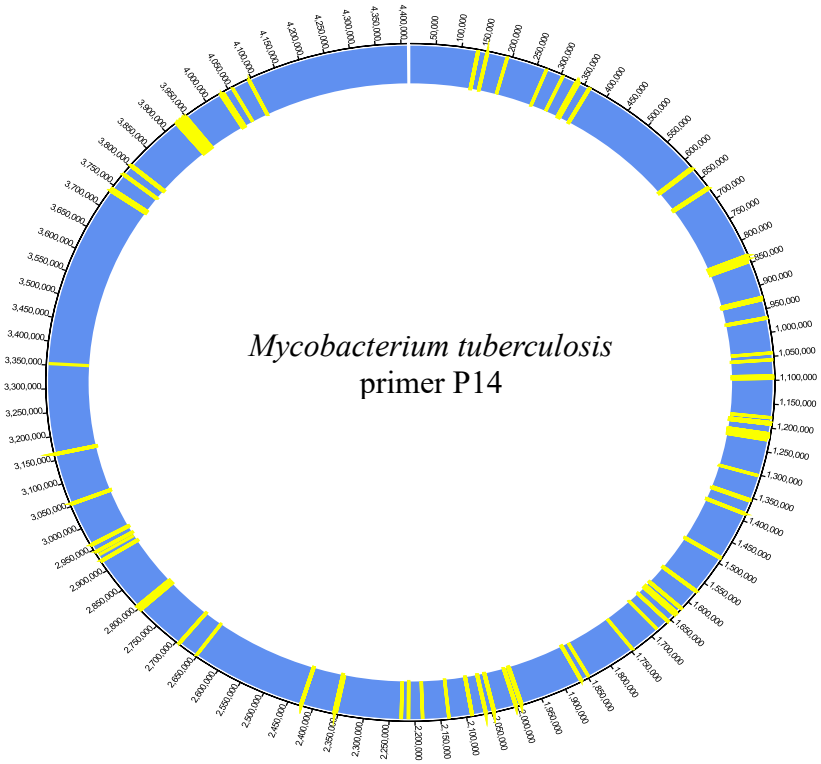
Appendix I Figure 1

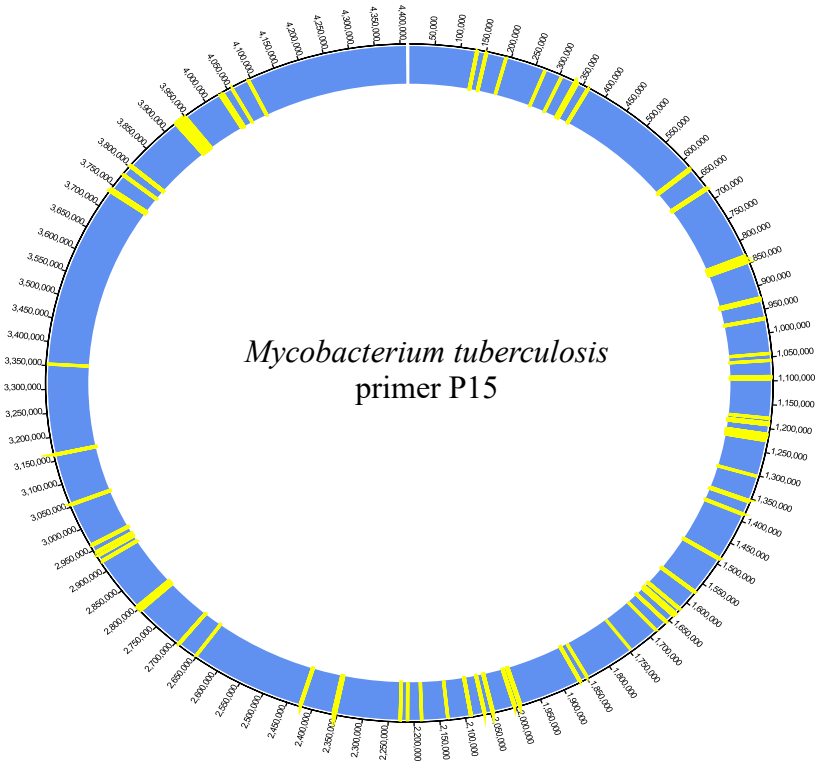




Appendix I Figure 2

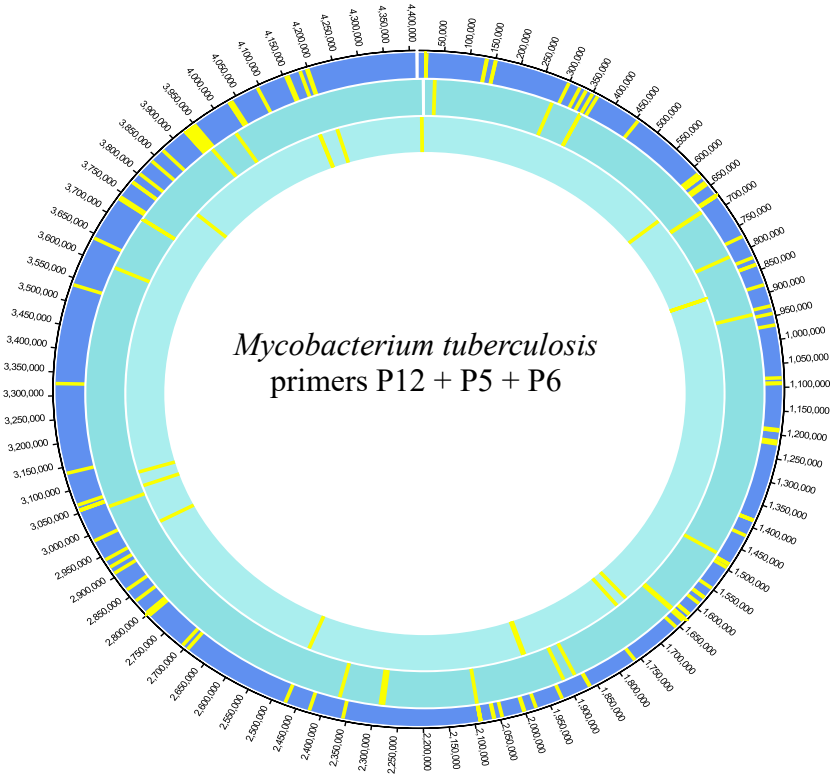
Appendix I Figure 3

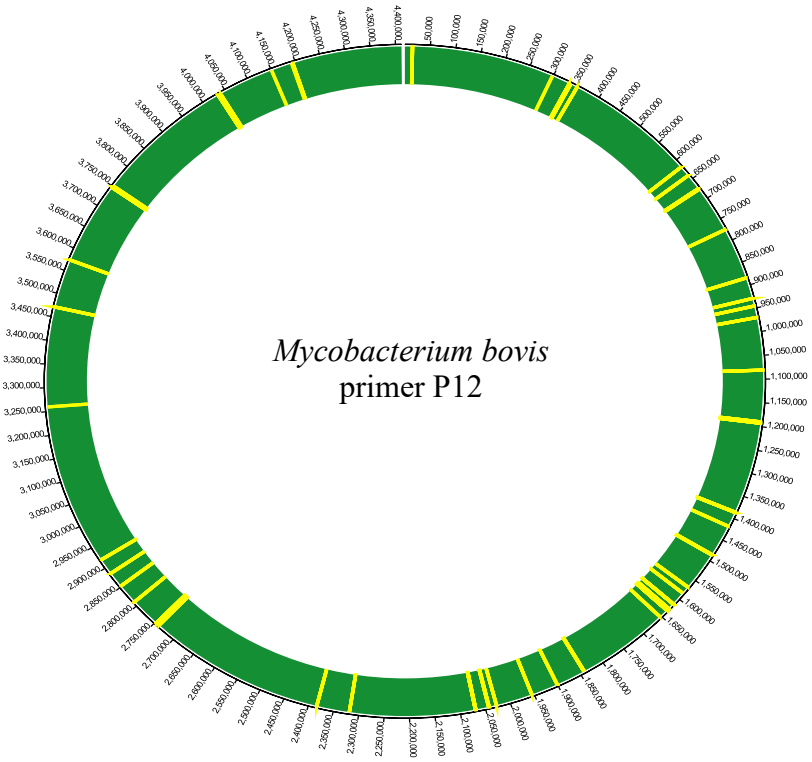




Appendix I Figure 4

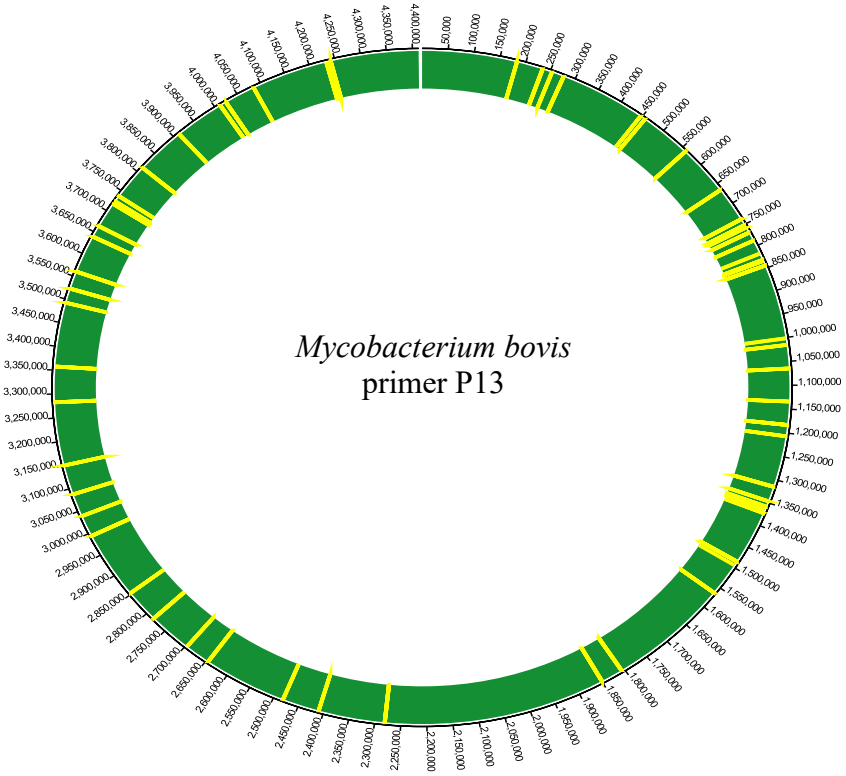
Appendix I Figure 5

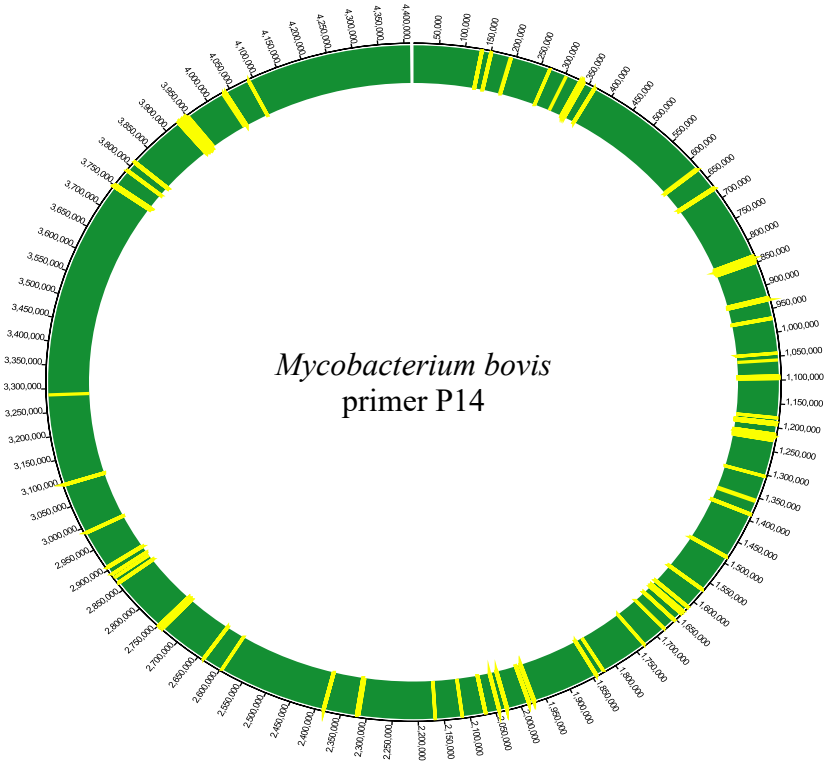




Appendix I Figure 6

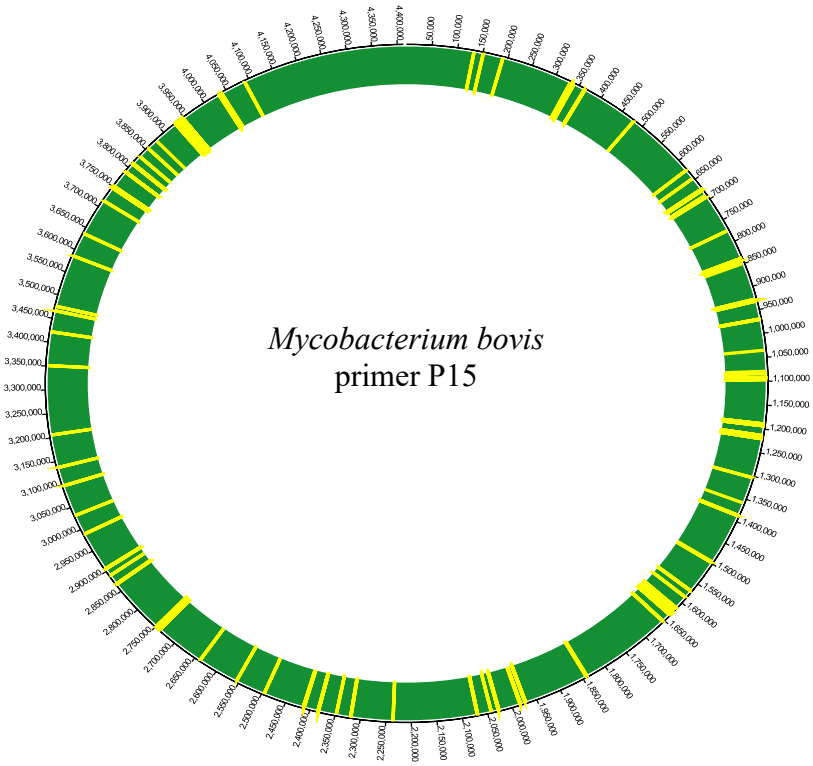
Appendix I Figure 7

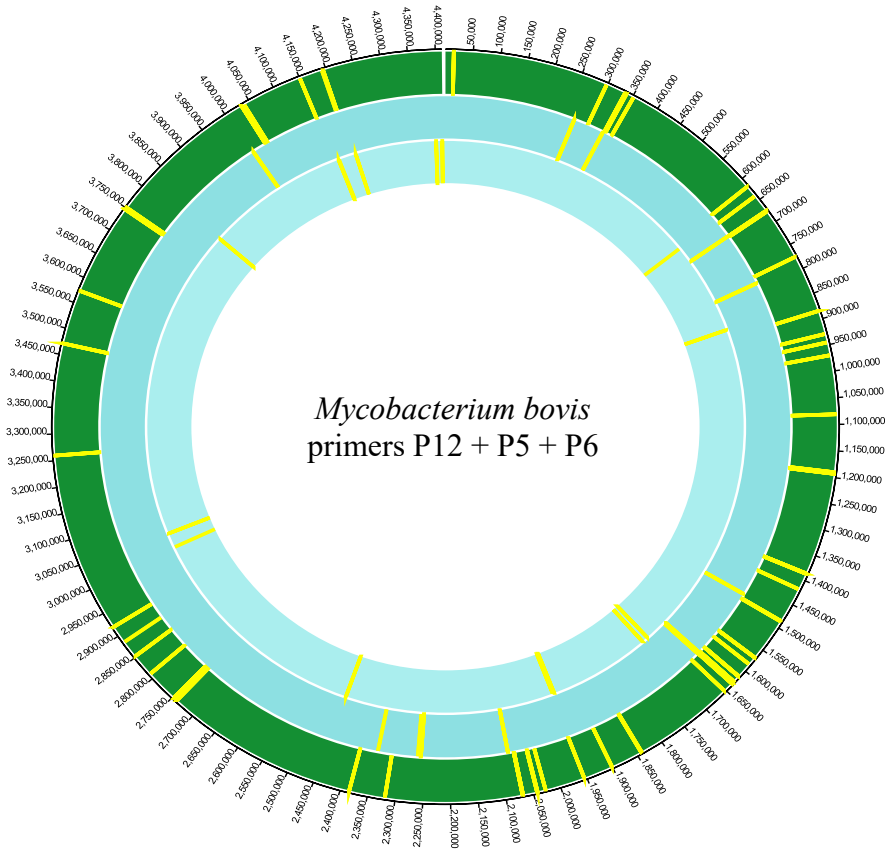




Appendix I Figure 8

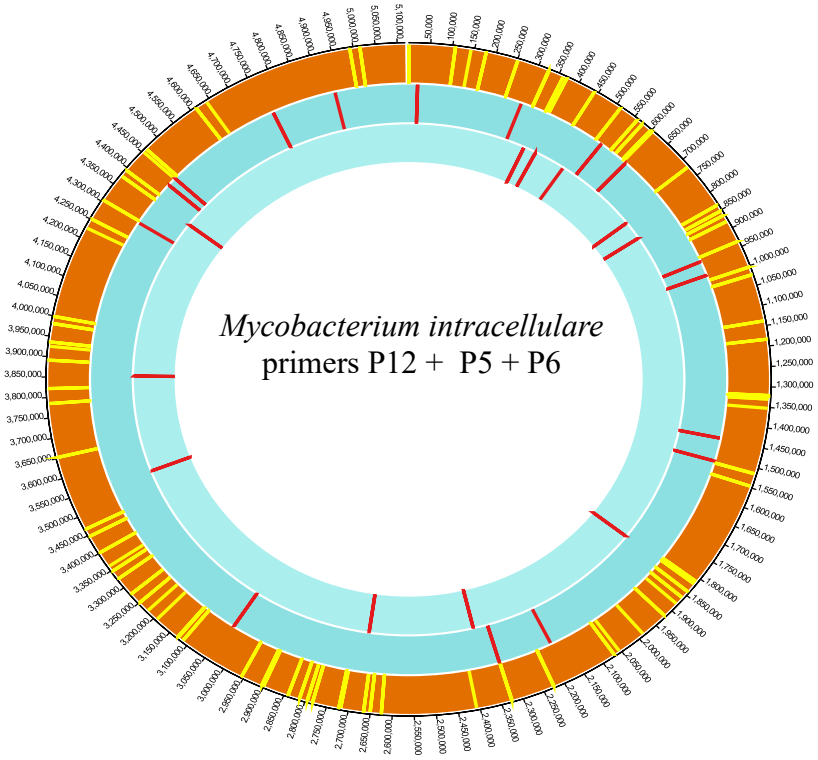
Appendix I Figure 9

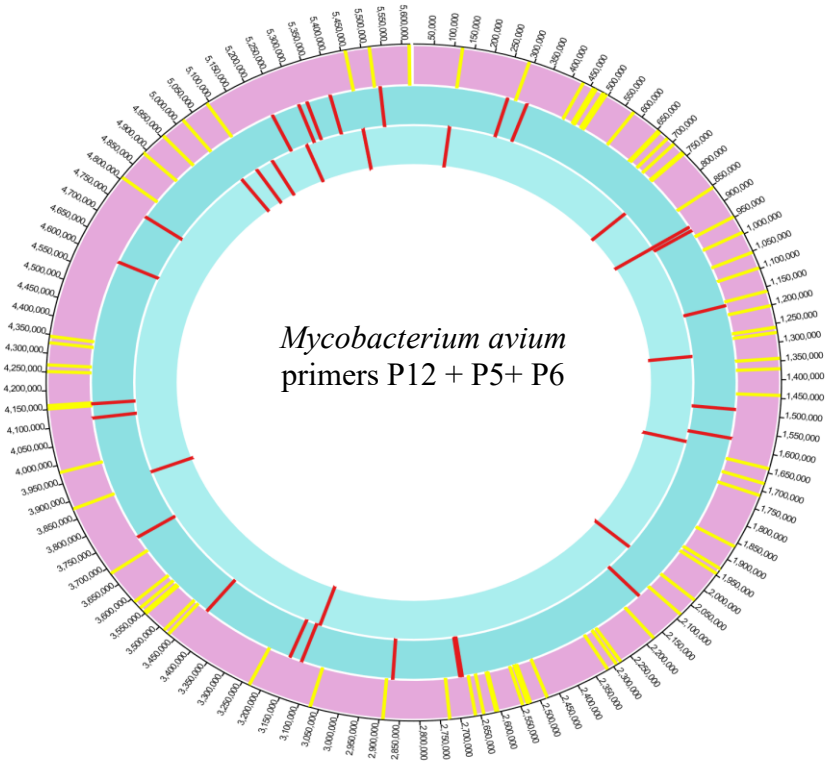




Appendix I Figure 10

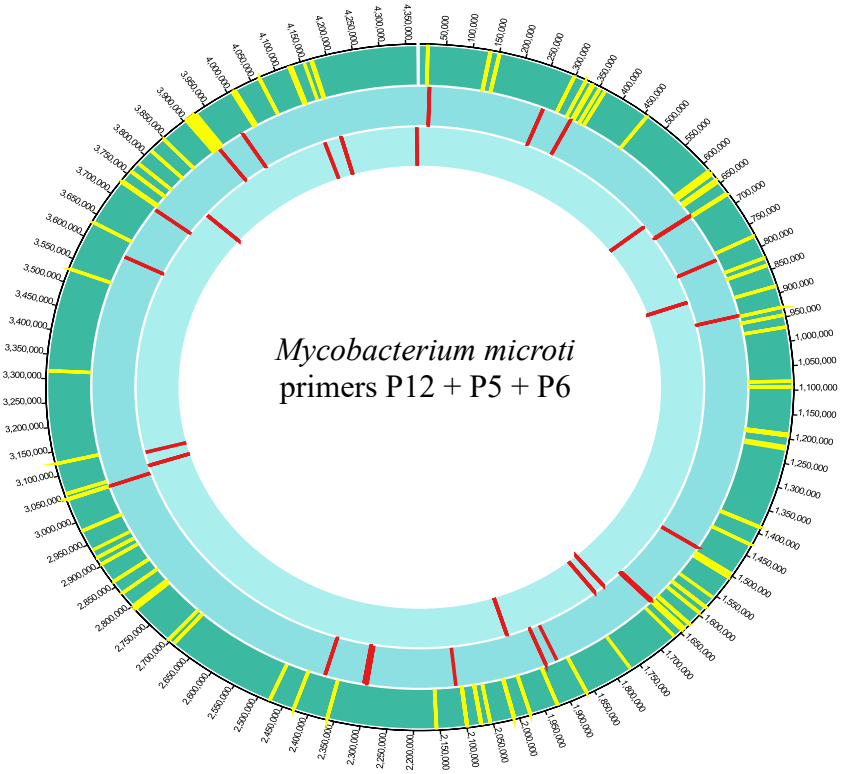
Appendix I Figure 11

