STUDIES IN THE DEVELOPMENT AND EVALUATION OF IN-HOUSE MOLECULAR TOOLS TO IDENTIFY AND CHARACTERISE THE *MYCOBACTERIUM TUBERCULOSIS* IN RESOURCE LIMITED SETTINGS

Thesis submitted by

Munish Puri

Masters by Research (Medical and Molecular Sciences)

Submitted in fulfilment of the requirements for the degree of Master of Philosophy

College of Public Health, Medical and Veterinary Sciences

James Cook University, Townsville

September 2021

DECLARATION

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. All the information derived from the published and unpublished work of other has been acknowledged in text and a list of references is given.

Munish Puri.

ACKNOWLEDGEMENTS

STATEMENT ON THE CONTRIBUTION OF OTHERS

Nature of assistance	Contribution of others
Supervision	This degree was supervised by A/Prof. Jeffrey Warner (primary advisor) and Dr. Graham Burgess (secondary advisor) for the full duration.
Intellectual support	I wrote all chapters of the thesis. Critical feedback and editorial support on all chapters and results was provided by A/Prof. Jeffrey Warner, A/Prof. Catherine Rush and Dr. Graham Burgess. Feedback on the project proposal was provided by Dr. Graham Burgess.
Project support	Project implementation was supported by A/Prof. Jeffrey Warner and Dr. Graham Burgess. The DNA samples and reagents for laboratory use were availed by few people, as follows: - DNA templates for Mtb testing: A/Prof. Jeffrey Warner and Dr. Tanya Diefenbach-Elstob. - *Mycobacterium* species DNA templates for assay specificity testing: Dr. Jacqueline Picard and Dr. Vanina Guernier - ATCC cultured isolates: Dr. Helma Antony - Laboratory based tools, kits, and reagents: Dr. Graham Burgess and Dr. Jennifer Elliman.
Chapter 3	All sample collection and processing procedures were provided by A/Prof. Jeffrey Warner and Dr. Tanya Diefenbach-Elstob. I solely performed all other laboratory work and recoded the results.
Financial support	I sponsored myself for this entire degree.

COPYRIGHT

Every reasonable effort has been made to gain permission and acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

DECLARATION OF ETHICS

The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council (NHMRC) National Statement on Ethical Conduct in Human Research, 2007.

Human research ethics approval was received from the James Cook University (JCU) Human Research Ethics Committee (Ethics Approval Number H6432).

ABSTRACT

Tuberculosis (TB) is one of the main causes of high morbidity and mortality rates worldwide (Ismail *et al.*, 2016). It infects 10 million people every year globally with annual mortality of 1.2 million. South-East Asia and West Pacific region alone contribute to 62% of the world TB cases. Early diagnosis and relevant treatment can prevent occurrence of deaths (WHO, 2021) with delay contributing to high incidence rates and nosocomial spread (Ao *et al.*, 2012).

The TB incidence rate in Papua New Guinea (PNG) is 432 per 100,000 people (WHO, 2021). Misdiagnosis, inadequate treatment programs and emergence of drug-resistant TB (DR-TB) lead to high morbidity and mortality rates (Gilpin *et al.*, 2008). Microscopy and recognising clinical signs and symptoms are the predominant ways of diagnosing TB in PNG. The insensitive nature of the smear microscopy and uncertainty of the physical signs and symptoms representing TB put healthy communities under risk and may aid in the emergence of the drug resistance due to inappropriate antimicrobial use (Guernier *et al.*, 2017).

Therefore, an economical, accurate, reliable identification tool needing minimal infrastructure is required. To meet the WHO ASSURED criteria(Mabey *et al.*, 2004), technologies to be used in resource-limited settings must be: Affordable, Sensitive, Specific, User-friendly, Rapid, and robust, Equipment-free and Deliverable (ASSURED). In-house molecular assays that are quality assured and established on the portable MIC™ qPCR platform meet this requirement.

The aim of the work presented in this thesis was to develop and validate in-house molecular-based identification tools to identify and characterise *Mycobacterium tuberculosis* (Mtb) which better meet the WHO ASSURED criteria for use in resource poor settings like rural PNG (Mabey *et al.*, 2004). Two TaqMan based assays targeting the multicopy insertion sequence IS6110 (this thesis) and the single copy *senX3-regX3* IR gene (Broccolo *et al.*, 2003) were evaluated using DNA extracted from both PNG patient samples and clinically and taxonomically related bacteria. An additional *rpoB* gene based TaqMan assay was used in parallel to verify the quality and amplifiability of all samples. The templates reacting in both the IS6110 and the published *senX3-regX3* IR assays (Broccolo *et al.*, 2003) with low C_t values for the *rpoB* TaqMan assay were selected for further analysis.

High-resolution melt analysis (HRM) was used to confirm and characterise Mtb from DNA samples that reacted in the novel duplex. Based on this technology, three different assays targeting 16S rRNA, *rpoB* and *katG* genes were designed, optimised, and tested on *Mycobacterium* spp. confirmed DNA samples. The HRM results were then verified by sequencing. The sequencing revealed that the strain definition and characterisation determined by HRM was accurate when using DNA extracted from MGIT™ culture tubes and control samples. However, variable results were attained when template were sourced from sputum-smeared slides and decontaminated sputum. Hence, sample source and quality greatly influence the outcome of HRM analyses.

The in-house TaqMan based assays however could sustainably be implemented and used to identify Mtb in resource-limited settings using the MIC™ qPCR platform. However, the findings of this study do not support the published reports that suggest that SNPs could be readily identified in the clinical samples using the HRM platform. Further studies are required to achieve optimal clinical specimen quality for HRM analysis.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A   Adenine

aDSM   active Drug Safety Monitoring

AFB   Acid-fast bacilli

API   Analytical Profile Index

ATCC   American Type Culture Collection

BDH   Balimo District Hospital

BLAST   Basic Local Alignment Search Tool

BMS   Bio Molecular Systems

BPQ   Bedaquiline

BSLII   Bio-Safety Level II cabinet

C   Cytosine

CI   Confidence interval

Cfu   Colony forming units

CSF   Cerebral Spinal Fluid

DLM   Delamanid

DNA	Deoxyribose nucleic acid
DR-TB	Drug-Resistant Tuberculosis
EPTB	Extra Pulmonary Tuberculosis
EMB	Ethambutol
ETH	Ethionamide
FRET	Fluorescent signal by energy transfer
FIND	Foundation for Innovative New Diagnostics
G	Guanine
GIT	Guanidine isothiocyanate
HCL	Hydrochloric acid
HGT	Horizontal gene transfer
HRM	High-Resolution Melt
HRMA	High Resolution Melt Analysis
HIV	Human Immunodeficiency Virus
IFNγ	Gamma interferon
IGRA	Interferon γ release assay

IPTG	Isopropyl β-d-1-thiogalactopyranoside
IUATLD	International Union Against Tuberculosis and Lung diseases
JCU	James Cook University
LAM	Latin-American-Mediterranean
LB broth	Luria-Bertani broth
LJ	Lowenstein-Jensen
LTB	Latent Tuberculosis
MDR-TB	Multi Drug Resistant-Tuberculosis
MGIT™ tube	Mycobacterium growth indicator
MIQE	Minimum Information for publication of quantitative real-time PCR experiments
mL	millilitre
MTBC	Mycobacterium tuberculosis complex
MTB	*Mycobacterium tuberculosis*
NAAT	Nucleic Acid Amplification Techniques
NaCl	Sodium chloride

NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
NRA	Nitrate Reductase Assay
NTM	Non-tuberculous mycobacterium
NTC	Non-Template Control
ONT	Oxford Nanopore Technologies
PBS	Phosphate buffer saline
PCR	Polymerase Chain Reaction
PC2 laboratory	Physical Containment Level 2 laboratory
PC3 laboratory	Physical Containment Level 3 laboratory
pDST	Phenotypic drug susceptibility testing
PTB	Pulmonary Tuberculosis
PNG	Papua New Guinea
PZA	Pyrazinamide
RIF	Rifampicin

RRDR	Rifampicin resistance determining region
RR-TB	Rifampicin Resistant Tuberculosis
REMA	Resazurin Microtiter Assay
SBS	Sequence by Synthesis
SD	Standard Deviation
STB	Smooth Tubercle Bacilli
STR	Streptomycin
SNP	Single Nucleotide Polymorphism
T	Thymine
TST	Tuberculin skin test
UV	Ultraviolet
WHO	World Health Organisation
XDR-TB	Extensively Drug Resistant-Tuberculosis
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
ZN	Zeihl Neelsen
μL	microliter

CHAPTER 1.

Introduction

1.1   Introduction

Tuberculosis (TB) is a predominant cause of high morbidity and mortality rates globally (Ismail *et al.*, 2016). Every year, around 10 million people contract TB leading to 1.2 million (HIV negative) deaths worldwide (WHO, 2021). Accurate identification in a timely manner makes TB treatable. However, any delay contributes to high morbidity and nosocomial spread, jeopardising the healthy population (Ao *et al.*, 2012). Southeast Asia and Western Pacific region are the lone contributor of world's 62% of TB cases. An approximate rate of 3.3% of new and 18% of previously treated cases suffer from multidrug-resistant (MDR)/rifampicin-resistant (RR) TB occurred in 2019 globally (WHO, 2021).

1.2   The causative organism

TB in humans is caused by the members of the *Mycobacterium tuberculosis* complex (MTBC). In this complex, *Mycobacterium tuberculosis* sensu stricto (Mtb) and *Mycobacterium africanum* are the most infectious strains leading to high mortality rates. (Coscolla and Gagneux, 2014). Identification of the causative organism is crucial so that patients can receive the most appropriate treatment.

There are three clinical manifestations of TB: (1) Pulmonary tuberculosis (PTB); (2) Extra Pulmonary tuberculosis (EPTB) and (3) Latent tuberculosis. The first type is confined to the lungs only, while the second type can include most organs such as pleura, lymph nodes, abdomen, genitourinary tract, skin, joints and bones or meninges (i.e., anywhere except lungs). A patient having infection in all organs plus lungs is categorised as PTB and the conditions such as tuberculous intrathoracic lymphadenitis (mediastinal and/or hilar) or tuberculous pleural effusion, without radiographic abnormalities in the lungs, comes under

EPTB category. Latent tuberculosis is asymptomatic and requires immunological tests for identification (Lee, 2015).

The WHO has set various diagnostic tests as gold standards, namely microscopic examination of sputa followed by its culturing with phenotypic drug susceptibility testing (pDST) (WHO, 2021). However, the unacceptable performance of these tests in terms of sensitivity and specificity and prolonged culture time, may lead to an inaccurate diagnosis and delay in the treatment initiation (Park *et al.*, 2014).

1.3 The rationale for the study

Tuberculosis in Papua New Guinea (PNG) is the third-highest cause of morbidity and mortality with a consistent increase in burden over many years (Ley *et al.*, 2014b). In 2019, TB incidence rate in PNG was 432 per 1,00,000 population [confidence interval (CI) 352-521 per 100,000 population] (WHO, 2021). Factors such as inadequate treatment programs and the emergence of drug resistance particularly in the Western province of PNG contribute to high morbidity and mortality rates (Gilpin *et al.*, 2008).

Diverse non-specific symptoms exhibited by TB patients may pose a problem in diagnosing the disease. Sole reliance on Ziehl-Neelsen staining and microscopy for Mtb detection or clinical diagnosis based only on clinical signs and symptoms of suspected patients, is the major mode of TB diagnosis in PNG. The insensitive nature of smear microscopy and uncertainty of physical signs and symptoms jeopardises the health of this community and may aid in the emergence of the drug resistance (Guernier *et al.*, 2017). Therefore, an economical, accurate, reliable diagnostic test needing minimal infrastructure is required.

1.4   WHO recommended molecular based diagnostics

Xpert (Cepheid) is an automated, cartridge-based diagnostic test that is accurate and rapid providing results in 1.5 hours. But the cost and reliability provide a sustainability challenge in resource-limited countries (Abebe *et al.*, 2011; Gupta and Anupurba, 2015) despite subsidised rates with the help of organisations like Foundation for Innovative New Diagnostics (FIND) (Nurwidya *et al.*, 2018). In addition, the short shelf life of the cartridges used for diagnosis, a requirement for continuous power and high quality water supply and regular instrument maintenance (recalibration), makes sustainable implementation a challenge in developing or under developed countries (Weyer *et al.*, 2013). Many sites are unable to maintain these facilities once international support is withdrawn.

More importantly, one study revealed the fluctuating specificity of this device on repetitive testing using same samples over time (Friedrich *et al.*, 2013). Another study demonstrated a significant variation in the sensitivity when different versions of cartridges were used (Sohn *et al.*, 2014). This questions the reliability of this device. Still, its availability continues to provide a major contribution in Mtb diagnosis as it can identify 99% of smear-positive and more than 80%of smear-negative cases. This makes it a beneficial add-on test to microscopy (Nurwidya *et al.*, 2018).

However, Xpert it is not as sensitive as some probe-based assays and the sole reliance on the *rpoB* gene limits its utility. The newer version, the Xpert Ultra targets three genes (*rpoB*, IS6110 and IS1081) and it has proven to be more sensitive (15.6 cfu/mL than the previous version (112.6 cfu/mL). However, its specificity is still concerning, as it varied over time of treatment completion (Chakravorty *et al.*, 2017; Dorman *et al.*, 2018). The recent version of the Xpert, the Xpert Omni has been devised. The WHO will conduct its evaluation once sufficient data on its performance is available (WHO, 2021).

Other WHO endorsed genotypic drug susceptibility tests (gDST) such as line probe assays (Hain GenoType MTBDRplus 2.0 and MTBDRsl 2.0) can be applied directly on sputum samples. These assays are accurate in providing the data on resistance to first and second-line drugs and results can be obtained in 1 to 2 days. Since these assays are expensive and target a limited number of mutations in the *rpoB* gene, they possess limited sensitivity as compared to the pDST (Heyckendorf *et al.*, 2018).

1.5   Other alternatives

Polymerase chain reaction (PCR) is a genome detection assay with an ability to detect even fewer than 10 bacilli per millilitre and is the method of choice for the Mtb identification (Khosravi *et al.*, 2017a). Taqman real-time assays are gaining popularity due to their proven performance. Several studies have demonstrated that the TaqMan assays can be sensitive, specific, cost-effective and easy to design and perform (Buh Gasparic *et al.*, 2008; Josefsen *et al.*, 2009; Tortoli *et al.*, 2012). Further, the TaqMan assays could be used on any available platform. This increases its versatility and easily adoptable in any laboratory. Even the positive and negative controls can be added easily. The commercial systems such as Xpert or Xpert ultra despite having provision of DNA extraction in same cartridge, two internal control reactions and three target genes, lack this feasibility. These commercial devices require their specific device to function, and these can't be applied/ used in any other PCR platform.

Armand *et al.* (2011) in their study, compared the TaqMan based assay with Xpert. They summarised that the cost per sample using the Xpert (50 euros/ US $65) was five times more than the TaqMan assay (11 euros/ US $ 15). In addition, the hands-on time required for TaqMan was less (2 minutes/sample) as compared to the Xpert (3minutes/sample). The other main characteristics are mentioned in the Table (Armand *et al.*, 2011). There are assays based on the TaqMan chemistry, which targets IS6110, a multicopy gene sequence (Broccolo *et al.*, 2003) or any other unique gene such as *senX3-regX3* IR region (Broccolo *et al.*, 2003) or the *rpoB* gene (Weyer *et al.*, 2013).

4

Table 1.1: Main characteristics of Xpert and IS6110 TaqMan assay (Armand *et al.*, 2011)

Characteristics	Xpert	IS6110 TaqMan
Workflow design	On demand	Batch mode
Sample format	Individual cartridge	Microplate, 96 wells
Sample preparation	Liquefaction and inactivation of sample (15mins/sample)	Three cycles of boiling-freezing and incubation with proteinase K (20 mins)
Time to results	2 hours	2 hour 45 mins
Hands-on time	3min/ sample	2min/ sample
Cost	50 euros (US $65)/ sample	11 euros (US $ 15)/ sample

Several studies claim that the High resolution melting (HRM) analysis is a cost-effective technique used to detect mutations (Sharma *et al.*, 2015; Galarza *et al.*, 2016; Bentaleb *et al.*, 2017; Sharma *et al.*, 2017), screening of presequence, SNP typing, methylation and quantification of different alleles. It is a technique to analyse a piece of DNA to identify the link between the temperature and level of denaturation using a real-time PCR instrument. A dye, which intercalates into the double-stranded DNA is added to the PCR product before melting and is a closed-tube system without requiring any specialized instrument. The DNA denaturation can be observed with an increase in temperature and the appearance of a sigmoid melt curve. (Vossen *et al.*, 2009; Tong and Giffard, 2012).

There are various platforms of varying brands that allow the operation of PCR and other relevant operations. The Magnetic Induction Cycler (MIC™) from Bio Molecular Systems (BMS) is a novel and portable qPCR device capable of running 48 samples in one PCR run. The device uses magnetic induction for heating and airflow for cooling. A provision of connecting this format to a computer via Bluetooth, allows use of multiple devices at one time. It can be operated on batteries as well. In addition, features like Relative Quantification and High-Resolution Melt (HRM) analysis might make it a suitable PCR platform that could be used in resource-limited settings. This could even comply with the

WHO ASSURED criteria (Mabey *et al.*, 2004). The ASSURED criteria stand for affordability, sensitive, specific, user-friendly, rapid, equipment free and deliverable. This criterion is for the resource limited settings where a complying device(s) could be used as a substitute to the expensive commercial systems such as Expert or Expert ultra are not feasible for routine use.

1.6   The intent and hypothesis

Is it TB? If yes, how much? What is the drug susceptibility profile of the infecting mycobacterial strain? These are the pivotal questions that a potential diagnostic assay should address. No doubt, Xpert (Cepheid) or its newer version the Xpert Ultra are fit for purpose for both Mtb diagnosis and a rifampicin susceptibility profile. Nevertheless, its maintenance costs (Weyer *et al.*, 2013), and variable sensitivity and specificity on repetitive testing using the same samples overtime presents challenges (Friedrich *et al.*, 2013; Sohn *et al.*, 2014).

Accurate diagnosis is the prerequisite for treating or initiating patient management. The easiest way to diagnose is based on clinical signs and symptoms. However, misdiagnosis due to heavy reliance on the clinical signs and symptoms and microscopy may contribute to high TB morbidity and mortality rates. The development and implementation of reliable and cost-effective assays with high sensitivity and specificity requiring minimal specialised infrastructure in resource-limited settings could potentially improve the diagnosis rates. This in turn, can lead to better patient management and treatment outcomes.

Therefore, the objective of this study was to develop and validate cost-effective, in-house molecular based identification tools to detect and characterise Mtb and drug susceptibility profiles using DNA from sputum samples collected in PNG. The development of TaqMan on BMS MIC™ qPCR platform for MTBC identification and HRM based in-house assays for drug resistance screening could be a useful alternative to currently available expensive commercial tests that are not optimal to use in resource-poor settings where TB is a serious problem.

CHAPTER 2

Literature review

2.1   Introduction

Tuberculosis (TB) is an old disease. Robert Koch identified the causal organism of TB, the *Mycobacterium tuberculosis* in 1882 (Fogel, 2015). With the epidemics of 18th and 19th century in Europe and North America, TB acquired the title of the Captain among the Men of Death (Daniel, 2006). According to a recent World Health Organisation (WHO) report, every year, 10 million healthy people worldwide become infected and 1.2 million people die of this disease (WHO, 2021).

2.1.1 The causal organism

TB in humans is caused by eight similar species of the *Mycobacterium* genus. They have wide range of natural hosts, namely human hosts (Mtb*, M. africanum*, *M. canettii),* bovine hosts (*M. bovis*), caprine hosts (M. *caprae*), rodent hosts (M. *microti*), and pinniped hosts (M. *pinnipedii*), along with the attenuated *M. bovis* strain BCG (Bacillus Calmette Guerin), the commonly used vaccine strain. These species are represented as Mtb complex (MTC or MTBC) (Reddington *et al.*, 2011). In this complex, Mtb and *M. africanum* are the obligate human pathogens without any animal reservoir (Gagneux, 2012) and are the most infectious strains with high mortality rates (Homolka *et al.*, 2008). *Mycobacterium orygis* has been recently reported to infect humans and has gained recognition as a member of MTBC (Coll *et al.*, 2014). The global phylogeography of MTBC is summarised in Table 2.1.

The TbD1 (Table 2.1) is a 2153bp long piece of DNA present in the Mtb that contains *mmpS6* and *mmpL6* genes in it. These genes encode members of the mycobacterial membrane protein families MmpL. This DNA piece upon deletion give rise to evolutionary modern Mtb strains that are highly virulent with high transmission potential (Bottai *et al.,* 2020).

Table 2.1: Different lineages of MTC members with genetic variation and their global occurrence (Brites and Gagneux, 2015; Ngabonziza *et al.*, 2020)

Lineage type	Other name	Genetic variation	Geographic Occurrence	Type
Lineage 1	Indo-Oceanic	No deletion	Indian Ocean and Philippines	Paraphyletic
Lineage 2	East -Asian	TbD1-deleted	East-Asian	Monophyletic
Lineage 3	East Africa	TbD1-deleted	East Africa, Central and South Asia	Monophyletic
Lineage 4	Euro-American	TbD1-deleted	Europe, America, Africa and Middle East	Monophyletic
Lineage 5	*M. africanum* West African1	No deletion	Western African countries	Paraphyletic
Lineage 6	*M. africanum* West African2	No deletion	Western African countries	Paraphyletic
Lineage 7	Ethiopian	No deletion	Ethiopia	Paraphyletic
Lineage 8	Africa	-	Greater lake region of Africa	-

-: No data available yet.

A recent study done by Ngabonziza *et al.* indicated the presence of a rare Mtb strain. This strain had its biochemical characteristics like the *Mycobacterium tuberculosis* sensu stricto. Therefore, this strain has been proposed (Table 2.1) as the lineage 8 representative (Ngabonziza *et al.*, 2020). The biochemical attributes of all the lineage strains have been complied in Appendix 1, Table 1.1. The data mentioned in that Appendix 1, Table 1.1 has entirely been taken from the Ngabonziza *et al.* (2020) (Ngabonziza *et al.*, 2020).

2.1.2.TB Treatment and emergence of drug resistance

Drug therapy plays a pivotal role in controlling and curing the infections caused by various pathogens. However, the emergence of drug resistance imparts a major threat, especially when the available drugs to cure a particular infection are limited (Hughes and Andersson, 2015). A combination of four drugs, namely rifampicin, isoniazid, ethambutol (EMB) and pyrazinamide (PZA) for 8 weeks duration and a further treatment with rifampicin and isoniazid for 16 weeks constitute the drug-based therapy to treat tuberculosis. This therapy is approximately 50 years old (Zumla *et al.*, 2013). The drugs are collectively called first-line drugs. However, the treatment time of 6 months nevertheless surges to at least 20 months if the suspected patient(s) incurs from multidrug resistance TB (MDR-TB) (Riccardi and Pasca, 2014).

2.1.3 Types of drug resistance and its mechanism

MDR-TB is the resistance of Mtb to isoniazid and rifampin anti-TB drugs. A patient with this kind of infection is treated with a separate range of drugs or second-line drugs (Figure 2.1). Resistance of the organism to first-line drugs plus to any fluoroquinolones is termed as pre-XDR-TB and resistance to first line drugs , any fluoroquinolone and at least one additional Group A drugs is called as extensive drug resistance (XDR-TB) (Dorman and Chaisson, 2007; WHO, 2021). This type of TB was first found in KwaZulu-Natal in South Africa in a population burdened with HIV infection (Lemos and Matos, 2013; Wang *et al.*, 2015) and is currently present in 100 countries (Cohen *et al.*, 2015). The worldwide mortality rate due to active TB in 2019 was 1.2 million. The treatment success rate of drug-susceptible TB is 85%, 57% for MDR-TB and 76% for TB patients with HIV positive (WHO, 2021) . Therefore, it can be said that the emanation of MDR-TB is a serious issue, which acts as a halt to the progressive goals of WHO to control and eliminate TB.

Figure 2.1 Classification of drugs used to treat drug-susceptible and drug resistant Tuberculosis. (Dorman and Chaisson, 2007; Zumla et al., 2013).

Antibiotics are the main remedy to infections caused by bacterial pathogens. These antibacterial agents act either by suppressing bacterial growth or by killing them. Different antibiotics target different parts of bacterial cells (Donkor, 2013). The action mechanisms of first line and second-line drugs are compiled in Figure 2.2 and 2.3.

Antibiotic	Chemical structure	Mechanism and target	Mutations associated with resistance	Half-life in humans (hours)
Isoniazid		Inhibits mycolic acid synthesis; primary target is InhA and secondary targets are KasA and DfrA	katG (required for drug activation); inhA (promoter mutations); and others	1–3
Rifampicin		Inhibits transcription; RNA polymerase β-subunit	rpoB	2–3
Ethambutol		Inhibits arabinogalactan synthesis; possibly EmbB	embB	3–4
Pyrazinamide		Unknown (possibly inhibits FAS-I or alters membrane energetics)	pncA (required for drug activation)	10
Streptomycin		Inhibits protein synthesis; 30S ribosomal subunit	rpsL and rrs	2–3

Figure 2.2 Mechanism of action, resistance and half-life of current first-line anti-TB drugs(Sacchettini *et al.*, 2008).
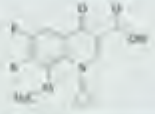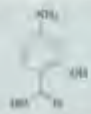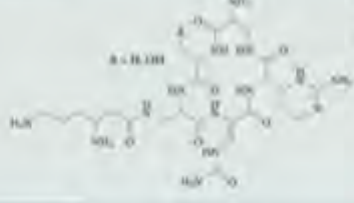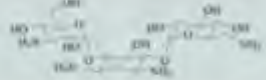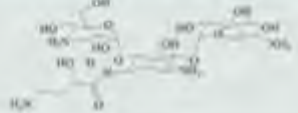
Antibiotic	Chemical structure	Mechanism and target	Mutations associated with resistance	Half-life in humans (hours)
Fluoroquinolones		Inhibits DNA gyrase	gyrB	Moxifloxacin: 12; Gatifloxacin: 8
Ethionamide		Inhibits mycolic acid synthesis; InhA	ethA (required for drug activation) and inhA (promoter mutations)	2
Cycloserine		Inhibits peptidoglycan synthesis by blocking the synthesis and use of D-alanine (Ala); Ala racemase and D-Ala-D-Ala ligase	alr (overproduction) and ddl (overproduction)	10
Para-aminosalicylic acid		Inhibits folate metabolism; possibly dihydropteroate synthase	thyA	0.75–1
Capreomycin		Inhibits protein synthesis; methylated nucleotides in both ribosomal subunits	tlyA and rrs	4–6
Kanamycin		Inhibits protein synthesis	rrs	2
Amikacin		Inhibits protein synthesis	rrs	3

Figure 2.3: Mechanism of action, resistance and half-life of current second-line anti-TB drugs(Sacchettini *et al.*, 2008).

Figure 2.4 demonstrates 22 new anti-TB drugs according to WHO. These drugs are in a different phase of trials. The drugs Bedaquiline (BDQ), delamanid (DLM) and pretomanid have already gained approval for prescription by the regulatory authorities. Six antimicrobial drugs such clofazimine, levofloxacin, linezolid, moxifloxacin, rifampicin (high dose) and rifapentine are getting testing against TB (WHO, 2021).

Phase I[a]	Phase II[a]	Phase III[a]
BTZ-043[b]	GSK-3036656[b]	Bedaquiline (TMC-207)[b]
Macozinone[b]	Telacebec (Q203)[b]	Delamanid (OPC-67683)[b]
OPC-167832[b]	TBA-7371[b]	Pretomanid
SPR720[b]	Delpazolid (LCB01-0371)	Clofazimine
TBAJ-876	SQ109	High-dose rifampicin for treatment of drug-susceptible TB
TBI-166	Sutezolid	Rifapentine for treatment of drug-susceptible TB
TBI-223	High-dose rifampicin for drug-susceptible TB (PanACEA)	Bedaquiline–delamanid–linezolid–levofloxacin–clofazimine (6-month oral regimen for RR-TB) or bedaquiline–delamanid–linezolid–clofazimine (6–9 month oral regimen for pre-XDR and XDR-TB) (BEAT TB trial)
ACTG A5312	Bedaquiline and delamanid (ACTG 5343 DELIBERATE trial)	Bedaquiline–pretomanid–moxifloxacin–pyrazinamide (BPaMZ) (SimpliciTB trial)
	Bedaquiline and pretomanid with existing and re-purposed anti-TB drugs for MDR-TB (TB PRACTECAL Phase II/III trial)	Bedaquiline–pretomanid–linezolid (NiX-TB trial)
	Shorter regimens including clofazamine and rifapentine for drug-susceptible TB (CLO-FAST trial)	Bedaquiline–pretomanid–linezolid (ZeNix trial) - Linezolid optimization
	Pretomanid-containing regimens to shorten treatment for drug-susceptible TB (APT trial)	Bedaquiline with two OBRs[c] (all-oral, 9 months; with injectable, 6 months) (STREAM trial)
	Delamanid–linezolid–levofloxacin–pyranzimamide for fluoroquinolone-susceptible MDR-TB (MDR-END trial)	Bedaquiline–linezolid–levofloxacin with OBR[c] for MDR-TB (NExT trial)
	Levofloxacin with OBR[c] for MDR-TB (Opti-Q)	Bedaquiline and delamanid with various existing regimens for MDR-TB and XDR-TB (endTB trial)
	4-month treatment for drug-susceptible TB (PredicTB trial)	Bedaquiline–delamanid–linezolid–clofazamine for fluoroquinolone-resistant MDR-TB (endTB-Q)
	High-dose rifampicin for TB meningitis (ReDEFINe)	Rifapentine–moxifloxacin for treatment of drug-susceptible TB (TB Trial Consortium Study 31/A5349)
	Multiple adjunctive host-directed TB therapies for drug-susceptible TB (TBHDT)	Several 2-month regimens for drug-susceptible TB (TRUNCATE-TB trial)

Figure 2.4: The worldwide development pipeline of new anti-TB drugs and drug regimes, August 2020 (WHO, 2021).

Bedaquiline (BDQ) drug was introduced as a part of the MDR-TB program (1 July 2015 till 31 December 2017) in Daru area of PNG in 2015. It was associated with active drug-safety monitoring (aDSM). Further, the confirmed rifampicin-resistant TB (RR-TB) was included as a part of MDR-TB. Of 277 MDR-TB cases, 77 received this drug. Of those, 8 patients had adverse effects with five deaths. Out of those five deaths, one was associated with BDQ. The QTcF interval prolonged on the electrocardiogram (ECG). Overall, the drug (BDQ) implementation was recorded as safe and feasible to implement in association with aDSM (Taune *et al.*, 2019).

Another approved anti-TB drug Delamanid (OPC-67683 OR Deltyba™), a prodrug requiring activation by the deazaflavin (F_{420}) dependent nitroreductase (*ddn*) has been approved by WHO against MDR-TB in 2014. Mutation(s) in this gene (*ddn*) or one of the five coenzymes (*fgd1, fbiA, fbiB, fbiC and fbiD*) has been linked to the delamanid resistance. The interesting part is the mutations in these cited genes have already been reported in the Mtb strains that were never even exposed to this drug (Reichmuth *et al.*, 2020). This naturally occurring

resistance among the Mtb against delamanid makes the implementation of this drug questionable.

The prescription of new anti-TB drugs after their successful testing and approval may aid in the reduction of the MDR-TB/RR TB cases. However, their misuse may give rise to the other novel mono or polyresistant (depending on the drug combination) strains with unknown repercussions. Therefore, it is quite imperative to accurately diagnose a TB suspected patient(s) before initiating the drug therapy. This would potentially aid in minimising the emergence of new drug resistant strains on using new drugs. Further, the new anti-TB drugs against which the Mtb strains are already acquiring natural resistance without prior exposure need to be prescribed very carefully. A comprehensive molecular characterisation of the Mtb strain(s) prevalent in a particular area may be required before the use of such drugs in that domain(s).

2.1.4 TB and HIV

Acquisition of resistance to the anti-TB drugs among Mtb strains and an intense relation of TB infection with human immunodeficiency virus (HIV) world widely serves a great threat to humanity (Al-Mutairi *et al.*, 2011). Globally in 2019, 456,426 TB with HIV positive people were recorded. A population of 208,000 (range, 177,000-242,000) death cases were due to HIV-associated TB were observed in the same year. Africa dominates in deaths (7.4 million) of HIV-associated TB infected people in 2019 (WHO, 2021).

HIV diminishes the CD_4T-cells population in the host. These cells play a central role in activating the macrophages and synthesise granuloma, an immune response to destroy the infection-causing pathogens. The activation of macrophages is critical to remove the TB causing bacteria. Therefore, people infected with HIV will not retain TB infection (Brites and Gagneux, 2012). There is a high risk up to 100 times in HIV positive patients already infected with latent TB to revert to active state TB and there is more threat to individuals who freshly acquire the TB infection as a result of HIV rather than other socio-economic factors actuating TB (Harries *et al.*, 2011).

TB infection is airborne (McNerney *et al.*, 2012). The congested areas with high populations have more chances of disease transmission from infected to healthy people. While in rural areas, less population due to undeveloped region, the presence of any infected case in that area would probably disperse the tubercle bacilli at a much lower rate than in urban areas due to less frequent person to person contact. In addition, the unavailability of anti-tuberculosis agents in urban areas and lack of use of standardised methods further imparts complication and make transmission rate high (Chadha *et al.*, 2003).

The prison population in developing nations with resource-limited settings represents a great risk of an infection outbreak. Even a single case of drug-resistant TB has the potential to disperse the infection at a high pace within prison premises due to a highly congested area packed with a rich population of prisoners. In such conditions, it would be appropriate to refer to the jail cells as incubators promoting the luxuriant growth of bacteria. During the release of such infected people, a serious threat to a healthy population is posed. This threat can be a disaster if one or more mono or poly resistant strains circulating in such premises. In the quest to control TB, it would be a critical pre-requisite to know in depth the repercussions of drug resistance on host-pathogen engagement and the factors responsible for the strain to transform to drug-resistant state.

2.2 Diagnostics

The WHO report published in 2019 indicated that only one in three patients suffering from drug resistant-TB are diagnosed and treated appropriately, while other undiagnosed cases suffer from higher morbidity and mortality (WHO, 2021). The high morbidity may serve as a good source of disseminating the drug resistant pathogen in the environment, hence jeopardising the healthy communities.

TB is of three types:
1. Pulmonary tuberculosis (PTB)
2. Extra Pulmonary tuberculosis (EPTB)
3. Latent tuberculosis (LTBI)

PTB targets lungs only, while EPTB affects different organs such as the pleura, lymph nodes, abdomen, genitourinary tract, skin, joints and bones or meninges excluding lungs. The LTBI on the other hand is asymptomatic.

A patient having infection in all organs and lungs is classified as PTB and conditions such as tuberculous intrathoracic lymphadenitis (mediastinal and/or hilar) or tuberculous pleural effusion, without radiographic abnormalities in the lungs, categorised under EPTB (Lee, 2015). While the LTBI is asymptomatic and requires immunological tests [Tuberculin skin test (TST) or Interferon γ release assay (IGRA)] for detection (Lin and Flynn, 2018).

The utility of diagnostics is to indicate whether a patient is suffering from a disease or not in terms of sensitivity, specificity, positive and negative predictive value. The sensitivity regarding diagnostics is the percentage of people having a particular abnormality, which is detected by an assay exhibiting positive for the concerned disease. The specificity on the other hand is termed as the proportion of people with no specific disease and is detected by a test giving negative report for the disorder under observation (Akobeng, 2007). Therefore, diagnosing a person is critical before the commencement of therapeutical treatment.

Since pulmonary TB being the most critical TB type (Palomino, 2005), the disease diagnosis is paramount for the effective treatment of the patient. However, a misdiagnosed or delayed diagnosed case can contribute to disease dispersal among healthy community (Ben-Selma *et al.*, 2009). There are many diagnostic tests to screen tuberculosis, each test having its advantages, which are discussed tersely in the following section.

Patients with smear negative PTB and EPTB are very difficult to diagnose. There are chances that other diseases such as Melioidosis may mask signs of TB (Warner and Currie, 2018). Hence, exposing the unreliability of diagnostic imaging, method of diagnostics. The patients with immunosuppression often present a challenge to diagnose PTB, especially when co-infected with HIV.

Those people exhibit an atypical radiographic pattern like middle and lower lung, absence of cavity formation, presence of lymphadenopathy and pleural effusions or military demonstration. This is dependent on the level of immunosuppression at the time of overt disease. The radiographic manifestations in patients with CD T cell count of $< 200/mm^3$ show a higher incidence of mediastinal or hilar lymph node enlargement and a value of more than $200/ mm^3$ indicates a lower prevalence of cavitation, and often extrapulmonary involvement compared with HIV patients. (Skoura *et al.*, 2015).

2.2.1   Microscopy

In resource-limited countries where the options (funding, instruments availability etc) are quite limited, approaches like microscopy are the most reliable method for diagnosing active TB. An instrument to look upon stained sputum smears (by Ziehl-Neelsen method) has proved to be a method of choice for decades to check for the presence or absence of tubercle bacilli in each clinical sample based on its cell morphology. Though cheap, it lacks high (up to 70 %) sensitivity (requiring $\geq$ 10 000 bacilli per mL^{-1} to display a positive test). It doesn't require any sophisticated infrastructure and trained personnel (Schofield *et al.*, 2012; Tortoli *et al.*, 2012; DeJesus *et al.*, 2013).

Sodium hypochlorite is used to digest the specimen and centrifuged to concentrate the cells (Palomino, 2005) or the Kinyoun method is used to process the specimen. Both methods show no significant improvements. The prepared smear also fails to distinguish between tuberculous and non-tuberculous strains without culturing the sample (Eichbaum and Rubin, 2002). Fluorescent microscopy is superior, but the high power consumption, instrument cost along its accessories (like bulbs, dyes etc.) price restrict its usage in routine work (Drobniewski *et al.*, 2013). A newer version of the fluorescent microscope- CellScope might help the microscopists to view digital images. Even though an LED microscope might be helpful, it is still labourious (McNerney *et al.*, 2015).

An automated system TBDx (Signature Mapping Medical Sciences, USA) consists of robotic loading of slides. This system provides high-resolution digital image analysis within minutes with a 200-slide capacity. However, it has low specificity, and a positive slides manual along its accessories is always required for comparison (McNerney *et al.*, 2015). Despite fallacies, light microscopy, a century-old method (Rachow *et al.*, 2011) is still the available and preferred method in low and middle-income developing countries.

2.2.2   Immunology based tests

Many people do not prefer to get their results back due to labourious examination making consequent delays from 2 to 7 days (Keeler *et al.*, 2006). As smear positive cases are very infectious, the smear negative cases may also contribute to the morbidity and mortality of the disease. Further, the approach becomes difficult in smear-negative TB patients having HIV with the suppressed immune system and /or extrapulmonary disease and DR TB patients (Weyer *et al.*, 2013).

In such cases, immunological assays are required. Nevertheless, assays like Tuberculin skin test fails to differentiate between tubercle and non-tubercle bacteria (Walzl *et al.*, 2011) and others such as the IFN–γ (IRGA) assay are expensive. Hence these cannot be implemented on a routine basis. Moreover, in some places mainly in developing countries, technical fallacies like improper functioning of the microscope, unskilled microscopist and inadequate supply of reagents further adds to the problem (Keeler *et al.*, 2006). The inability to differentiate between latent and active TB without robustness further demonstrates the unsuitability of immunological tests (TST and IGRA). Therefore, WHO doesn't recommend these assays in TB high burden countries (McNerney *et al.*, 2015; Walzl *et al.*, 2018).

2.2.3   Antigen based tests

The Alere Determine LAM TB test is another WHO endorsed test. Its low cost with fast turnover time makes it suitable to opt for (Walzl *et al.*, 2018). Proteomics based analyses with the potential to discriminate the active from latent TB are expensive. This test is recommended for TB patients with HIV(WHO, 2021). Acquisition of samples from the disease site to measure the immune response markers might help. However, sample collection may represent an issue in some situations (McNerney *et al.*, 2015).

2.2.4   Culture

Culturing allows the organism to be isolated from a variety of samples, pulmonary (sputum, bronchial washing) and non-pulmonary (urine, gastric aspirate, synovial fluid), tissue specimens, bone marrow aspirate and peripheral blood (Verettas *et al.*, 2003). Being more sensitive than microscopy with the ability to identify 100 bacilli in sputum, it also helps in the revival of bacteria for further antibiotic susceptibility and genotype investigations (Palomino, 2005).

Conventionally, Lowenstein-Jensen and Middlebrook media are used to isolate the pathogen from the clinical specimens. The bacteria take several weeks to grow and further a sequence of biochemical tests are done to confirm the pathogen, making the overall process laborious and time-consuming. The estimated overall time for tubercle bacilli identification is 3 to 8 weeks. Further, this method is prone to contamination (Bachmann *et al.*, 2008).

Apart from this, the system requires special equipment and lab facilities like incubators, biosafety cabinets with continuous power supply due to biosafety concerns (Weyer *et al.*, 2013) and takes a long turnover time to put out results (Schofield *et al.*, 2012; Weyer *et al.*, 2013). In addition, the biosafety risk is similar to the test performers when compared with microscopy, therefore a requirement of proper lab infrastructure for the safety of the workers (Keeler *et al.*, 2006; Boehme *et al.*, 2010; Weyer *et al.*, 2013).

In other phenotypic methods like phage-based assay or colorimetric assays [resazurin microtiter assay (REMA) and nitrate reductase assay (NRA)], there is always a risk of contamination and false positives and negative results (Palomino, 2005; Schofield *et al.*, 2012). MB/BacT has turnover period of weeks and the method requires trained staff (Palomino, 2005; Naveen and Peerapur, 2012). The heavy reliance of culturing on the sophisticated infrastructure and reagents seems to make it difficult to fit for purpose in resource-limited settings. Even though they provide vital information, the cost to run these methods is not affordable.

2.3   Molecular Tests

The nucleic acid amplification techniques (NAAT) are far better than the above-described methods in terms of sensitivity, specificity, and swiftness. Further, their applicability even on sterilized material makes these tests unique over culture methods (Deshpande *et al.*, 2010). In addition, nucleic acid (DNA or RNA) based detection and characterisation of bacteria and viruses put these techniques above in preference list, out of which PCR,

enzymatic DNA restriction, nucleic acid hybridization and methods based on fluorescence form the basis of these techniques (Versalovic and Lupski, 2002; Barken *et al.*, 2007).

Currently, nucleic acid amplification-based assays are the most preferred methods to identify Mtb. These assays are highly sensitive and specific that improves the identification rates in smear-negative samples (Amin *et al.*, 2011). Molecular methods are routinely used to distinguish between tuberculous and non-tuberculous species of Mycobacteria (Eichbaum and Rubin, 2002). The Mtb confirmation is quite helpful in initiating patient management due to its rapid output of results.

The Xpert is a diagnostic assay to identify TB endorsed by WHO in December 2010. The assay utilises five unique nucleic acid hybridisation probes (molecular beacons) in an automated hemi-nested PCR assay or simply the assay is PCR based that used five different bacon probes. Each probe is labelled with a coloured fluorophore responding to its complementary target sequence inside the *rpoB* gene of the Mtb. The production of all five fluorescent colours during amplification informs the presence of rifampicin susceptible Mtb. Conversely, any mutation in the target does not allow the binding of the probes to the site under examination, resulting in the absence of colour formation and indicating rifampicin resistance (Weyer *et al.*, 2013). The method has high sensitivity to detect the rifampicin resistance in both high and low-risk population (Abebe *et al.*, 2011). The assay is based on the amplification of mycobacterial DNA straight from the sputum samples.

 The Xpert targets the 81-bp region of the *rpoB* gene only. This rifampicin resistance-determining region (RRDR) is responsible for more than 95% mutations conferring resistance to the rifampicin drug (Toosky and Javid, 2014). However, 41.9% mutations outside this 81 bp hot spot (codons 507 to 533) have also been reported (Adikaram *et al.*, 2012; Elbir and Ibrahim, 2014). Therefore, it is unable to differentiate between the genetic variations linked to drug resistance and silent mutations (Ajbani *et al.*, 2015). Moreover, cost of each test at US $ 10/cartridge is very high in resource-limited settings (Toosky and Javid, 2014).

An intensive study by Friedrich et al. in which the sensitivity and specificity of Xpert assay were assessed by taking 2741 sputum samples from 221pateints at different intervals of time (week 0, 1 up to 8, 12, 17 22 and 26). On comparing with the combined reference standard of smear and culture methods, the Xpert had excellent sensitivity of 97% but poor specificity of just 49% (Friedrich *et al.*, 2013). Another finding by Sohn *et al*. (2014) indicated the significant variation in sensitivity on using different types of cartridges. With cartridge G3, the observed sensitivity was 67%, but when the G4 version was used, sensitivity dropped to half (33%), which is unacceptable (Sohn *et al.*, 2014).

Around 120 countries use Xpert. It fails to demonstrate its applicability on smear negative or extrapulmonary TB, patients with HIV to diagnose children. For the *rpoB* gene-based drug characterisation of the strains carrying phenotypically silent mutations, the possibility of false-positive results arises. The advent of its newer version Xpert Ultra has been proved to be better than its previous version. An addition of two additional target genes IS6110 and IS1081 increases its sensitivity. However, the specificity of this new version is still a concern (Dorman *et al.*, 2018). Despite limitations, Xpert Ultra may provide a significant contribution to the detection of Tuberculosis.

Taqman real-time assay, on the other hand is gaining popularity due to its proven performance. It is a single step closed tube assay (Kulesh *et al.*, 2004) that amplify target DNA along with allele discrimination or genotyping ability, thereby narrowing the chances of contamination (Gupta and Anupurba, 2015). A number of studies have unanimously demonstrated that the TaqMan assay is sensitive, specific, cost effective and easy to design and perform (Buh Gasparic *et al.*, 2008; Josefsen *et al.*, 2009; Tortoli *et al.*, 2012). Some of the PCR based novel diagnostics are summarised in Table 2.2.

Table 2.2: Novel diagnostics based on NAAT (Broccolo *et al.*, 2003; McNerney *et al.*, 2015)

Technology	Amplification reaction	Operational features	Target	Time (mins)	Stage of product development
Easy NAT	Cross priming amplification	Isothermal 65°C Instrument free visual output and DNA extraction	IS6110	<90	Released to market
Xpert	PCR	Automated sample extraction Resistant to Rifampicin	*rpoB*	<90	CE mark and US FDA approval. WHO endorsement
NEAT	Nicking enzyme amplification reaction	Isothermal from 55°C to 59°C	-	-	-
RPA	Recombinase polymerase amplification	Isothermal 39°C	IS6110 and IS1081	<20	Proof of concept Study published
Truenat	PCR	Miniaturized chip based Semi-automated DNA extraction	Ribonucleoside diphosphate reductase gene	<60	Released to market. CE mark
VerePLEX Lab-On-Chip	PCR	Microarray technology Isoniazid and Rifampicin resistant plus nine non-TB mycobacteria	IS6110 and 16S RNA	<180	Released for research use
Genedrive	PCR	Paper based DNA extraction technology Rifampicin resistant	REP 13 E12 *rpoB*	60	Filed trials
TaqMan	PCR	Uses primer set and a probe	Any gene	90	Published study

- No data given.

The Xpert or Xpert Ultra no doubt can aid in the Mtb identification and characterisation. However, their low specificity (Dorman *et al.*, 2018) may lead to false- positive results. Comparing them to the PCR, they are expensive to implement and sustain in low-income countries. PCR with relative performance as compared to Xpert or Xpert Ultra seems feasible in those settings. Moreover, the PCR has the flexibility of adding more target genes for the Mtb identification and characterisation. The Xpert is limited to its one target and Xpert Ultra to three targets(Dorman *et al.*, 2018). The assays mentioned in Table 2.2 may also be useful and even fit for purpose. However, the cost is the main concern. Developing nations may not be able to afford them and use them as routine Mtb identification commercial assays.

The BMS MIC™ qPCR allows the provision of TaqMan and other analysis such as mutation detection in one run. Even it provides the platform to perform multiplex assays thereby saving time. The machine is available in either two or four channel modes without requirement of any calibration at all. The other commercial machines (such as Xpert) on the other hand requires it routinely (https://biomolecularsystems.com/mic-qpcr/). It could be said that the BMS MIC™ qPCR cycler can be used for most of the required analysis with minimum handling. This may be a suitable option for resource-limited setting. The detailed comparison of this device with other commercial systems has not been done yet.

2.3.1 TB biomarkers in the diagnosis

Bacteria generally acquire resistance by horizontal gene transfer (HGT) and de novo mutation. It has been documented that MTC lacks the HGT activity (Gagneux, 2012; Trauner *et al.*, 2014; Castro *et al.*, 2020). Therefore, the mutation is the only driving factor for attaining resistance in MTC. Mutations in the genetic material, particularly in genes encoding drug targets or mechanisms of drug metabolism confer loss of sensitivity to antibiotics (Sacchettini *et al.*, 2008; Seifert *et al.*, 2015). These genetic mutations can serve as biomarkers to identify drug-resistant infections in hours rather than days, weeks or months (Nebenzahl-Guimaraes *et al.*, 2014).

The mutation rates for each of the available anti-TB drugs are different (Figure 2.5). Therefore, it is pivotal to determine which mutations lead to drug resistance in the Mtb. This will aid in the development of more efficient and high precision molecular diagnostic tests for better patient management. The rate of emergence of drug resistance depends on many elements such as the mutation supply rate, relative fitness of the drug-resistant mutant as a function of drug concentration, selective pressures, clonal interference, compensatory evolution and the epistatic interaction between resistance genes or between drugs (Hughes and Andersson, 2015; Castro *et al.*, 2020).

Rifampicin	1 drug-resistant mutant for every 10^{7-8} bacilli
Isoniazid	1 drug-resistant mutant for every 10^{5-6} bacilli
Ethambutol	1 drug-resistant mutant for every 10^{5-6} bacilli
Pyrazinamide	1 drug-resistant mutant for every 10^{2-4} bacilli
Streptomycin	1 drug-resistant mutant for every 10^{5-6} bacilli
Ethionamide	1 drug-resistant mutant for every 10^{3-6} bacilli

Figure 2.5: Frequency of drug-resistant mutants to anti-TB drugs (Lemos and Matos, 2013)

Figure 2.5 illustrates the mutation frequency with respect to all anti-TB drugs is quite low(Lemos and Matos, 2013). The frequency of mutations conferring to delamanid resistance range from 2.51×10^{-5} to 6.44×10^{-6} in vitro (Reichmuth *et al.*, 2020). The generation of a genetically variant organism after the propagation of bacterial cells represents a minute heterogeneity in a genetically homologous bacterial cell population. A large population size with a very low mutation rate minimises the chance of clonal interference (competition between resistant clones to let dominant clone survive).

In such case, susceptible strains may sweep the resistant form(s). Surprisingly, that the mutant strains are still emerging and becoming resistant to the available anti-TB drugs. The MDR-TB/RR-DR cases are on the consistent increase every year (WHO, 2021). Hence, the gene(s) mutation(s) that confers resistance to the anti-TB drugs could be used as targets to identify the drug resistant Mtb strains.

Lineages 2 (East-Asian lineage) and 4 (Euro-American lineage) are regarded as evolutionary modern due to deletion of the TbD1 region. These strains are considered highly transmittable and virulent. The strains of these lineages also have a high replication tendency in the host (human and animal models) macrophages (Brites and Gagneux, 2015). Bacterial cells with high replication capacity inside the host cells will probably have less doubling time. This will in turn give rise to a luxuriant population of bacteria in a short time inside the host cells. When the macrophage(s) full of bacterial cells burst, the new bacterial progeny are released.

The presence of any mutant cells undergoing spontaneous mutation is quite possible. On further acquiring a fresh host cell, this mutant cell(s) will have an opportunity to amplify its population. This will probably increase the population of resistant bacteria. An evolution may further cause the resistant bacteria to acquire more mutation(s), which may further boost the survival rate of these mutated organisms. Therefore, the transmission of those mutated strain(s) to new human hosts will already be resistant to the drugs may have added biological fitness. Such bacterial cells will help sustain the drug-resistant bacteria in the community hence sustains the continuity of DR-TB emergence.

The latent state of TB has been linked to the pathogen's strategy to sustain and transmit itself from one host to another (Brites and Gagneux, 2015). Again lineage 2 strains have five times more tendency to develop the active state of disease compared to strains belonging to lineage 6. The lineage 6 representatives are the least likely to develop an active TB state compared with other MTBC lineages (de Jong *et al.*, 2008). Therefore, there are fewer chances that a patient with latent TB infected by a strain of lineage 6 would have their disease state transformed to an active state. This is the opposite for lineage 2 representatives. However, if the person is immunosuppressed then the condition is certain to prove fatal if infected with lineage 2 strains. Even immunosuppression may provide an opportunity for lineage 6 strains to transform the latent state to an active state more progressively than in a healthy human host.

The acquisition of attributes like high virulence, transmissibility, ability to progress from the latent to active state of disease among the modern strains of lineage 2 and 4 could be due to the TbD1 deletion. This genetic region may contain any gene(s), which encodes a protein(s) that function to suppress these features when present. As an example, the strains of ancient lineages 5, 6 and 7, the strains of these lineages contain TbD1 region and have relatively low virulence, less transmissibility, and less tendency to progress to an active state. Therefore, sequence analysis of lineage 5, 6 and 7 representative strains by sequencing the TbD1 region may provide information on the presence of any gene(s) encoding protein(s) that may have a role in influencing the above-discussed characters. The TbD1 including other signature genes of a particular lineage could be used as a target for the identification of lineage specific Mtb representative strains.

The IS6110 is another gene that is present as multiple copies (up to 25) and is specific to MTBC. Therefore, it is used as an important biomarker for diagnosis and strain typing of the *Mycobacterium* spp. The presence of the multiple copies per genome could increase the sensitivity of an assay (Roychowdhury *et al.*, 2015). The *senX3-regX3* intergenic region is present in the Mtb as a single copy gene. This gene is claimed to be Mtb specific. Therefore, it could an target for highly specific assay (Broccolo *et al.*, 2003).

2.3.2   The *rpoB* gene and its function

More than 90% of rifampicin-resistant TB strains are also isoniazid-resistant and these isolates mainly (higher than 95%) possess mutations within cluster I of the *rpoB* gene, which encodes the β-subunit of the RNA polymerase, making this gene worthy of study (Pietzka *et al.*, 2009). Referring to *E. coli* β subunit amino acids numbering, this region is composed of 4 units:

N-terminal cluster (amino acids 146 to148)

Cluster I (amino acids 507 to 534)

Cluster II (amino acids 563 to 574)

Cluster III (amino acid 687) (Brandis *et al.*, 2012).

Therefore, the mutation rate in amino acids 507- 534 making cluster I of the *rpoB* gene is highest. More recently, Andre *et al.* (2017) proposed an alternative numbering system based on the Mtb *rpoB* gene (Figure 2.6). Referring to the *E. coli rpoB* sequence annotation might lead to misunderstanding according to him especially when technologies like whole genome sequencing is used (Andre *et al.*, 2017). Therefore, mutations based on the *rpoB* genes will be named according to this novel numbering system for more precision (Table 2.3).

Figure 2.6: Aligned *rpoB* sequences of *E. coli*, Mtb*, M. leprae* and *M. kanasii*. All sequences are the same, but the location of mutation prone codons is different (Andre *et al.*, 2017).

Table 2.3: Conversion table between the two numbering systems for the *rpoB* mutations among MTBC (Whole data taken from Andre *et al.)* (Andre *et al.*, 2017).

Mtb H37Rv wildtype (GenBank AL123456)	Codon position (Conversion from *E. coli* numbering)
GTC (Val)	170(+24)
GGC (Gly)	426(-81)
CTG (Leu)	430(-81)
CAA (Gln)	432(-81)
GAC (Asp)	435(-81)
CAC (His)	445(-81)
TCG (Ser)	450(-81)
CTG (Leu)	452(-81)
ATC (Lle)	491(-81)

Alignment of the *rpoB* sequence from the different organisms (Mtb, NTM and *E. coli*) gives a hint that this sequence might be conserved across other organisms due to the common function of the gene. This may provide cross-reaction when an assay(s) based on this gene is designed.

Many bacteria including the Mtb have a single RNAP enzyme consisting of a multi-subunit αββ'ω core making a crab claw-like structure (Figure 2.7). The β and β' subunits make a site or a pocket of 27Å. This site is positively charged due to the presence of two Mg^{2+} ions that receives the template DNA and provides a catalytic locale for phosphodiester bond formation, a second channel for incoming nucleotides and a distinct exit for the elongating RNA transcript. Generally, RIF binds to the β subunit, which is *rpoB* gene encoded, on propitious proximity, hence halting transcription (Comas *et al.*, 2012; Alifano *et al.*, 2015).
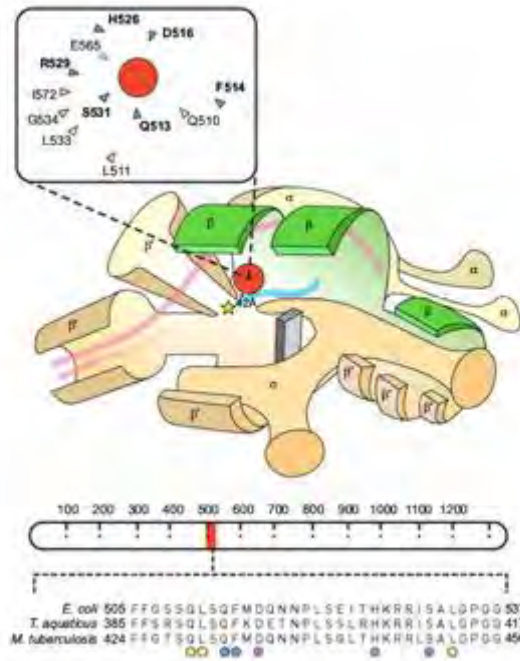
Figure 2.7: Schematic representation of the interaction between RNAP enzyme and RIF molecule. The enzyme functions to transcribe mRNA from the template DNA (pink strands). The RIF molecule (red circle) interacts with the active site of the enzyme (yellow star) inhibiting transcription. On acquiring mutation, the drug molecule fails to bind the active site of the enzyme and the transcription process occurs successfully. Below the model is a representation of the *rpoB* gene fragment encoding the β subunit of RNA polymerase. The amino-acid sequence of the RRDR (blue region) is magnified below. The alignment contains the amino-acid sequences of *E. coli*, *T. aquaticus* and Mtb*.* Amino acids that interact directly with RIF are indicated by circles and the colours correspond to the inset diagram. Circles highlighted in red indicate residues that are most frequently observed in RIFR isolates (Comas et al., 2012).

2.3.3   Effect of mutations on biological fitness of the bacteria

The functional genes encoding different proteins are generally the target points of the antibiotics. The mutations in such genes may also affect the pathogenicity of the pathogen, which can be termed as fitness cost (Koch *et al.*, 2014). The biological fitness of bacterial cells generally decreases on achieving resistance. However, occasionally, this may be accompanied by another genetic variation that may improve the fitness cost instead. Such

mutations are called fitness-compensatory mutations (Figure 2.8)(Andersson and Hughes, 2010; Hughes and Andersson, 2015).
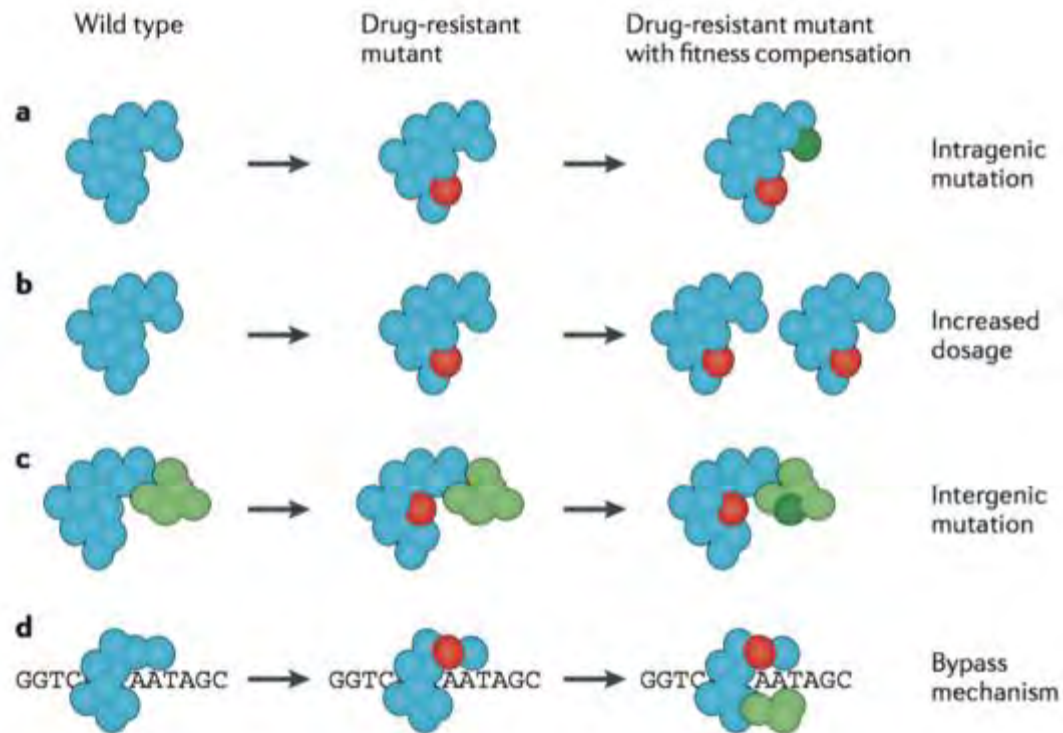
Figure 2.8: The proteins derived from drug resistant mutations (red circles) decreasing in organism fitness and different mechanisms of fitness compensation. a Compensation by intragenic mutation (dark green circle). b Compensation by increased dosage of a low-fitness mutant protein. c Compensation by intragenic mutation (dark green circle) occurs during protein interactions or complex formation. d an additional regulatory factor compensates for the low functionality of altered protein (Hughes and Andersson, 2015)

A study done by Brandis *et al.* (2012) used *Salmonella enterica* as a model to determine the location and response of compensatory mutations regarding improved growth rate and RIF susceptibility change. A well-characterized *RifR* mutation (*rpoB* R529C) was targeted to emanate independent lineages. The study reported compensatory mutations in every emerged lineage allocated in *rpoB A, rpoB B, rpoB C* genes. These mutations improved the fitness of a primary *RifR* mutation. Apart from this, resistance to rifampicin was observed due to lone mutations in the *rpoB A or rpoB C* gene (Brandis *et al.*, 2012).

A similar finding was observed in another study by Comas *et al.* (2012) which reported a set of compensatory mutations was reported in rifampicin-resistant strains of Mtb*.* These mutations increased the fitness cost of the resistant strains in vitro and were found in a major proportion of more than 30% in the clinical isolates from high burden countries (Comas *et al.*, 2012). Based on the studies, compensation mutations may add to the biological fitness of the bacteria. Such mutations may or may not in combination with other factors like selective pressures aid to sustain and further disseminate the resistant strains. On achieving mutation(s) independent of compensatory ones, the biological fitness of bacteria decreases. This decreased fitness with other factors may eliminate the emerged strain.

Three types of resistance have been reported in Mtb:

1. Natural resistance: resistance due to spontaneous mutation independent of any prior drug exposure.
2. Primary resistance: resistance in people who were never exposed to TB curative drugs.
3. Acquired resistance: a resistance acquired in individuals treated with TB drugs (Dye, 2009; Lemos and Matos, 2013).

2.3.4 Implications of drug resistance on an individual

Consider a scenario in which a TB infected person (say, person A) transmits TB causing bacteria to a healthy person (person B) via coughing. There are two chances; person B will not contract the infection if the bacterial cells are inadequate in number (just 1 or 2). However, if an abundant bacterial population is transmitted, there are increased chances that person B will become infected. Once the person starts exhibiting typical symptoms of TB, he may start receiving medication. The prescribed drug(s) type and their relative concentration will apply a selective pressure inside the patient (person B) body.

Success in eliminating the infection will depend on the type of cell in the patient's lungs. If person B acquires a homogenous population of susceptible bacterial cells, then there is a possibility of killing the whole infecting flora due to synergistic association of host immune response and prescribed anti-Tb drug(s). However, bacterial cells may remain in the lungs if a mixture of susceptible and drug resistant strains or a neat population of cells resistant to the given drug(s) is transferred. In such a case, the elimination of infection depends solely on the patient's immune system.

 In the worst case, if person B is immunosuppressed, e.g., HIV infected or diabetic, the patient's weak immune system and the presence of ineffective drug(s) due to the presence of drug-resistant strains, will fail to tackle and eliminate the bacteria. This condition can be fatal. Within the host, bacterial cell diversity acquired either by acquisition of mixed population or by mutation(s) inside the host cells represents a challenge to identify the cell diversity (Hatherell *et al.*, 2016). In the case of a mixed population, there are more chances that different bacterial cells with or without any genetic variation may not be evenly present and uniformly distributed in the host lungs. In such conditions, a particular cell type with a higher population may be identified easily with any of the dedicated molecular assays causing the detection of the other cell type to be relatively difficult. Repetitive patient sampling and robust assay(s) of high resolution such as sequencing may prove beneficial.

Bacterial cells evolved from either natural or acquired resistance may differ in their transmission rate. Despite low mutation rates in MTBC, the emerged resistant strains via natural resistance might have higher transmission rates and virulence as compared to the strains that rose from the acquired resistance, or vice versa can also be possible. Acquired drug resistance occurs mainly by abridgement in internal drug concentration, target modification and drug inactivation or inability to stimulate (Hughes and Andersson, 2015). A comparative analysis of rates of transmission and degree of virulence between the strains evolved by natural and acquired resistance might help identifying which resistant strain source promotes continual drug-resistant TB in epidemic areas.

2.3.5   Implications of drug resistance on a community- PNG context

Tuberculosis is the third major cause of mortality in PNG. The TB annual incidence rate in PNG is 432 /100,000 population nationwide (WHO, 2021). The incidence rate is surprisingly more, 674/100,000 population in the Western Province. Further, from 2014 till 2016 the rate in Balimo and Gogodala region was 727/100,000 population (Diefenbach-Elstob *et al.*, 2019). These are nevertheless an approximate figure due to lack of precise case reporting to health care facilities especially in rural domains where transport facilities are quite limited and represent a major obstacle (Cross *et al.*, 2014a).

There is no detailed information about the TB lineage pattern in the whole of PNG. However, different studies in different regions or provinces have been conducted to determine the lineage prevalence. According to Gilpin *et al*. study (2008), patients from the Western province residing in the Torres Straits were found infected with multi drug resistant strains of Lineage 2 (Beijing type) (Gilpin *et al.*, 2008). The results obtained from sequencing the samples from Kikori, Gulf province indicated the correlation with strains of Lineage 2 (in maximum cases) as well. The high prevalence of lineage 2 circulating in the provinces of PNG represents its attribute of high virulence and transmission rate. This observed feature matches the Casali *et al*. (2014) statement (Casali *et al.*, 2014). Therefore, it can be said the presence of lineage 2 could be the reason for drug resistance in PNG due to the high replication rate of the representative strains.

Ballif *et al. (*2012) in their study applied molecular tests on 173 isolates. Of those 173, 172 had phenotypic DST (pDST) results. The study indicated the prevalence of lineage 4 (39/173), lineage 2 (39/173) and lineage 1 (1/173). Overall, 27/172 isolates were reported as resistant to more than one drug, 15/172 were monoresistant and 3/172 were MDR. Ten strains were rifampicin (RIF) resistant, 21 were isoniazid (INH) resistant and four as ethionamide (ETH) resistant. Nine strains were reported to be Streptomycin (STR) resistant (Ballif *et al.*, 2012).

A study by Ley *et al. (*2014) indicated the high prevalence (84.4%) of Lineage 2 in Alotau, followed by Eastern highland province, and Goroka, with 60.5% incidence in infected patients. Lineage 4 was prominent (76.6%) in the Madang region, while Lineage 1 was quite sparse across all three provinces. The samples belonging to lineage 4 found in Madang were of the LAM spoligotyping family, while the samples of lineage 2 were of the Beijing family and were further subtyped to identify specific regions of difference (RD). Of 67 Beijing strains from all three sites, one belonged to sublineage 1(no deletion in RD 181, RD150 or RD 142) and another was of sublineage 2 (having only the RD 181 deleted), and nearly all (63/67) Beijing samples belonged to sublineage 3, with RD 180 and RD 150 deleted (Ley *et al.*, 2014a).

A recent study done by Guernier-Cambert *et al.* (2019) demonstrated the presence of the mixed Mtb strains of different lineages. In this study, 345 smeared sputum samples were tested. Of those 345, 206 reacted in the TaqMan based qPCR assays indicated the presence of the *Mycobacterium* DNA in them. Further good quality DNA (Ct< 36) templates (n=162) were sorted for spoligotyping. Of 162, successful spoligotyping was done in 80 samples. All of those were AFB positive. The spoligotyping results revealed that 44% of the samples (35/80) belonged to lineage 2 and 30% (24/80) were representatives of Lineage 4. One (1/80) template indicated presence of the lineage 1 and the remaining 20 were a mixture of lineage 2 and 4. Further, the sequencing of few samples revealed the presence of the S450L mutation in the lineage 2 representative strains (Guernier-Cambert *et al.*, 2019).

Almost all studies unanimously conclude the presence of the mixed Mtb strains belonging to lineage 1, 2 and 4. Out of these three, lineages 2 and 4 are regarded as modern lineages that are more virulent with a high transmission rate than lineage 1 (ancient) (Brites and Gagneux, 2015). In addition, the representative strains of lineage 2 and 4 both make high proportion that reflects their widespread in PNG. In section 2.2.7, a study by Andersson and Hughes has been mentioned. The study claims that the mutation(s) in the bacterial cells generally decreases their biological fitness. However, seldom fitness-compensatory mutations may contribute to improving the overall fitness cost (Andersson and Hughes, 2010; Hughes and Andersson, 2015).

Sustained presence of these strains (of lineage 1, 2 and 4) for nearly a decade (based on the studies conducted in PNG from 2008 till 2019) demonstrates their high fitness cost. This might have led them to remain longer. The prevalence of the S450L mutation furthermore indicates that some strains (of lineage 2) have already acquired the mutation in their *rpoB* gene. This mutation confers resistance to the rifampicin drug. Now the question arises, is there any compensatory mutation(s) in those mutated Mtb strains that improved their overall fitness cost? Where did they acquire the mutation(s), in the environment or the host lungs? If in the host lungs, how? Was it a result of spontaneous mutation or mutation that occurred under the influence of internal selective pressure(s) such as in the presence of drugs?

Ballif *et al.* (2012) study indicated the presence of mutations in the other key genes that conferred resistance against rifampicin, isoniazid, streptomycin and ethionamide drugs (Ballif *et al.*, 2012). These mutations might have a direct or indirect role in adding fitness to the mutated Mtb strains. It can be said that the Mtb strains in PNG are a mixture of the lineage 1, 2 and 4. These strains may further be divided into two categories: wildtype(s) (without any mutation) and mutant(s) (harbouring mutation such as S450L). However, it was interesting to note that all the strains that had S450L mutation were of lineage 2 (Guernier-Cambert *et al.*, 2019).

This further raises a series of questions such as: does the Mtb belonging to lineage 2 have more mutation supply rate as compared to the lineage 4 Mtb? If yes, what makes those Mtb more receptive to undergo frequent genetic variation(s)? The lineage 2 strains are present in most of the PNG. Are those the wildtype Mtb undergoing natural mutation(s) leading to new mutants or the mutated Mtb are simply disseminating across the country due to their high transmission rate? Is there any difference between the mutation rates of the mutated lineage 2 Mtb strains (with S450L mutation) and the wildtype Mtb strains? What is the fitness cost of the mutated Mtb cells emerging from the mutated (with S450L mutation) lineage 2 strains because of spontaneous mutation? Genomic insights of these strains may provide the answers to these questions.

Consider three scenarios:

1. The host acquires mutated strains. During infection initiation, there is no clonal interference at all. The mutated strains simply undergo division thereby increasing their population due to high replication and mutation supply rate.

2. The host had an infection with the wildtype strains. When those cells underwent primary resistance during propagation, they gave rise to few mutants. Those mutated cells (if of lineage 2 and 4 strains) on repetitive divisions acquire population size that might have swept most of the wildtype cell population instead. High clonal interference occurred. This could be possible due to the attribute of lineage 2 and 4 strains that have high transmission and replication rate within the host cells.

3. Acquired mutation (on exposure to anti-TB drugs) among the infecting Mtb strains in the host lungs occurs. Again, high clonal interference takes place between the wildtype and the emanated mutated cells. The lineage 2 and 4 strains due to their attributes have a population comparable to the wildtype cells.

The first possibility appeared to happen in Madang PNG on considering the Ballif *et al.* (2012) study (Ballif *et al.*, 2012). Now the question is, why only the lineage 2 Mtb acquired the S450L mutation when both lineages 2 and 4 are evolutionary modern lineages? Both are attributed to be highly virulent with a high replication rate and transmissibility (Brites and Gagneux, 2015). Further, is there any variation(s) in the clinical manifestation of TB caused by the Mtb of different lineages? Does lineage 1 representative produce symptoms different to the lineage 2 and 4 Mtb and so on? What would be the scenario of clinical signs and symptoms if the Mtb strains of different lineages are present in a single host? Will there be any difference in TB signs and symptoms comparable to infection by a monoclonal Mtb cell population?

In a study conducted by Guernier-Cambert *et al. (*2019), some samples had a DNA mixture of lineage 2 and 4 representatives. Their proportion was 25% (20/80) (Guernier-Cambert *et al.*, 2019). Mixed infection might equate to the presence of different TB strains in a single host. This suggests the following possibilities:

1. The host received a mixed population infection of Mtb belonging to lineage 2 and 4 at the time of infection. A mixed infection occurred. Both cell types had a population in the host lungs disseminated in a way that cells of both lineages were present in the aliquot of sputum sample taken. Hence, they were identified.

2. The patient was infected with the lineage 2 Mtb strain (or lineage 4) initially. For treatment/ interstate migration due to work or personal reasons, this individual may have visited an area where the other lineage was predominant.

It could further be said that the clonal interference between the Mtb strains of lineage 2 and 4 might be minimum. This might have allowed them to undergo clonal expansion. The presence of both cell types made it possible to get detected as mixed infection on referring to the mixed samples in Guernier-Cambert *et al. (*2019) study (Guernier-Cambert *et al.*, 2019). This interference might be variable (ranging from low to high) when the population of either lineage 2 or 4 Mtb and the monoclonal population of the wildtype Mtb co-exist in the host lungs.

In such a case, identification of both may depend on the state of their existence. Whether they co-exist synergistically or antagonistically. An antagonistic relationship may cause one cell type to dominate over the other type. This competition may contribute to the identification of the dominant strain only. Population size may matter if the relationship is synergistic. Large population size may mean that there are more chances of that cell population getting easily identified.

Bainugnosia *et al.* (2018) in their study characterised an extensive drug-resistant (XDR) Mtb strain of the Beijing sublineage 2.2.1.1. This strain caused a drug-resistance outbreak on Daru Island in the Western Province of PNG. The Oxford Nanopore Technologies (ONT) MinION sequencing technology was used to determine molecular-based mechanisms that are responsible for the pathogenesis and virulence of the strain. Genes such as proline-glutamate (PE) with 99 loci and proline-proline-glutamate (PPE) having 69 loci were expected to be the main contributors to virulence and pathogenicity. The strain's whole genome including above-cited genes was analysed and compared to the phenotypic drug susceptibility testing data. Illumina sequencer reads were also involved for comparison (Bainomugisa *et al.*, 2018).

The presence of a strain that is already resistant to a range of drugs might be an indication of a potential outbreak. Further, its high transmission rate may lead to its widespread dissemination. Such dispersion might lead to either fresh infection, reinfection, or superinfection. This is an alarming state where these drug-resistant strains need to be accurately identified and eradicated to halt their further spread. The reversion of resistance is naturally possible. The time needed to curtail a large population of resistant bacteria is inversely proportional to the cost of resistance (Andersson and Hughes, 2010).

2.4 Single nucleotide polymorphisms and their detection

Bacterial genotyping contributes to TB control (Sengstake et al., 2014). The genetic analysis in which different polymorphic markers are used to locate a domain inside a genome associated with disease phenotype is one of the most potent tools in the disease process (Mhlanga and Malmberg, 2001). The polymorphic markers are referred to as Single Nucleotide Polymorphisms (SNPs), a genetic locus where two or more alternative bases prevail with noticeable frequency (>1%) (Csako, 2006). The term single nucleotide polymorphism (SNP) represents genetic change. This change arises either by the addition or deletion of nucleotide base(s). The methods to identify SNP are mentioned in the following section.

2.4.1 High resolution melting analysis

High resolution melting (HRM) analysis is a technique used to identify the link between the temperature and level of denaturation using a real-time PCR instrument. The DNA denatures on increase in temperature and a sigmoid melt curve is observed (Figure 2.9) (Tong and Giffard, 2012). The melt curve is the denaturation of a piece of DNA with increasing temperature. The melting temperature is associated with sequence length and proportion of GC content (G+C %). This is due to the presence of additional hydrogen bond

between GC pairs on comparing AT pairs. An estimated relationship between sequence length and content is as follows

$T_m = 81.5 + 16.69 \times (\log [Na]) + 0.41 \times (\%G\ C) = (500/\text{sequence length})$.

A change in T_m (ΔT_m) corresponds to a single G-C $\Leftrightarrow$ A-T change. A transition from A to T or G to C is possible under stringent conditions but less discernible (~<0.2°C). However, a change of base A to G or A to C could be identified (Tong and Giffard, 2012). This can potentially differentiate between wildtype and genetic variant DNA samples. There are chances that multiple SNPs may compensate and do not get identified. This can be overcome by using positive controls derived from wildtype organisms.
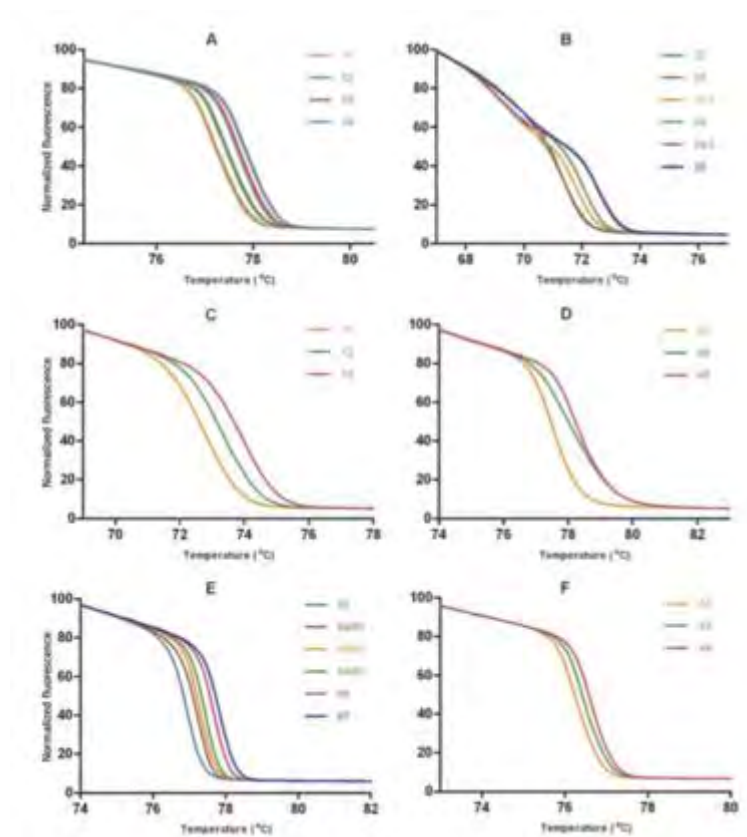
Figure 2.9: HRM curves for six internal fragments from the multilocus sequence typing loci of *Staphylococcus aureus*. The number of G+C residues contained in the fragment labels the curves. Panel A shows the melting curve variants of a 181-bp internal fragment of arcC demonstrating an increased melting temperature (T_m) as the number of G+C residues increases from 51 to 54. Panel B exhibits evidence of two melting domains in this 140-bp internal aroE fragment, leading to discriminate the variants of the same %G C. These

additional curves have been labelled 23.5 and 24.5. Panel E shows melting curves from variants with 64 and 65 G+C residues that could not be consistently discriminated for this large 219-bp internal fragment of tpi and have been labelled 64/65. The other fragments are from gmk (C), pta (D), and tpi (F) (a region of tpi different from that represented in panel E) (Tong and Giffard, 2012).

The HRMA is cost effective technique and easy to perform. A dye is added to double stranded PCR product before melting and is a closed-tube system without requiring any specialized instrument. This technique is used to detect mutations, screening of presequence, SNP typing, methylation and quantification of different alleles (Vossen *et al.*, 2009; Tong and Giffard, 2012). A recent study done by Sharma *et al*. reported the application of HRMA to screen drug resistance in tuberculous meningitis, a form of EPTB. The real-time assays based on *rpoB*, IS6110 and MPB64 were applied on 110 patients. The resistance to rifampicin was observed in 3 cases and mutations in the 516 codon and 531 codons were found (Figure 2.10) (Sharma *et al.*, 2015).
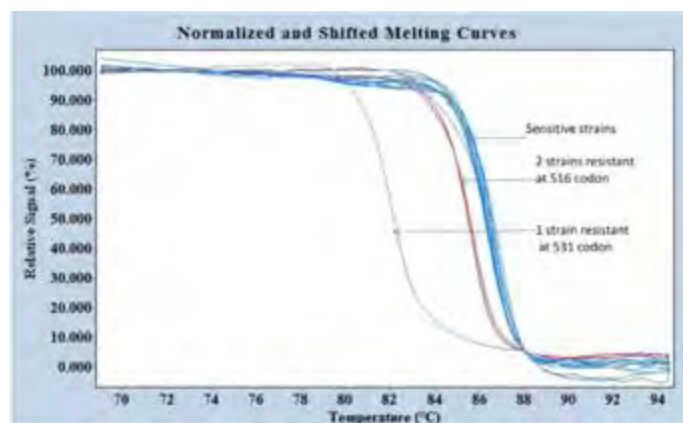
Figure 2.10: Melting curves derivative of fluorescence over the temperature of CSF samples containing mutations in different codons. Each trace shows the change in fluorescence ratio over time with respect to temperature. Susceptible refers to *Mycobacterium tuberculosis* that does not contain any drug resistance point mutations. Derivative melting curves derivative of fluorescence over temperature of CSF samples containing mutations in various codons (Sharma *et al.*, 2015).

Based on the above cited study, it can be said that HRM analysis has the potential to screen the extrapulmonary samples. If this claim is replicated and verified, it may be quite useful in diagnosing the EPTB patients. This would be an additional advantage of real-time PCR.

2.4.2   Probe-based detection

Hybridisation methods are the preferred ways of identifying genetic variation(s). The specificity of the hybridization, however, presents a major problem. The factors relative to mass transport and kinetics are controlled to achieve the required level of probe binding efficiency. Parameters affecting the binding directly such as temperature, ionic strength sequence and oligonucleotide concentration are to be in the controlled state as the relative binding of the probe determines the identification of SNPs. The need of defining the parameters demands a great effort in assay optimisation for every target DNA. The advent of assays based on TaqMan and beacon probe chemistry minimizes the effect of specific conditions on hybridisation thereby making these assays more sensitive and robust (Knez *et al.*, 2014). This study will devise assays based on the TaqMan chemistry, for their claimed features and utilities.

2.4.3   Extension primers

The addition of the end of the forward primer to SNP and the ability of primer to bind and extend determines the success of an assay. The evaluation of SNP allele frequency in pools of individual DNA samples is an alternative approach to genotyping methods, hence reducing the reaction numbers while retaining the information for screening. Moreover, SNPs exhibit no homoplasy which means there are no chances of misclassification of MTBC strains into robust phylogenetic lineages (Stucki *et al.*, 2012).

Two comprehensive SNPs based methods have been developed :
1. Multiplex Oligonucleotide Ligation-PCR (MOL-PCR) and
2. Taqman real-time PCR

The former is based on allele-specific ligation to differentiate alleles. It requires a Luminex device, which permits simultaneous interoggation of SNPs. Rather than stuffer sequences, the method uses flouresecent beads and two tag-sequences for the multiplexed analysis.The signal is identified by the flow cytometer or Luminex device (Figure 2.11). The Taqman real-time PCR on the other hand, a sensitive and specific, commercially available method considered as a standard for SNP-typing with ease to perform it in a 96-well plate within 2 hours. The MOL-PCR shows its inferiority to Taqman real-time PCR in analysing just a single SNP per reaction (Stucki *et al.*, 2012)
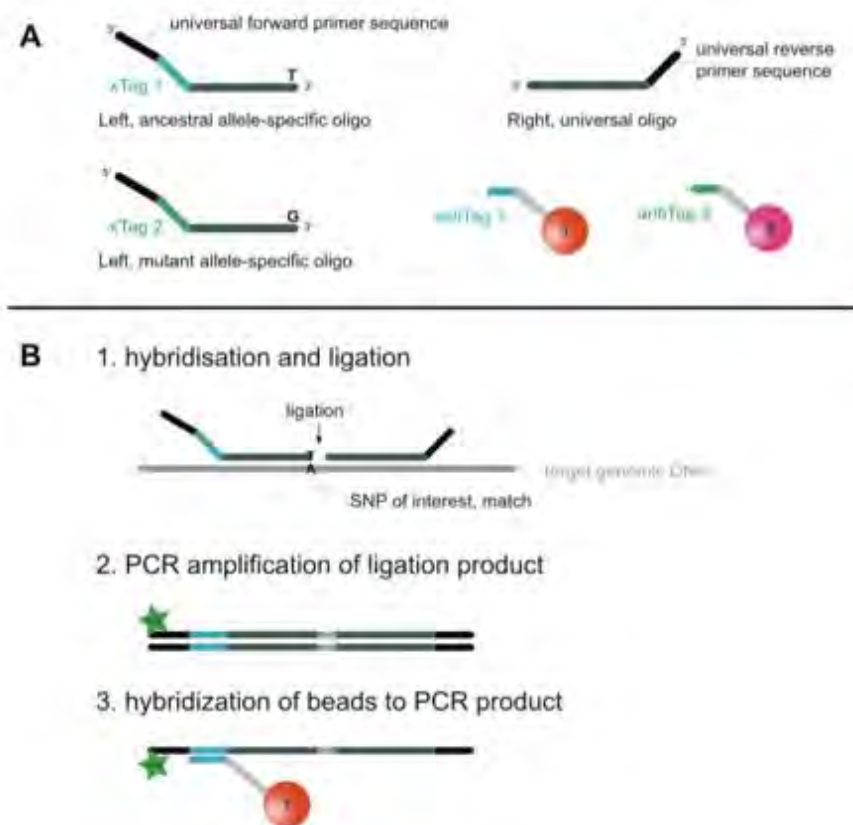
Figure 2.11: Illustration of MOL-PCR. A. Three oligonucleotides and two fluorescently labelled beads are used for the interrogation of one SNP. B. MOL-PCR consists of three steps. 1. One of two competitive allele-specific left-hand probe oligonucleotides (LPO) and one universal right- hand probe oligonucleotide (RPO) are hybridized to the template DNA and ligated. 2. With PCR and a reporter-labelled forward primer, the ligated oligonucleotides (LPO+RPO) are amplified. 3. After denaturation of the PCR product, allele-specific fluorescent beads carrying an allele-specific antiTag sequence are hybridized to the

amplicons. This will result in beads carrying reporter fluorescence (bead 1 in the example) and beads not carrying fluorescence (bead 2). Reporter fluorescence per bead is measured with a flow cytometric device (Luminex) (Stucki *et al.*, 2012).

2.4.4 Sequencing

"…[A] knowledge of sequences could contribute much to our understanding of living matter" – Frederick Sanger (Heather and Chain, 2016).