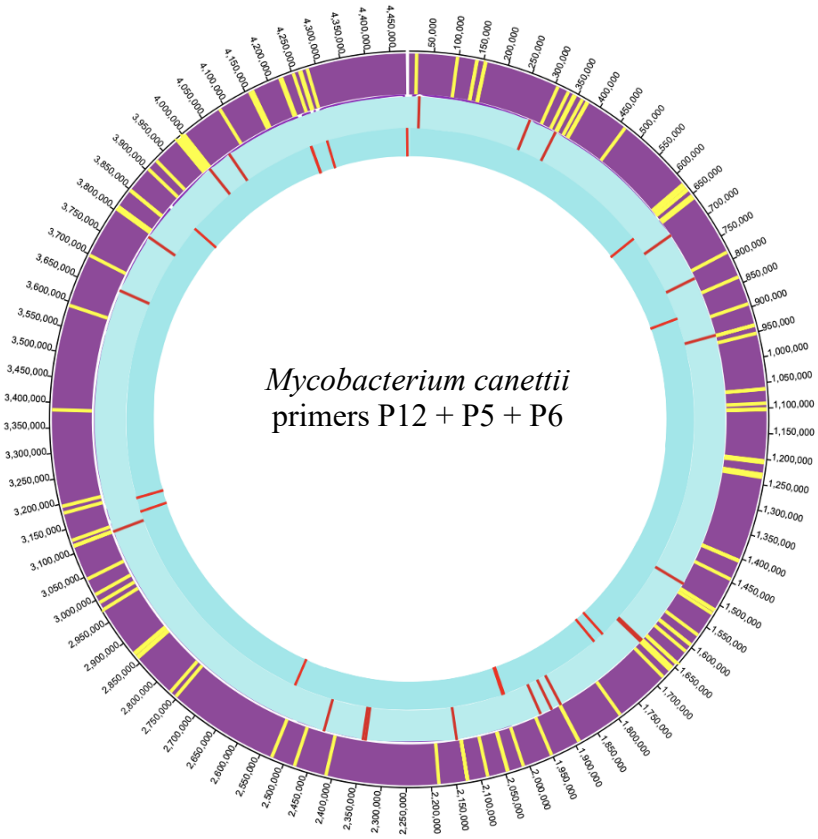




Appendix I Figure 12

Appendix I Figure 13





Appendix I Figure 14

Legend for figures Appendix 1: 1 -14:

The figures illustrate the coverage of different Massey primers across various *Mycobacterium* genomes. The circular layout allows for a holistic view of the genomic landscape and the specific regions covered by each primer. The yellow or red lines on each of the plots indicates regions where there is coverage of the *Mycobacterium* species genome from the specified primer. Where more than one track is present, Primer P5 is the middle track and Primer P6 is the innermost track.

Discussion:

These species share many genetic similarities, especially in regions that are associated with antibiotic gene resistance loci. Mapping of the Massey primers to these various *Mycobacterium* genomes has revealed that a majority of the primer binding regions have a high degree of sequence conservation in the primer-targeted regions across each of the *Mycobacterium* species. There are minor variations in some regions of each Circa plot, though most primer binding regions remain conserved across each *Mycobacterium* species genome.

The conservation of the targeted regions and the use of the novel Massey primers for the whole MTB genome and closely related species genomes have significant functional implications for TB diagnosis. This includes broad spectrum detection, rapid and early detections and drug resistance testing and detection. This also shows the broad specificity of the Massey primers and their applications outside of MTB detection and diagnosis.

Appendix II

Python Scripts for making Circa Plots

Python Script for Forward Sequence

```
# Open the sequence file
with open("sequence.fasta", "r") as seq_file:
    # Read the contents of the file and remove any newline characters
    seq = seq_file.read().replace("\n", "")

# Define the motif to search for
motif = " PRIMER SEQUENCE "

# Find the positions where the motif occurs in the sequence
positions = [i for i in range(len(seq) - len(motif) + 1) if seq[i:i+len(motif)] == motif]

# Write the positions to a file
with open("Forward_position.txt", "w") as output_file:
    for position in positions:
        output_file.write(str(position) + "\n")
```

Python Script for Reverse Sequence

```
# Open the sequence file
with open("sequence.fasta", "r") as seq_file:
    # Read the contents of the file and remove any newline characters
    seq = seq_file.read().replace("\n", "")

# Define the motif to search for
motif = " PRIMER REVERSE COMPLEMENT SEQUENCE "

# Find the positions where the motif occurs in the sequence
positions = [i for i in range(len(seq) - len(motif) + 1) if seq[i:i+len(motif)] == motif]