The rapid acquisition of information regarding the constituents of nucleic acids accurately is pivotal to understand various mechanisms in the field of biological sciences (Pettersson *et al.*, 2009). Sequencing is different from the other SNP detection methods in the fact that it is direct and definite. Sequencing variants are as follows:

2.4.4.1 *Sanger sequencing*

The development of chain-termination or dideoxy technique uses chemical analogues of deoxyribonucleotides (dNTPs) and dideoxynucelotides (ddNTPs) devoid of 3'OH hydroxyl group to halt the DNA extension. As it is depended on the activity of DNA polymerase, this technique is also called enzymatic sequencing or Sanger sequencing (Heather and Chain, 2016). Sanger sequencing even though is still in practice, there are some limitations associated with this technique. Due to an enzymatic reaction and the need to prepare the template, the throughput is low. Moreover, it is confined to up to 1000 bp length only. The templates of bigger size are unable to get sequenced with this technique. The cost per base is still high. (Morey *et al.*, 2013).

It is quite clear that Sanger sequencing will be suitable to sequence the segments (conserved or variable regions) of the target sequence(s) selected in this project. This may help to find novel SNP(s) or any amino acid substitution(s) conferring drug resistance in the plasmid containing the *rpoB* gene insert derived from DNA extracted from the sputum samples in PNG.

2.4.4.2 *Next-generation sequencing*

The second-generation sequencing has the provision of using a number of the same or different clonal templates in parallel with sequencing chemistry based on DNA polymerase assisted replication (McGinn and Gut, 2013). But rather than using radio or fluorescent labelled dNTPs or oligonucleotides, it is based on the luminescent method of determining the amount of pyrophosphate production during the replication process. The reliance on enzymatic replication in both Sanger and second-generation sequencing makes them sequence by synthesis (SBS) techniques (Heather and Chain, 2016).

Being available in different platforms, this generation sequencing provides high throughput, with suitability to sequence even the whole genome of the organisms. Also, next-generation technologies have dragged down the DNA sequencing cost from $1000 per megabase to mere ten cents per megabase. This drastic price reduction has extended the accessibility and use of these technologies in a wide range of fields providing crucial data in a comprehensive manner (Shendure and Lieberman Aiden, 2012).

However, the amount of data attainment and its management along with accurate interpretation is the main challenge. The other limitation of NGS is its short-read length, which is mainly due to the decrease in the efficiency of the sequencing chemistry while running. The short reads are often difficult to align up against the reference genome. In addition, NGS is based on the use of PCR and PCR-incorporated bias complicates the potential of NGS (Morey *et al.*, 2013).

2.4.4.3 *Targeted sequencing*

Targeted sequencing on the other hand relies on targeting the specific region(s) in the genome(s). This overcomes the complex nature of sequencing whole genomes. The amount of data with this strategy will be sufficient. However, the role of PCR is crucial. The correct normalization of PCR helps to sustain the coverage uniformity. In addition, it provides an opportunity to sequence multiple reads of the same DNA segment in parallel, thereby increasing the sequencing depth, which is important to identify the mutation(s). Facilitated targeted next-generation sequencing provides cost-effective full-length (more than 1kb) gene analysis (Daum *et al.*, 2014). The DNA fragments of more than 1Kb are appropriate for this technique. This technique might be beneficial in resource-limited settings to identify SNPs.

2.5   Conclusion

Mycobacterium tuberculosis has structural complexity. Despite having a low mutation rate and lacking HGT, it exhibits great genomic diversity. This genetic diversity among different Mtb strains makes them phenotypically variant. This causes them to differ in their virulence and transmissibility. The further acquisition of a drug resistant state presents a real challenge for TB control. There are many diagnostics to identify Mtb. However, factors such as cost, robustness, turnover time, convenience and whether they are fit-for-purpose need to be considered.

In resources-constrained countries like PNG, there is a requirement for cost-effective molecular based tools to identify and characterise the Mtb. Early accurate diagnosis may lead to prompt patient management. This in turn may reduce the misuse or overconsumption of anti-TB drugs that otherwise could lead to Mtb strains acquiring drug-resistance against those drugs. This project, therefore, aims at addressing the need of developing the Mtb identification and characterisation tools, so that, those tools could be used as in-house routine tests in these settings and provide an economical and sustainable

alternative to expensive commercial devices. This project aims to develop molecular based in-house identification tools to detect, quantify and characterise Mtb from DNA samples. These tools may also support the laboratories that grow bacteria from the patient sputa in liquid cultures (MGIT™ tubes) for Mtb confirmation after DNA extraction from culture supernatant or DNA from the microscopic slides.

These real-time PCR based assays will help address some important questions including: is it TB? If yes, how many organisms are present? This thesis will describe the utility an IS6110-based assay to provide qualitative information regarding the presence or absence of Mtb DNA.. Therefore, this assay could be quantitative in nature and will provide information on the number of copies of DNA present in the sample, as it is a single copy gene. High *rpoB* gene copies mean high copies of the Mtb genome. Furthermore, the *rpoB* assay will provide quality assurance of the quality and amplifiability of DNA from clinical material.

The HRM assay will identify outliers and has the potential to differentiate between drug susceptible and resistant bacteria. In the Mtb/PNG context, the use of sustainable and cost-effective Mtb qPCR controls is essential. Typically, genomic DNA or plasmids carrying target gene(s) are used as controls in these assays. However, the overall cost of the genomic DNA extraction or constructing a plasmid(s) is challenging in these settings and hence may influence the accuracy of Mtb testing. Control reagents that could be used in these settings were also developed as part of the current research. The BMS MIC™ qPCR cycler will be used in this study. The device could comply with the WHO ASSURED criteria (Mabey *et al.*, 2004) based on its characteristics mentioned in the Chapter 1, section 1.5.

Utilising such tools in these types of settings may, in the future, contribute to answering other key questions including: How can strain diversity within a single host be found without using expensive commercial systems? Is there an economical alternative that could be applied in resource-constrained settings? What gene targets could be used to determine the lineages of the Mtb in PNG?

CHAPTER 3

Materials and Methods

3.1   DNA samples available for analysis

DNA templates available for the current study are described in a previous study (Guernier *et al*. 2018). These samples were prepared from a range of clinical and culture material imported from Balimo, PNG, stored for different period at JCU. They included: stored sputum, NaOH decontaminated sputum, archived stained sputum smears on glass slides and MGIT™ culture tubes inoculated in Balimo and cultured at JCU in the PC3 laboratory. These samples were from presumptive TB patients.

It is important to note here that the sample acquisition, processing, and DNA extraction from the imported samples was entirely done by the other members of the JCU TB research group. The DNA extraction from the MGIT™ tubes and the decontaminated sputum was done in the PC3 laboratory using commercial kits without changing manufacturer's instructions. The smeared slides were used and processed in the PC2 facility. The protocol is described in Guernier *et al.* (2018) (Guernier *et al.*, 2018). The extracted and sterility checked DNA templates (from the PC3 set up) were sorted and provided for this project. Therefore, the provided DNA templates for this project were ready to use in the PC2 laboratory.

The samples provided (Table 3.1) for this study were old (remained stored over time), potentially contaminated during transit and were of low quality. There was a need for an identification tool(s) to identify the Mtb from those imported samples. So that, it could be confirmed whether those collected samples contained the Mtb in them.

Table 3.1: Sample types provided for this project.

DNA name	Type of sample	Type of sputum
248 DS 1	DNA extracted from archived sputum smear slides	Direct smear
361 DS 1	DNA extracted from archived sputum smear slides	Direct smear
322 DS 1	DNA extracted from archived sputum smear slides	Direct smear
26 DS 1	DNA extracted from archived sputum smear slides	Direct smear
68 NaOH DS 2	DNA extracted from archived sputum smear slides	NaOH decontaminated and concentrated sputum
323 DS 1	DNA extracted from archived sputum smear slides	Direct smear
274 DS 2	DNA extracted from archived sputum smear slides	Direct smear
321 DS 1	DNA extracted from archived sputum smear slides	Direct smear
84 NaOH 2	DNA extracted from archived sputum smear slides	NaOH decontaminated and concentrated sputum
291 DS 1	DNA extracted from archived sputum smear slides	Direct smear
297 DS 1	DNA extracted from archived sputum smear slides	Direct smear
51 DS 1	DNA extracted from archived sputum smear slides	Direct smear
52 DS 1	DNA extracted from archived sputum smear slides	Direct smear
53 DS 1	DNA extracted from archived sputum smear slides	Direct smear
54 DS 1	DNA extracted from archived sputum smear slides	Direct smear
JCU 565	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 567	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 578	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 582	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 594	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 598	DNA extracted from sputum	NaOH decontaminated and concentrated sputum

JCU 602	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 609	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 611	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 624	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 625	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 632	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 634	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 648	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 661	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 667	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
JCU 670	DNA extracted from sputum	NaOH decontaminated and concentrated sputum
57-2	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum
78-1	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum
78-2	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum
96	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum
105	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum
530	MGIT™ culture supernatant	NaOH decontaminated and concentrated sputum

3.2   The Mtb reference strain

DNA from the Mtb reference strain, H37Rv strain (NC_000962.3; American Type Culture Collection (ATCC)) was extracted by others following MGIT™ tube culture in the PC3 laboratory and made available for this project. The DNA was diluted in nuclease-free water and stored at -80°C prior to use.

3.3   DNA from clinically and taxonomically associated bacteria

A set of 40 different genomic DNA samples isolated from 40 different bacterial organisms was also made and used to test the specificity of the developed assays (Table 3.2 and Appendix 15). This set comprised clinically and taxonomically associated bacteria along with some members of *Mycobacterium tuberculosis* complex (MTBC). Except for the MTBC members, all organisms were grown, tested biochemically, and processed in the PC2 laboratory. All the samples were stored at -80°C freezer. Details of those are mentioned in Appendix 15.

Table 3.2: List of organisms and their respective source of acquisition.

Organisms	Source
Clinically associated organisms	
Acinetobacter baumannii	JCU Microbiology Culture Collection
Aeromonas hydrophila	JCU Microbiology Culture Collection
Bacillus cereus	JCU Microbiology Culture Collection
Bacillus mycoides	JCU Microbiology Culture Collection
Brachybacterium sp.	JCU Microbiology Culture Collection
Enterobacter cloacae	JCU Microbiology Culture Collection
Enterococcus faecalis (ATCC)	JCU Microbiology Culture Collection
Erysipelothrix rhusiopathiae	JCU Microbiology Culture Collection
Escherichia coli (ATCC)	JCU Microbiology Culture Collection
Alcaligenes sp.	JCU Microbiology Culture Collection
Klebsiella pneumoniae	JCU Microbiology Culture Collection
Listeria monocytogenes	JCU Microbiology Culture Collection
Micrococcus luteus	JCU Microbiology Culture Collection
Moraxella catarrhalis	JCU Microbiology Culture Collection
Neisseria gonorrhoeae	JCU Microbiology Culture Collection
Neisseria meningitidis	JCU Microbiology Culture Collection
Nocardia sp.	JCU AITHM PC3 lab
Paenibacillus polymyxa	JCU Microbiology Culture Collection
Proteus mirabilis	JCU Microbiology Culture Collection
Pseudomonas aeruginosa (ATCC)	JCU Microbiology Culture Collection
Salmonella typhimurium (ATCC)	Dr. Jenny Elliman, JCU
Staphylococcus aureus (ATCC)	JCU Microbiology Culture Collection
Staphylococcus epidermidis (ATCC)	JCU Microbiology Culture Collection
Staphylococcus saprophyticus	JCU Microbiology Culture Collection
Streptococcus agalactiae	JCU Microbiology Culture Collection
Streptococcus dysgalactiae	JCU Microbiology Culture Collection

Streptococcus pneumoniae (ATCC)	JCU Microbiology Culture Collection
Streptococcus pyogenes	JCU Microbiology Culture Collection
Streptococcus viridians	JCU Microbiology Culture Collection
Ureaplasma plasmid	DNA 2.0
Corynebacterium diphtheriae	JCU Microbiology Culture Collection
Bacillus cereus	JCU Microbiology Culture Collection
Taxonomically associated organisms	
M abscessus	JCU AITHM PC3 lab
M. avium	Dr. Jackie Picard, JCU
M. bovis	JCU AITHM PC3 lab
M. intracellulare	Dr. Jackie Picard, JCU
M. fortuitum	Dr. Jackie Picard, JCU
M. peregrinum	Dr. Jackie Picard, JCU
DNA 12 (M. intracellulare)	Dr. Jackie Picard, JCU
M. tuberculosis H37Rv	JCU AITHM PC3 lab

3.4   Plasmid construction for use as qPCR controls

Two different plasmids were constructed. One had an 1045bp long Mtb IS6110 sequence and the other contained a 1013bp fragment of the Mtb *rpoB* gene. They were to be used as positive controls in Chapter 4.

3.4.1   Extraction of IS6110 and *rpoB* gene for plasmid construction

The H37Rv genomic DNA was used as template for qPCR. The amplification of the IS6110 and the *rpoB* sequence was carried out separately using dedicated sequencing primer sets both designed at JCU. The intent was to amplify the target sequences to construct control plasmids for TaqMan assays. Amplification of IS6110 and *rpoB* gene fragments was done using the Bio molecular systems (BMS) MIC™ qPCR cycler. Primer sets (Mtb-IS6110-1045-F and Mtb-IS6110-1045-R) for the IS6110 sequence and (*rpoB*-PL4-1013–F and *rpoB*-PL4-1013–R) for the *rpoB* gene were used. The conventional PCR was performed in BMS MIC™

qPCR 48 well PCR plates, with 20µL reaction volume. The details of these reaction mixes are detailed in Appendix 6.

The amplified target sequence (IS6110) was approximately 1045bp in length. The IS6110 TaqMan primer set, and the probe was used to verify the amplified IS6110 sequence. Similarly, the *rpoB* amplicon has 1013bp length. It was confirmed by using the designed *rpoB* TaqMan primer and probe. The PCR products were further purified using Wizard SV Gel and the PCR Clean-Up system (Promega, Madison, WI) individually. No gel electrophoresis was used for the confirmation of the target gene product.

3.4.2 Insert preparation using Promega Wizard SV Gel and PCR Clean-Up system

3.4.2.1 *Reagent preparation*

All reagents were prepared according to the manufacturer's instructions (Appendix 2).

3.4.2.2 *Processing of PCR amplified products*

The target of choice was amplified using standard amplification conditions. To 20 µL of cleaned PCR product, an equal volume of Membrane Binding Solution was added.

3.4.2.3 *DNA Purification by Centrifugation*

One SV Minicolumn was placed in a Collection Tube for each PCR product. Prepared PCR product was transferred to the SV Minicolumn assembly and incubated for 1 minute at room temperature. The SV Minicolumn assembly was centrifuged in a microcentrifuge at 16,000 × g (14,000rpm) for 1 minute. The SV Minicolumn was removed from the Spin Column assembly discarding the liquid in the Collection Tube. The SV Minicolumn was again placed into the Collection Tube.

The column was washed by adding 700µL of Membrane Wash Solution, previously diluted with 95% ethanol to the SV Minicolumn. The SV Minicolumn assembly was centrifuged for 1 minute at 16,000 × g (14,000rpm). The liquid in the collection tube was discarded and the SV column was inserted into the collection tube. The 500µLof Membrane Wash Solution was

added to the column for washing. Centrifugation of the SV Minicolumn assembly was done for 5 minutes at 16,000 × *g*.

The SV Minicolumn assembly was removed from the centrifuge with caution not to wet the bottom of the column with the flowthrough. Discarded the contents in the Collection Tube, and recentrifuge the column assembly for 1 minute to evaporate any remaining ethanol. The SV Minicolumn was transferred to a labelled 1.5ml microcentrifuge tube. 50µLof Nuclease-Free Water was added to the centre of the column without touching the membrane with the pipette tip and incubated at room temperature for 1 minute. The column and tube assembly were centrifuged for 1 minute at 16,000 × *g* (14,000rpm). The SV Minicolumn was discarded. The microcentrifuge tube containing the eluted DNA was stored at –80°C.

3.4.2.4 *Ligation procedure*

The pGEM®-T or pGEM®-T Easy Vector and Control Insert DNA tubes were briefly centrifuged to collect the contents at the bottom of the tubes. Established up ligation reactions as described below. The reactions were incubated overnight at 4°C.

Table 3.3 Ligation reaction

Reaction component	Standard reaction	Positive control	Background control
2X Rapid Ligation Buffer, T4 DNA Ligase	5µL	5µL	5µL
pGEM®-T or pGEM®-T Easy Vector (50ng)	1µL	1µL	1µL
PCR product	XµL*	–	–
Control Insert DNA	–	2µL	–
T4 DNA Ligase (3 Weiss units/µL)	1µL	1µL	1µL
Nuclease-free water to a final volume of	10µL	10µL	10µL

3.5   Insert: Vector molar ratio optimisation

A 3:1 molar ratio of the control DNA insert to the vector was used. The pGEM®-T and pGEM®-T Easy Vectors were approximately 3kb and were supplied at 50ng/µL. To calculate the appropriate amount of PCR product (insert) to include in the ligation reaction the following equation was used:

ng of vector × kb size of insert × insert: vector molar ratio= ng of insert
Kb size of the vector

For the IS6110 sequence
A 1045bp long sequence was targeted and extracted (Appendix 4, Table 4.1).

$$\frac{50 \times 1.0}{3} \times \frac{3}{1} = \ 50ng$$

For the *rpoB* insert
The sequencing primers extracted 1013bp long gene slice (Appendix 4, Table 4.2).

$$\frac{50 \times 1.0}{3} \times \frac{3}{1} = \ 50ng$$

Therefore, 50ng of the IS6110 and *rpoB* PCR product/DNA insert was used in the ligation reaction.

3.5.1   Transformations Using the pGEM®-T and pGEM®-T Easy Vector Ligation Reactions

High-efficiency JM109 competent cells (≥1 × 10^8 cfu/µg DNA) were used for transformation.

3.5.1.1 *Transformation Protocol*

Two LB/ampicillin/IPTG/X-Gal plates for each ligation reaction were made and kept at room temperature. The tubes containing the ligation reactions were centrifuged to collect the contents at the bottom. 2µL of each ligation reaction was then added to a sterile 1.5ml microcentrifuge tube on ice. The frozen JM109 High-Efficiency Competent Cells tube was removed from the freezer and was thawed by placing on ice.

After around five minutes, the cells were mixed by gently flicking the tube. Excessive pipetting was avoided to provide minimal damage to the fragile cells. 50µL of cells were carefully transferred into each tube prepared in Step 2. Tubes were mixed by gentle flicking and placed on ice for 20 minutes. Cells were heat-shocked for 45–50 seconds in a water bath set exactly at 42°C (without shaking) and then immediately returned to ice for 2 minutes.

950µL of LB broth was added to the tubes containing cells transformed with ligation reactions and 900µLto the tube containing cells transformed with uncut plasmid. Tubes were incubated overnight at 37°C with gentle shaking (~150 rpm). 100µL of each transformation mix was plated onto duplicate LB/ampicillin/IPTG/X-Gal plates. After incubation, the tubes were centrifuged at 1,000 × g for 10 minutes to form the cells pellets at the bottom of the tube. Resuspended each pellet in 200µL LB both, and 100µLplated on each of two plates. Plates were incubated overnight (16–24 hours) at 37°C. White colonies were selected for plasmid preparation and expected to contain plasmid insert.

3.5.2 Plasmid extraction protocol

Plasmid DNA was extracted using Promega PureYield™ Plasmid miniprep system (A#1222). No amendments were done in the manufacturer's instructions during the extraction.

3.6 DNA concentration determination using the Nanodrop® Spectrophotometer

The concentration of the extracted plasmid suspensions was determined by using Nanodrop® Spectrophotometer P 360 and purity assessed using the A_{260}/A_{280} ratio. Nuclease-free water was used as blank sample. One drop of each plasmid suspension was placed on the cuvette and the sample concentration, A_{260}/A_{280} and A_{260}/A_{230} ratios were recorded for each sample. The A_{260}/A_{280} ratio indicated the sample purity. The ratio of A_{260}/A_{230} indicated about presence/absence of impurities (such as proteins, reagents etc) in the DNA sample(s).

CHAPTER 4

Development and evaluation of novel TaqMan assays to identify *Mycobacterium tuberculosis*

4.1   Introduction

The annual incidence rate of TB in the Western Province of PNG is 674/100,000 population. The rate (from 2014 till 2016) is 727/100,000 people in the Balimo region (Diefenbach-Elstob *et al.*, 2019). TB diagnosis in the Balimo region of PNG relies on clinical signs and symptoms and sputum smear microscopy. Several other conditions such as melioidosis may mimic the typical signs and symptoms of TB in this region (Warner and Currie, 2018). Furthermore, the low sensitivity of microscopy (Ziehl-Neelsen staining) may enhance the chances of misdiagnosis. This means that individuals may be unnecessarily commence on TB therapy which may contribute to the acquisition of drug resistance among Mtb strains in this region (Guernier *et al.*, 2017).

The WHO recommends phenotypic drug susceptibility testing (pDST) as a gold standard to detect the Mtb from different samples. This method requires time (8-12 weeks), investment, infrastructure (Biosafety level laboratories) and trained technical staff (Guernier *et al.*, 2017). This represents a potential challenge in resource-constrained areas to carry out accurate diagnosis and effective patient management. Therefore, a cost-effective, convenient, rapid, and reliable identification tool that is fit-for-purpose is required in this type of setting. A wide range of novel diagnostics are currently being tested or are in the development stage and each have their own advantages and disadvantages (Walzl *et al.*, 2018).

Due to its relative simplicity, high specificity, sensitivity and faster turnover time, real-time PCR is now widely used in the clinical microbiology laboratory for diagnosis of diverse infectious diseases (Barken *et al.*, 2007). Numerous real time-PCR diagnostic platforms are now available for MTBC detection including Xpert (Cepheid), Fluorotype MDRTB (Hain Lifescience), RealTime MDR TB (Abbott Molecular), BDProbeTec (Becton Dickinson), MeltPro

Drug-Resistant TB RIF and drug-Resistant INH kits (Xiamen Zeesan Biotech). These have already been approved by the major regulatory bodies like the WHO and the Food and Drug Administration. High-throughput identification of mutation(s) and high levels of automation make these systems advantageous in many settings. However, their limited target range and high cost of implementation and maintenance especially in resource-limited settings presents a major challenge (Peng *et al.*, 2016).

Every year in Balimo, PNG, new drug-resistant cases emerge (3.4%) with 26% of cases associated with previously treated cases (Diefenbach-Elstob *et al.*, 2019). Several studies conducted by different research groups have concluded the widespread prevalence of DR-TB in different areas of PNG. The level of drug resistance was different in each study (Gilpin *et al.*, 2008; Simpson *et al.*, 2011; Ballif *et al.*, 2012; Cross *et al.*, 2014b; Ley *et al.*, 2014a; Bainomugisa *et al.*, 2018; Guernier-Cambert *et al.*, 2019). High DR-TB prevalence and minimum availability of sustainable routine diagnostic tools may pose a serious threat in the transmission of the tubercle bacilli to healthy communities resulting in severe outbreaks. TB treatment based on the clinical signs and symptoms and microscopy alone may lead to misuse of anti-TB drugs. This in turn may give rise to drug resistant Mtb strains.

The overall objective of this project is to develop molecular based in-house identification tools to detect, quantify and characterise Mtb in samples from Balimo, PNG. As described in Chapter 2, there are numerous target genes that have been exploited by both commercial systems and in-house molecular assays. The IS6110 (or IS986) insertion sequence has been used as a marker for epidemiological studies and many new diagnostic assays are based on this gene (Walzl *et al.*, 2018). IS6110 is present as multiple copies (up to 25) per genome and therefore markedly increases the sensitivity of any assay targeting this gene (Abraham *et al.*, 2013; Alonso *et al.*, 2013; Huyen *et al.*, 2013).

However, some South Asian and South East Asian TB causing strains possess very low copies or even no IS6110 sequence at all (Meghdadi *et al.*, 2015). There is a strong possibility that Mtb strains lacking IS6110 may be present in PNG and the application of IS6110 TaqMan on the genomic extract of those strains may lead to misidentification, hence other gene targets are needed to minimise false negative diagnoses.

Misidentification of Mtb could be minimised by adding additional assay(s) to any proposed testing panel. Broccolo *et al.* (2003) designed a combined assay that targets a highly conserved region inside the IS6110 sequence and the *senX3-regX3* intergenic region(IR) (Broccolo *et al.*, 2003). These authors report that this assay is highly sensitive (because of the multicopy IS6110 target) and specific, as their *senX3-regX3 IR* target was reported to be only present in MTBC. Incorporating a *senX3-regX3 IR* targeted assay may also overcome the issues with identification of Mtb strains lacking IS6110. The *senX3-regX3* IR sequence is only present as a single-copy and its use alone may compromise assay sensitivity. Therefore, as per the claim of the published study (mentioned above), an IS6110 assay in combination with a *senX3-regX3 IR* assay could be used instead of using other genes such as IS1081.

Alternative Mtb diagnostic targets include the *rpoB* gene which is already used in commercial systems including Xpert. The *rpoB* gene is only present as a single copy gene per genome (Yu *et al.*, 2013), so assays utilising this target alone may have issues with diagnostic sensitivity. In fact, the recently released Xpert Ultra assay incorporates molecular beacon probes for the multicopy IS6110 and the IS1081 genes to increase assay sensitivity. A *rpoB*-based assay might also have utility to assess DNA quality or suitability for further sequence-based analysis.

This triage process might ensure that subsequent analysis is not wasted on poor quality templates. The C_t value of the amplified sequence may equate to the overall quality of the DNA being tested. This could be a useful information in further decision-making. High quality Mtb DNA(s) with low C_t value would be eligible for further testing that may include the melt and the sequencing analysis. The objective of the studies described in this chapter was the design and evaluate novel TaqMan assays that could identify Mtb from different types of clinical material derived from PNG sputum samples.

The specific aims of the studies described in this chapter were:
1. To use a stepwise process incorporating TaqMan primer and probe design and validation to develop and utilise qPCR assays that may have utility in TB diagnosis.

2. To use these tools to determine whether clinical material from individuals in Balimo, PNG with presumptive TB contained Mtb DNA.

4.2   Materials and Methods

4.2.1   Sample collection

Thirty-eight samples of DNA extracted from the sputa, supernatants of MGIT™ cultures and sputum smear slides of suspected TB patients were made available for the current study by A/Prof. Jeffrey Warner. These 38 different DNA samples were already processed, determined to be sterile and were ready to use in the PC2 laboratory. All DNA samples were stored in a -80°C freezer to maintain their integrity. These samples are described in Chapter 3, section 3.1.

4.2.1.1 *Plasmid construction*

Two different plasmid suspensions were prepared and used as positive controls. The plasmid suspension pGEM-IS6110-1045 contained a 1045bp Mtb IS6110 target sequence and the other, Mtb-*rpoB* -1013, contained a 1013bp segment of the Mtb *rpoB* gene. Both were verified by Sanger sequencing to ensure the presence of correct target sequences. The Mtb reference strain (H37Rv) genomic DNA was used as template to amplify these target sequences separately using two different sets of specific primers. Details of this process are described in Chapter 3, section 3.4-3.7.

4.2.1.2 *Non-TB organisms*

A set of 40 different genomic DNA samples isolated from 40 different bacterial organisms was also used to test the specificity of the developed assays. The details are mentioned in Chapter 3, section 3.3, and Appendix 15.

4.2.2 Primer design

Broccolo *et al*. (2003) described two assays based on the IS6110 and *sensX3-regX3* IR gene. The former gene-based assay was sensitive (due to multiple copies of IS6110 per genome) but lacked specificity and could detect other Mycobacteria. According to this study, the *sensX3-regX3* assay was MTBC specific, but with, with compromised sensitivity. Both assays were used in combination in their study to identify the Mtb from the samples (Broccolo *et al.*, 2003).

This published IS6110 based assay has been used by our team previously was in use by the TB research team at JCU (Broccolo *et al.*, 2003) . In the current study, AlleleID® v 7.7 (Premier Biosoft International, Palo Alto, CA) was used to design new primers and a probe for comparison with the published Broccolo et al. IS6110 assay. Rating scores (0-100) were also determined for new and existing assays. These ratings predicted the possible performance of an assay based on the factors such as sequence length, melting temperature, GC content, self-dimer, cross dimer, GC clamp etc.

The existing Broccolo IS6110 primer set and probe achieved scores of 0.6 and 65.4 respectively (Broccolo *et al.*, 2003). Therefore, an alternative IS6110 assay was developed and evaluated using the same program. The detail of that primer set, and probe is in the following section with ratings of 62.6 and 73.6 for the forward and reverse primer respectively and 87.9 for the probe indicating that this assay may be superior to the published assay (Appendix 3).

4.2.2.1 *The IS6110 sequence primers*

Two pair of primers and a probe were designed using the AlleleID® program. One pair was used to amplify part of the IS6110 gene to make a plasmid control (Appendix 4 and 5). The other pair in combination with the probe was used to identify Mtb from the DNA samples.

Twenty-one Mtb IS6110 sequences were downloaded from the National Centre for Biotechnology Information (NCBI) nucleotide database, GenBank®. These highly identical sequences were then aligned using Geneious software. Muscle option was opted to carry out the multiply alignment of these sequences. GeneDoc program was further used to view and test the designed primer sequences on the aligned published sequences. All the guidelines to make a good primer set were taken into consideration while designing each primer set and probe.

The primer set (IS6110-108-F and IS6110-108-R) and the probe (IS6110-108-P) was evaluated as a diagnostic assay (Table 4.1) and targeted a 108bp segment of the IS6110 gene. This gene sequence was within the conserved region of the IS6110 gene and was identical across all Mtb IS6110 genes used for primer design.

A dual-labelled probe (IS6110-108-P) consisting of a single-stranded oligonucleotide labelled with two different dyes: a reporter (5'end) and a quencher (3' end) was designed. The TaqMan probe IS6110-108-P had a 6-Carboxyflourescein (6-FAM™) fluorophore as a reporter that was detectable by the green channel of any real-time PCR device. It contained a carboxylic acid that reacts with the primary amines (https://www.thermofisher.com/order/catalog/product/C1360). Black Hole Quencher 1 (BHQ®-1), a non-fluorophore with an absorption maximum of 534nm was used as a quencher. This quencher emits heat instead of the light that leads to less background fluorescence. In addition, it provides higher sensitivity due to the high signal-to-noise ratio (https://www.sigmaaldrich.com/technical-documents/articles/biology/choosing-the-right-probe.html).

Table 4.1 Primer sequences targeting IS6110 gene for Mtb identification

Details of the IS6110 diagnostic primer set with the probe:		
Primer set	Sequence detail	Length(bp)
IS6110-108-F	5'-GAC CTA CTA CGA CCA CAT C-3'	108
IS6110-108-R	5' -CCG TAA ACA CCG TAG TTG-3'	108
IS6110-108-P	5'-FAM TGT GCT CCT TGA GTT CGC CAT BHQ1-3'	108

4.2.2.2 *The rpoB gene primers*

As in section 4.2.2.1, AlleleID® program was used to design two sets of primers and a probe. One pair was to amplify a section of the *rpoB* gene to make a plasmid control (Appendix 4 and 5). The other pair along with the probe was to assess whether template was amplifiable i.e., of good quality (Table 4.2).

The diagnostic primer (*rpoB*-TM-108-F and *rpoB*-TM-108-R) and the probe (*rpoB*-TM-108-P) were designed as described for the IS6110 assay. The target site (108bp) of this assay was a conserved region with no polymorphisms across many different *rpoB* genes from many different bacteria. The probe-annealing site was inside the target region to enhance the specificity of the assay. The utility of this assay was to measure the quality of the DNA templates based on the C_t value of the amplicon. The DNA samples positive by the IS6110 TaqMan assay were then subsequently tested with this *rpoB* assay.

Table 4.2 Primer sequences targeting *rpoB* gene segment to assess the DNA quality of templates being tested

Details of the *rpoB* diagnostic primer set with the probe:		
Primer name	Sequence detail	Length(bp)
rpoB-TM-108-F	5'-CGC TAT AAG GTC AAC AAG AA-3'	108
rpoB-TM-108-R	5'-CAA GCG GAC CAG ATA TTC-3'	108
rpoB-TM-108-P	5'- CAL Flour CGT CTT CTT CGG TCA GCG TCG BHQ1 - 3'	108

As in section 4.2.2.1, the TaqMan probe *rpoB*-TM-108-P was designed as a dual labelled probe with 5'-CAL Flour ® Orange 560 and 3'-BHQ-1 as a quencher. The reporter dye is a novel xanthene fluorophore with relatively high stability and is cost-effective (https://www.biosearchtech.com/support/education/fluorophores-and-quenchers/cal-fluor-dyes). This reporter dye was detectable by the yellow channel of any real-time PCR machine. The two different reporter dyes used for the IS6110 and *rpoB* assays allowed them to be used in a duplex assay format.

4.2.2.3 *The SenX3-regX3 IR primer and probe set*

Broccolo *et al.* (2009) reported that a 146bp *senX3-regX3* intergenic region assay was highly specific to MTBC. The details of the assay are described in Table 4.3. This published assay was used to minimise any misidentification of Mtb due to the possibility of a Mtb strain(s) having a very low copy number or no copies of IS6110.

Table 4.3 Primer set, and probe details of the published assay based on the *senX3-regX3* IR gene.

Details of the *senX3-regX3* IR primer and the probe set		
Primer set	Sequence detail	Length (bp)
TAQregT2	5'-GTA GCG ATG AGG AGG AGT GG-3'	146
TAQreg2L	5'-ACT CGG CGA GAG CTG CC-3'	146
Probe	5'-FAM ACG AGG AGT CGC TGG CCG ATC C BHQ1-3'	146

The published assay's *senX3-regX3* IR probe had a reporter dye 6-carboxyfluorescein and the quencher dye 6-carboxytetramethylrhodamine at 5' and 3' end of the oligonucleotide respectively (Broccolo *et al.*, 2003).

All primer sets and probes were synthesised by Macrogen (Seoul, South Korea). Each freeze-dried tube of primer was briefly centrifuged and reconstituted in nuclease-free water to produce a stock solution of 100µM concentration. The working stock was prepared by

further diluting the stock solution to the concentration of 10µM prior to using for PCR. All work was done in a Biosafety cabinet to minimise any chances of contamination. All the solutions were stored in the freezer at- 80°C. The selection and incorporation of different fluorophores (6-FAM™ in the selected IS6110 TaqMan probe and CAL Fluor® Orange 560 in the designed *rpoB* TaqMan probe) were done to use them together as a duplex assay.

4.2.3 Extraction of the IS6110 and *rpoB* gene from the H37Rv genome for plasmid construction

All assays were established on the Bio molecular systems (BMS) MIC™ qPCR cycler which, because of its portability, may have utility in resource-poor settings. https://biomolecularsystems.com/mic-qpcr/. The amplification of the target sequences was separately carried out in the Bio molecular systems (BMS) MIC qPCR machine. H37Rv genomic DNA was used as template for the amplification of the IS6110 and the *rpoB* sequences for plasmid cloning (Appendix 4). The primer set Mtb-IS6110-1045-F and Mtb-IS6110-1045-R for the IS6110 sequence and *rpoB*-PL4-1013–F and *rpoB*-PL4-1013–R for the *rpoB* gene was used to carry out the amplification. Each reaction was carried in a dedicated 48 well PCR plate, each well containing a 20µL reaction mixture with the reagents has been detailed in Appendix 6. The dynamic mode setting was chosen as suggested in the MIC™ PCR cycler manual for all TaqMan assays.

4.2.4 Construction of JCU IS6110 and the *rpoB* control plasmids

The amplified target sequence (IS6110) was approximately 1045bp in length. The IS6110 TaqMan primer set, and a probe were used to verify the amplified IS6110 sequence. The final PCR product was then purified using Wizard SV Gel and the PCR Clean-Up system (Promega, Madison, WI) (Chapter 3, section 3.4.2). No gel electrophoresis was used for the confirmation of the target gene product. To construct a plasmid containing the target sequence, the pGEM® –T vector System (Promega, Madison, WI) kit was used, and the protocol, according to the manufacturer's instruction, was followed.

Competent JM109 *E. coli* cells were then transformed with a prepared construct, of which an aliquot was then plated over the LB plates containing ampicillin, IPTG and X-gal, for

effective blue/white screening (Appendix 7, Table 7.1). The clones containing the plasmid with the IS6110 sequence produced white colonies. Thus, the five different white colonies were selected, and each was inoculated into 10mL of ampicillin LB broth to culture the transformed bacteria at 37°C overnight in a shaking incubator. The overall process has been described in Chapter 3, section 3.6-3.7.

Plasmid DNA was isolated using the Wizard® Plus SV Minipreps DNA purification System (Promega, Madison, WI), following the manufacturer's instructions. The concentration of the extracted plasmid was determined by using Nanodrop® Spectrophotometer and the purity of the plasmid was observed by using A_{260}/A_{280} ratio (Chapter 3, section 3.7). For the *rpoB* plasmid, the same technique was used to amplify, verify, and incorporate the target sequence into a plasmid. The same process of blue-white screening was also carried out to eventually extract the *rpoB* plasmid. Both IS6110 and the *rpoB* plasmid suspension vials were stored separately at -80°C after determining their purity.

4.2.5   Verification of cloned inserts and sequencing

Prepared plasmid suspensions (IS6110 and *rpoB*) were tested separately with the IS6110 *and rpoB* TaqMan assays respectively to verify the presence of the right insert in both plasmids prior to sequencing. GoTaq® Probe qPCR master Mix (Promega A#6101) was used for verification. Aliquots of each of the five-plasmid suspensions were sent to Macrogen (Seoul, Korea) for sequencing to confirm the presence of the target sequence in each of the plasmid suspensions. The primer pair Mtb-IS6110-1045-F and Mtb-IS6110-1045-R for the IS6110 plasmid and *rpoB*-PL4-1013–F and *rpoB*-PL4-1013-R for the *rpoB* plasmid was used for sequencing.

The sequence reports were supplied as ab1 files, which were further processed in the Geneious software to carry out the *de novo* assembly. The resultant contigs were edited and were further used as a query sequence to carry out the nucleotide BLAST. Published sequences with the highest homology (100% or 99%) were downloaded and further aligned with their respective contigs using the Geneious program. Both the aligned files (IS6110 and

rpoB) were saved and processed in the GeneDoc (a viewing utility) software to edit if required. The insert length and sequenced data were confirmed using this workflow.

4.2.6   Standard curve preparation

H37Rv genomic gDNA was used to establish a standard curve for the quantification of the copy numbers of both the IS6110 and *rpoB* DNA in each sample. For this, ten tenfold dilutions of H37Rv gDNA were prepared using nuclease-free water. The overall volume of each dilution prepared was 100μL (90μL nuclease-free water and 10μL of the template). Reactions were performed in triplicate.

4.2.7   Calculation of sequence copy numbers

URI Genomics and Sequencing Centre website was used to calculate the copy number. A Nanodrop® spectrophotometer was used to determine the concentration of stock H37Rv gDNA for preparation of dilutions: 35ng/ μL was used for copy number calculations. The entire length of the genome including the target sequence was 4411532bp and the resultant number of copies of the genomic DNA was calculated. It is important to note here that the sequence copy number for the IS61110 gene was not determined. As it is present as multiple copy per genome.

4.3   Validation of novel JCU IS6110 and *rpoB* Taqman assays

All TaqMan assays were validated in accordance with Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE guidelines) (Bustin *et al.*, 2009). The assays were evaluated for sensitivity, specificity, robustness/ repeatability, and efficiency. Each assay was tested for its sensitivity, specificity, repeatability, and efficiency individually and then compared by drafting a table.

4.3.1 Sensitivity

Ten tenfold dilutions of H37Rv gDNA were prepared with triplicates of each of the dilutions. The dilution at which 50% of replicates were detected was the detection limit of each TaqMan assay. This method provided the analytical sensitivity of the assays. In addition, ten tenfold dilutions of two DNA samples extracted from MGIT™ culture supernatants (clinical samples 56-1 and 96) were used to check the relative sensitivity of both assays.

4.3.2 Specificity

Assay specificity was determined in two ways: *in-silico* analysis and analytical specificity. The *in-silico* specificity of the primer set used in each assay was evaluated *in silico* using a nucleotide BLAST® (BLASTn suite) in which the primer sequences (one at a time) were used as query sequences to search for homologous published GenBank sequences. The *rpoB* and IS6110 gene sequences identified were downloaded, aligned using Geneious program choosing Muscle option and viewed in GeneDoc software. The approximate location of primer binding sites was then determined against the aligned genomic database for both assays individually.

The analytical specificity of the assays based on the IS6110 and the *rpoB* gene was determined by using a panel of 40 different bacterial genomic DNA samples consisting of clinically and taxonomically associated bacteria containing non-tuberculous and tuberculous mycobacteria. The details of the organisms have been described in Microbial testing data excel sheet (Appendix 18). Each qPCR run was accompanied by a series of positive and negative controls. Three different dilutions of the JCU IS6110 plasmid suspension (in duplicate) for IS6110 assay and three of *rpoB* plasmid suspension for the *rpoB* assay were used as positive controls. Nuclease free water was used in the negative control wells to check for any contamination. The appearance of sigmoid curves indicated successful amplification of the target gene(s).

Similarly, the specificity of the published assay *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was tested to verify the reported MTBC specificity.

4.3.3   Robustness/ repeatability or reproducibility

Intra assay repetitive testing was performed to check the repeatability or robustness of both (JCU IS6110 and *rpoB)* TaqMan assays. Each time the assay was run, and the C_t value of the controls was recorded and plotted with standard deviation (SD) according to MIQE guidelines to indicate variation and trends. To limit variability across assays runs, reagents were aliquoted into single use aliquots, stored at -80°C.

4.3.4   Efficiency

The replicates of at least three control dilutions and a non-template control (NTC) were used for every experiment. The MIC™ software calculated the efficiency as a percentage and the R^2 value showing compliance with the MIQE guidelines.

4.4   Testing of the developed JCU IS6110-*rpoB* TaqMan duplex on the H37Rv DNA

The H37Rv genomic DNA was subjected to real-time PCR for the amplification of the *rpoB* and IS6110 sequence. Bio molecular systems (BMS) MIC™ was used to amplify the target sequences. The primer and probe set specific for IS6110 and *rpoB* (Table 4.4) were used to carry out the amplification. Each reaction was carried in a dedicated 48 well PCR plate, each well containing a 20µL reaction (18 µL mastermix from readymade stock and 2µL DNA template) mixture with the following components (Table 4.4):

Table 4.4 Reaction mix details of the JCU IS6110-*rpoB* TaqMan duplex assay using the BMS MIC™ qPCR cycler.

Stock Concentration	Reagent	Name	Final Concentration	Reaction (μL)
2X	Master mix	GoTaq® Probe qPCR (Promega A#6101)	1X	10
IS6110				
10 μM	Forward Primer	IS6110-108-F	0.8 μM	1.6
10 μM	Reverse Primer	IS6110-108-R	0.8 μM	1.6
10 μM	Probe	IS6110-108-P		0.6
rpoB				
10 μM	Forward Primer	*rpoB*-TM-108-F	0.8 μM	1.6
10 μM	Reverse Primer	*rpoB*-TM-108-R	0.8 μM	1.6
10 μM	Probe	*rpoB*-TM-108-P		0.6
DNA template				2
Nuclease free water				0.4
Total volume of one reaction				20

Cycling parameters were 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds and 40 cycles of PCR using the MIC™ qPCR cycler. Green (for JCU IS6110 TaqMan) and yellow (for *rpoB* TaqMan) channels were used for simultaneous detection of both amplicons.

4.5   Use of JCU IS6110-*rpoB* TaqMan duplex assay using the BMS MIC™ qPCR cycler on DNA samples from different sources (sputum, slides and MGIT™ tube supernatant)

After the successful testing of the duplex assay on the H37Rv reference strain DNA, it was further tested on a panel of 38 DNA samples and dilutions of plasmid controls, all in duplicates (Table 4.4). As previously the green and yellow channels were used for JCU

IS6110 and *rpoB* assays respectively. Duplicate negative controls were used to verify the absence of contamination. The PCR cycling parameters were used as described in section 4.4.

4.6   Use of *senX3- regX3* assay on DNA samples from different sources (sputum, slides and MGIT™ tube supernatant)

As described previously, IS6110 is absent or has a low copy number in some South Asian Mtb strains. Hence, an identification tool(s) was required to minimise the likelihood that Mtb strains lacking IS6110 would be missed. The *senX3-regX3* published assay by Broccolo *et al.* (Broccolo *et al.*, 2003) was used for this purpose. The qPCR reaction mix is shown in Table 4.5. All PCR cycling parameters for this assay were as described in section 4.4.

Table 4.5 Reaction mix details of the published *senX3-regX3* IR gene assay

Stock Concentration	Reagent	Name	Final Concentration	Reaction (μL)
2X	Master mix	GoTaq® Probe qPCR (Promega A#6101)	1X	10
10 μM	Forward Primer	TAQregT2	0.8 μM	1.6
10 μM	Reverse Primer	TAQreg2L	0.8 μM	1.6
10 μM	Probe			0.6
Template				2
Nuclease free water				4.6
Total volume of one reaction				20

4.7   Interpretation of combined JCU IS6110, *rpoB* and *senX3-regX3* IR assays

Differential reactions between the three PCR assays will aid in the determination of positive Mtb, NTB or poor-quality template. Table 4.6 depicts the theoretical expected results when using the three assays (JCU IS6110, JCU *rpoB* and *senX3-regX3* IR) on the DNA samples.

Table 4.6 Theoretical interpretation of combined JCU IS6110, *rpoB* and *senX3-regX3* IR assays

JCU IS6110 TaqMan	*senX3-regX3* IR	*rpoB* TaqMan	Mtb presence
+	+	+	+
-	+	+	*
+	-	-	?
-	-	-	-
-	-	+	-
+	-	+	#

? = Probably *Mycobacterium* spp., insufficient DNA to detect single copy gene

* =Probably TB, DNA could be from a strain devoid of IS6110

= DNA template sufficient, possibly Non-Tuberculous Mycobacteria

Reaction in all three assays may indicate the Mtb presence, while reaction in none could mean absence of any DNA template. The reaction in only IS6110 assay might suggest the presence of non-*Mycobacterium* spp. However, inadequate genomic DNA copies that the single copy gene-based assays (*rpoB* and *senX3-regX3* IR) could amplify the target genes. While reaction in the single copy gene assays only might indicate that the DNA template is sufficient and of good quality, but the strain might be devoid of the IS6110 gene. The reaction in IS6110 and *rpoB* assay may mean that the template may be quality sample having non-tuberculous mycobacteria due to the presence of the IS6110 across the *Mycobacterium* spp. and other bacteria (Appendix 16). A redesign of primers targeting Mtb specific sequences may increase the specificity.

4.8   Results

The objective of this study was to develop and evaluate two novel TaqMan assays for detection of Mtb DNA and to apply these assays to clinical samples from Balimo, PNG. Furthermore, we aimed to evaluate the performance of these novel JCU IS6110 and *rpoB*

gene-based assays alongside previously published IS6110 and *senX3-regX3* IR assays (Broccolo *et al.*, 2003).

4.8.1   Construction of IS6110 and *rpoB* plasmids

The transformed competent *E. coli* cells were plated on LB plates for blue/ white colony screening. The white colonies (Figure 4.1) were expected to have the plasmid with a DNA insert. Both the white and blue colonies were observed on the plates after overnight incubation.
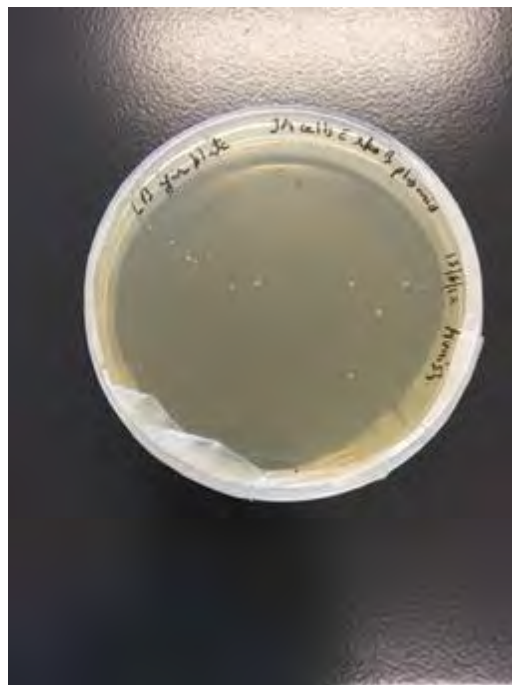
Figure 4.1: The plate showing blue/white bacterial colonies. The white colonies had a target sequence incorporated into a pGEM plasmid. The blue colonies lacked the target sequence in the plasmid, or they were devoid of the plasmid.

The presence of a high proportion of white colonies indicated that cloning was successful. These colonies were suitable for their further inoculation in LB broth containing ampicillin. The concentration and purity of plasmid DNA extracted from 5 colonies each from *rpoB* and IS6110 cloning, and transformation experiments is shown in Table 4.7 and 4.8. Based on the $A_{260/280}$ ratio, it was concluded that all the samples contained DNA of good quality. However,

carryover from miniprep was potential reason for relative less $A_{260/280}$ ratio of the *rpoB* plasmids.

Table 4.7: Concentration and purity ($A_{260/280}$ ratio) of the purified *rpoB* plasmids.

Sample number	Concentration (ng/ μL)	$A_{260/280}$ ratio
S1	29	1.611
S2	50	1.538
S3	33	1.692
S4	26.5	1.606

Table 4.8: Concentration and purity ($A_{260/280}$ ratio) of the purified IS6110 plasmids.

Sample number	Concentration (ng/ μL)	$A_{260/280}$ ratio
S1	31	1.824
S2	32.5	1.857
S3	36.5	1.825
S4	39.5	1.837
S5	54	1.714

4.8.2   Verification that plasmids contained correct inserts

All plasmid preparations of both IS6110 and *rpoB* were sent to Macrogen (South Korea) for sequencing. For each sample, two sequence results were obtained. Each pair of the sequences was then *de novo* assembled individually using the Geneious program and the resultant sequence or the contig was entered as a query sequence in NCBI BLAST to search for similar published sequences.

Many published sequences with 99% homology were identified during this search. However, when viewed using GeneDoc (Figure 4.2 and 4.3), it was found that S1 of the *rpoB* and S2 of the IS6110 plasmid suspension had mixed sequencing results. Therefore, these were discarded.

```
                 660         *        680         *        700         *        720         *        740         *        760         *
LC005454    : ATGCACCGTCGAACGGCTGATGACCAAACTCGGCCTGTCCGGGACCACCCGCGGCAAAGCCCGCAGGACCACGATCGCTGATCCGGCCACAGCCCGTCCCGCCGATCTCGTCCAGCGCCGCTTCGGCA :  779
LC005455    : ........................................................................................................................................ :  779
LC005456    : ........................................................................................................................................ :  779
LC005457    : ........................................................................................................................................ :  779
LC005458    : ........................................................................................................................................ :  779
LC005462    : ........................................................................................................................................ :  779
LC005465_( : ........................................................................................................................................ :  779
LC005472    : ........................................................................................................................................ :  779
LC005475_( : ........................................................................................................................................ :  779
LC005476_( : ........................................................................................................................................ :  779
LC005477_( : ........................................................................................................................................ :  779
LC005482    : ........................................................................................................................................ :  779
CP008744    : ..............................................................................................................................--........ :  529
CP012095    : ..............................................................................................................................--........ :  529
LC005473    : ........................................................................................................................................ :  779
LC005459    : ........................................................................................................................................ :  779
LC005471    : ........................................................................................................................................ :  779
LC005481    : ........................................................................................................................................ :  779
LC005463    : ........................................................................................................................................ :  779
AF357173    : ........................................................................................................................................ :  779
Mtb-pGEM-I : ...........................................................T............................................................--........ :  529
Mtb-pGEM-I : ...........................................................T............................................................--........ :  529
Mtb-pGEM-I : ...........................................................T............................................................--........ :  529
Mtb-pGEM-I : ...........................................................T............................................................--........ :  529
Mtb-pGEM-I : ...........................................................T............................................................--........ :  529
ISPlasmid3 : ...........................................................T............................................................--........ :  529
ISPlasmid4 : ...........................................................T............................................................--........ :  529
ISPlasmid5 : ......................................................................................................................--........ :  529
ISPlasmid1 : ...........................................................T............................................................--........ :  529
```

Figure 4.2: Comparative analysis of 1045 bp IS6110 sequence within pGEM-T vector with the other published IS6110 sequences available on NCBI. One transversion (C to T) at 711 bp position was observed in the sequence of prepared plasmid reference. The dots indicated identical nucleotide bases among the selected sequences, with 99% homology. Abbreviations: A- Adenine, T-Thiamine, C-Cytosine, G- Guanine. The aligned sequences with accession numbers starting from LC, CP and AP were the published IS6110 sequences. More information in Appendix 14, Table 14.1.

```
                220        *        240        *        260        *        280        *        300        *        320        *        340        *
rpoB_Hybri : AACTTGTTCTTCAAGGAGAAGCGCTACGACCTGGCCCGCGTCGGTCGCTATAAGGTCAACAAGAAGCTCGGGCTGCATGTCGGCGAGCCCATCACGTCGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCGAAT :  356
BX842574_r :  ........................................................................................................................................... :  356
CP000611_r :  ........................................................................................................................................... :  356
CP003248_r :  ........................................................................................................................................... :  356
AE000516_r :  ........................................................................................................................................... :  356
AM408590_r :  ........................................................................................................................................... :  356
AM412059_r :  ........................................................................................................................................... :  356
AP010918_r :  ........................................................................................................................................... :  356
BX248336_r :  ........................................................................................................................................... :  356
CP002095_r :  ........................................................................................................................................... :  356
MSGRPCB_OR :  ........................................................................................................................................... :  356
AP012340_r :  ........................................................................................................................................... :  356
CP001662_O :  ........................................................................................................................................... :  356
CP001662_r :  ........................................................................................................................................... :  356
CP002992_r :  ........................................................................................................................................... :  356
HE608151_r :  ........................................................................................................................................... :  356
JX303328   :  ........................................................................................................................................... :  356
JX303329   :  ........................................................................................................................................... :  356
JX303332   :  ........................................................................................................................................... :  356
JX303331   :  ........................................................................................................................................... :  356
JX303330   :  ........................................................................................................................................... :  356
JX303307   :  ........................................................................................................................................... :  356
JX303308   :  ........................................................................................................................................... :  356
JX303311   :  ........................................................................................................................................... :  356
JX303314   :  ........................................................................................................................................... :  356
JX303313   :  ........................................................................................................................................... :  356
CP000717_r :  ........................................................................................................................................... :  356
JX303312   :  ........................................................................................................................................... :  356
JX303321   :  ........................................................................................................................................... :  356
JX303318   :  ........................................................................................................................................... :  356
JX303326   :  ........................................................................................................................................... :  356
JX303316   :  ........................................................................................................................................... :  356
JX303317   :  ........................................................................................................................................... :  356
JX303319   :  ........................................................................................................................................... :  356
JX303325   :  ........................................................................................................................................... :  356
JX303327   :  ........................................................................................................................................... :  356
JX303324   :  ........................................................................................................................................... :  356
JX303315   :  ........................................................................................................................................... :  356
JX303320   :  ........................................................................................................................................... :  356
JX303322   :  ........................................................................................................................................... :  356
JX303309   :  ........................................................................................................................................... :  356
JX303310   :  ........................................................................................................................................... :  356
JX303323   :  ........................................................................................................................................... :  356
CP023591   :  ........................................................................................................................................... :  301
CP023639   :  ........................................................................................................................................... :  301
CP023708   :  ........................................................................................................................................... :  301
rpoBPlasmi :  ........................................................................................................................................... :  301
rpoBPlamis :  ........................................................................................................................................... :  301
rpoBPlasmi :  ........................................................................................................................................... :  301
rpoBPlasmi :  ........................................................................................................................................... :  301
```

Figure 4.3: Comparative analysis of 1013 bp *rpoB* sequence within pGEM-T vector with the other *rpoB* sequences available on NCBI. The dots indicate identical nucleotide bases among the selected sequences, with 100% homology. Abbreviations: A- Adenine, T-Thiamine, C-Cytosine, G-Guanine. The aligned sequences with accession numbers starting from BX, CP, AE, AM, AP, MX and JX, were the published *rpoB* sequences. Detailed in Appendix 14, Table 14.2.

For the above figures, it was clear that both the IS6110 and the *rpoB* plasmids contained the correct inserts, however, the IS6110 plasmids had one nucleotide base variation (Figure 4.2) when compared to published sequences. This variation was quite consistent in nearly all the extracted plasmids. There were two possible explanations: (1) the IS6110 sequence from the H37Rv strain already had this nucleotide change which may have been because of repeated passages of this reference strain in the laboratory or (2) that this change was introduced during PCR cycling. Both the target sequences had a high degree of similarity with the published sequences and therefore, it was concluded that they could still be used as positive controls for subsequent assays.

4.8.3   Standard curve preparation

4.8.3.1 *Standard curve for novel JCU IS6110 TaqMan assay*

An efficiency of 100% with an R^2 value of 0.9947 (Figure 4.4) was observed. The equation of the linear regression line was y=-3.32x+38.29. This range of values was expected, as 100% efficiency reflects the potential of an assay to double the copy number of target gene(s) every cycle. The correlation coefficient of 0.9947 represented perfect linearity of the standard curve. Therefore, the primer and probe set designed at JCU to target the IS6110 sequence had the ability to increase the target copies in a standard manner with each cycle.
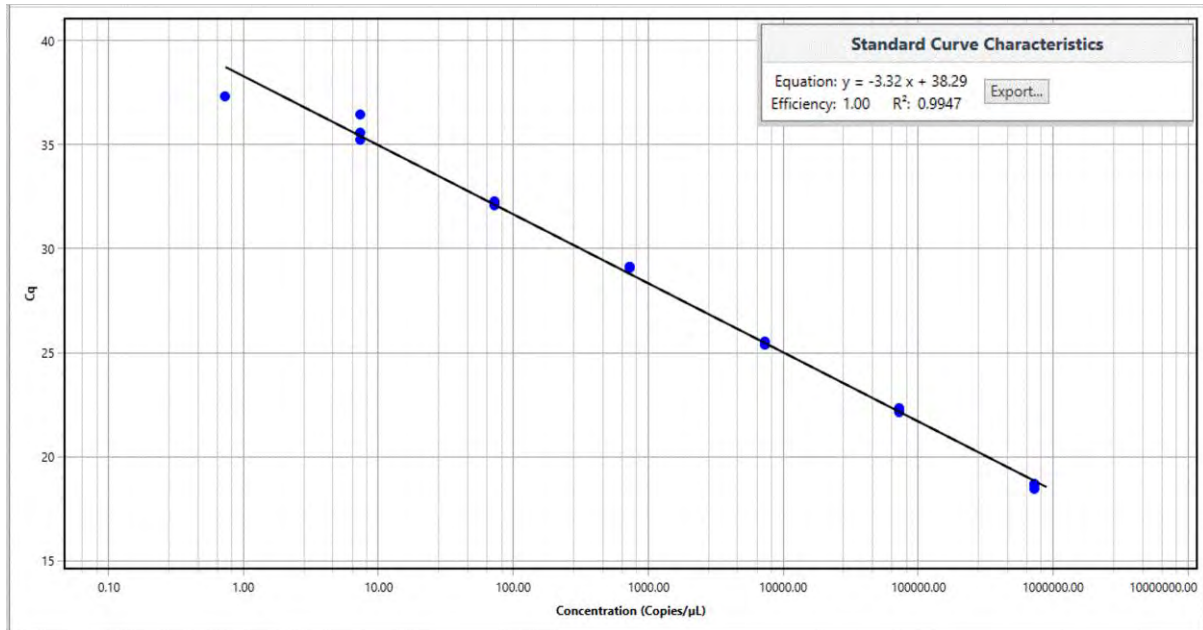
Figure 4.4: Standard curve of JCU IS6110 TaqMan assay with an efficiency of 100% and 0.9947 R^2 value. The DNA concentration in copies/ µL on the X axis and C_t value on the Y-axis. The blue dots over the slope represented the target amplicon.

4.8.3.2 *Standard curve for the published IS6110 TaqMan assay*

A value of 98% efficiency and 0.9620 R^2 was observed (Figure 4.5). The equation of the linear regression line was y=-3.36x+38.87. This indicated that the assay was not precisely able to double the copy number of the target gene in each PCR cycle. This indicated that, the primer pair and the probe for the published IS6110 assay underperformed when compared to the novel JCU IS6110 assay. The results proved that the designed JCU IS6110 assay was more efficient than the published IS6110 assay with less C_t values for the same dilution.
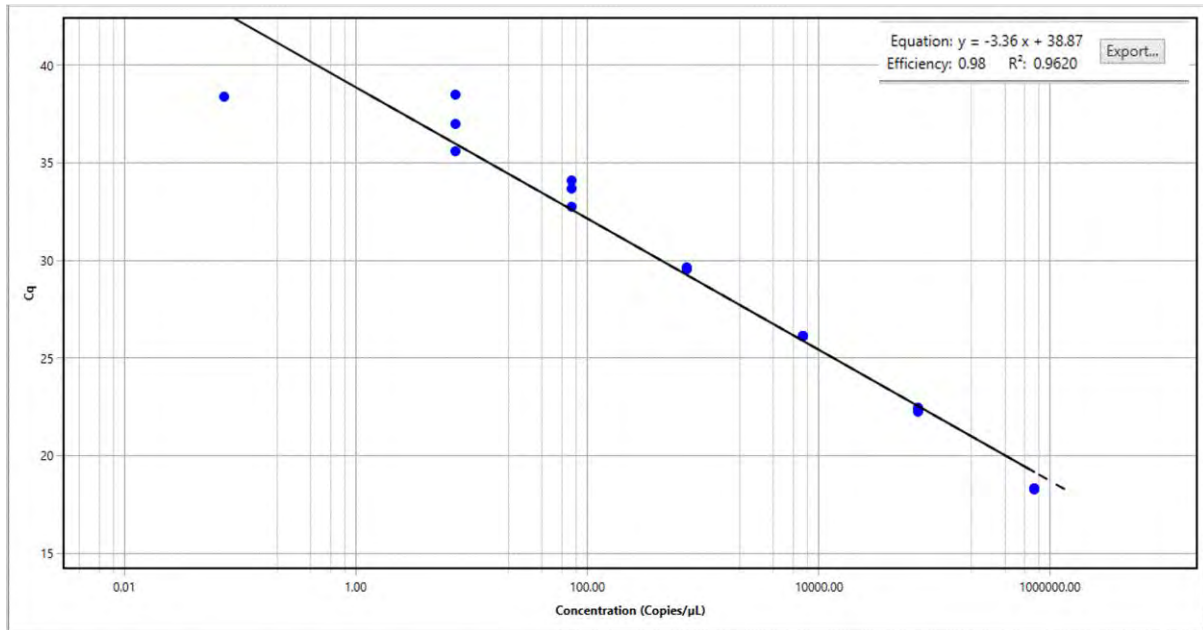
Figure 4.5: Standard curve of published IS6110 TaqMan with 98% efficiency and 0.9620 R^2 value. The DNA concentration in copies/ μL on the X axis and C_t value on Y-axis. The blue dots over the slope represented the target amplification.

4.8.3.3 *Standard curve for the novel rpoB TaqMan assay*

A standard curve for the *rpoB* TaqMan assay was prepared (Figure 4.6) with 84% efficiency and an R^2 value of 0.9968 (Figure 4.6) which was below the MIQE guidelines optimal range (90-110%) (Bustin *et al.*, 2009) for publication. However the observed R^2 value was 0.9968, which was very close to 1, the optimal value (Bustin *et al.*, 2009). Therefore, the assay was repeatable. The *rpoB* assay was not that efficient as the JCU IS6110 assay.
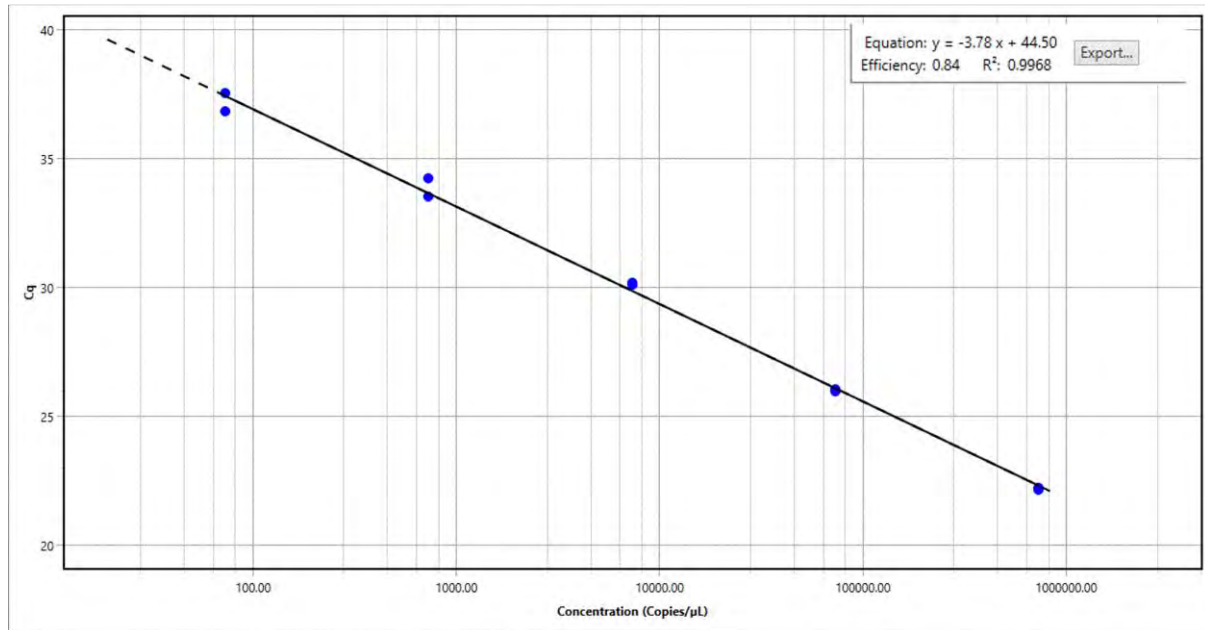
Figure 4.6: Standard curve for *rpoB* TaqMan assay with 84% efficiency and 0.9968 R^2 value. The DNA concentration in copies/µL on the X axis and C_t value on the Y-axis. The blue dots over the slope indicated the target amplification.

A summary of the standard curves for the three TaqMan assays is shown in Table 4.9.

Table 4.9: Comparison of the standard curves of three TaqMan assays.

TaqMan assay name	Efficiency in %	R^2	Equation of the linear regression
JCU IS6110 assay	100	0.9947	y=-3.32x+38.29
Published IS6110 assay	98	0.9620	y=-3.36x+38.87
JCU *rpoB* assay	84	0.9968	y= -3.78x+44.50

The IS6110 TaqMan designed at JCU had the highest efficiency (100%) with a good R^2 value followed by the published IS6110 TaqMan assay, with 98% efficiency and 0.9620 R^2 value. The efficiency of the *rpoB* TaqMan assay requires improvement based on these results and use of fresh reagents in subsequent experiments was required. The *rpoB* TaqMan had the highest R^2 value of 0.9968 of all TaqMan assays.

4.8.4   Calculation of sequence copy numbers

The URI Genomics and Sequencing Centre website (Figure 4.7) with a Mtb genome length of 4.41×10^6 bp used to calculate copy numbers. It was determined that the undiluted H37Rv genome used for construction of standard curves had a concentration of 35ng/µL which is equivalent to 7.35×10^6 copies/µL.
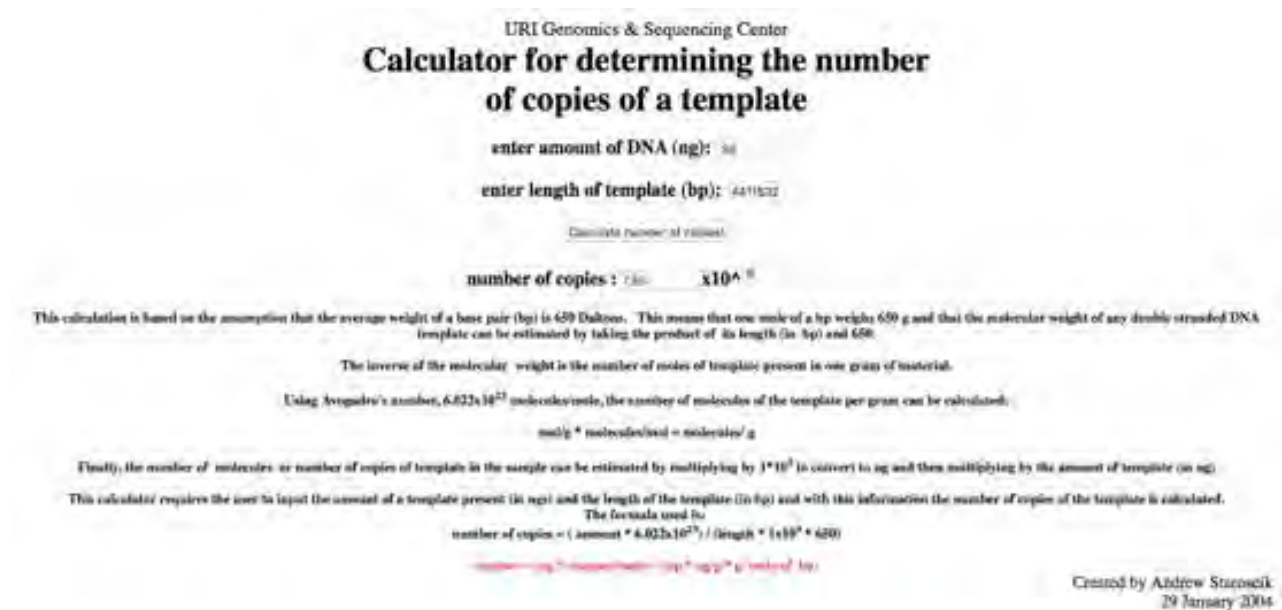

Figure 4.7: Determination of copy number of H37Rv genomic DNA with 35ng/µL concentration and 4.41×10^6 bp Mtb genome length.

The detection limit of both IS6110 based assays (JCU IS6110 and published IS6110 TaqMan) was at the 10^{-6} dilution of the H37Rv stock DNA. However, the corresponding number of DNA copies/µL could not be determined due to the unknown copy number of the IS6110 sequence in the genome of the reference strain H37Rv.

The *rpoB* assay amplified the *rpoB* sequence up to the 10^{-5} dilution of H37Rv template and the calculated number of DNA copies was therefore 73.5/µL and as 2 µL of template DNA per reaction was used, the number of copies per reaction were 147 copies/ µL (Figure 4.7).

4.8.5   Validation of IS6110 assays

4.8.5.1 *Analytical sensitivity of IS6110 assays*

Ten tenfold dilutions of H37Rv genomic DNA were used to determine the analytical sensitivity of the two TaqMan assays, both targeting the IS6110 gene.

Published TaqMan assay

H37Rv genomic DNA dilutions were prepared and tested with the published TaqMan assay (Figure 4.8). Template dilutions from 10^{-7} dilution did not react, while all the three replicates of the 10^{-6} dilution reacted. Therefore, it was concluded that 10^{-6} dilution was the limit of detection of this assay.
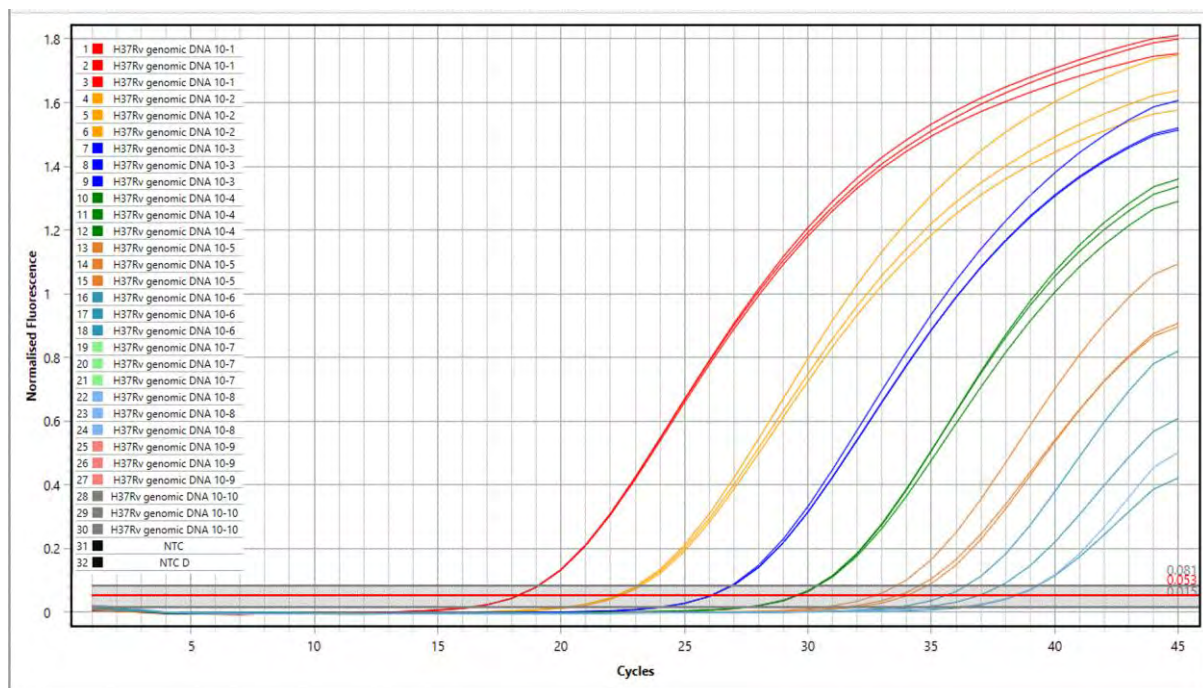

Figure 4.8: Analytical sensitivity of published IS6110 TaqMan assay. All the replicates of all dilutions up to 10^{-6} reacted. Only one replicate of 10^{-8} dilution, but none of 10^{-7} reacted. There was an irregularity in the reaction pattern of dilutions.

JCU IS6110 TaqMan

As previously, H37Rv genomic DNA dilutions were used to determine the analytical sensitivity using the primer and probe set designed at JCU (Figure 4.9). It is clear from the figure that only one replicate of the 10^{-7} dilution reacted, while all three distinct dilution replicas of the 10^{-6} dilution reacted. Therefore, 10^{-6} was the detection limit of this assay.
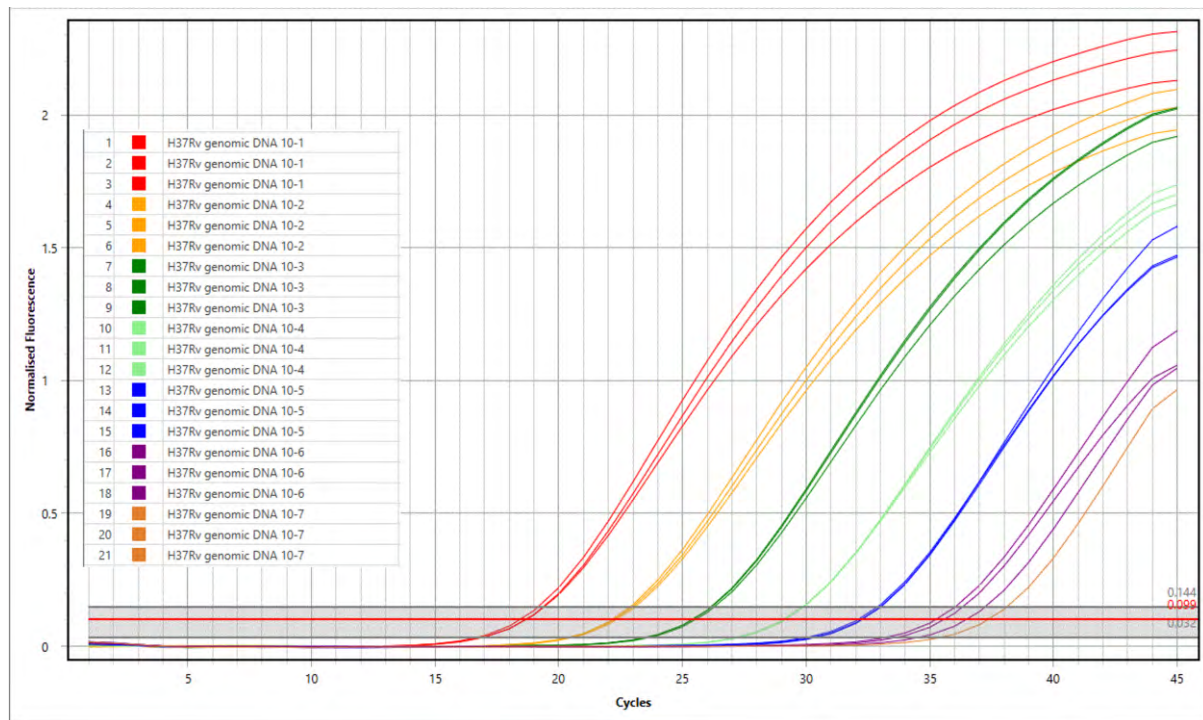

Figure 4.9: Analytical sensitivity of JCU IS6110 TaqMan assay. All the replicates of all dilutions up to 10^{-6} reacted. Each dilution was in triplicate. Only one replicate of 10^{-7} dilution reacted.

Upon comparing the detection limit of both published and JCU IS6110 TaqMan, both assays had a similar limit of detection. Both were able to identify and amplify the target sequence up to its 10,00,000 times. Relative sensitivity of these two IS6110 TaqMan assays was also done using DNA from two different isolates from PNG as described below.

4.8.5.2 *Relative sensitivity using DNA from two different isolates from PNG*

DNA extracted from the supernatant of cultured isolates 57-2 and 96 were used to compare the relative sensitivity of both IS6110 TaqMan assays. These isolates were derived from two sputum samples from two different TB suspected patients in PNG. Ten tenfold dilutions of each decontaminated sample were prepared in nuclease-free water for further use.

Published IS6110 assay sensitivity for isolate 57-2 DNA

It is clear from the Figure 4.10 that only two duplicates of dilution 10^{-4} reacted. This means that 10^{-3} was the detection limit of this published assay.
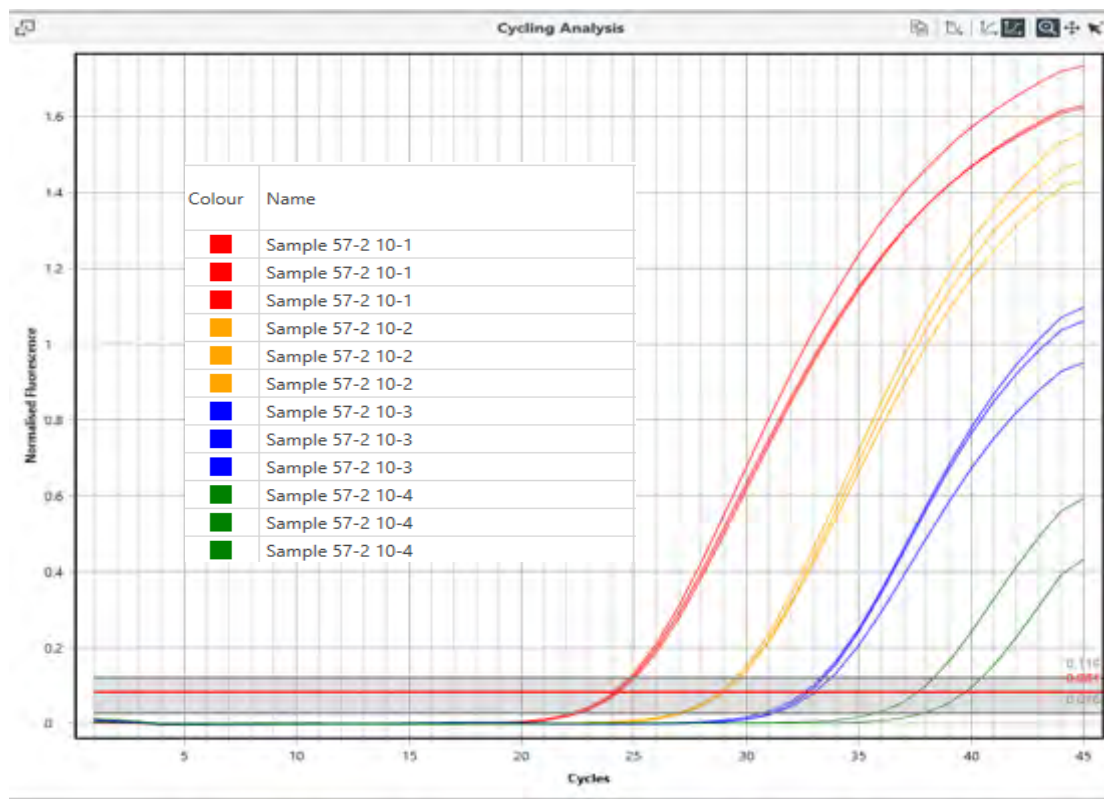

Figure 4.10: Sensitivity of published IS6110 assay using ten tenfold dilutions of the isolate 57-2 DNA. Reaction in all until 10^{-4} dilution was observed. Since only two replicates of 10^{-4} reacted, the suggested limit of the detection was 10^{-3} dilution.

JCU IS6110 assay sensitivity for isolate 57-2

All replicates up to 10^{-5} dilution reacted using the assay designed at JCU (Figure 4.11). Hence this was its limit of detection.
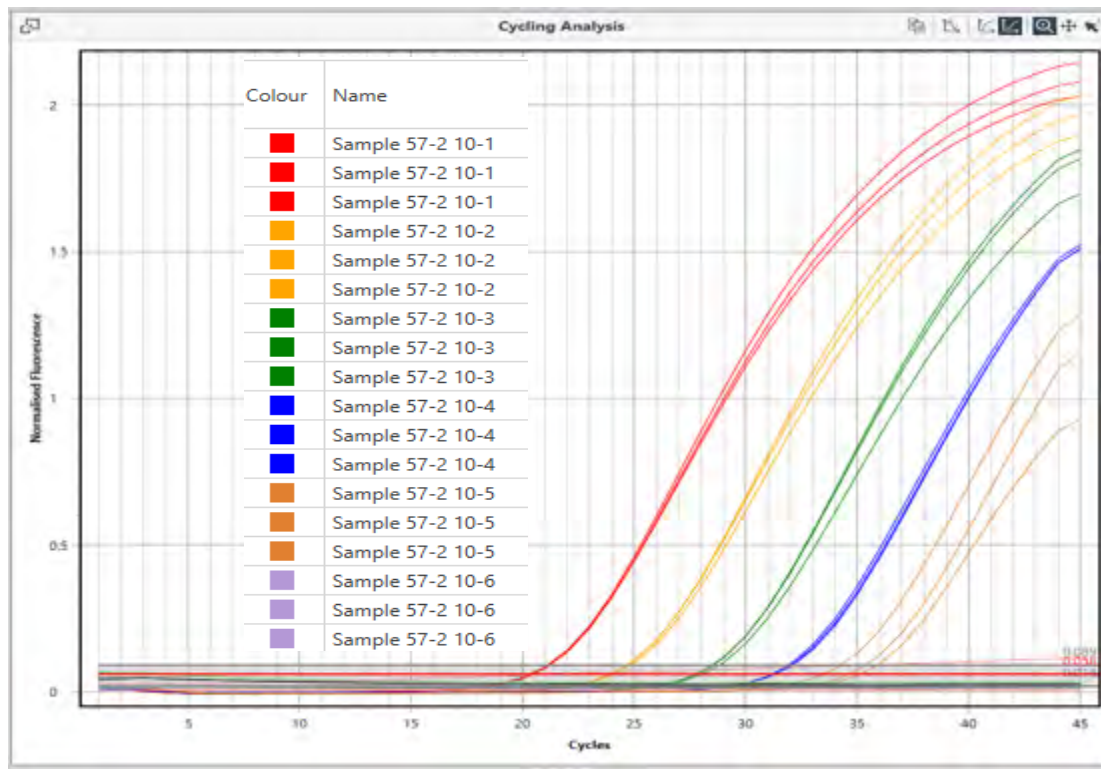

Figure 4.11: Sensitivity of JCU IS6110 assay using ten tenfold dilutions of the isolate 57-2 DNA. Reaction up to 10^{-5} dilution was observed. Therefore, the suggested limit of detection was 10^{-5} dilution.

Comparing the above figures (Figure 4.10 and 4.11), it appeared that the assay designed at JCU was more sensitive. Since one dilution factor roughly imparts change(s) in 6 C_t value. Though it differed in just one tenfold dilution, it was 10 times more sensitive than the published assay. This increased sensitivity suggests that the JCU IS6110 assay has superior ability to amplify the target sequence even from fewer copies of source DNA.

Published IS6110 assay sensitivity on DNA from 96 isolate

Isolate 96 DNA was diluted ten times by tenfold dilution. Both TaqMan assays were then applied separately, and the results were compared.

It can be seen in the following figures that both assays were able to amplify the target sequence up to the 10^{-4} dilution (Figure 4.12 and 4.13). Only two replicates of 10^{-4} dilution reacted in case of the published assay. In contrast, JCU IS61110 TaqMan was able to amplify the IS6110 sequence in all replicates of 10^{-4} dilution.
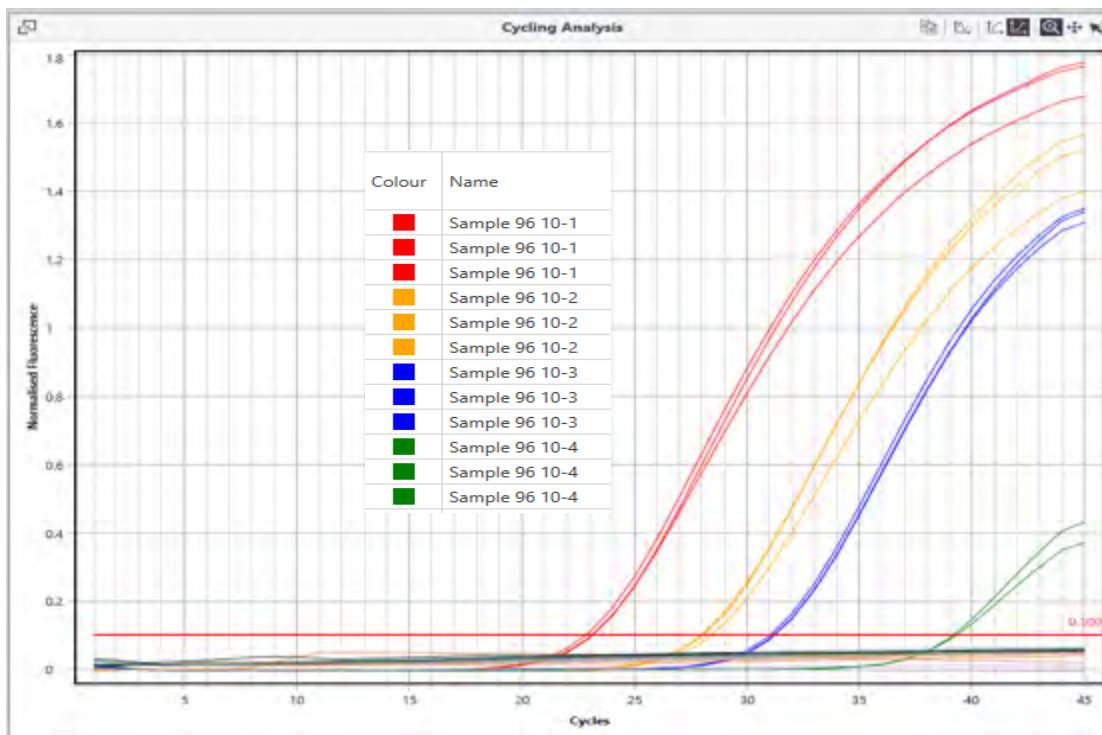

Figure 4.12: Sensitivity of published IS6110 assay using ten tenfold dilutions of the isolate 96 DNA. Reaction in all dilutions up to 10^{-4} dilution was observed. Since only two replicates of 10^{-4} reacted, the suggested limit of detection was 10^{-3} dilution.
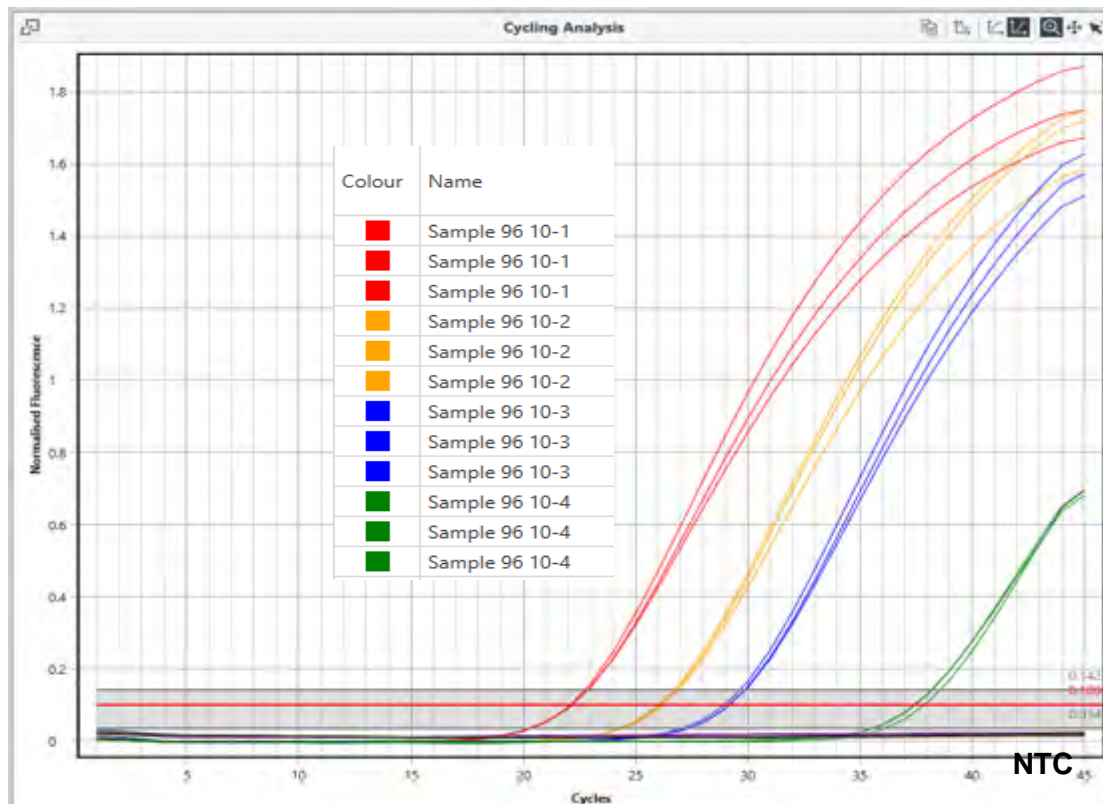
JCU IS6110 TaqMan assay on DNA of the 57-2 isolate

Figure 4.13: Sensitivity of JCU IS6110 assay using ten tenfold dilutions of the isolate 96 DNA. Reaction in all until 10^{-4} dilution was observed. This was the limit of detection. Dynamic mode of reaction was chosen as suggested in the PCR device manual for the TaqMan assays.