# Write the positions to a file
with open("Reverse_position.txt", "w") as output_file:
    for position in positions:
        output_file.write(str(position) + "\n")
```

Python Script for One Mismatch

```
# Function to find positions of a motif with one mismatch in a sequence
def find_positions_with_mismatch(sequence, motif):
    positions = []
    motif_len = len(motif)

    for i in range(len(sequence) - motif_len + 1):
        mismatches = sum(a != b for a, b in zip(sequence[i:i+motif_len], motif))
        if mismatches <= 1:
            positions.append(i)

    return positions

# Read the FASTA file
fasta_file = "sequence.fasta"

# Define the motif
motif = "CCGCCGTCGCCG"

# Initialize the results table
results_table = []

# Search for the motif with one mismatch in each sequence
for record in records:
    sequence_id = record.id
    sequence = str(record.seq)
    positions = find_positions_with_mismatch(sequence, motif)

    if positions:
        # Add positions to the results table
        results_table.extend([(sequence_id, pos, pos + len(motif), len(motif)) for pos in
positions])

# Write the results to a file
with open("results_table.txt", "w") as output_file:
    output_file.write("Sequence ID\tPosition\tEnd Position\tMotif Length\n")
    for entry in results_table:
        output_file.write("\t".join(map(str, entry)) + "\n")

print("Search completed. Results saved in results_table.txt")
```