Based on the above comparisons (Figure 4.12 and 4.13), it was concluded that JCU TaqMan was more sensitive than the published TaqMan. This result suggested that the JCU IS6110 TaqMan assay may be more suitable when applied to clinical samples.

Table 4.10: Comparison of the detection limit of JCU IS6110 and published IS6110 assay.

DNA type	TaqMan assay detection limit	
	JCU IS6110	Published IS6110
H37Rv genomic DNA	10^{-6}	10^{-6}
Isolate 57-2	10^{-5}	10^{-4}
Isolate 96	10^{-4}	10^{-3}

Table 4.10: shows that both the JCU IS6110 and the published IS6110 assays had variation in their respective sensitivities when using DNA from clinical isolates. Both assays had 10 times less sensitivity with the isolate 57-2 DNA. The sensitivity further decreased when the DNA of the 96 isolate was used. Inter-assay comparison with the isolates DNA demonstrated the higher sensitivity of the JCU IS6110 assay than the published IS6110 (Broccolo *et al.*, 2003).

This variation could be related to the source, from which the DNA was extracted. MGIT™ supernatant of the isolates 57-2 and 96 were sourced from the sputum of the two TB suspected patients. The sputum harbours more than one organism. Therefore, the DNA from those isolates might have DNA from contaminants that eventually led to decreased sensitivity of the assays. Furthermore, as primers and probes were designed against H37Rv any sequence variation in the IS6110 genes from clinical sample isolates many influence this sensitivity.

4.8.5.3 *Analytical specificity of IS6110 TaqMan assays*

In-silico primer specificity was carried out along with the analytical specificity. It was important to check how and to what extent the designed primer sets bind to the target site of the genomic DNA (the downloaded published sequences aligned together). This would in turn reflect the possibility of amplifying the correct target gene for plasmid control construction. It is important to note here that all the assays were tested for their specificity individually and then; their results were compared to each in tabular form (Table 4.13).

***In silico* primer set evaluation**

Published IS6110 sequences with 100% homology were downloaded from the NCBI nucleotide website. They were aligned together using the Geneious program. The aligned sequences were then viewed using the GeneDoc program.

Figure 4.14 demonstrated the binding pattern of the IS6110 diagnostic primers against the reference sequences. The forward (IS6110-108-F) and reverse primer (IS6110-108-R) are highlighted in blue, and the grey region shows the annealing site of the probe (IS6110-108-P). The probe bound in a region that was located inside the area where both primers annealed. This was the target region of 108bp length. However, a mismatch of the forward primer on one of the plasmid controls was noted. This single nucleotide change was approximately in the middle that imparted no considerable difference in the binding pattern of the primer sequence. Therefore, the diagnostic primers were also able to bind and amplify the target sequence with high specificity.
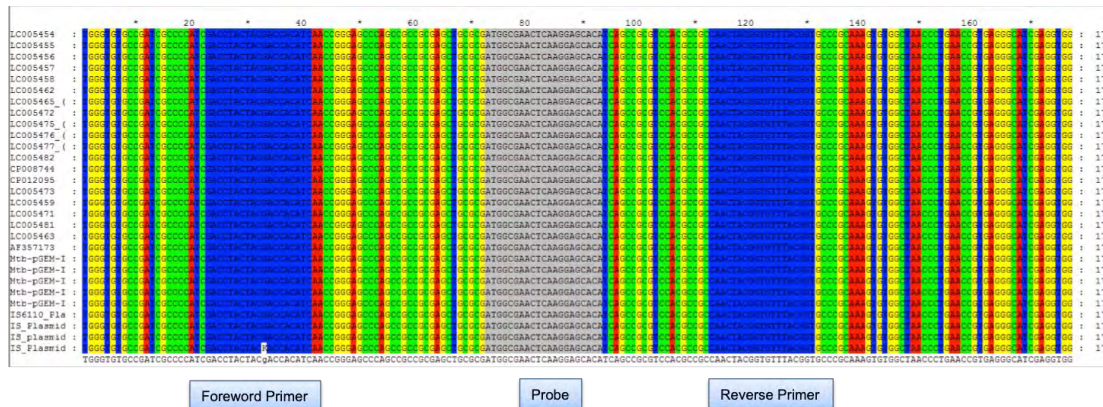

Figure 4.14: Representation of the binding sites of IS6110 diagnostic primers set and the probe against a genomic database including the isolated sequence in the study. The sequence of interest is in between the flanking primers set, while the probe binds in the region where the target sequence was located. The published sequences used are detailed in Appendix 14, Table 14.1. The *in-silico* specificity testing of the primer pair used to extract the IS6110 sequence for plasmid construction was also done (Appendix 5).

Analytical specificity

Genomic DNA from 40 different organisms was extracted and tested to check the specificity of both IS6110 TaqMan assays. It was anticipated that the assay designed at JCU would be more specific towards its target sequence.

Published IS6110 TaqMan assay

The published IS6110 primer and probe were used with DNA from 40 different organisms to determine its analytical specificity (Figure 4.15). Three different dilutions of the IS6110 plasmid tested in duplicate were used as positive controls. The absence of amplification in NTC was observed. It was expected. Of the 40 different DNA sample tested, only four had amplifications. These were H37Rv, *Mycobacterium bovis, Mycobacterium intracellulare and Mycobacterium abscessus.* All 40 DNA was extracted from pure culture. Therefore, they all were quality samples.
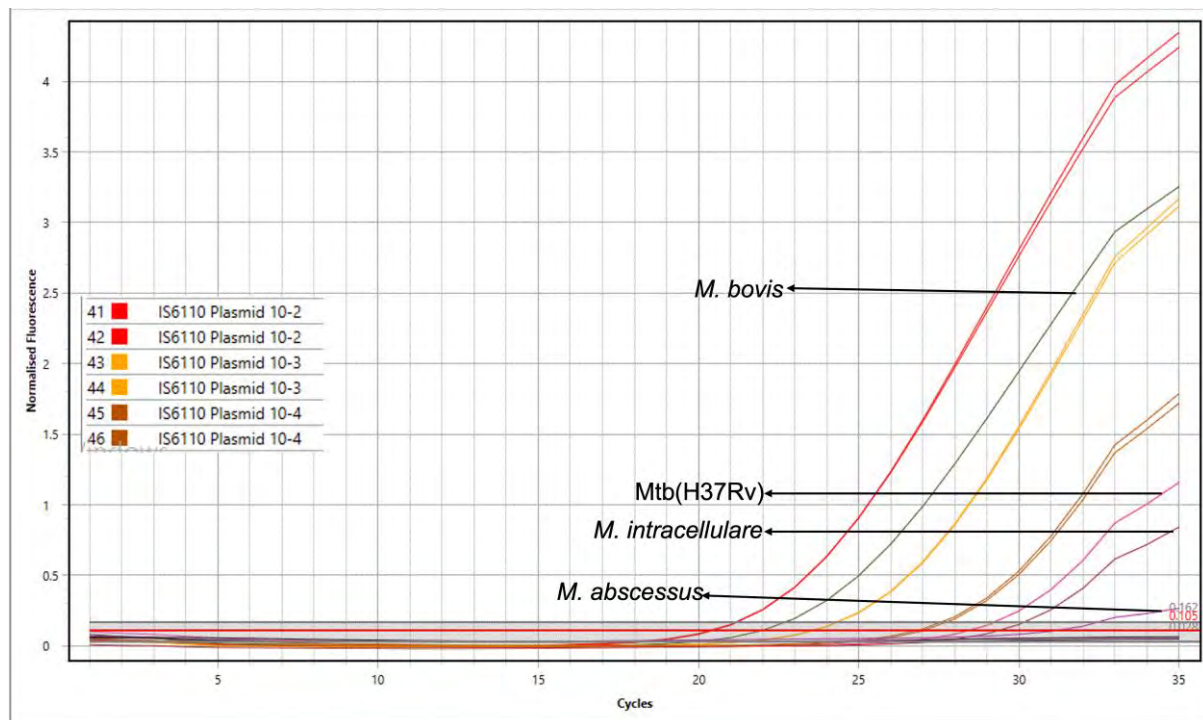

Figure 4.15: Specificity data of published TaqMan assay. The assay was used on a panel of DNA samples from 40 different organisms to determine its specificity. Only *Mycobacterium* sp. were amplified, indicating good specificity of this assay. Positive and negative controls were as expected.

JCU IS6110 TaqMan

The analytical specificity of the primer and probe set designed at JCU was determined using the same panel of 40 DNA samples (Figure 4.16). Positive controls consisted of three different dilutions of one IS6110 plasmid each in duplicate. Non-template control (NTC) was included to ensure the absence of any contamination. As was the case for the published IS6110 assay, H37Rv, *M. bovis* and *M. intracellulare and M. abscessus* DNA were detected by this novel assay.
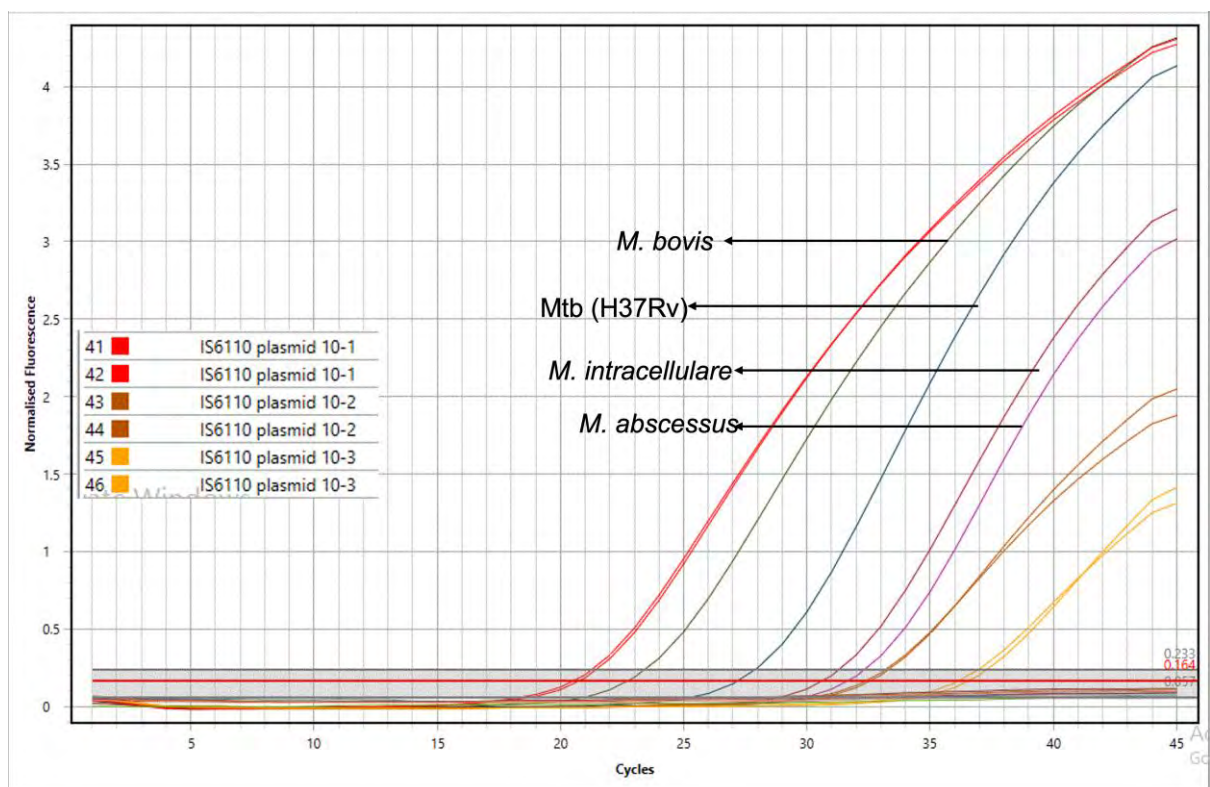

Figure 4.16: Specificity data of JCU IS6110 TaqMan assay. The assay was used on a panel of DNA samples from 40 different organisms to evaluate its specificity. Only *Mycobacterium* spp. were amplified, indicating good specificity of this assay. Positive and negative controls were as expected.

Based on the specificity data of both IS6110 TaqMan assays, both assays had similar specificity for mycobacterial species. Amplification of the non-tuberculous mycobacteria (NTM), *M. intracellulare* was not unexpected, as Broccolo *et al.* (2003) also reported that their IS6110 assay could detect mycobacterial species outside the MTBC i.e., *M. intracellulare* which is a member of the *M. avium* complex. Interestingly, this study revealed that the phylogenetically distant *M. abscessus* is also detected by both IS6110 assays. Both assays had quite similar C_t values for *M. intracellulare* and *M. avium.*

A further *in silico* evaluation of both IS6110 assays on these two NTM was done using nucleotide BLAST searches. It was found that the published IS6110 primer set and probe (Broccolo *et al.*, 2003) (Appendix 8, Figure 8.1) had homology with various regions in the genome of the *M. abscessus.* A similar pattern was observed for the genome of *M. intracellulare* (Appendix 8, Figure 8.2). Similarly, the primer pair and probe of the JCU IS6110 assay had 100% identity to regions of the *M. abscessus* genome (Appendix 8, Figure 8.3). The primer set mismatch may be due to identical target gene segment present in the *M. abscessus*. The primers were designed to achieve maximum sensitivity.

Furthermore, 11 bases of JCU IS6110 reverse primer, 13 bases of forward primer and 14 bases of probe bound to the various DNA segments of the *M. intracellulare* genome (Appendix 8, Figure 8.4). It was therefore concluded that both published IS6110 and JCU IS6110 TaqMan assays, although specific to the *Mycobacterium* genus, were not highly specific to MTBC. This could be a limitation of this study.

Similarly, the analytical specificity of the *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was determined and compared with the above assays (Table 4.13). *In-silico* specificity testing contradicted the published claims. The *senX3-regX3* IR gene was not only present in a wide range of *Mycobacterium* spp*.*, but also different bacterial genera such as *Pseudonocardia sp.* and Saccharothrix spp*.* (Appendix 9). These findings question the specificity of this gene to MTBC. Therefore, based on this testing it was concluded that the *senX3-regX3* IR gene is also not completely specific to the MTBC.

4.8.5.4 *Assay reproducibility of the IS6110 assays*

Inter-assay reproducibility experiments were performed using IS6110 plasmid/H37Rv dilutions as templates. The dilutions of the genomic DNA/plasmid suspension were made before use. They were used immediately on same day, followed by next weeks (one attempt in one week using same dilution stock) to get reproducibility data. The vials were kept under ice in an esky all the time during work. The C_t value of each dilution were recorded in three independent attempts (Table 4.11 and 4.12, Figures 4.17 and 4.18).

Published IS6110 TaqMan assay

Table 4.11 and Figure 4.17 demonstrate good reproducibility for the published IS6110 assay with only minor variation in the C_t values of all dilutions in all three different PCR runs. Each run was performed in one individual week. The graph was constructed with sample number and their mean C_t values with standard deviation. The average coefficient of variance was 2.832. As, it was less than 10, it was therefore reproducible (https://salimetrics.com/calculating-inter-and-intra-assay-coefficients-of-variability/). Therefore, it was concluded that this assay could potentially provide consistent and reliable data and could be applied to DNA extracted from clinical samples/cultures. The details of all C_t values are tabulated in Appendix 10-Table 10.2 and calculations in Appendix 17.

Table 4.11: Reproducibility data of published IS6110 TaqMan. Three different genomic DNA dilutions were run in duplicates three times and the C_t values were compared.

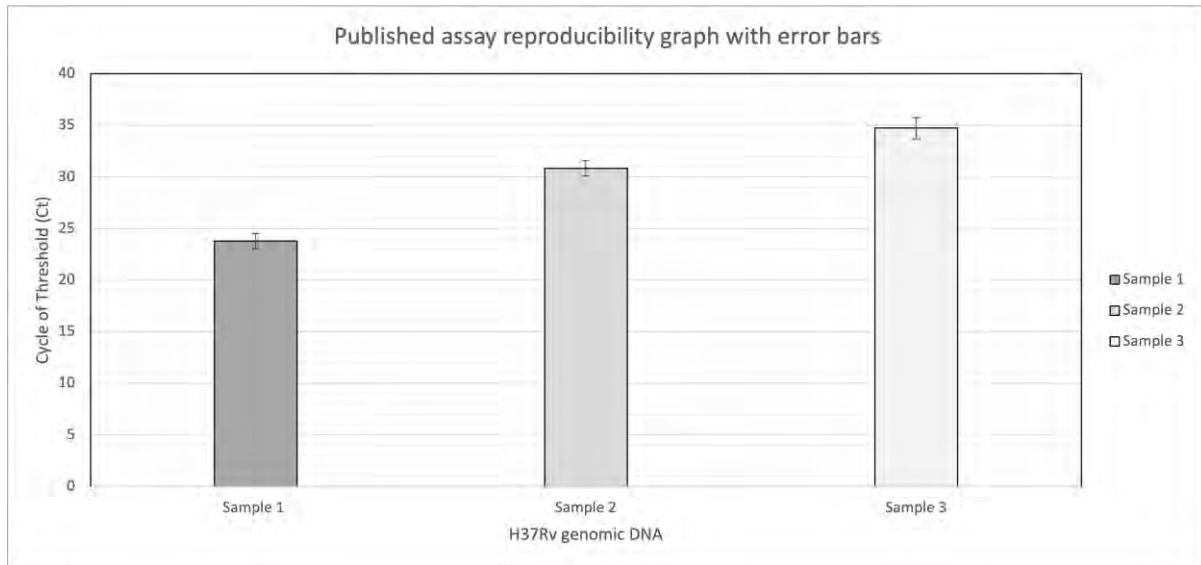| H37Rv genomic DNA	Mean C_t value on 1st attempt	Mean C_t value on 2nd attempt	Mean Ct value on 3rd attempt	Mean	Standard Deviation (SD)	Coefficient of Variance (CV%)
10^{-1}	22.885	24.155	24.255	23.765	0.763	3.213
10^{-2}	30	31.26	31.235	30.831	0.720	2.336
10^{-3}	33.545	35.36	35.275	34.726	1.024	2.949

Figure 4.17: The reproducibility data of the published IS6110 assay. The error bars show fading pattern. They darkest Sample 1 was the H37Rv genomic DNA of 10^{-1} dilution. The Sample 2, mid colour indicates the H37Rv genomic DNA of 10^{-2} dilution and the lightest shade means DNA of 10^{-3} dilution. Based on the CV of the data; the assay was reproducible in nature.

JCU IS6110 TaqMan

As previously, the IS6110 plasmid was used at three different dilutions to check the assay's reproducibility. As shown Table 4.12 and 4.19, the JCU IS6110 TaqMan assay was reproducible as there were only minor differences in the C_t values of all dilutions in all three different PCR runs. Each run was performed in one individual week. The graph was constructed with sample number and their mean C_t values with standard deviation. The average coefficient of variance was 3.288. Since it was less than 10, it was therefore reproducible (https://salimetrics.com/calculating-inter-and-intra-assay-coefficients-of-variability/). Therefore, this assay was also suitable to provide reliable data on clinical material. The details of all C_t values are tabulated in Appendix 10-Table 10.3 and calculations in Appendix 17.

Table 4.12: Reproducibility data of JCU IS6110 TaqMan. Three different plasmid dilutions were run in duplicates three times and the C_t value were recorded and compared.

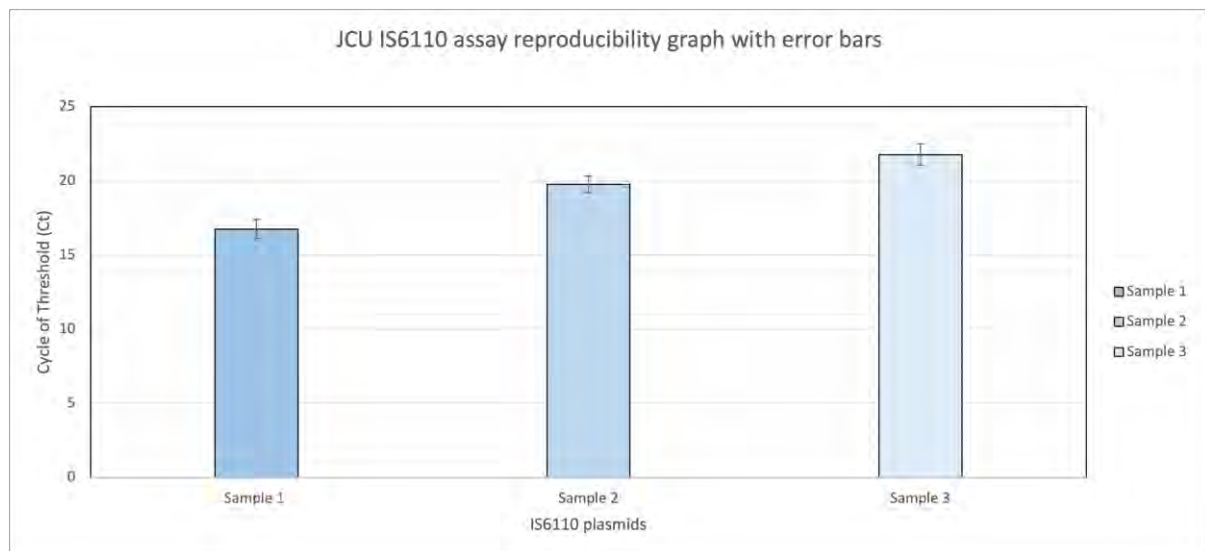| H37Rv genomic DNA	Mean C_t value on 1st attempt	Mean C_t value on 2nd attempt	Mean Ct value on 3rd attempt	Mean	Standard Deviation (SD)	Coefficient of Variance (CV%)
10^{-1}	17	17.22	16.03	16.75	0.633	3.780
10^{-2}	19.866	20.24	19.163	19.756	0.546	2.767
10^{-3}	22.053	22.303	20.946	21.767	0.722	3.317

Figure 4.18: The reproducibility data of the designed JCUIS6110 assay. The error bars show fading pattern. They darkest Sample 1 was the IS6110 plasmid suspension of 10^{-1} dilution. The Sample 2, mid colour indicates the IS6110 plasmid of 10^{-2} dilution and the lightest shade means plasmid of 10^{-3} dilution. Based on the CV of the data; the assay was reproducible in nature.

4.8.6   Validation of the *rpoB* assay

4.8.6.1 *Analytical sensitivity of the rpoB assay*

Ten tenfold dilutions of the H37Rv genome were prepared in nuclease free water (Figure 4.19). It is clear from the figure that only one replicate of dilution 10^{-6} reacted, while successful amplification occurred in each replicate of dilution 10^{-5}. This means that 10^{-5} was the detection limit of this assay.

Therefore, based on the successful amplification of the target gene in each reaction of 10^{-5} dilution, it was deduced that the *rpoB* TaqMan assay could reliably detect the target DNA even if it was 10^{5} times diluted.

Figure 4.19: Sensitivity data of *rpoB* TaqMan assay. All replicates up to 10^{-5} dilution reacted.

4.8.6.2 *Analytical specificity of designed rpoB sequencing primer set*

***In-silico* primer specificity evaluation**

The published *rpoB* gene sequences were downloaded from NCBI nucleotide. They were aligned using Geneious software and further viewed using the GeneDoc program for specificity testing.
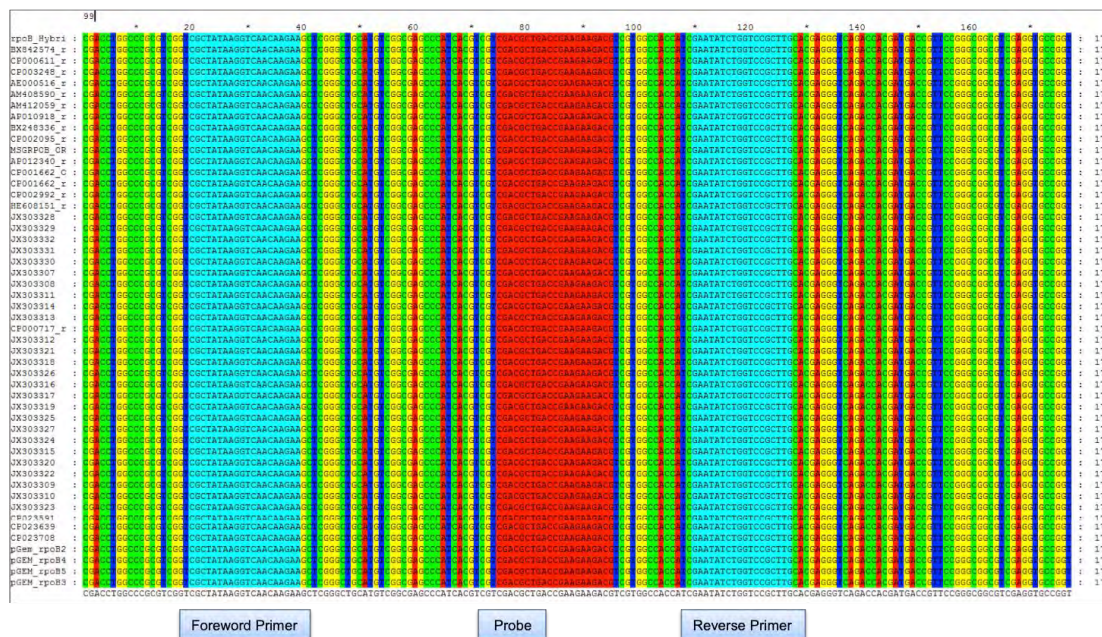
Figure 4.20: Binding pattern of *rpoB* TaqMan assay against all aligned published *rpoB* sequences of Mtb. The published *rpoB* sequences used are described in Appendix 14, Table 14.2.

In Figure 4.20, the forward (*rpoB*-TM-108-F) and the reverse primer (*rpoB*-TM-108R) bound at position 20 to 40bp and 110 to 128bp respectively. Both are highlighted in blue. The probe *rpoB*-TM-108-P (red) bound to the target region that was located at the position 75 to 98bp between forward and the reverse primer. Binding to all the reference sequences demonstrated high specificity towards the target gene. Therefore, the sequencing and diagnostic set were able to locate and amplify the *rpoB* gene segment.

The *in-silico* specificity analysis of the primer set (*rpoB*-PL4-1013-F and *rpoB*-PL4-1013-R) was also determined (Appendix 5, Figure 5.3 and 5.4) to ensure the designed primer pair was specific to its target.

Analytical specificity of *rpoB* assay

The *rpoB* TaqMan primers and probe were used for determining analytical specificity. All the taxonomically relevant *Mycobacterium* species and nearly all the clinically relevant DNA reacted (Figure 4.21) i.e., 39/40 DNA samples.

The *rpoB* gene encodes the β-subunit of bacterial polymerase enzyme and is required to carry out DNA replication. Although the *rpoB* primers/probes were designed to target *M. tuberculosis*, the high degree of conservation of this gene across all bacteria means that this result is not surprising. As *rpoB* is a single copy gene its utility here is to provide evidence that samples contain sufficient DNA are "amplifiable".
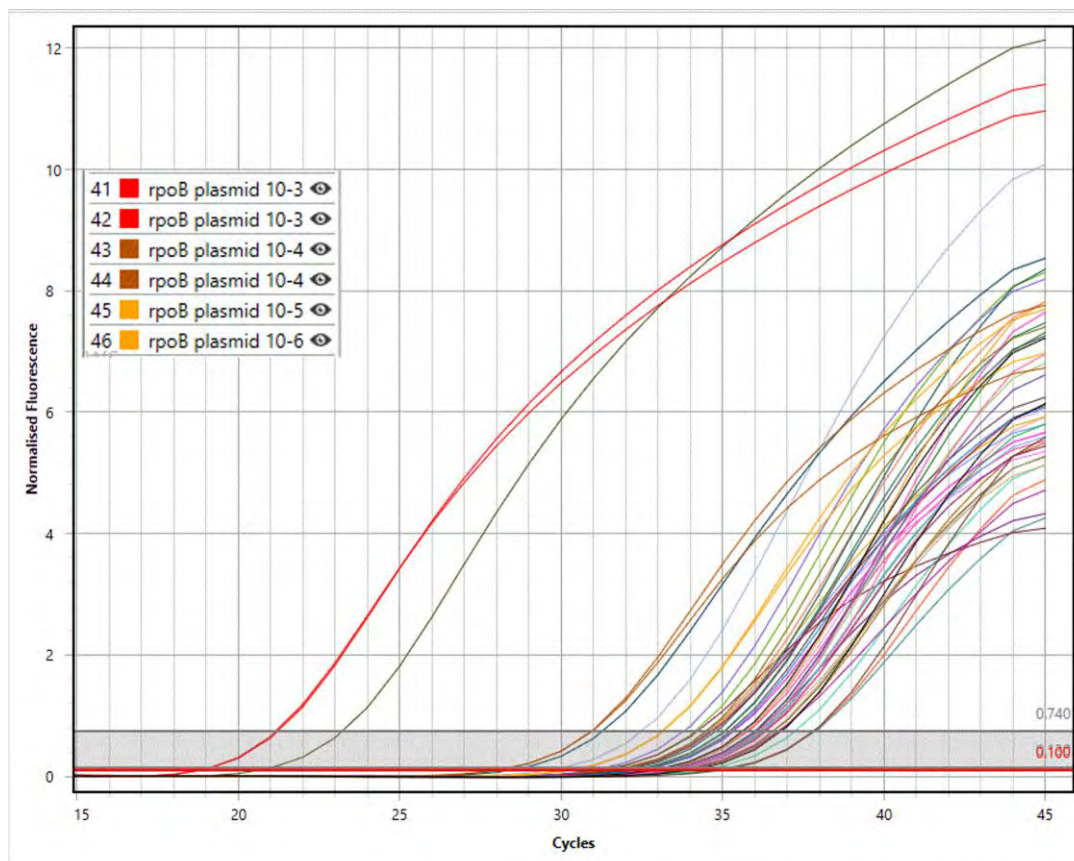

Figure 4.21: Application of *rpoB* TaqMan assay to 40 different organisms. 39/40 DNA samples were detected along with the positive controls.

Table 4.13: Comparison of published and designed assays for their analytical specificity.

Organisms	No. of strains reacting by TaqMan with				
	Published IS6110 assay	*senX3-regX3* IR	JCU IS6110	*rpoB*	C_t value from *rpoB* assay
Clinically associated organisms					
Acinetobacter baumannii	-	-	-	+	37.18
Aeromonas hydrophila	-	-	-	+	34.06
Bacillus cereus	-	-	-	+	36.22
Bacillus mycoides	-	-	-	+	38.63
Brachybacterium sp.	-	-	-	-	-
Enterobacter cloacae	-	-	-	+	36.83
Enterococcus faecalis (ATCC)	-	-	-	+	34.37
Erysipelothrix rhusiopathiae	-	-	-	+	35.51
Escherichia coli (ATCC)	-	-	-	+	33.19
Alkaligenes sp.	-	-	-	+	34.73
Klebsiella pneumoniae	-	-	-	+	34.03
Listeria monocytogenes	-	-	-	+	36.02
Micrococcus luteus	-	-	-	-	-
Moraxella catarrhalis	-	-	-	+	31.91
Neisseria gonorrhoeae	-	-	-	+	36.01
Neisseria meningitidis	-	-	-	+	34.78
Nocardia sp.	-	-	-	+	36.12
Paenibacillus polymyxa	-	-	-	+	33.52
Proteus mirabilis	-	-	-	+	36.36
Pseudomonas aeruginosa (ATCC)	-	-	-	+	34.14
Salmonella typhimurium (ATCC)	-	-	-	-	-

Staphylococcus aureus (ATCC)	-	-	-	+	32.51
Staphylococcus epidermidis (ATCC)	-	-	-	+	35.07
Staphylococcus saprophyticus	-	-	-	+	35.32
Streptococcus agalactiae	-	-	-	+	33.39
Streptococcus dysgalactiae	-	-	-	+	35.17
Streptococcus pneumoniae (ATCC)	-	-	-	+	35.41
Streptococcus pyogenes	-	-	-	+	35.20
Streptococcus viridians	-	-	-	+	33.72
Ureaplasma plasmid	-	-	-	+	35.03
Corynebacterium diptheriae	-	-	-	+	34.45
Bacillus cereus	-	-	-	+	36.27
Taxonomically associated organisms					
M. abscessus	+	-	+	-	-
M. avium	-	-	-	+	36.35
M. bovis	+	+	+	+	22.30
M. intracellulare	+		+	+	33.95
M. fortuitum	-		-	+	35.67
M. peregrinum	-		-	+	36.63
H37Rv	+	+	+	+	32.52
DNA 12 (M. intracellulare)	-	-	-	+	36.19

Table 4.13 demonstrates the overall comparison of specificity data of all TaqMan assays described in this project. The published *senX3-regX3* IR assay had the highest specificity as previously reported (Broccolo *et al.,* 2003). Surprisingly the *in-silico* specificity of this assay was quite different than this. (Appendix 9) The published IS6110 and JCU IS6110 TaqMan had similar specificity showing cross-reaction with *M. abscessus* and *M. intracellulare.* The *rpoB* TaqMan on the other hand was able to amplify the *rpoB* gene from nearly all DNA samples tested. The *rpoB* TaqMan assay was essentially used to determine the overall quality of the DNA samples rather than for Mtb identification so that HRM analysis (Chapter 5) and drug resistance-associated mutation analysis (Chapter 6) could be carried out.

4.8.6.3 *Reproducibility of the rpoB assay*

Intra assay repetition was done using H37Rv genomic DNA as the positive control. Three different dilutions were used and tested in duplicates along with the incorporation of the negative control (nuclease-free water instead of DNA as template). The details of all C_t values are tabulated in Appendix 10-Table 10.3 and calculations in Appendix 17. Each run was performed in one individual week. The graph was constructed with sample number and their mean C_t values with standard deviation. The average coefficient of variance was 2.216. It was less than 10 hence it was therefore reproducible (https://salimetrics.com/calculating-inter-and-intra-assay-coefficients-of-variability/).

Table 4.14: Reproducibility data of the *rpoB* assay. Three different plasmid dilutions were run in duplicates three times and the C_t value was noticed and compared.

H37Rv genomic DNA	Mean C_t value on 1st attempt	Mean C_t value on 2nd attempt	Mean Ct value on 3rd attempt	Mean	Standard Deviation (SD)	Coefficient of Variance (CV%)
10^{-1}	27.135	26.355	27.14	26.876	0.451	1.680
10^{-2}	34.305	32.86	33.335	33.5	0.736	2.198
10^{-3}	38.835	36.775	37.495	37.701	1.045	2.772

From Table 4.14 and Figure 4.22, every time an experiment was run, the C_t value of each dilution used was mostly identical with little variation. This suggests the potential of this assay to provide similar results that can be acquired again and again. Therefore, it was concluded that the assay gives reliable results and was repeatable in nature.
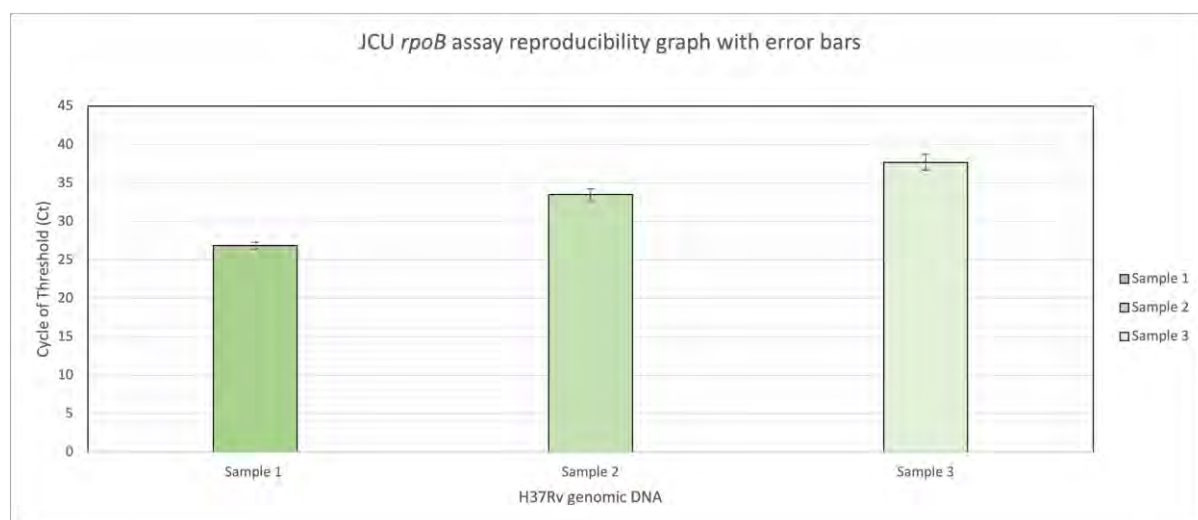

Figure 4.22: The reproducibility data of the designed JCU *rpoB* assay. The error bars show fading pattern. They darkest Sample 1 was the H37Rv genomic DNA of 10^{-1} dilution. The Sample 2, mid colour indicates the H37Rv genomic DNA of 10^{-2} dilution and the lightest shade means DNA of 10^{-3} dilution. Based on the CV of the data; the assay was reproducible in nature.

4.8.7   Utility of IS6110-*rpoB* duplex TaqMan assay on DNA extracted from clinical material

   4.8.7.1 *IS6110 detection in IS6110-rpoB duplex assay*

Out of 38 DNA samples, the JCU IS6110 specific primer and probe set was able to amplify the target sequence from 31 DNA templates (i.e., 82%) (Figure 4.23). All replicates of three plasmid controls reacted, while there was no amplification in NTC as per expectation.
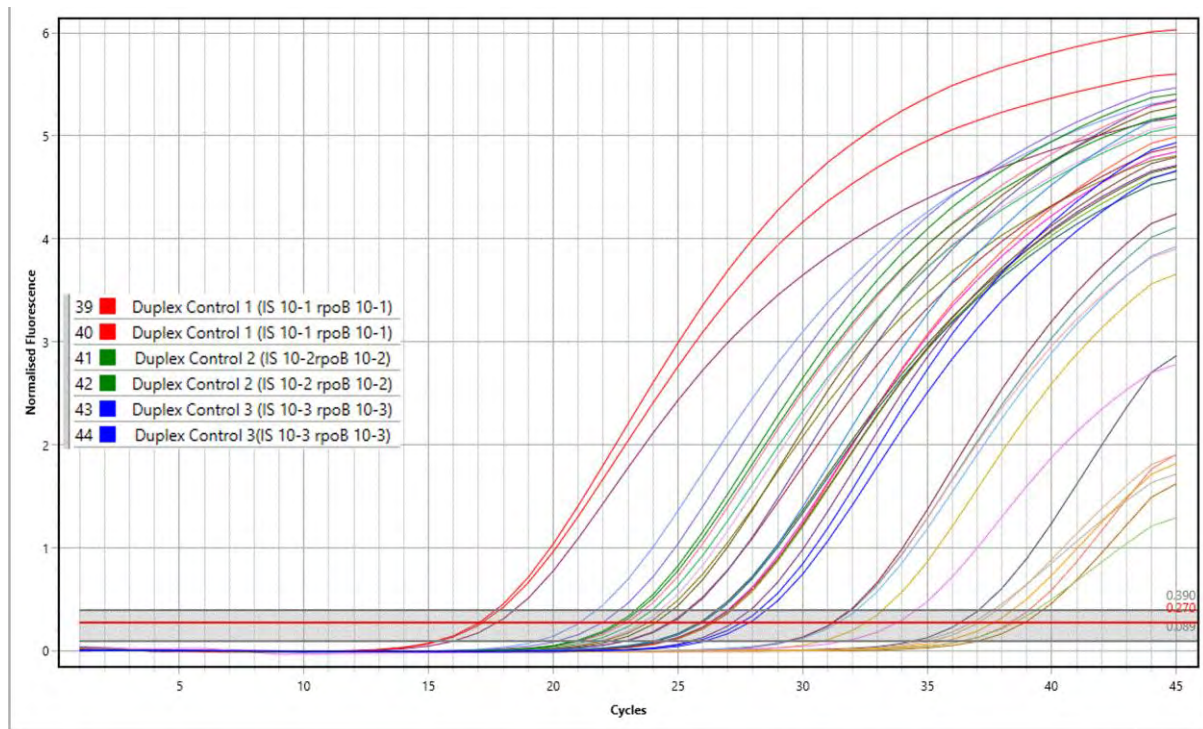
Figure 4.23: Duplex assay JCU IS6110 TaqMan assay amplification curves from 38 DNA samples. JCU IS6110 was detected within the green channel (MIC™ cycler). The absence of any amplification in NTC confirmed no contamination. 31 out of 38 isolates reacted with IS6110 TaqMan.

4.8.7.2 *rpoB detection in IS6110-rpoB duplex assay*

Twenty-four of 38 templates (i.e., 63%) reacted with the *rpoB* TaqMan assay (Figure 4.24). All positive controls were amplified. The absence of any amplification in NTC confirmed no contamination.
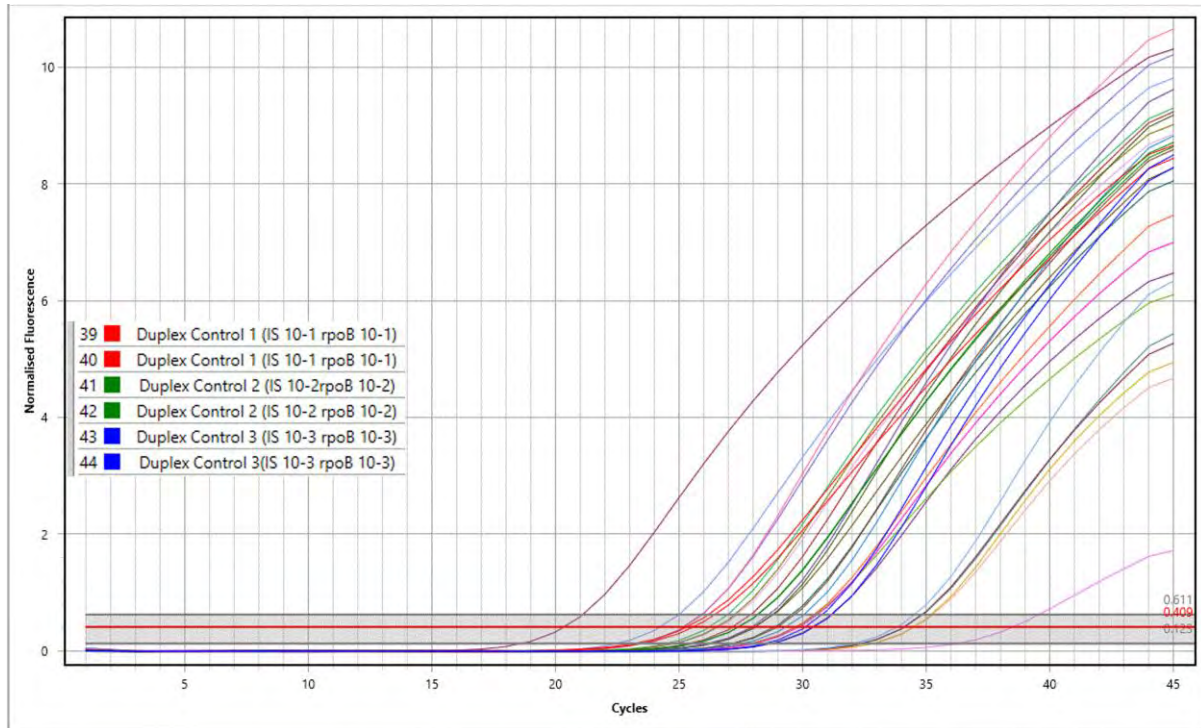
Figure 4.24: Duplex assay *rpoB* TaqMan assay amplification curves from 38 DNA samples. The *rpoB* amplification was detected in the yellow channel (MIC™ cycler). The absence of any reaction in NTC confirmed no contamination 24 out of 38 DNA samples reacted with *rpoB* TaqMan.

Table 4.15 shows that out of 38 DNA samples, 24 reacted in both the *rpoB* and IS6110 assays. Based on the theoretical interpretation of assays presented in section 4.7, this suggests that all these 24 DNA samples tested had *Mycobacterium* DNA present and all had sufficient DNA to carry out further analysis (Chapter 5 and 6). The other seven DNA samples reacted in the IS6110 TaqMan assay only. No amplification of the *rpoB* gene occurred in those samples. This result suggests that these DNA samples contained insufficient template for amplification of single copy genes (like *rpoB*) although multicopy genes like IS6110 could be detected. Hence, these samples were not analysed further in the experiments described in Chapters 5 and 6).

The remaining seven of 38 DNA samples didn't react in either of the assays. This suggested there was insufficient DNA in these samples. The other possibility of non-amplification in some could be the presence of inhibitors. An incorporation of internal control such as whale sperms etc could assist in determining such possibility. In addition, RNAseP test could be

used to check if there was any human DNA material was present in the template. As all bacteria possess the single copy *rpoB* gene, amplification would be expected for this assay even if mycobacterial DNA was not present. Failure of a reaction in the *rpoB* assay simply means the template used for testing was devoid of DNA.

4.9   Application of the published *senX3-regX3* assay on a panel of DNA samples

Out of the 38 DNA samples, 25 reacted (65%) (Table 4.15). All six DNA isolated from the supernatant of MGIT™ tubes had their target gene amplified, as expected.

Table 4.15 demonstrates the overall results of the designed duplex assay with the *sensX-regX3* assay (Broccolo *et al.*, 2003). 24 samples out of total 38 DNA (63%) reacted in the duplex assay. There were six DNA samples extracted from positive MGIT™ cultures (57-2,78-1,78-2, 96,105 and 530). All six Mtb isolates reacted in the duplex assay and the *senX3-regX3* assay. In addition, when DNA extracted from decontaminated sputum (JCU marked DNA samples such as JCU 565) was used as a template in the duplex assay, some positive results were found suggesting that this assay can also be applied to more crude clinical samples.

All *senX-regX3* positive are indicative of the presence of the *Mycobacterium* DNA. This is because our *in- silico* specificity testing of the *sensX-regX3 IR* gene primers/probe revealed that these tools have the potential to recognise other mycobacterial and related species. The developer of this published assay claimed that it is Mtb specific (Broccolo *et al.*, 2003). Further studies will be required to determine whether these positive samples truly contain MTBC organisms.

Seven DNA samples had no amplification with any of the assays. This ruled out the possibility of Mtb DNA in any of these samples. Alternatively, the samples were devoid of any DNA. The DNA of samples JCU 567, 578,598,609, 611 and 670 reacted with the JCU IS6110 assay only without amplification in the other two assays. This could be due to the presence of multiple copies of the IS6110 sequence and the low quality and amount of the

DNA template. However, it is still unclear whether those seven templates contain Mtb DNA or DNA of any species of the *Mycobacterium* genus family. Further DNA extraction of those samples might be helpful.

The DNA of the 297DS1 reacted with JCU IS6110 and the *senX3-regX3* assay. This was indicative of *Mycobacterium* DNA present in the template. The absence of reaction with the *rpoB* assay suggests that this sample doesn't warrant further testing due to inferior quality and low amounts of DNA. Except for one sample, JCU 594 (38.47 C_t value), all the DNA had C_t value indicative of good quality/yield of DNA that in turn was suggestive of the fact that these DNA were suitable candidates to conduct further investigation in Chapters 5 and 6.

Broccolo *et al.* (2003) reported that the *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was specific for MTBC and this assay was included in the current study to provide definitive confirmation of Mtb. However, the results presented in this chapter question the specificity of this *senX3-regX3* IR gene assay and challenge the published data. Our *in-silico* analyses highlight that it may recognise a range of bacteria (Appendix 9). A thorough investigation of the specificity of the published *senX3-regX3* IR assay's (Broccolo *et al.*, 2003) against other *Mycobacterium* species is needed and further studies are required in find an assay that is truly MTBC specific. The designed duplex in addition to such a confirmative assay(s) could be used in combination with the clinical signs and symptoms and microscopy to aid with TB diagnosis.

Table 4.15: Results of JCU IS6110-*rpoB* duplex assay and *senX3-regX3* IR assay from 38 DNA samples

DNA name	Type of sample	Number of strains reacting with TaqMan			
		JCU IS6110	*senX3-regX3 IR*	*rpoB*	C_t value associated with the *rpoB* TaqMan
248 DS 1	DNA extracted from archived sputum smear slides	+	+	+	30.37
361 DS 1	DNA extracted from archived sputum smear slides	+	+	+	29.69
322 DS 1	DNA extracted from archived sputum smear slides	+	+	+	29.69
26 DS 1	DNA extracted from archived sputum smear slides	+	+	+	34.10
68 NaOH DS 2	DNA extracted from archived sputum smear slides	+	+	+	28.04
323 DS 1	DNA extracted from archived sputum smear slides	+	+	+	34.47
274 DS 2	DNA extracted from archived sputum smear slides	+	+	+	24.21
321 DS 1	DNA extracted from archived sputum smear slides	+	+	+	29.83
84 NaOH 2	DNA extracted from archived sputum smear slides	+	+	+	34.09
291 DS 1	DNA extracted from archived sputum smear slides	+	+	+	34.46
297 DS 1	DNA extracted from archived sputum smear slides	+	+	-	0
51 DS 1	DNA extracted from archived sputum smear slides	-	-	-	0

52 DS 1	DNA extracted from archived sputum smear slides	-	-	-	0
53 DS 1	DNA extracted from archived sputum smear slides	-	-	-	0
54 DS 1	DNA extracted from archived sputum smear slides	-	-	-	0
JCU 565	DNA extracted from sputum	+	+	+	26.58
JCU 567	DNA extracted from sputum	+	-	-	0
JCU 578	DNA extracted from sputum	+	-	-	0
JCU 582	DNA extracted from sputum	+	+	+	26.21
JCU 594	DNA extracted from sputum	+	+	+	38.47
JCU 598	DNA extracted from sputum	+	-	-	0
JCU 602	DNA extracted from sputum	+	+	+	25.16
JCU 609	DNA extracted from sputum	+	-	-	0
JCU 611	DNA extracted from sputum	+	-	-	0
JCU 624	DNA extracted from sputum	+	+	+	28.83
JCU 625	DNA extracted from sputum	+	+	+	20.32
JCU 632	DNA extracted from sputum	-	-	-	0
JCU 634	DNA extracted from sputum	-	-	-	0
JCU 648	DNA extracted from sputum	+	+	+	26.49
JCU 661	DNA extracted from sputum	+	+	+	33.80
JCU 667	DNA extracted from sputum	-	-	-	0
JCU 670	DNA extracted from sputum	+	-	-	0
57-2	MGIT™ culture supernatant	+	+	+	28.08
78-1	MGIT™ culture supernatant	+	+	+	27.15
78-2	MGIT™ culture supernatant	+	+	+	28.97
96	MGIT™ culture supernatant	+	+	+	27.92
105	MGIT™ culture supernatant	+	+	+	29.29
530	MGIT™ culture supernatant	+	+	+	25.20

4.10   Discussion

This objective of the current study was to establish in-house molecular tools to meet WHO ASSURED (affordable, sensitive, specific, user-friendly, rapid, robust, equipment free and deliverable) criteria (Mabey *et al.*, 2004) and to assist in the laboratory diagnosis of TB from rural PNG. Previous studies from our research group utilised a published TaqMan assay (Broccolo *et al.*, 2003) based on the multicopy IS6110 sequence. The current study evaluated this assay by *in silico* analysis with AlleleID® and the resultant low rating suggested that alternative assays could be developed. Therefore, an attempt was made to design a new set of IS6110 primers and probe that may show superior performance compared to the existing assay (Appendix 3).

An unacceptable rating of 0.6 to both the published primers suggested that the assay was suboptimal, whereas the probe rating of 65.4 was satisfactory. The melting temperature (T_m) of the forward and the reverse primer was found to be 68.4°C and 71.3°C respectively, while the probe had the highest temperature of 77.4°C. The recommended melting temperature for attaining optimal results is 60°C (Fraga *et al.*, 2008). This indicated that T_m of existing primers was not optimum and may result in secondary annealing. Furthermore, it was observed that the T_m of the reverse primer was close to the probe T_m. Ideally this should be higher than that of primers (preferably 10°C) so that the probe binds to the template prior to initiation of primer annealing (Mackay *et al.*, 2002) Also, the length of the oligonucleotide sequences were 17 nucleotides, which is below the recommended 18-22 nucleotide length (Breslauer *et al.*, 1986).

Experimentally, the R^2 (indicative of reproducibility of the assay) value of the existing assay was 0.9620 and any value less than 0.98 is indicative of an assay that requires further optimisation (Muller *et al.*, 2004). The standard curve equation for this existing assay (y= 38.87 to 3.36 log (x)) was like that previously published (y= 38.85 to 3.422 log (x)). The minor variation maybe was due to different PCR machines used (Broccolo *et al.*, 2003).

Although experimentally the probe appeared to work well and received a relatively higher score than the primer pair *in-silico* evaluation the diagnostic primer set was not great and

this was the main reason for not getting a perfect reaction in which at the very single cycle in which the DNA gets double giving a slope of value -3.32 (Mackay *et al.*, 2002). This suggested that a new primer set, and probe needed to be designed, with the potential to perform well practically and be able to achieve high scores when evaluated *in-silico* using any evaluation software.

Broccolo *et al.* (2003) used 11 taxonomically relevant organisms to evaluate the specificity of their IS6110 assay whereas seven members of the *Mycobacterium* genus used in this project. The current study used a broader range of clinically relevant organisms (30 versus 9) respectively (Broccolo *et al.*, 2003). Most of the organisms included in the current study were pathogens infecting the lower respiratory tract. Inclusion of all members of the MTBC members would be essential for future studies. Broccolo *et al.* (2003) reported that their *senX3-regX3 IR* assay was specific to the MTBC (Broccolo *et al.*, 2003). A thorough *in-silico* search was conducted in the current study, and it was found that *senX3-regX3 IR* sequence was not specific to MTBC. The sequence was found in a wide range of *Mycobacterium* species (Appendix 9). Hence, both the Broccolo IS6110 assay and the *senX3-regX3* IR assay can detect other members of the *Mycobacterium* genus.

The novel JCU IS6110 TaqMan assay, scored AlleleID® ratings of 62.6 and 73.6 ratings the forward and reverse primer respectively. The new probe had the highest rating of 87.9. Furthermore, the T_m of the primer set, and the probe were different (5.9°C for forward and 8°C for reverse) suggesting that these reagents may have superior performance in TaqMan assays.

Experimentally, the JCU IS6110 TaqMan assay had 100% efficiency with R^2 value of 0.9947, including with DNA samples from MGIT™ supernatants. The JCU IS6110 TaqMan assay appeared to be more sensitive than the published IS6110 assay. However, both assays were equally specific and reproducible. Both primers and probe were homologous to sequences from the *M. abscessus* and *M. intracellulare* genomes and were able to detect these organisms experimentally. However because the NTM is rare in PNG (Guernier *et al.*, 2017) the novel assay still has the potential to be used to identify Mtb particularly in combination with another more Mtb- or MTBC-specific assays. Therefore, the designed assays are still

valuable if they are further optimised to improve their specificity. This has been described below.

An ideal value of 0.9968 for R^2 was obtained for the *rpoB* TaqMan assay. However, the assay failed to attain optimal efficiency (acceptable range 90-110% (Bustin *et al.*, 2009)). The assay therefore requires further optimisation or replication using fresh PCR reagents. In summary, this assay was able to detect 147 target DNA copies/µL, was reproducible and has utility in assessment of the quality of DNA templates.

H37Rv genomic DNA was superior to plasmids when used in experiments for standard curves. Inconsistency in the C_t values of different dilutions was found for the plasmid controls with ten-fold plasmid dilutions showing variable C_t values rather than the expected ~3 C_t values between dilutions, despite multiple repeat experiments. Despite plasmid templates being heated at 65°C for 10 minutes prior to cycling, plasmid conformation/ supercoiling may have influenced these results. Future studies using linearised plasmids as controls are warranted.

The IS6110 sequence is a multiple copy sequence and hence, this limit of detection of this assay could not be determined directly due to the unknown copy number of IS6110 in the H37Rv strain used in this project. The *rpoB* gene is present in the genome as a single copy and the limit of analytical sensitivity for the *rpoB* TaqMan assay was at the 10^{-5} dilution of H37Rv stock gDNA, which is equivalent to 147 copies/µL.

The TaqMan duplex assay (JCU IS6110 and *rpoB* TaqMan) worked well. Out of 38 DNA samples, 24 samples reacted in both *rpoB* and IS6110 assay, while 7 DNA samples reacted in the JCU IS6110 TaqMan only. The remaining 7 samples didn't react in either assay. Six DNA samples of MGIT™ supernatants reacted in both assays. The published *senX3-regX3* IR assay was used in combination with the duplex assay to minimise misidentification of the Mtb (Broccolo *et al.*, 2003). These same 24 samples reacted in both the duplex and the Broccolo IS6110/ *senX3-regX3* IR assay. One DNA template was amplified by the JCU IS6110 and *senX3-regX3* IR assays but not the *rpoB* assay. This result suggested that the quality of the

DNA template was low. Acquisition of fresh genomic DNA from the crude sputum sample (if available and possible) would be required for retesting.

Based on the results presented above, the *rpoB* TaqMan still requires further improvement. The JCU IS6110 TaqMan as expected was superior to the published IS6110 assay, particularly on more crude samples (i.e., MGIT™ supernatants). Importantly, the incorporation of both JCU IS6110 and *rpoB* TaqMan assays in a duplex worked well. Broccolo *et al.* (2003) used *senX3-regX3* IR for increased MTBC specificity, however the results of the current study suggest that it is not MTBC-specific (Broccolo *et al.*, 2003). Incorporation of a third -MTBC-specific assay into this multiplex is the next important step. Therefore, it could be said that the designed assays (IS6110 and *rpoB*) were in accordance with the WHO ASSURED criteria.

This study used standard TaqMan probes. Other types of probes such as Minor Groove Binding (MGB) and Lock Nucleic Acid (LNA) are available. Comparative studies suggest that MGB and LNA are superior to standard TaqMan probes with higher sensitivity and specificity. Even a difference of a mere one nucleotide in the target sequence could be identified. Moreover, their ease to design and being shorter in sequence length than the standard probe may increase the performance of the assay described in this study (Buh Gasparic *et al.*, 2010).

Wada *et al*. (2004) designed eight MGB probes: six targeting the Mtb *rpoB* gene and the remaining two *katG* and *embB.* Taking advantage of the ability of MGB probes to differentiate even a single nucleotide change in a sequence, they designed five probes targeting inside an 81bp hotspot within the *rpoB* gene and one probe targeting the outside this hotspot region to serve as a TB control. Probes with specificity for codon numbers 315 and 306 of *katG* and *embB* respectively were also designed. This set of assays was able to detect genetic variations in the *rpoB*, *katG* and *embB* genes in DNA samples of 45 lab strains and 27 clinical samples derived from patients with PTB. This study reported the potential for using MGB probes and ΔC_t to detect drug-resistant *rpoB* and showed that wildtype strains had ΔC_t less than 3.5 whereas a difference of more than 6 suggested the presence of the mutation(s) in the target gene(s) (Wada *et al.*, 2004).

Ramirez *et al.* (2010) used LNA probes in addition to the HRM for the identification of multi-drug resistant TB. These authors reported that HRM alone was unable to detect all transversion type mutations (i.e., where purine is exchanged with purine and pyrimidine with pyrimidine). The isolates were analysed in a multiplexed reaction consisting of primers specific for the IS6110 and the hot spot of the *rpoB* gene from the MTBC. The appearance of in summary, the aim of this chapter was to determine whether clinical material from individuals with presumptive TB based on clinical diagnosis were infected with Mtb. This required the implementation of identification tool(s) to detect Mtb in these imported samples and two TaqMan assays based on the IS6110 and the *rpoB* genes were designed and evaluated.

The IS6110 assay aided in identifying the presence/absence of the *Mycobacterium* DNA. Almost all MTBC isolates have an IS6110 element, and it was interesting to note that there was no evidence of MTBC lacking the IS6110 sequence, as has been reported in neighbouring South-East Asian regions including Vietnam. The *rpoB* assay evaluated the overall DNA quality i.e., amplifiability. Future studies could also use MTBC-specific regions of the *rpoB* gene for in combination with IS6110 for diagnosis. Two peaks (IS6110 and *rpoB* fragments) while using HRM confirmed MTBC. The presence of only one peak (from *rpoB* gene) corresponded to NTM, due to the absence of IS6110 in NTM. This study also targeted *katG* and *inhA* genes (Ramirez *et al.*, 2010).

Therefore, a transition from using the standard TaqMan probes to more specific and sensitive MGB or LNA probes may improve assay performance especially in the case of JCU *rpoB* TaqMan. Moreover, based on the principle of the Wada *et al*. study, the identification of mutation(s) associated with drug-resistance based on the differences in C_t values in association with another assay may provide more conclusive results for MTBC vs. NTM. A duplex or multiplex assay targeting at least two or more genes would be particularly informative.

4.11   Conclusion

In summary, the aim of the studies described in this chapter was to use a stepwise process incorporating TaqMan primer and probe design and validation to develop and utilise qPCR assays that may have utility in TB diagnosis. A further aim was to use these tools to determine whether clinical material from individuals with presumptive TB based on clinical diagnosis were infected with Mtb. Two TaqMan assays based on the IS6110 and the *rpoB* genes were designed and evaluated. The IS6110 assay aided in identifying the presence/absence of the *Mycobacterium* DNA. Almost all MTBC isolates have an IS6110 element, and it was interesting to note that there was no evidence of MTBC lacking the IS6110 sequence, as has been reported in neighbouring South-East Asian regions including Vietnam. We also found that the *senX3-regX3* IR gene is not specific to MTBC, and caution must be used in relying on this assay alone for TB diagnosis.

Our *rpoB* assay assessed the overall DNA quality i.e., amplifiability in samples. It is widely known that DNA template quality influences diagnostic qPCR results and hence it is essential to know that clinical material is appropriate for analysis and minimise false-negative molecular TB diagnosis. Future studies could also include MTBC or drug-resistance determining regions of the *rpoB* or other genes in the IS6110/ generic *rpoB* multiplex. Implementation of such an assay in a resource-limited setting like PNG would be beneficial for obtaining prompt results so that the suspected TB patient(s) could commence treatment. This type of real-time multiplexed platform has the benefit of availability of real-time PCR machines and reagents at competitive price and the flexibility to change assays relatively easily compared to standalone self-contained platforms like GenXpert and others.

CHAPTER 5

High resolution melt analysis for *Mycobacterium* species identification

5.1 Introduction

The studies presented in Chapter 4 revealed that additional molecular tools are required to differentiate *Mycobacterium* species and furthermore to characterise the Mtb strains as drug susceptible versus drug resistant. In low-resource settings, cost, convenience, and fit-for-purpose considerations are essential for sustainable implementation of any diagnostic and strain characterisation assays.

DNA sequencing is now widely used in high-resource regions for bacterial identification and drug resistance characterisation. However, this technology is unavailable in PNG due to its prohibitive cost. Moreover, the template preparation requirements and transport to overseas sequencing facilities further adds to the cost, with cost per base high for Sanger sequencing (Morey *et al.*, 2013). A cost-effective assay(s) that is implementable in resource poor settings as an alternative to sequencing is required (Mabey *et al.*, 2004)

Molecular based tools that can identify genetic variations in DNA could potentially be used in these settings to differentiate between different strains and species. Moreover, these are preferred over culture for their prompt results (within 1-2 days). Techniques such as molecular line probe assays, Genotype MTBDR/MTBDR*plus* and Genotype MTBDR*sl*, fully automated Xpert and Xpert Ultra are recommended by the WHO to identify and characterise the Mtb. Even though they all have high sensitivity and specificity, they target a limited number of mutations in the *rpoB* gene. Moreover, the high cost of implementation and sustainability of these platforms in resources limited countries like PNG currently limits their long term use (Bentaleb *et al.*, 2017).

Several studies have claimed that High Resolution Melt analysis (HRM) may have utility as a molecular based identification tool for characterisation of drug susceptibility and resistance. It is a powerful method for SNP genotyping, mutation scanning in DNA samples. It is cost-effective and could be used as a substitute to the expensive commercial diagnostics (Pietzka *et al.*, 2009; Vossen *et al.*, 2009; Ramirez *et al.*, 2010; Wang *et al.*, 2011; Lee *et al.*, 2012; Nagai *et al.*, 2013; Pholwat *et al.*, 2014; Singh *et al.*, 2014; Galarza *et al.*, 2016; Anthwal *et al.*, 2017; Bentaleb *et al.*, 2017; Sharma *et al.*, 2017). HRM is a closed tube post PCR method to identify any difference in the PCR product by monitoring changes in the melting behaviour of DNA duplex upon heating the DNA sample (Arthofer *et al.*, 2011). Intercalating dyes such as SYBR Green or LCGreen incorporate into the amplicon during the PCR reaction and fluorescence increases with successive PCR cycles. During post-PCR HRM analysis, amplicon melting across a range of temperatures results in decreased fluorescence with the change in the fluorescence dependent upon the length and the composition of the gene (Druml and Cichna-Markl, 2014).

A genetic variation in a homozygous sequence alters the T_m of the sequence. A transitional change from C to A or G to T and A to T or T to A in the sequence changes the T_m in an approximate range of 0.8 to 1.4 °C. However, the T_m change is difficult to identify if the base pairs exchange their position moving to the complementary strand (C to G or A to T) while remaining bonded (Taylor, 2009). It is still possible with a robust assay under strict conditions. This technique has been used to discriminate different bacteria and species, as different organisms and different species of the same bacterial genera may potentially have substantial genetic variations that could be measured during DNA melt curve analysis.

The aim of the studies described in this Chapter is to determine if HRM can be used to discriminate *Mycobacterium* species.

5.2 Materials and Methods

5.2.1 DNA samples used

A set of 40 different genomic DNA samples isolated from 40 different bacterial organisms was also used to test the specificity of the developed assays and these are described in Chapter 3, section 3.3, and Appendix 15. The results of TaqMan assays using these samples are described in Chapter 4, Table 4.13.

5.2.2 Primer design

HRM analysis requires an amplicon(s) for analysis and therefore, three different genes were selected for evaluation: 16S rRNA gene, *rpoB* gene and *katG* gene. Although the ideal size of the target gene for the HRM is 80-100bp, target genes of variable lengths (147-312 bp) were chosen in this study. It was done to cover the conserved area (in 16S assay) and to study the gene sections that harbour frequent mutations in the *rpoB* and *katG* genes.

A primer pair specific for each gene were designed using dedicated computer programs. All primer sets were designed in a way that they all were more biased towards the Mtb identification. This means they all were constructed so that they could target the Mtb DNA. For each primer pair, relevant published target gene sequences were acquired from the National Centre for Biotechnology Information (NCBI) nucleotide database, GenBank®. These published gene sequences were then aligned using the Geneious software. Muscle option was opted to carry out to multiply align those downloaded sequences.

GeneDoc program was further used to view and test the annealing capacity of the designed primer strings on the aligned published gene sequences specific to them. All the necessary recommendations to make robust primers were taken into consideration while designing. All the designed primer pairs were able to amplify their specific target region of varying length.

An *in-silico* specificity testing of all three designed primer sets was carried out. Only highly specific primer sequences were selected. The full description is in Appendix 10. Primers

sequences were purchased from Macrogen (Seoul, South Korea). Each freeze-dried tube of primer was reconstituted in nuclease-free water to produce a stock solution of 100-µM concentration. The working stock was further prepared by diluting the stock solution to the concentration of 10-µM prior to use for PCR. All the solutions were stored at -80°C.

The mastermix used in this study was GoTaq® qPCR (Promega A # 6002). This mix already contains the BRYT Green, a fluorescent dye that binds to DNA duplex leading to fluorescence (Source: https://us.vwr.com/store/product/17692499/gotaq-qpcr-master-mix-promega).

5.2.3   16S rRNA HRM assay

The rationale for choosing the 16S rRNA gene was: (1) to check the DNA quality of the templates being tested for further analysis; DNA from all bacteria should be detected and if not detected template was considered unsuitable for other assays. This had a similar utility to the *rpoB* TaqMan assay described in Chapter 4 and (2) to determine whether it could be used to discriminate different bacteria using the HRM platform.

The panel of 40 genomic DNA samples mentioned in section 5.2.1 were first subjected to molecular testing to verify its source organism to its species level through sequencing. For this, published 16SrRNA universal primers (sourced from http://lutzonilab.org/16s-ribosomal-dna/) capable of amplifying 1.39kb long 16SrRNA gene were used for this purpose. An amplicon of 1.39 kb length was sufficient to track its source organism to its species level. The details of those primers are provided in Appendix 10.

Further, the 1.3kb long amplicons of those 40 DNA samples tested were further used as a template to determine their quality using the designed 16S rRNA assay (Table 5.1). A 312bp long segment of the 16S rRNA gene was the target site of the designed primer set. The primer sequence detail is shown in Table 5.1.

Table 5.1 Primer set details of 16SrRNA HRM assay

Primer name	Sequence details	Target region (312Kb)	Length of amplicon (bp)
RibRNA-312-F	5'-GGATTAGATACCCTGGTA-3'	16S rRNA region	312
RibRNA -312-R	5'-GACTTAACCCAACATCTC-3'	16S rRNA region	312

The Bio molecular systems (BMS) MIC ™real-time qPCR cycler using 48 well tube rotor was used for qPCR and HRM. The primer set RibRNA-312-F and RibRNA-312-R was used. The DNA from the Mtb reference strain H37Rv and negative controls were included. DNA isolated from a panel of 40 organisms (described in Chapter 3, section 3.3, and Appendix 14) were tested in this assay. A 20μL reaction volume was used for all samples. That reaction mix and PCR cycling conditions are detailed in Appendix 11, Table 11.1. The data was analysed with the program that comes with the PCR cycler machine.

5.2.4 The *rpoB* HRM assay

The primer set specific to a 147 bp *rpoB* gene segment was designed (Table 5.2). The amplification site was the hotspot of the *rpoB* gene that harbours most of the drug-resistant based mutations. This primer pair had two utilities; (1) to differentiate between the *Mycobacteria* from other bacteria and to discriminate the different species of the *Mycobacterium* genus from one another; (2) the assay could be applied to the DNA tested in the previous chapter to further characterise strains with respect to drug susceptibility/resistance (this data is presented in Chapter 6).The *rpoB* gene was chosen because; (1) it is the most widely used gene to study Mtb.(2) More than 95 % of mutations conferring drug resistance in Mtb occur in this gene.

Table 5.2 Details of primer set for the *rpoB* HRM assay.

Primer name	Sequence	Target gene	Length of the amplicon (bp)
Mtb-*rpoB*-147-F	5'-ATCAAGGAGTTCTTCGGCACCAG-3'	*rpoB* hyper variable region	147
Mtb- *rpoB* -147-R	5'-CACGTCGCGGACCTCCAG-3'	*rpoB* hyper variable region	147

The primer set Mtb-*rpoB*-147-F and Mtb-*rpoB*-147-R was used for amplification. The DNA from the Mtb reference strain H37Rv and negative controls were included. DNA isolated from a panel of 40 organisms (described in Chapter 3, section 3.3, and Appendix 14) were tested in this assay. Each reaction was carried in a dedicated BMS 48 well PCR plate for the MIC™ qPCR cycler, each containing a 20μL reaction mixture. That reaction mix and PCR cycling conditions are described in Appendix 11, Table 11.2.

5.2.5 The *katG* HRM assay

The novel *katG* HRM assay amplifies a 130bp sequence of the Mtb *katG* gene (Table 5.3). The region selected for amplification is known to harbour multiple mutations associated with drug resistance. The *katG* is another important gene that is considered when Mtb is studied. The genetic variation(s) in this DNA confers drug resistance to the organism. Therefore, it was used to devise an assay that could first be verified for its organism differentiating ability and further provide insight to the drug susceptibility profile of the Mtb.

Table 5.3 Primer pair details of the *katG* HRM assay.

Primer name	Sequence	Target gene	Length of amplicon (bp)
Mtb-*katG*-130-F	5'-ATGGGCTTGGGCTGGAAG-3'	*KatG*	130
Mtb-*katG*-130-R	5'-TCGAGATCCTGTACGGCT-3'	*KatG*	130

The primer set Mtb-*katG*-130-F and Mtb-*katG*-130-R was used along with three positive controls dedicated to the *katG* gene only. Negative control in the form of NTC was also included to check on any contamination. DNA isolated from a panel of 40 organisms (described in Chapter 3, section 3.3, and Appendix 15) were tested in this assay. Assays were performed in 48 well PCR plates in a 20µL reaction mixture. The details of the reaction mix and PCR cycling is in Appendix 11, Table 11.3.

5.3   Results

5.3.1   Application of the developed assays on the panel for HRM

HRM assays (16S, *rpoB* and *katG*) were used to determine whether this technique has the potential to discriminate *Mycobacterium* species and clinically relevant organisms.

5.3.1.1 *The 16S rRNA gene assay*

As expected, all 40 DNA samples reacted (Figure 5.1). Each template had its own unique melt curve. Nearly all melt curves were sharp peaks and most of the curves overlapped. Because of this overlap, the assay did not have sufficient resolution to clearly differentiate between the different organisms. The members of MTBC and NTM were selected for further comparison (Figure 5.2).
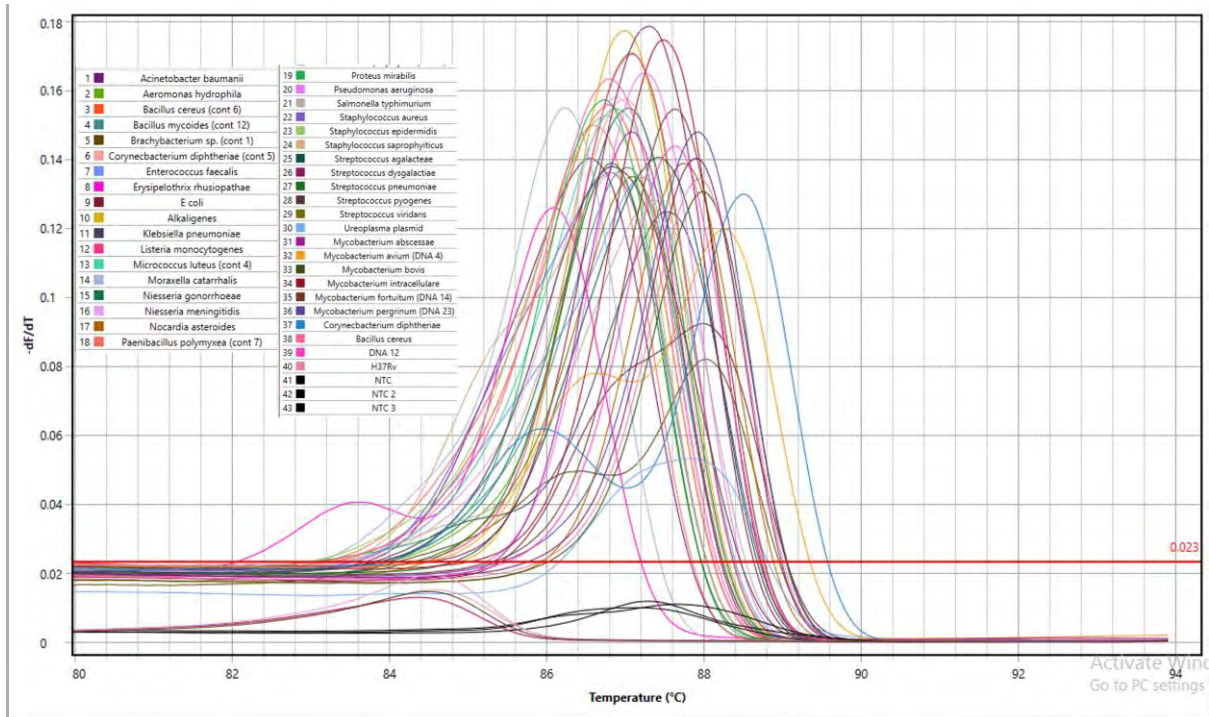
Figure 5.1: 16S HRM assay on the panel of DNA samples. The 312kb 16S rRNA gene sequence was able to be amplified in all the tested templates using the designed 16S specific assay. Some isolates had double peaks that might be due to low DNA or primer dimer formation.
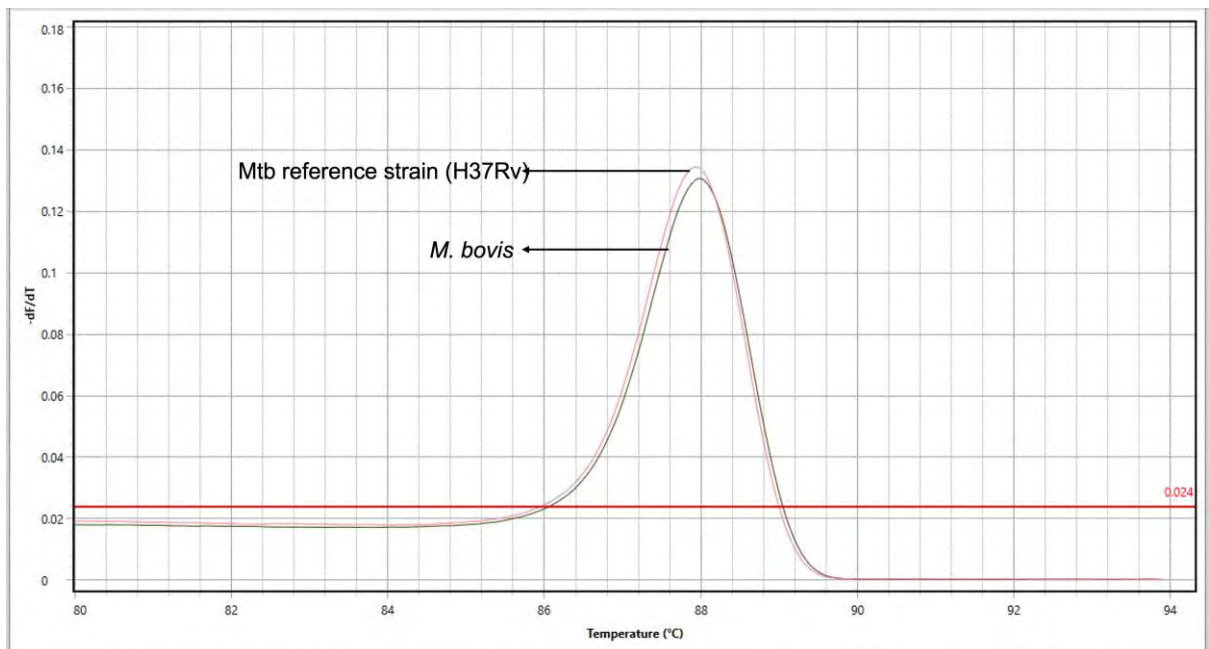
Figure 5.2: Comparison of H37Rv and *Mycobacterium bovis* melt curves. Both appeared identical. Their melting temperature was the same.

It is clear from Figure 5.2 that DNA melt curves for H37Rv and *M. bovis* were almost identical (T_m=~88°C). This suggested that the amplified regions of 16S rRNA genes are likely to have very similar, if not identical sequences. This result was expected, as Mtb H37RV and *M. bovis* are both part of the MTBC.

Figure 5.3 shows the comparison of other members of the *Mycobacterium* genus. All these amplicons had very sharp melt curves that differed in their respective melting temperatures and all were distinguishable from each other. However, there was still significant overlap between peaks and thus the resolution and discriminatory power of the 16S rRNA gene HRM assay was not considered optimal to be used as a diagnostic or species discrimination assay. The melting temperatures of *M. bovis* (87.98°C), *M. peregrinum* (87.91°C) and the H37Rv (87.93°C) amplicons were quite similar and respectively and their respective melt curves were almost identical (Figure 5.3). All other species had curves that were distinct from Mtb H37Rv.
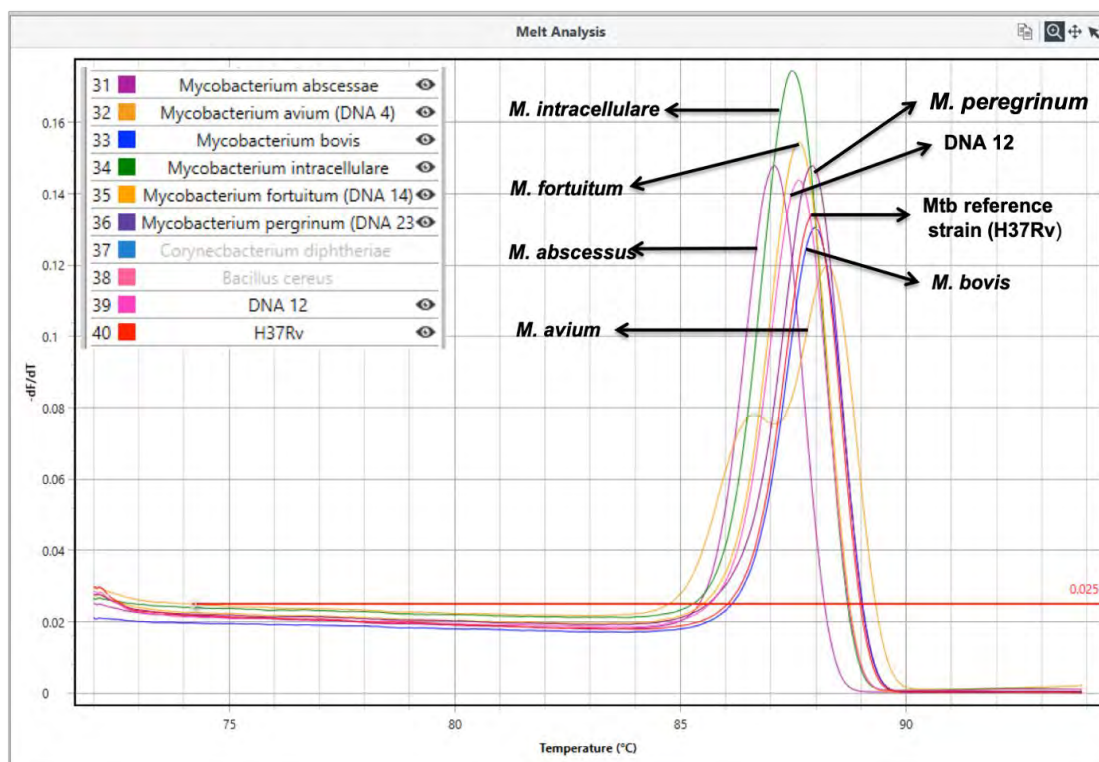
Figure 5.3: Comparison of 16S rRNA gene melt curves of the members of *Mycobacterium* genus. Each template had a slightly different melting temperature. The Mtb reference strain H37Rv is showed in red.

The T$_m$ of *M. abscessus* was 87.07°C which was a lower melting temperature than the other species (range 87.49-88.24°C). This 0.42°C difference may have been associated with changes at four positions (Figure 5.4) and was probably due to transitions, the substitution of a purine with purine and pyrimidine with a pyrimidine. A larger version of Figure 5.4 is presented in Appendix 12. Interestingly, the *M. bovis* 16S rRNA sequence appeared to harbour numerous transitions, however this resulted in only a 0.02°C difference in T$_m$ compared to H37Rv. These mutations would remain undetected if sequencing hadn't been done.

This led to the conclusion that the 16S rRNA assay may fail to detect these transitions. Mutations involving shift(s) of cytosine or guanine are more likely to be detected. Furthermore, *M. intracellulare* and *M. fortuitum* had quite different melt curves (more towards left side comparing to the H37Rv). The T$_m$ of the *M. intracellulare* and *M. fortuitum* was 87.49°C and 87.62°C when compared to Mtb H37Rv T$_m$ of 87.93°C, however sequencing revealed identical sequences to H37Rv. The variable DNA concentrations or the presence of salts or phenol could be the reason.
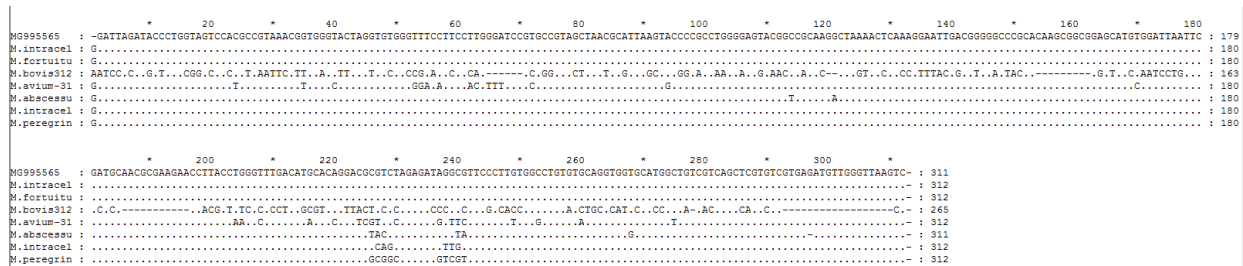
Figure 5.4: The 16S rRNA gene sequence alignment of different members of the *Mycobacterium* genus compared to Mtb H37Rv (MG995565), the top sequence in the figure.

5.3.1.2 *The rpoB assay*

As the 16S HRM assay, the panel of 40 DNA samples was analysed using the novel *rpoB* HRM assay. Of 40 samples, 38 reacted (Figure 5.5). No amplification of *Alcaligenes* and *Streptococcus pneumoniae* DNA was recorded. Surprisingly, nearly all, except the *Micrococcus luteus and M. fortuitum* had similar melt curves. This observation indicated that most organisms likely share extensive sequence homology in this *rpoB* gene amplicon. The differentiating potential of HRM was further analysed by selecting the members of *Mycobacterium* genus only.
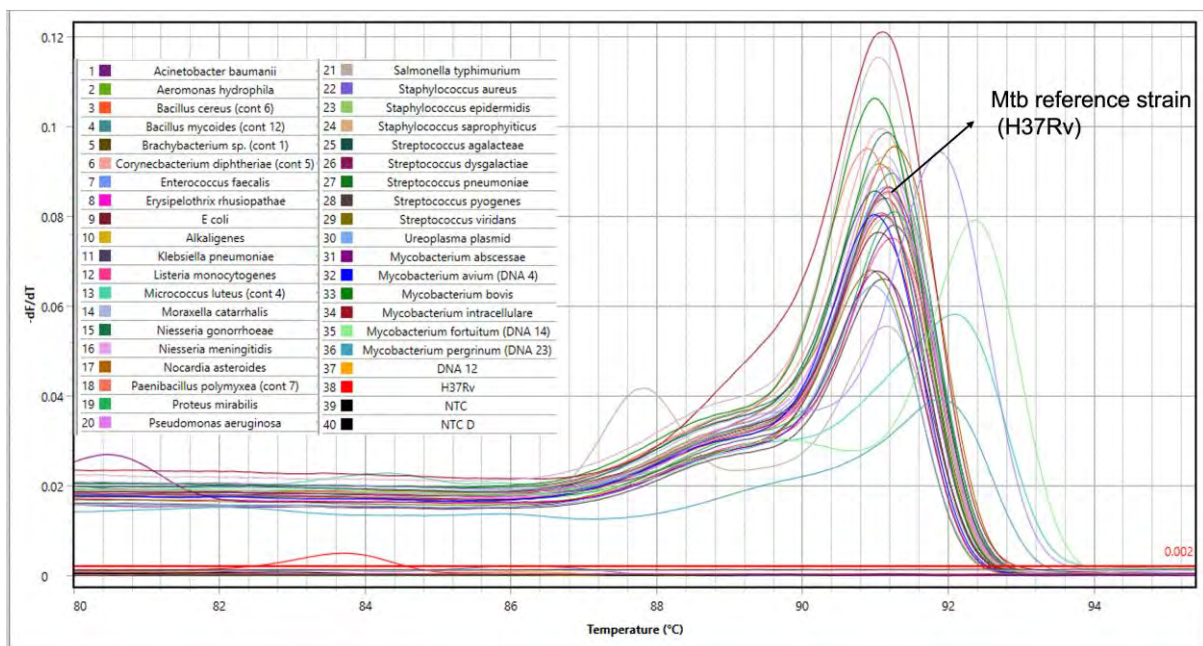
Figure 5.5: Melt curves of DNA from diverse bacteria following *rpoB* amplification. Most of the isolates had curves in proximity to each other. This indicated similarities in the *rpoB* amplicon across many different bacterial genera.

Figure 5.6 shows the melt curves of the *Mycobacterium* genus members used in this study. All seven were able to be differentiated. *M. intracellulare, M. bovis, and M. avium* had similar peaks. *M. fortuitum* had two peaks which may indicate a low-quality DNA sample. Because the protein encoded by the *rpoB* gene has a conserved function in all bacteria it is not surprising that similar melt curves were seen across this group of diverse organisms. It was therefore concluded that the *rpoB* HRM assay might not be particularly useful as a

diagnostic assay, despite some clear distinctions within the *Mycobacterium* genus (Figure 5.6).
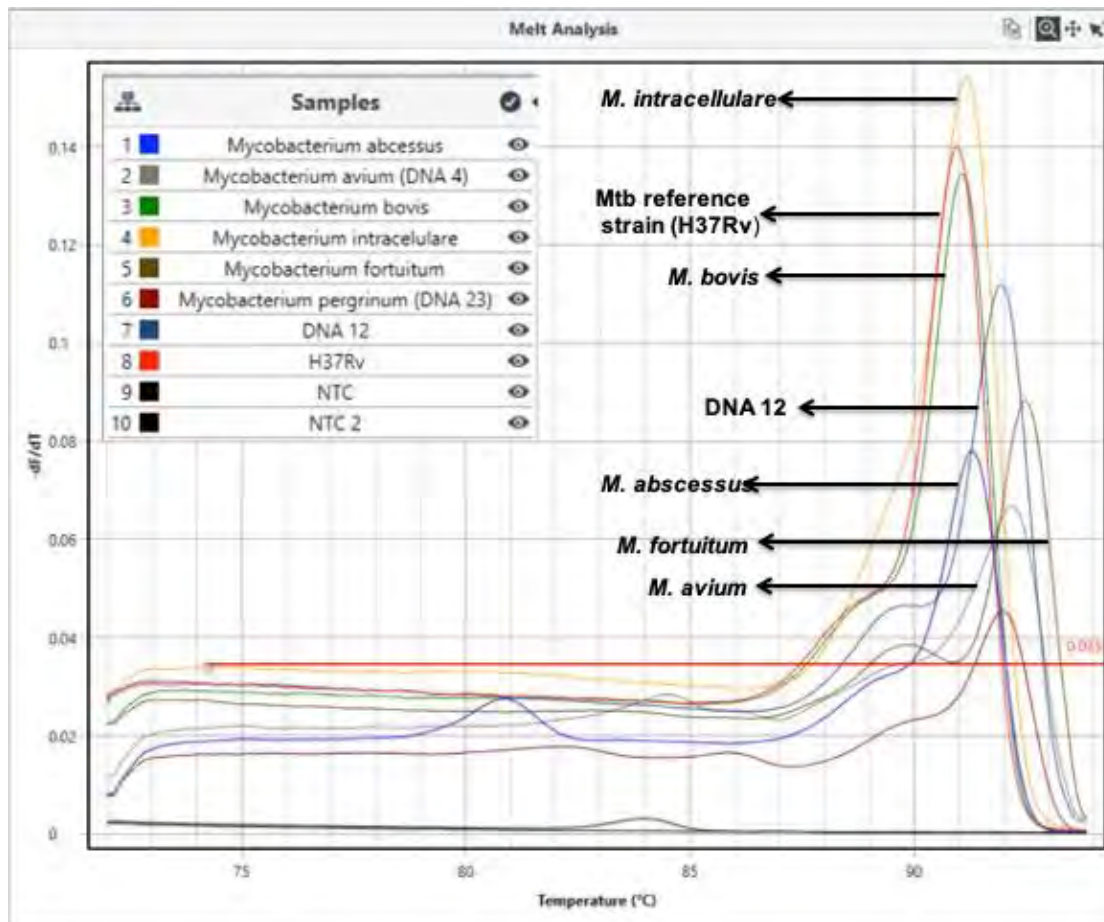
Figure 5.6: Melt curves of members of *Mycobacterium* genus in the *rpoB* HRM assay. DNA 12 was *M. intracellulare.*

Figure 5.7 shows *Mycobacterium* genus *rpoB* sequence alignment of published sequences acquired from the NCBI from the different species of used in this study. These all were the except the BC30 and M68 DNA. These two were incorporated for comparison, as both DNA samples had a genetic variation in them. BC30 had an SNP (CAC-CGC) around 74bp location. This SNP was unique and found in just one isolate so far.

```
                99
                      *        20         *        40         *        60         *        80         *        100        *        120        *        140
BX842574_x : ATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTG : 147
BC-30      : ..............................................................................................G.............................................. : 147
M68        : ...............................................................................................................T............................... : 147
JQ411539   : ................................................G..............C..TC..............C..C........C.............C.........C.....C...........T.....- : 146
CP029332   : ...............................TC...G.........................C.C..............C...........C..G..T.......C..G....G............................. : 147
AY544893   : ........................................................-------.............................................................................. : 140
```

Figure 5.7: Sequence alignment of published *rpoB* sequences. H37Rv (BX842574) and the
M. bovis (AY544893) were quite similar with some missing bases in the *M. bovis* sequence.
M. intracellulare (JQ411539) and *M. avium* (CP029332) had some differences when
compared to H37Rv. BC30 and M68 were the DNA sequences used just for comparison.
These two were DNA samples each with a different SNP that were not including in the
testing panel due to very limited availability. Hence their sequenced data was used for
comparison. An enlarged version of this figure is presented in Appendix 12.

The sequenced *M. intracellulare* (JQ411539) and *M. avium* (CP029332) segments differed in
their genetic composition at different locations when compared to the H37RV sequence
(BX842574). This may explain the different melt curves and temperatures. BC30 DNA
sample wasn't reanalysed. M68 was another sample's DNA possessing a second SNP (TCG-
TTG) isolated from PNG at approximately 89bp position. Figure 5.7 could be reviewed in
Appendix 12.

Because the protein encoded by the *rpoB* gene has a conserved function in all bacteria it is
not surprising that similar melt curves were seen across this group of diverse organisms. It
was therefore concluded that the *rpoB* HRM assay might not be particularly useful as a
diagnostic assay, despite some clear distinctions within the *Mycobacterium* genus (Figure
5.6).

5.3.1.3 *The katG assay*

Melt analysis of a panel of 40 different organisms' DNA was carried out using the *katG*
primer set. Out of 40 DNA samples, only 3 DNA reacted (Figure 5.8). The peaks of triplicate
samples of Mtb H37Rv were identical. The *M. bovis* amplicon had a similar melting
temperature to that of Mtb H37Rv.This was expected. However, *M. intracellulare* had a
quite distinct curve (to the right of the H37Rv), thereby revealing the potential of this

129

technique of differentiating organisms to their species level. Hence the *katG* HRM appeared to be quite specific for the *Mycobacterium* genus.
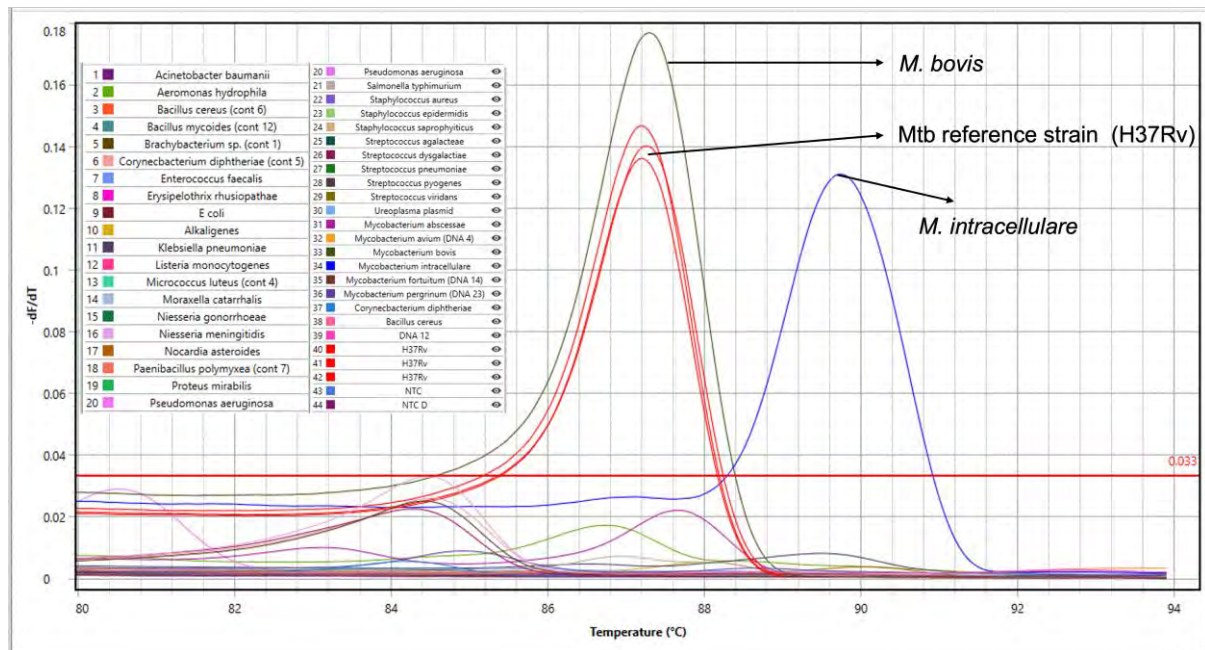

Figure 5.8: Melt analysis of the panel of 40 organisms using *katG* HRM primer set.

Further comparison (Figure 5.8) of the *Mycobacterium* genus members was done to compare their genetic similarity. The *M. intracellulare* and *M. abscessus* had quite distinct peaks that lied in extreme right and left to the positive controls (H37Rv and *M. bovis*) respectively. This means they both had their melting temperatures significantly different when compared to the melting temperature of the positive controls. Based on this substantial difference, it could be said that their genetic composition was not as the H37Rv control. This observation led to the conclusion that even the members of the *Mycobacterium* genus were genetically variant, possessing diversity in their *katG* gene.

They could easily be differentiated from each other to their species level based on their melting behaviour. While the H37Rv and *M. bovis* both being Mtb reference strains, were almost genetically identical with the same melting temperature as per expectations. The differentiation based on the melting temperature due to genetic composition might provide the insights to categorise the Mtb strains (wildtype versus DR).

5.4   Discussion

Successful TB treatment starts with an accurate diagnosis. WHO currently recommends a range of commercial molecular-based TB diagnostics and even in resource-limited countries effective, low cost medicines for drug-susceptible TB may be available. However, the availability and use of TB diagnostics and treatment resources differ markedly between and within countries with these obstacles resulting in ongoing transmission in these settings so the disease is still one of the main causes of mortality worldwide (Aye *et al.*, 2016). Misdiagnosis and the emergence of drug-resistant Mtb strains are the key factors that limit significant reductions in new TB incidence and mortality rates (WHO, 2021).

Each molecular diagnostic test has their own advantages and disadvantages. The availability of resources to establish and most importantly sustain these assays and their associated equipment can be particularly challenging in developing nations. Hence, rapid, and accurate profiling of Mtb genotypes in complex diagnostic samples remains a challenge for existing molecular detection technologies. HRM is a rapid, relatively inexpensive, closed-tube DNA sequence characterisation technique. Small differences in PCR melting curves in HRM analysis may lead to sensitive identification in a turnaround time fast enough to impact clinical decision making. Several studies have confirmed the utility of HRM to successfully identify and characterise the Mtb from sputum samples collected from the TB suspected/confirmed patients. HRM could be a relatively inexpensive alternative to the commercial based diagnostics suggested by WHO (Sharma *et al.*, 2015; Galarza *et al.*, 2016; Bentaleb *et al.*, 2017; Sharma *et al.*, 2017)

The objective of the experiments described in this chapter was to determine the potential of using HRM on different DNA sample types. Before carrying out HRM analysis on 31 DNA samples containing *Mycobacteria* DNA, assay specificity was evaluated using DNA from 40 different taxonomically and clinically related organisms. Three different assays were designed that targeted sequences within the 16S rRNA, *rpoB* and *katG* genes. The differing DNA melt curves reflected their genetic composition. All three assays were able to

differentiate between the tested organisms even to their species level. However, the 16S rRNA gene assay appeared to be less useful, as the curves of all 40 DNA samples tested overlapped each other resulting in poor resolution and discriminatory power. A similar trend was found with the *rpoB* assay although it was able to differentiate mycobacteria species. Tong and Giffard in their review on HRM mentioned its application of identifying organisms to their species-level (Tong and Giffard, 2012).

HRM analysis has been used previously to speciate mycobacteria. Perng *et al.* (Perng *et al.*, 2012) designed HRM assays based on the hypervariable regions of 16S and 16S-23S internal transcribed spacer region (ITS) for the identification of NTM. 134 isolates were confirmed as NTM by 16S rRNA gene sequencing and of these, 101 were accurately identified by their assay with a reported assay specificity of 96.4 to 100%. These authors also suggested the possibility of having different NTM species generating similar melting curves and closely related NTM species with different curves.

Yang *et al.* (2016) reported similar findings when testing a panel of 100 different organisms. Their assay targeted three hypervariable (V1, V3 and V6) regions of the 16S rRNA gene, each flanked by highly conserved sequences (Yang and Rui, 2016). In the current study, the DNA melt curves of H37Rv, *M. bovis* and *M. peregrinum* (Figure 5.3) were distinguishable from others, however overlapping melt curves of several *Mycobacterium* species potentially limits the utility of our 16S rRNA HRM assay. It was surprising that MTBC and even different NTM members produce similar DNA melt curves. This clearly reflects their genetic similarity. The H37Rv, *M. bovis* (MTBC members) and *M. peregrinum* (NTM member) all had very similar curves with melting temperatures of were 87.81°C, 87.99°C and 87.82°C respectively. This high degree of similarity between NTM and MTBC members was not expected based on analysis of published sequences. These observations led us to the conclusion that our 16S rRNA HRM might not be suitable as a diagnostic assay.

(Khosravi *et al.*, 2017b)) designed a HRM assay targeting the *rpoBC* locus for differential identification of *Mycobacterium* species. In that study, 98 clinical mycobacterial isolates that had been phenotypically and biochemically characterised were tested and a phylogenetic tree was constructed. The amplicons of different NTM species and the members of MTBC

(except Mtb) had the same or similar melting curves. Members of the *M. avium* complex, *M. abscessus* and *M. chelonae* that had quite different *rpoB* sequences had similar melt curves. Consistent with these published findings, the *rpoB* assay described in this chapter produced similar melt curves for *M. abscessus*, *M. intracellulare, M. bovis, and M. avium* DNA templates. The DNA melting temperatures of *M. abscessus, M. avium, M. intracellulare and M. bovis* were 91.03°C, 90.98°C, 91.09°C and 90.99°C respectively. A clear resolution of the *Mycobacterium* species melting curves led to a conclusion that the HRM based on the *rpoB* gene could be used for species differentiation.

The *katG* assay however was more specific to the MTBC (MtbH37Rv and *M. bovis)* (Figure 5.7). This assay clearly differentiated MTBC from other taxonomically and the clinically relevant organisms. Furthermore, this assay may have the potential to differentiate wildtype strains (H37Rv reference) and *katG* mutants (e.g., drug resistant).

CHAPTER 6

High-resolution melt analysis to aid in the identification of wildtype MTB versus non-wildtype MTB

6.1   Introduction

Papua New Guinea (PNG) has a high burden of drug-resistant TB (DR-TB) although the true prevalence is unclear (Aia *et al.*, 2016b). Several studies in different locations, have identified heterogeneity in prevalence across the country with the highest proportion of MDR-TB found in the Western Province (Gilpin *et al.*, 2008; Simpson *et al.*, 2011; Ballif *et al.*, 2012; Cross *et al.*, 2014b; Ley *et al.*, 2014a; Bainomugisa *et al.*, 2018; Diefenbach-Elstob *et al.*, 2019; Guernier-Cambert *et al.*, 2019).

Resistance-conferring mutations in the 81-base pair region of the rpoB gene (<u>R</u>ifampicin <u>r</u>esistance <u>d</u>etermining <u>r</u>egion [RRDR]) are associated with >90% of resistance to rifampicin. Similarly, any variation(s) in the Mtb *katG* gene reduces the activation of the front line antibiotic prodrug INH, while a genetic change in the promoter of *inhA* gene causes INH resistance (Bengtson *et al.*, 2017). Previous studies (Diefenbach-Elstob *et al.*, 2019; Guernier-Cambert *et al.*, 2019) identified DR-TB in samples collected at the Balimo District Hospital, with RIF resistance-associated mutations identified in 10% (5/50, 95% CI 4–21) of the MTB or MTBC/NTM samples where rpoB sequencing was obtained. Two different SNPs in the *rpoB* gene were identified in this study; *rpoB* S450L and *rpoB* I480V. Furthermore, novel katG mutations were found, although S315T, the most common INH resistance-associated *katG* mutation, was not identified.

It is essential that nucleic acid-based assays for detection of drug resistance in Mtb must be able to differentiate wildtype (drug susceptible) strains from any genetic variant conferring drug resistance. In the previous chapter (Chapter 5), HRM was used to differentiate organisms to their species level and the results presented suggested that both the *rpoB* and

the *katG* HRM assays may have further utility for detection of polymorphisms associated with drug resistance.

The aim of the studies presented in Chapter 6 is to use HRM as a triage tool to identify samples that have altered melt curves in Mtb *rpoB* and *katG* HRM assays and select these samples for further characterisation by DNA sequencing. DNA samples that reacted in the novel IS6110/ *senX3-regX3* IR duplex assay (Chapter 4) are then analysed using the HRM assays developed in Chapter 5.

6.2   Materials and Methods

6.2.1   Samples used in this study

The DNA samples that reacted in the duplex TaqMan assay described in Chapter 4 (i.e. novel IS6110 and published *senX3-regX3* IR assays (Broccolo *et al.*, 2003)) were selected for further testing. Reaction in this assay was indicative of the presence of the *Mycobacterium* DNA in the samples and amplification of the *rpoB* gene suggested template was of sufficient quality for further analysis. Of the 24 DNA samples that had *Mycobacterium* DNA, four DNA samples (JCU 594, JCU 602, JCU 624, and JCU 661) had insufficient template remaining, therefore 20 DNA templates were analysed in the studies described in this Chapter.

6.2.2   Design of synthetic oligonucleotide controls for *rpoB* and *katG* HRM assays

Synthetic oligonucleotides were designed as controls to allow the differentiation of wildtype and outlier sequences in HRM assays targeting the *rpoB* and *katG* genes. This type of control is convenient and cost-effective when compared to the genomic DNA preparation and plasmid construction methodologies typically used for controls in qPCR assays. Hence, it may have utility for routine use in resource-constrained settings.
For each target gene (*rpoB* and *katG*), a total of three oligos were designed: one based on wildtype sequence and two based on previously described SNPs for each gene. The *rpoB*

control oligos were prepared based on SNPs previously identified in sputum samples from TB patients from Balimo, PNG (Diefenbach-Elstob *et al.*, 2019). The control oligonucleotides were synthesised by Macrogen. Previous *katG* gene sequencing of DNA extracted from sputum samples from Balimo TB patients did not identify the most common, INH resistance-associated *katG* mutation S315T, despite being seen in other studies from PNG, including in the Western Province (Diefenbach-Elstob et al., 2019). Therefore, the sequences of the *katG* control oligos used in the current study were based on published reports of common *katG* SNPs.

6.2.2.1 *Generation of the rpoB controls*

Three different oligos representing wildtype rpoB sequence (H37Rv) and SNPs S450L and H445R were designed (Table 6.1). The *rpoB* hybrid primer set Mtb-*rpoB* -147H-F and Mtb-*rpoB* -147H-R primer pair was used as an extension primer set to extend the length of these oligos from 130bp to 147bp (Table 6.2 and Appendix 13). Macrogen synthesises oligos of up to 130bp length. The length extension was conducted to incorporate extra nucleotides that may harbour a genetic variation(s). The oligos sequence detail is in Table 6.1.

Table 6.1: Sequence detail of the *rpoB* oligos

Control name	Sequence detail
H37Rv (130 mer)	5'-CAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTG-3'
S450L (130 mer)	5'-CAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGGTTGACCCACAAGCGCCGACTGTTGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTG-3'
H445R (130 mer)	5'-CAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGGTTGACCCGCAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTG-3'

The oligos (Table 6.1) were used as templates for the primer pair (Mtb-*rpoB* -147H-F and Mtb-*rpoB* -147H-R). The resultant 147bp amplicons were further used as the templates to further generate sequences that could be used as controls in HRM analysis.

Table 6.2: Primer details of the *rpoB* assay

Primer name	Sequence	Target gene	Length of the amplicon (bp)
Mtb-*rpoB*-147-F	5'-ATCAAGGAGTTCTTCGGCACCAG-3'	*rpoB* hyper variable region	147
Mtb-*rpoB*-147-R	5'-CACGTCGCGGACCTCCAG-3'	*rpoB* hyper variable region	147

Hence, the above primer set (Mtb-*rpoB* -147-F and Mtb-*rpoB* -147-R) (Table 6.2) was designed to amplify three separate 147bp long *rpoB* sequences using three different synthesised oligos as a template. The resultant products were serially diluted to ensure that C_t values of between 15 and 20 cycles were achieved for all controls. The Biomolecular systems (BMS) MIC™ qPCR cycler was used for all experiments in 48 well PCR plates with 20μL reaction volumes as described in Table 6.3.

137

Table 6.3: Description of reaction mix for *rpoB* gene-based oligo controls

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	qPCR Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer Mtb-*rpoB*-147-F	0.8 µM	1.6
10 µM	Reverse Primer Mtb-*rpoB*-147-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total volume of one reaction mix			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds for 20 cycles with green channel. Melt condition: 72-97°C, 0.3°C/sec.

6.2.2.2 *Generation of katG controls*

The same protocol was followed as described in section 6.2.2.1. The most prevalent mutation found to occur in *katG* gene is AGC to ACC at codon 315, or Ser315Thr (S315T) substitution. Other mutations on the same codon (AGC) due to different nucleotide base substitution have been reported. Therefore, three oligos were used to generate these controls: one was wildtype having no SNP (*katG*-H37RV). Each of the other two controls (*katG*-JX303275 and *katG*-CP001642) had a different SNP, shown in red in Table 6.4. The sequence AGC at codon 315 in wildtype sequence *katG*-H37RV is underlined in black in Table 6.4. The canonical *katG* S315T AGC→ACC mutation is present in the control oligo *katG*-JX303275 and underlined in red. Therefore, it was named as S315T2. Another *katG* mutation, *katG* Ser315Gly AGC→GGC is present in the control oligo *katG*-CP001642 or

S315T1. No oligo sequence extension step was required. The Mtb-*katG*-130-F and Mtb-*katG*-130-R primer pair (Table 6.5) was used for the *katG* HRM assay.

Table 6.4: Sequence detail of the *katG* control oligos

katG-H37RV	ATGGGCTTGGGCTGGAAGAGCTCGTATGGCACCGGAACCGGTAAG GACGCGATCACC<u>AGC</u>GGCATCGAGGTCGTATGGACGAACACCCCG ACGAAATGGGACAACAGTTTCCTCGAGATCCTGTACGGCT
katG-JX303275 S315T2	ATGGGCTTGGGCTGGAAGAGCTCGTATGGCACCGGAACCGGTAAG GACGCGATCACCA<u>CC</u>GGCATCGAGGTCGTATGGACGAACACCCCG ACGAAATGGGACAACAGTTTCCTCGAGATCCTGTACGGCT
katG-CP001642 S315T1	ATGGGCTTGGGCTGGAAGAGCTCGTATGGCACCGGAACCGGTAAG GACGCGATCACC<u>G</u>GCGGCATCGAGGTCGTATGGACGAACACCCCG ACGAAATGGGACAACAGTTTCCTCGAGATCCTGTACGGCT

Table 6.5: Primer pair of the *katG* assay to analyse HRM

Primer name	Sequence	Target gene	Length of the amplicon (bp)
Mtb-*katG*-130-F	5'-ATGGGCTTGGGCTGGAAG-3'	*KatG*	130
Mtb-*katG*-130-R	5'-TCGAGATCCTGTACGGCT'3'	*KatG*	130

6.3   Application of designed assays on the selected DNA samples containing *Mycobacterium* DNA

The DNA samples that reacted in the duplex TaqMan assay described in Chapter 4 (i.e. novel IS6110 and published *senX3-regX3* IR assays (Broccolo *et al.*, 2003)) were analysed using these *rpoB* and *katG* gene HRM assays.

6.3.1 The *rpoB* assay

The Mtb-*rpoB* -147-F and Mtb-*rpoB* 147-R primer set was used for PCR reactions with the three positive controls (described above) and Mtb DNA-containing samples as templates (Table 6.6). No template controls were also used in each PCR run.

Table 6.6: Details of the reaction mix for *the rpoB* assay

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer Mtb- *rpoB* -147-F	0.8 µM	1.6
10 µM	Reverse Primer Mtb- Mtb- *rpoB* - 147-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total Volume			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds for 40 cycles with green channel. Melt condition: 72-95°C, 0.3°C/sec.

6.3.2 The *katG* assay

The primer set Mtb-*katG*-130-F and Mtb-*katG*-130-R primer set was used for PCR reactions with the three positive controls (described above) and Mtb DNA-containing samples as templates (Table 6.7). Duplicates of all controls and no template controls were included in each PCR run.

Table 6.7: Details of the *katG* assay mix components

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer Mtb-*katG*-130-F	0.8 µM	1.6
10 µM	Reverse Primer Mtb-*katG*-130-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total Volume			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds for 40 cycles with green channel. Melt condition: 72-95°C, 0.3°C/sec.

6.4   Verification of HRM results with sequencing

For both target genes, HRM analysis was used to identify DNA samples that had melt curves: (1) like wildtype control oligos (i.e., presumably contained wildtype sequence); (2) similar to SNP-containing control oligos or (3) different to all control oligos. Templates that were suggestive of containing mutant or outlier sequence by HRM were selected for sequencing. Sequencing was therefore used to confirm whether the HRM *rpoB* and *katG* assays could be used as an in-house mutation-screening tool for identification of mutations associated with antimicrobial resistance in Mtb.

6.5   Results

6.5.1   HRM assays for discrimination of wildtype and mutant sequences

6.5.1.1 *The rpoB assay*

All 20 DNA samples reacted in the *rpoB* assay, and each had a unique melt curve (Figure 6.1-
6.6). The melt curves of the three positive controls showed sharp peaks, as expected. The
medium filtering option was used for the sharper peaks. The melt curves for most DNA
samples were in the vicinity of the positive control curves.
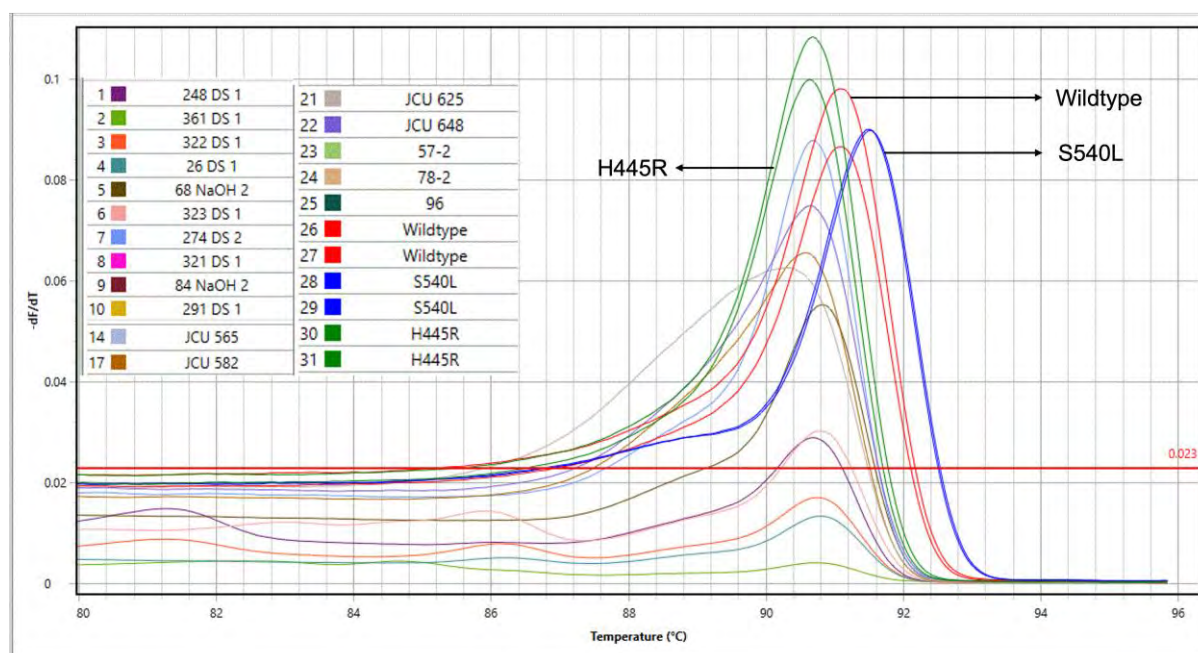
Figure 6.1: Melt curves of control oligos (wildtype and mutant) and DNA samples in the *rpoB*
HRM assay. Melt rates (y-axis) are represented as change in fluorescence units with
increasing temperatures (dF/dT).

Figures 6.1 and 6.2 show the melt curves and HRM analysis of all DNA samples and control
oligos on the same plots. In Figure 6.2 the *rpoB* wildtype (red curve) is the reference against
which the other two controls-*rpoB* S450L (blue) and *rpoB* H445R (green) were compared.
Figure 6.2a shows a difference plot normalised to wildtype control oligos, whereas Figure

6.2b shows a standard normalised melt curve, both across the 89.4-95 °C temperature range. Curves from most of the DNA samples were in the vicinity of the H445R control. DNA templates from the supernatant of some MGIT™ cultured tubes (57-2, 78-2, 96) had curves closer to the wildtype control (Figure 6.5 and 6.6). While the remaining two DNA samples had their curves near to the S450L control. For improved visualisation or curves and analysis, controls and selected samples are presented on separate plots (Figure 6.3-6.6).
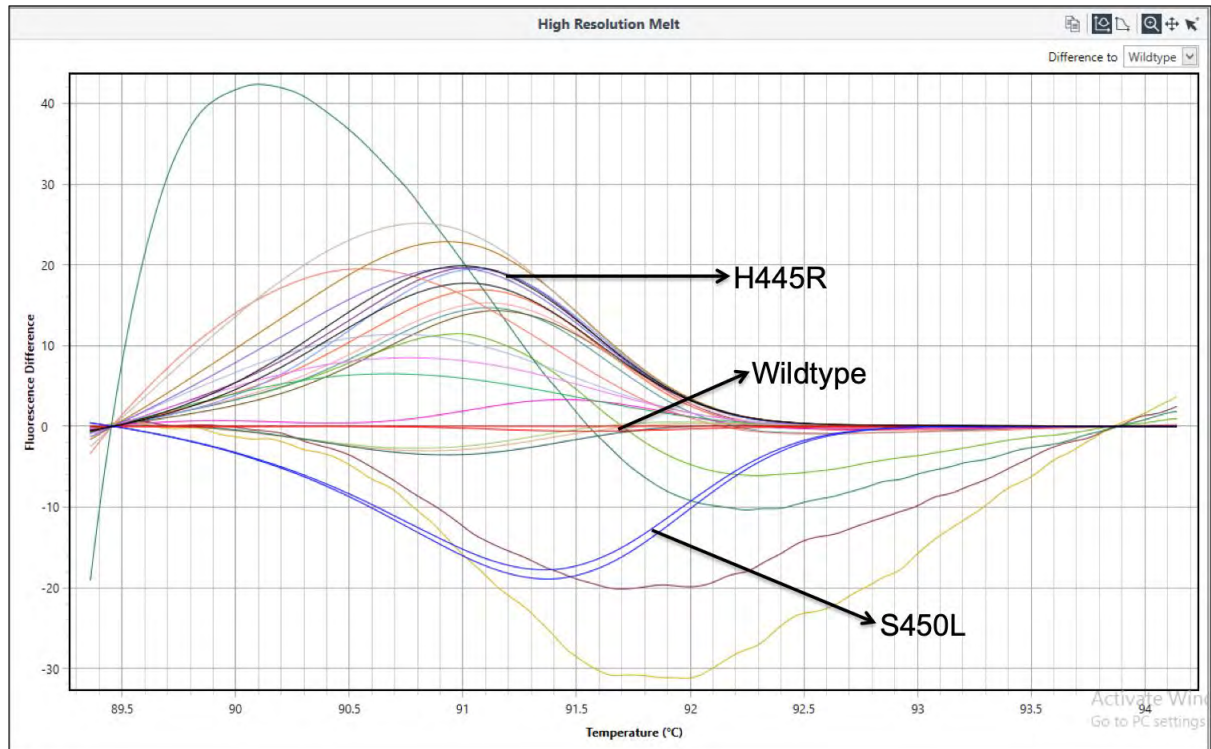
(a)
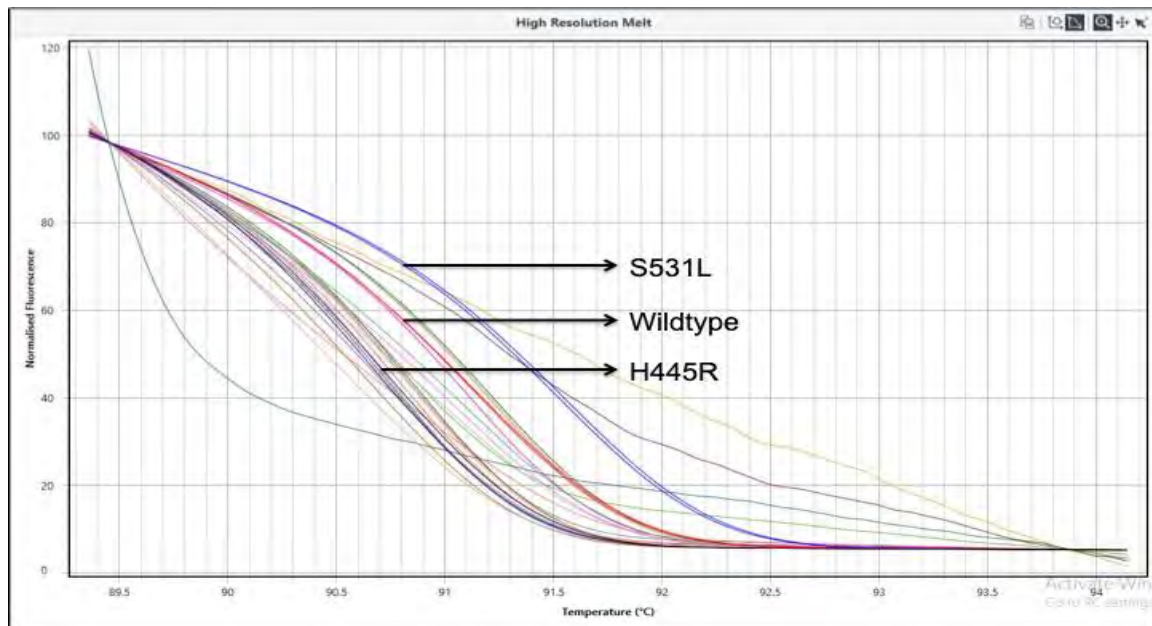
(b)

Figure 6.2: (a-b) HRM analysis of *rpoB* HRM assay on the selected DNA. There are two views (a) where Y-axis represents Fluorescence difference compared to wildtype; and (b) shows a standard normalised melt curve. Most of the DNA appeared closer to the H445R control. Few were near to the wildtype and two, 84 NaOH 2 and 291 DS 1 were near to the S450L control.

It is clear from the Figure 6.3 and 6.4 that all positive controls (duplicates) used had different melt curves. The H445R curve (green) was shifted to the left of the wildtype control (red), while the S450L (blue) shifted to the right. Hence, these single nucleotide changes had indeed influenced DNA melting temperatures.
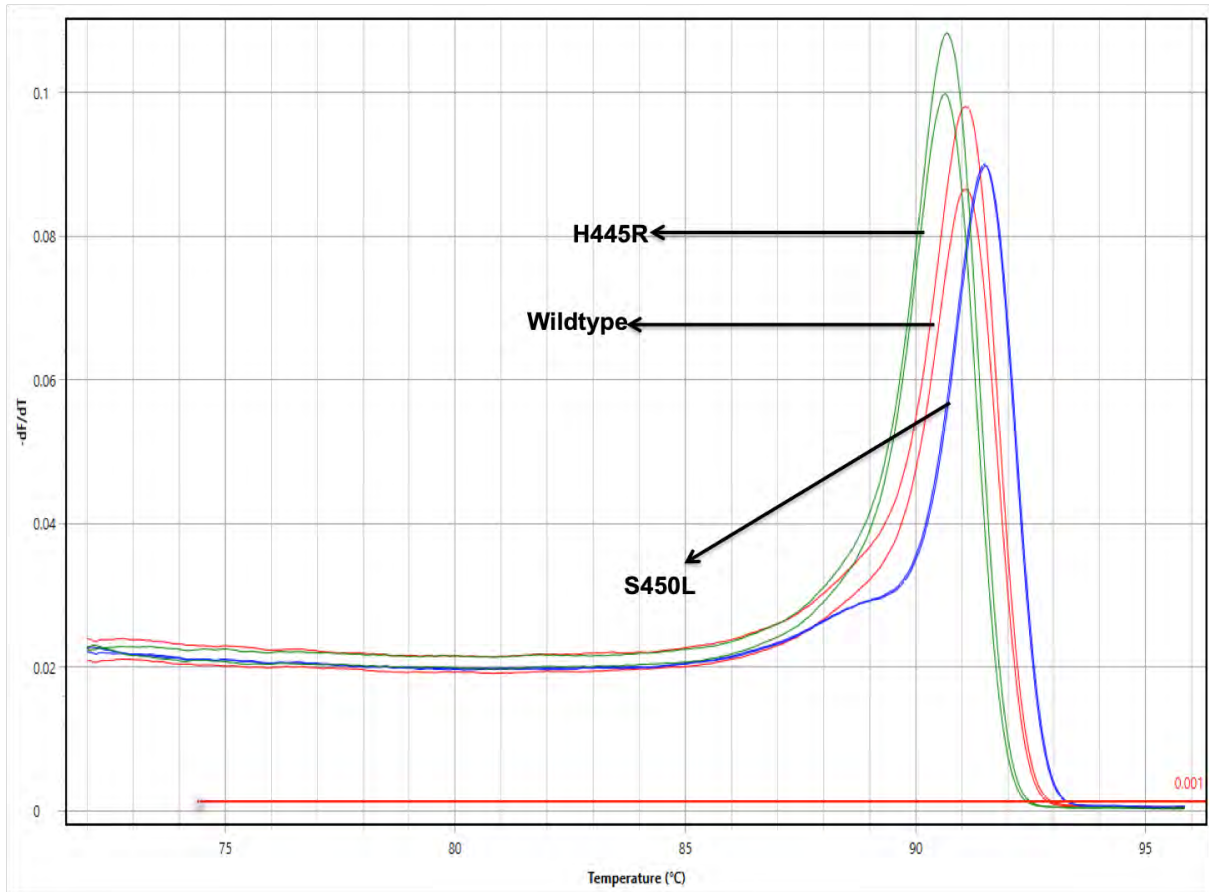
Figure 6.3: Melt curves of the positive controls used in the *rpoB* HRM assay. All three were able to be discriminated based on their peak melting temperatures.
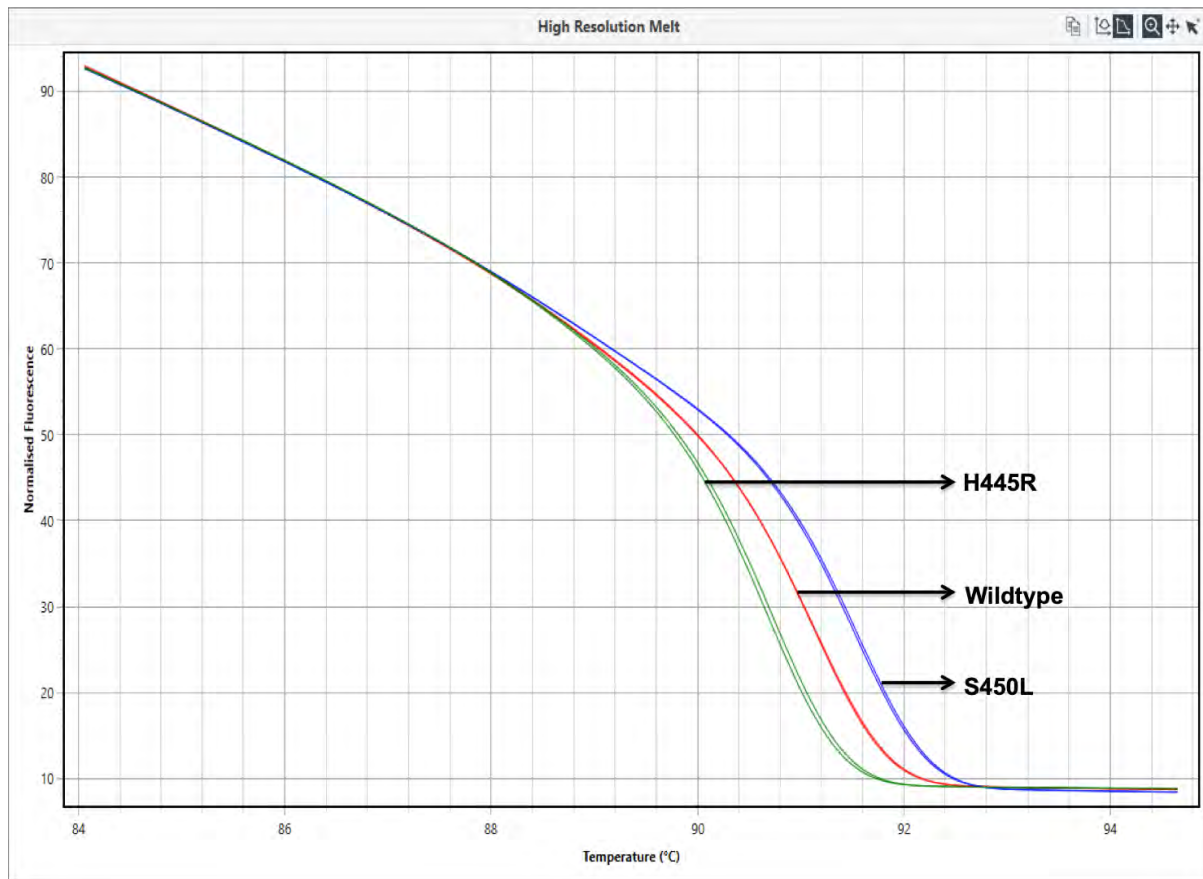
Figure 6.4: Normalised HRM of the three positive controls used in the *rpoB* HRM assay. All three were able to be discriminated based on their peak melting temperatures.

Figures 6.5 and 6.6 show the melting behaviour of the DNA samples from the MGIT™ tubes (57-2, 78-2 and 96) compared to the wildtype and mutant control sequences. These samples are comparable to the wildtype H37Rv reference strain control with identical melting temperatures. This result suggested that these samples were also *rpoB* wildtype i.e., susceptible. However, sequencing information was required to confirm this finding.
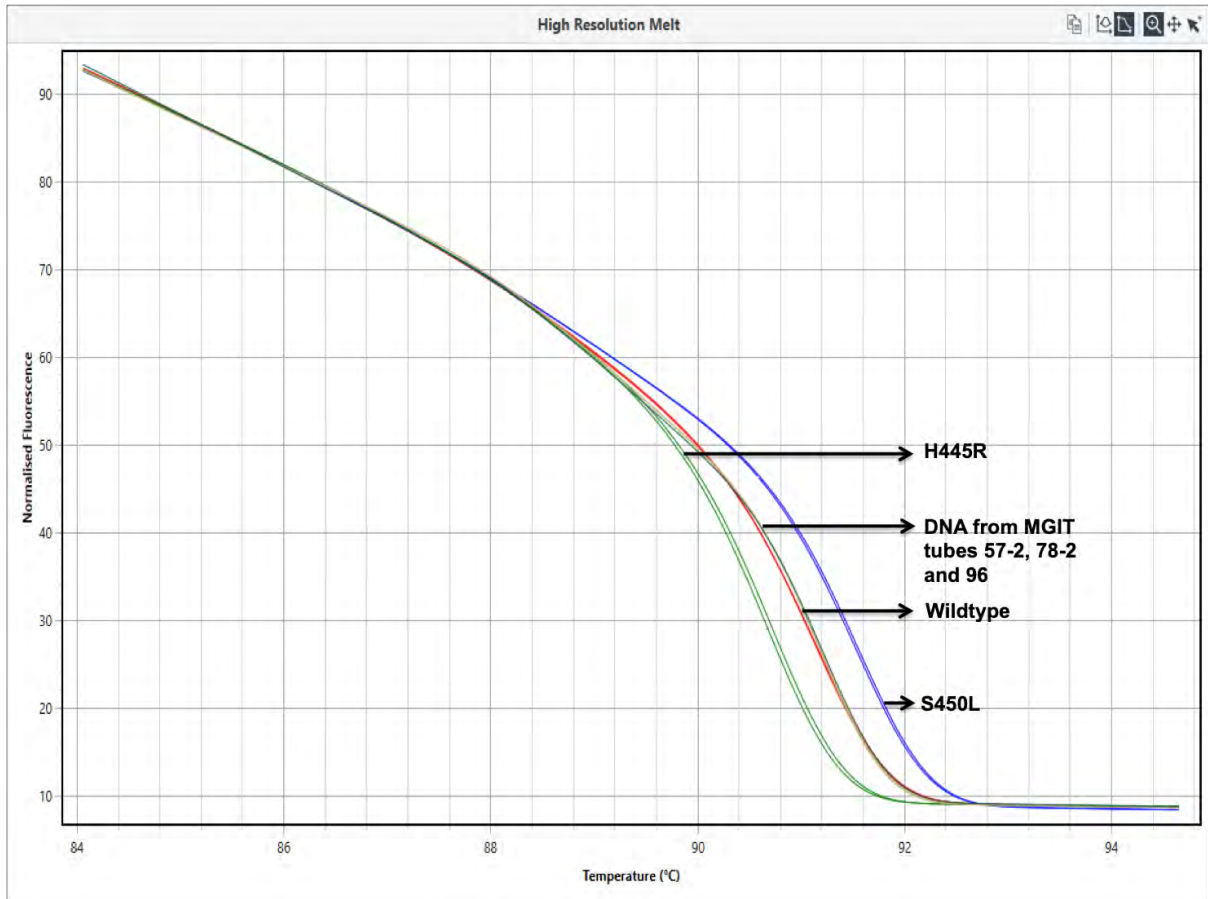
Figure 6.5: Normalised HRM of DNA from three MGIT™ cultures (57-2, 78-2 and 96) samples with the positive controls. All melt curves were closer to the wildtype control than mutant control curves, indicating that all may contain wildtype *rpoB* sequence.
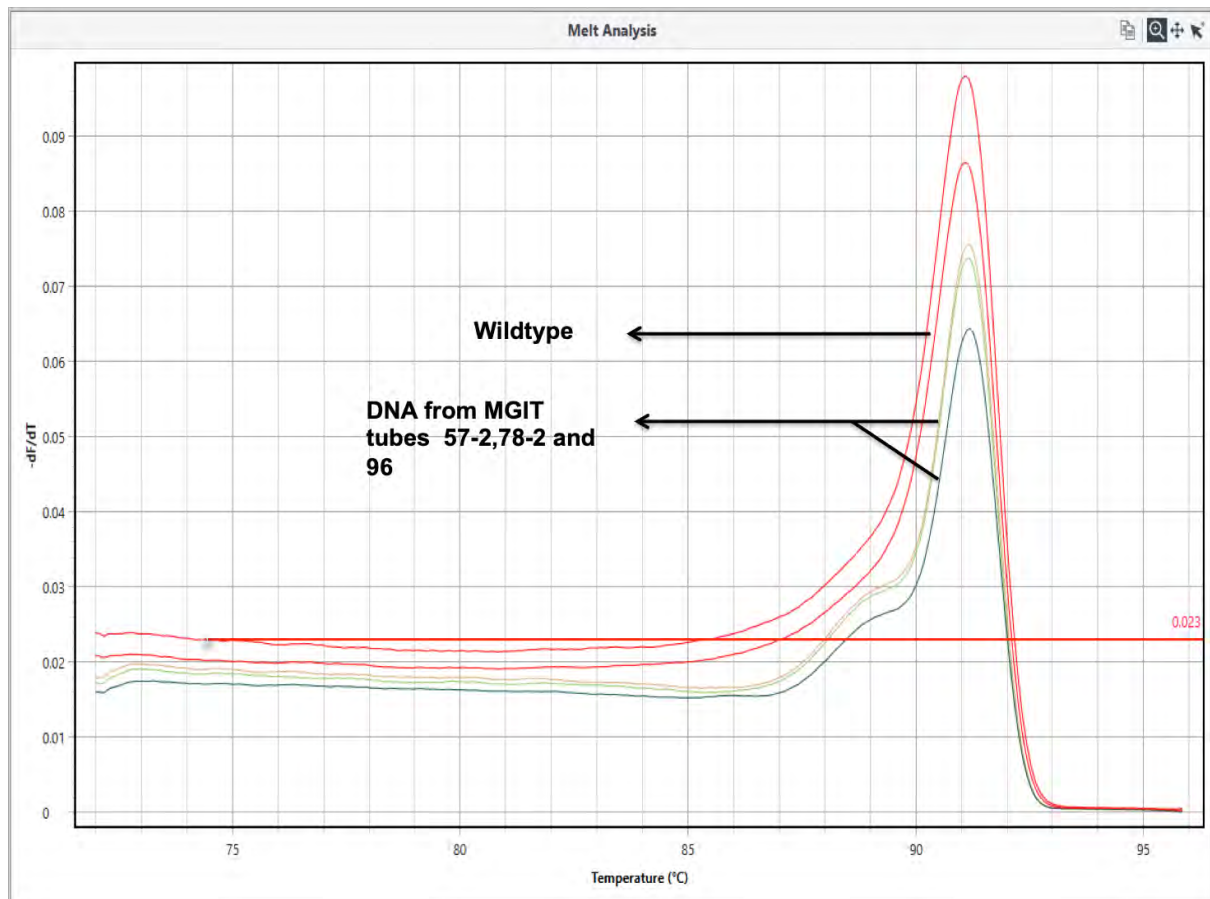
Figure 6.6: Melt curve of DNA from three MGIT™ cultures compared with the H37Rv curve. They all have identical melting temperature suggesting wildtype genotype.

Single nucleotide changes differentiated the three oligo controls and therefore it was anticipated that their respective T_m values would be quite similar. MIC™ qPCR HRM analysis software was used to allocate DNA samples to wildtype, H445R or S450L genotypes using the control oligos as reference genotypes. The results of this prediction are presented in Table 6.8. From Table 6.8, of 20 DNA samples, 12 (60%) were mutants. The remaining 8 (40%) were predicted to be wildtype. The confidence level for all DNA samples was 93-100%.

Therefore, it was concluded that the *rpoB* HRM assay was able to classify the DNA from MGIT™ tubes supernatant as wildtypes. The positive controls designed and used for this assay worked well. The next step in this analysis was to validate these HRM results using sequence data (section 6.6 below).

Table 6.8: Predictive *rpoB* genotypes based on HRM results.

Sample number	Wildtype	H445R	S450L
57-2	+		
78-1	+		
78-2	+		
96	+		
105	+		
530	+		
321DS1	+	-	-
84 NaOH 2	-	-	+
291 DS 1	-	-	+
26 DS 1	-	+	-
68 NaOH DS 1	-	+	-
323 DS 1	-	+	-
274 DS 2	-	+	-
248 DS 1	-	+	-
361 DS 1	-	+	-
322 DS 1	-	+	-
JCU 582	-	+	-
JCU 625	-	+	-
JCU 648	-	+	-
JCU 565	+	-	-

DNA from MGIT™ culture

DNA from slides

DNA from sputum

6.5.2   The *katG* assay

The same 20 DNA samples that were tested with the *rpoB* assay were also analysed using the *katG* HRM assay (Figure 6.7 and 6.8). As expected, all three controls (wildtype *katG* (red) and 2 common mutants (S315T1 (blue) and S315T2 (green) showed subtlety distinct melt curves (Figure 6.7-6.10). There was no reaction in NTC. These distinctions were most obvious when represented using normalised fluorescence and fluorescence difference relative to the wildtype control (Figure 6.8a and b).
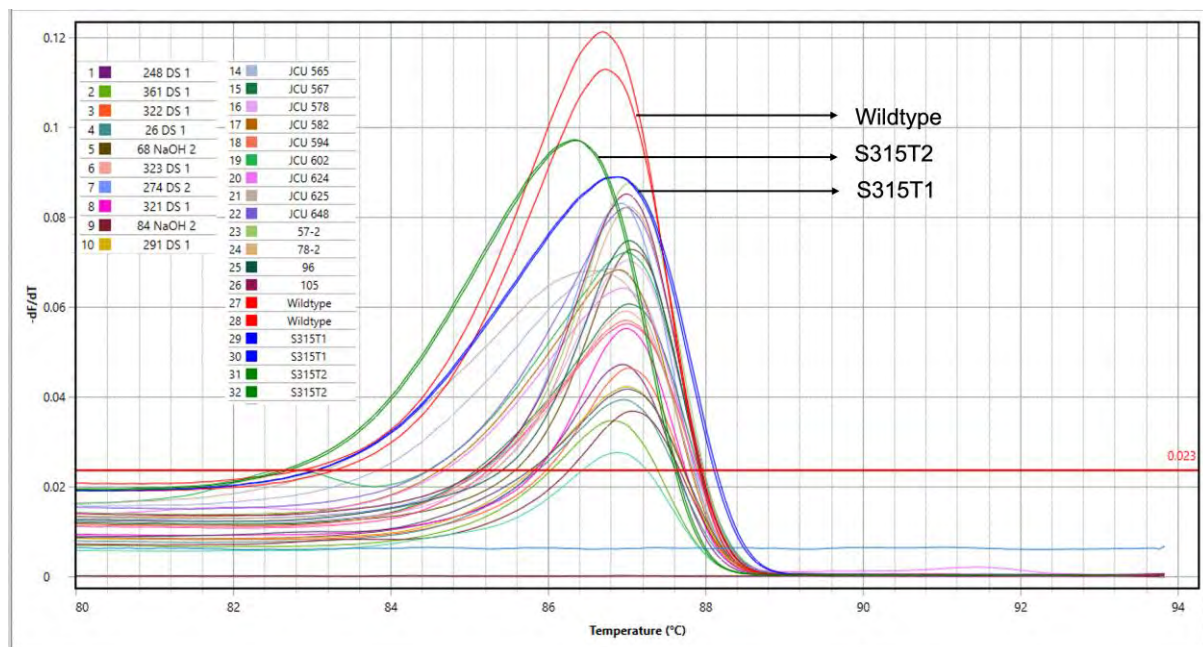
Figure 6.7: Melt rate analysis of *katG* amplicons in HRM analysis. Melt rates (y-axis) are represented as change in fluorescence units with increasing temperatures (dF/dT).
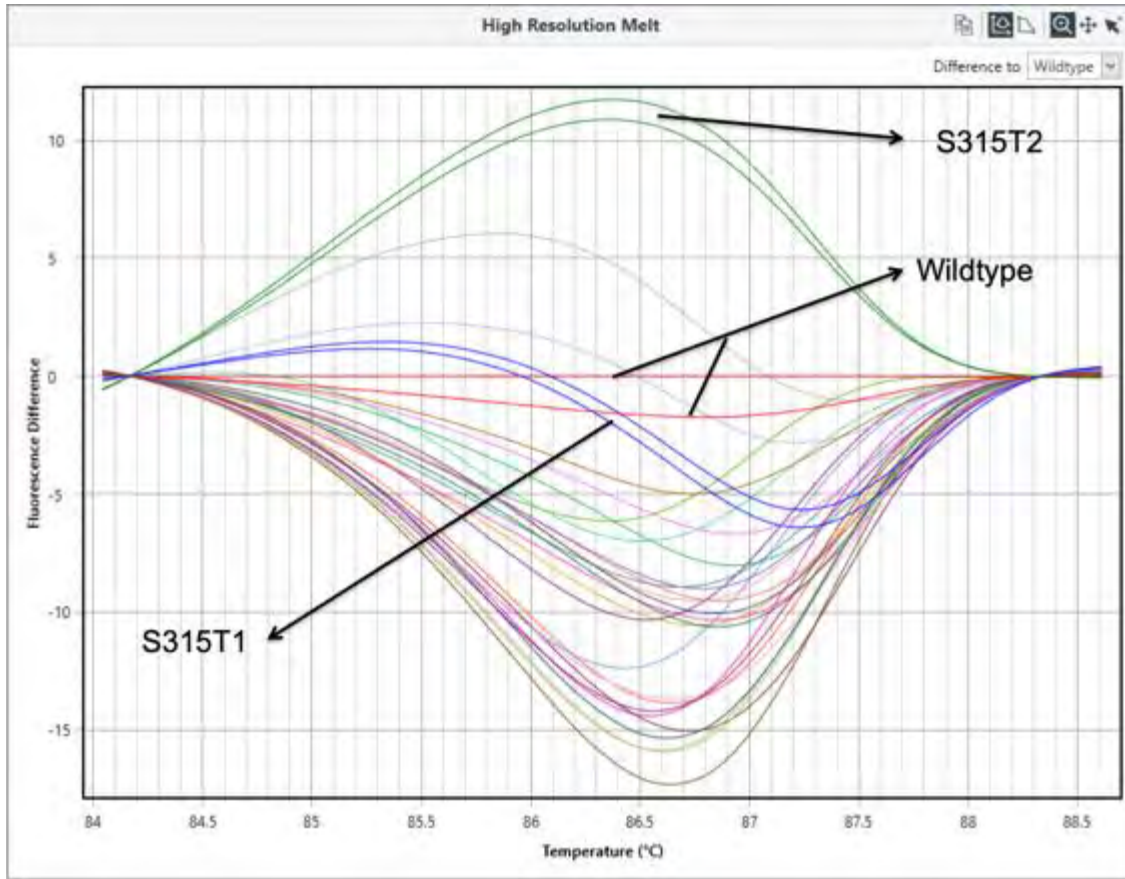
(a)
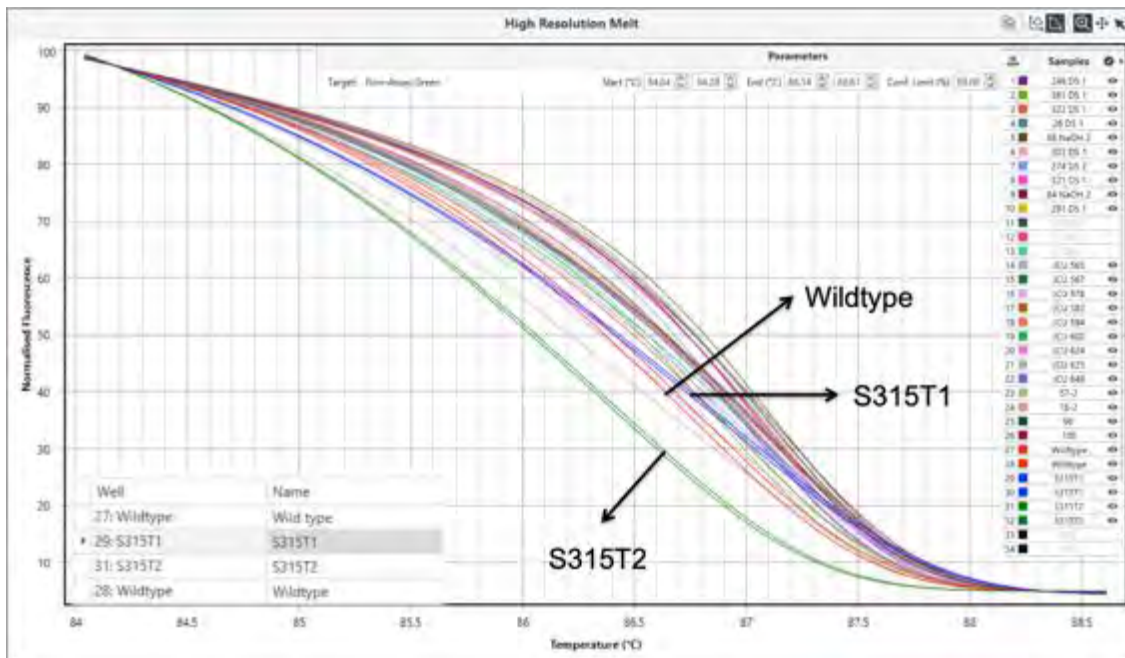
(b)

Figure 6.8: *katG* HRM analysis of various DNA samples (a) represented as fluorescence difference and (b) normalised fluorescence, both relative to the wildtype *katG* control (red).

There was only minimal difference between the curves of the wildtype *katG* oligo (red) and the S315T1 mutant oligo.

Closer analysis of wildtype and mutant oligo controls is shown in Figure 6.9 and 6.10. All three controls (in duplicate) showed sharp peaks and HRM was able to differentiate them, although the curves were closer together. The *katG* oligo wildtype control (red baseline) (Figure 6.9) was chosen as a reference against which the other two controls and the DNA were compared. The *katG* oligo S315T1 or CP001642 (the blue line just over the baseline) was almost indistinguishable from the *katG* oligo wildtype control. This suggested that they might have a small difference in their T_m values. While the *katG* oligo S315T2 or S315T1(green line) could be discriminated from the two other controls. All DNA templates tested where distinct from the three controls (Figure 6.8); therefore, it was unlikely that these *katG* sequences were present in any of the samples.
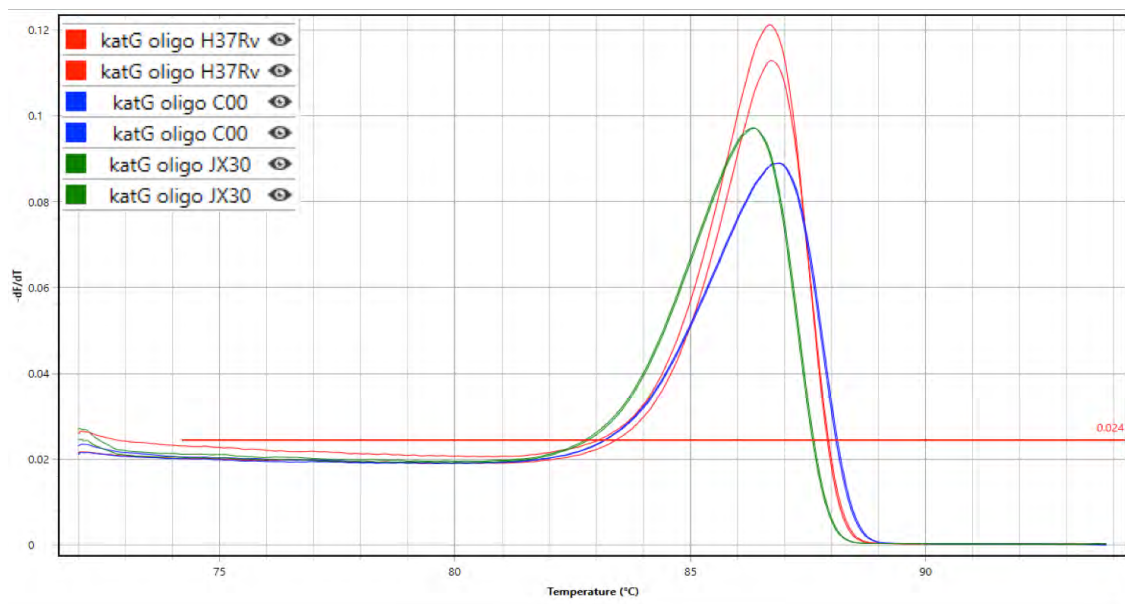
Figure 6.9: Melt curves of the three *katG* controls represented as change in fluorescence units with increasing temperatures (dF/dT) on the y-axis.
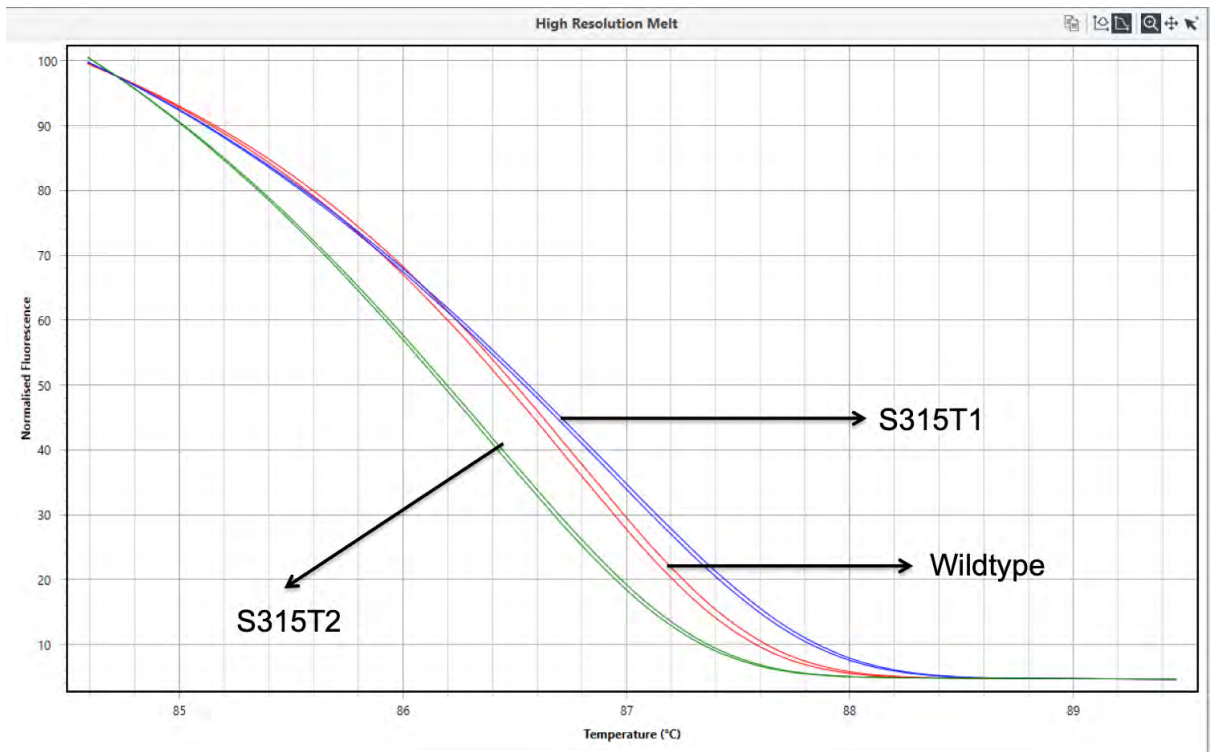
Figure 6.10: *katG* HRM analysis of the three controls shown as normalised fluorescence.

Figure 6.11 is the overall representation of all reacted DNA samples compared with the *katG* oligo wildtype control. Peaks from all samples were right shifted as compared to the wildtype control suggestive of slightly different melting temperatures and raises the possibility that they are not wildtype sequence, nor are they the harbouring the common S315T *katG* mutations.
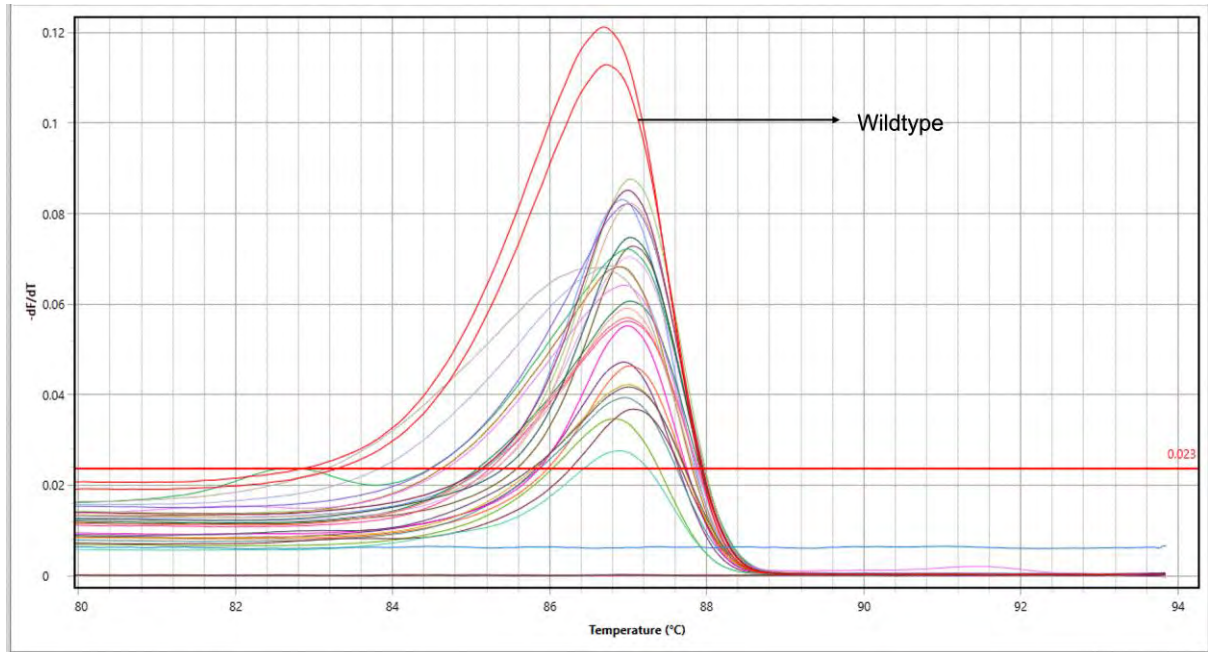
Figure 6.11: Melt analysis of all DNA samples compared with the Wildtype control duplicates (red). All the DNAs' peaks were right shifted as compared to the Wildtype control.

Use of all the positive controls were done to achieve the required results of the DNA. Table 6.9 was the predictive HRM results based on the *katG* gene. All templates failed to receive any nomination. This was unexpected. Further verification was required.

MIC™ qPCR HRM analysis software was used to allocate DNA samples to wildtype, or mutant using the control oligos as reference genotypes. The results of this prediction are presented in Table 6.9. None of the templates were assigned to the control genotypes. An explanation for this might be that the DNA samples tested might possess a novel SNP that differed from the positive controls used in this assay.

Table 6.9: Predictive *katG* genotypes based on HRM results. The assay failed to assign genotypes to all sample.

Sample number	Wildtype	S315T1	S315T2
57-2	-	-	-
78-1	-	-	-
78-2	-	-	-
96	-	-	-
105	-	-	-
530	-	-	-
321DS1	-	-	-
84 NaOH 2	-	-	-
291 DS 1	-	-	-
26 DS 1	-	-	-
68 NaOH DS 1	-	-	-
323 DS 1	-	-	-
274 DS 2	-	-	-
248 DS 1	-	-	-
361 DS 1	-	-	-
322 DS 1	-	-	-
JCU 582	-	-	-
JCU 625	-	-	-
JCU 648	-	-	-
JCU 565	-	-	-

6.6   Verification of HRM with sequencing

Amplicons from *rpoB* and *katG* assays were sequenced to verify HRM predictions (Table 6.8 and 6.9). Overall, 12 out of 20 were classified as outliers/mutants using the *rpoB* HRM assay. Sequencing of *rpoB* gene regions from these samples (JCU and DS marked in the Figure 6.12) confirmed that indeed they all were mutants (Figure 6.12). Two samples (84 NaOH 2 and 291 DS 1 in Figure 6.12) were predicted to possess the S450L mutation. However, sequencing confirmed those as wildtypes. The remaining 10 samples harboured different mutations than those predicted using the mutant control oligos. The remaining 8 DNA samples were predicted to be wildtype by the *rpoB* assay; this was confirmed by DNA sequencing. Table 6.10 shows a summary of *rpoB* HRM assay results and sequencing data.
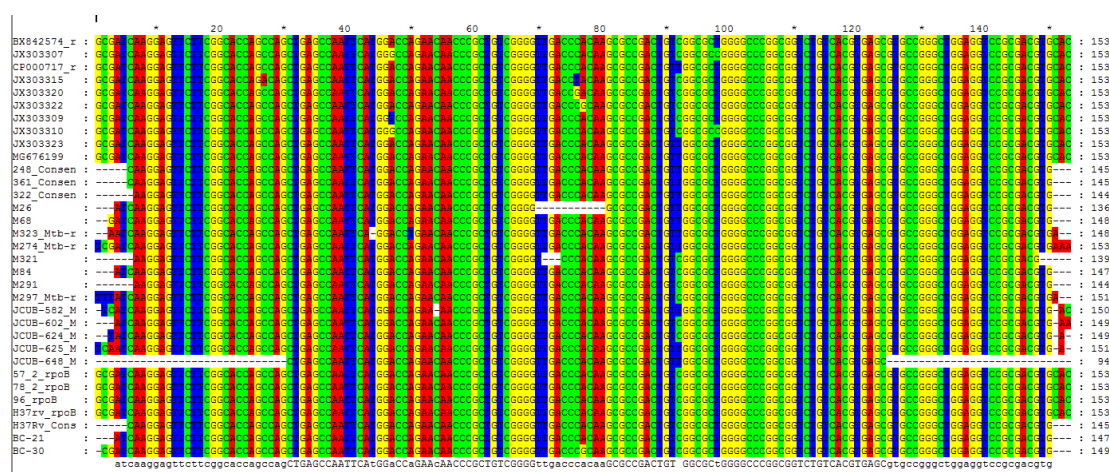
Figure 6.12: Sequencing results of the *rpoB* gene amplified from various DNA samples. Wildtype *rpoB* sequence is labelled as "H37Rv_refr" located on the top of all aligned sequences. Following are the published sequences (CP, JX and MG) Of 20, 10 isolates had a single nucleotide change at (92 bp) showing the TCG-TTG change (blue colour) associated with the prevalent *rpoB* S450L mutation. One sequence, BC-30 had different SNP (CAC-CGC) relevant to the H445R oligo control. This sample wasn't included in the DNA samples panel due the very limited amount available. The published *rpoB* sequences (top 10 sequences in this alignment) are described in more detail in Appendix 14, Table 14.3.
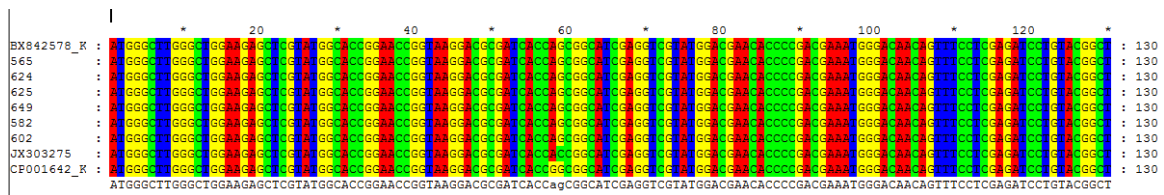
Figure 6.13: Sequencing results of the isolates that were screened as mutants by the *katG* HRM assay. All the isolates were wildtypes. The published *katG* gene sequences used are shown in Appendix 14, Table 14.4.

Despite DNA samples having HRM melt curves that were somewhat distinct from wildtype and S315T mutant oligo controls, sequencing revealed that all samples contained wildtype *katG* sequences (Table 6.11).

Therefore, although the *rpoB* HRM assay could detect mutations, the *katG* assay combined with the controls developed for this study, failed to accurately detect these single nucleotide changes. As most of the predominant SNPs linked with DR-TB in PNG to date are associated with single nucleotide changes, this may preclude the use of these types of assays as screening tools in their current form.

Table 6.10: Comparison of the *rpoB* based HRM predictive results and sequencing.

Sample number	Wildtype	H445R	S450L	Sequencing	Result agreement
				Wildtype	
57-2	+			Wildtype	
78-1	+			Wildtype	
78-2	+			Wildtype	
96	+			Wildtype	YES
105	+			Wildtype	
530	+			Wildtype	
321DS1	+	-	-	Wildtype	
84 NaOH 2	-	-	+	Wildtype	
291 DS 1	-	-	+	Wildtype	
26 DS 1	-	+	-	RIF Mono-resistant	
68 NaOH DS 1	-	+	-	RIF Mono-resistant	
323 DS 1	-	+	-	RIF Mono-resistant	
274 DS 2	-	+	-	RIF Mono-resistant	
248 DS 1	-	+	-	RIF Mono-resistant	NO
361 DS 1	-	+	-	RIF Mono-resistant	
322 DS 1	-	+	-	RIF Mono-resistant	
JCU 582	-	+	-	RIF Mono-resistant	
JCU 625	-	+	-	RIF Mono-resistant	
JCU 648	-	+	-	RIF Mono-resistant	
JCU 565	+	-		Wildtype	YES

Table 6.11: Comparison of the *katG* gene HRM assay results and sequencing data.

Sample number	Wildtype	S315T1	S315T2	Sequencing	Result agreement
57-2	-	-	-	Wildtype	
78-1	-	-	-	Wildtype	
78-2	-	-	-	Wildtype	
96	-	-	-	Wildtype	
105	-	-	-	Wildtype	
530	-	-	-	Wildtype	
321DS1	-	-	-	Wildtype	
84 NaOH 2	-	-	-	Wildtype	
291 DS 1	-	-	-	Wildtype	
26 DS 1	-	-	-	Wildtype	
68 NaOH DS 1	-	-	-	Wildtype	
323 DS 1	-	-	-	Wildtype	
274 DS 2	-	-	-	Wildtype	NO
248 DS 1	-	-	-	Wildtype	
361 DS 1	-	-	-	Wildtype	
322 DS 1	-	-	-	Wildtype	
JCU 582	-	-	-	Wildtype	
JCU 625	-	-	-	Wildtype	
JCU 648	-	-	-	Wildtype	
JCU 565	-	-	-	Wildtype	

6.7   Discussion

HRM is widely used post-PCR as a screening tool to identify genetic variants in target genes. Once optimised, these assays are simple, don't require sophisticated instruments, are suitable for high throughput with fast turnaround times, are relatively inexpensive and minimal personnel training is required. Previous studies have used it as a tool for providing molecular evidence of drug resistant TB (Pietzka *et al.*, 2009; Vossen *et al.*, 2009; Ramirez *et al.*, 2010; Wang *et al.*, 2011; Lee *et al.*, 2012; Nagai *et al.*, 2013; Pholwat *et al.*, 2014; Singh *et al.*, 2014; Galarza *et al.*, 2016; Anthwal *et al.*, 2017; Bentaleb *et al.*, 2017; Sharma *et al.*, 2017). The utility of these types of assays for characterising resistance-associated gene targets in a low-resource setting like PNG has not been explored previously.

The aim of the current study was to evaluate the utility of HRM assays based on *rpoB* and *kat*G genes to screen for mutant genotypes in archived DNA from clinical and TB culture material. Previous studies using clinical material from Balimo TB patients, reported *rpoB* Ser450Leu (T**C**G-T**T**G) and His445Arg (C**A**C-C**G**C) mutations. Of the 20 selected DNA templates, the *rpoB* HRM assay developed in the current study identified 12 samples as potentially containing RIF-resistance-associated SNPs.

Controls were developed that allowed for detection of a single Cytosine (C) to Thiamine (T) change from the H37Rv wildtype sequence (T**C**G) to the common mutant S450L (T**T**G). This change is called as the transition in which a purine/pyrimidine changes to another purine/pyrimidine respectively. These changes are classified as class 1 SNPs. (Liew *et al.*, 2004). A similar pattern was observed in His445Arg where Adenine (A) changed to Guanine (G) taking wildtype control as reference. Liew *et al.* (2004) demonstrated that HRM could easily identify mutant(s) that belong to class 1 SNPs (Liew *et al.*, 2004). However, in the current study, *rpoB* based HRM failed to correctly identify all SNPs. Furthermore, the *katG* assay was unable to identify any samples as containing wildtype sequence. Use of the HRM to identify single nucleotide changes may be limited for the purpose of identifying the SNPs. The assay could further be optimised and verified with current findings.

Table 6.10 detailed the overall comparison of the *rpoB* based HRM results and the sequencing data of the 20 DNA tested. All templates (n=6) extracted from MGIT™ tube supernatant had 100% concordance between the HRM prediction and the corresponding sequencing results. However, when DNA from sputum smears (n=10) or decontaminated sputum (n=4) were used as template in HRM assays, only 10% and 25% of samples respectively, showed concordance. This variation between the HRM predictive results and the sequencing data could potentially be due to differences in sample quality. The different sources of the DNA used might have led to DNA of varying quality. This might have limited the potential of the HRM. The presence of inhibitors such as salts and phenol might have contributed as well. The HRM works fine when quality samples are used. The cultured DNA is used for the standardisation process to test the DNA samples.

All the necessary considerations were considered when designing the *rpoB* and *katG* HRM primers and control oligos e.g., short oligos targeting less than 150bp were used. They were tested on wildtype DNA prior to testing for their specificity and eventually their application on the 20 DNA samples. The appearance of single sharp melt curves for all controls suggested that reagents were appropriate and functional, and that sample quality was the most likely explanation for lack of concordance for HRM and sequencing data for some samples.

As described in Chapter 3, section 3.1, sputum samples included in this analysis were old, and were stored in Balimo for an extended period before importation to Australia. Repeated sample thawing during power interruptions in Balimo are likely to have affected sample integrity. Similarly, archived sputum smear slides were, in some cases, stored for many years prior to DNA extraction. Therefore, these samples were likely to have very low quality.

Herrmann *et al*. (2006) compared different PCR machines and dyes for mutation and genotype analysis. In their study, they concluded that the power of DNA melting analysis depends directly on the resolution of the melting instrument. Melting resolution in turn is dependent on the temperature control, temperature and fluorescence measurement (Herrmann *et al.*, 2006). A recently launched PCR machine (MIC™ qPCR cycler) was used in

this project. Its performance data for HRM is not yet available. Therefore, it is possible that the HRM results reflected the resolution of the MIC™ qPCR for HRM. Its advantages have been mentioned later in this discussion.

Slomka *et al.* (2017) conducted extensive research on factors influencing HRM and identified its main limitations: the DNA source was most significant. The group incorporated 190 formalin-fixed and paraffin-embedded tissue samples targeting two genes using HRM. They were unable to identify the exact number of variants due to poor clustering of the isolates, as the reference curves overlapped the polymorphic sample clusters. These authors reported that their reference sample DNA was of good quality and worked appropriately. They also showed that HRM was prone to inconsistent melt curve differentiation when DNA of variable quality was used (Slomka *et al.*, 2017). This scenario was like this study. In this project, synthetic oligos were designed and were used as controls. Upon acquisition, they were tested using specific pair of primer and they worked perfectly. However, the melt curves failed to lie in the expected vicinity (on comparing to the controls used).

Slomka *et al.*, (2017) also described the influence of DNA preparation method on melt curves. DNA extracted using magnetic beads had quite distinctive melt curves when compared with the curves of the samples isolated from isopropanol-ethanol purification analysed in parallel (Slomka *et al.*, 2017). This finding is like the current study where melt behaviour varied according to initial sample type.

The choice of HRM PCR reagents, dyes and reaction conditions may also influence HRM analysis. The current study used Promega qPCR Mastermix containing LCGreen dye for all HRM analyses. This dye has been reported to be highly suitable for genotyping and mutation scanning (Herrmann *et al.*, 2006). Therefore, the impact of these particular technical issues is likely to be negligible (Slomka *et al.*, 2017).

A study conducted by Galarza *et al.* (2016) used a panel of 167 Peruvian isolates (89 susceptible and 78 multi-drug resistant) in 4 different HRM assays for comparison with existing drug susceptibility data. In this study, IS6110 (for Mtb confirmation), *rpoB*, *katG* and *16S* gene segments of varying lengths were chosen as the targets for their HRM assays with

sequencing used to verify HRM results (Galarza *et al.*, 2016). Only the phenotypically resistant isolates were sent off for the sequencing. Sequencing of a sample of susceptible isolates would have been beneficial for the comparative analysis. In addition, no real-time PCR, or any other pre-screening tool(s) prior to the HRM analysis was reported. This means that either the study involved HRM assays only or they didn't provide PCR data. Most of the studies conducted on the HRM have involved a sort of screening tool (real-time PCR in most cases) before carrying out the actual HRM analysis. Lastly, the target regions used for the HRM didn't cover all the resistance conferring mutations. This might limit its in-house utility.

Another study by Bantaleb *et al*. used 67 isolates sourced from TB-confirmed Moroccan patients to detect *rpoB* gene mutations using HRM. Among those 67 isolates, 22 cases were RIF-susceptible while the other were phenotypically RIF-resistant. Unlike, Galarza *et al.* they incorporated real-time PCR prior conducting HRM followed by sequencing to verify the results. They constructed five plasmids as positive controls. Four plasmids contained a documented SNP (S531L, S531W, H526Y and D516V) whilst the fifth was a wildtype control. Out of the 45 mono- resistant cases, 40 were successfully identified by HRM. The remaining 5 isolates were devoid of any mutation within the amplified *rpoB* region. Therefore, they were classified as phenotypically RIF-resistant isolates (Bentaleb *et al.*, 2017).

In both of the studies described above, mutation at 531 codon (S531L TCG-TTG) was the dominant mutation (Galarza *et al.*, 2016; Bentaleb *et al.*, 2017). This mutation was also observed in the current study. An assay(s) that can reliably identify this type of globally dominant mutation(s) with an ability to also screen out the undocumented genetic variation(s) is what is needed.

The BMS MIC™ qPCR was used for the HRM analysis. Its price is around A$20,000. The device was able to be used as a suitable platform for the Taqman assays (Chapter 4) and the HRM. Based on the run profile, the device was able to provide results for both real-time PCR and HRM within one hour. This suggested that the device is rapid. Further, the handling and operation of this cycler was quite easy and simple. It was indeed user-friendly. The device is portable due to its small size (two kg) and can be connected to the computer/laptop simply via Bluetooth. The device program is quite easy to use. Even this cycler can be operated in

absence of power with batteries. This makes it deliverable. Therefore, it can be said that this qPCR cycler complies with the WHO ASSURED criteria.

In the PNG context, so far, the JCU TB research team has found two SNPs in the *rpoB* gene. The implementation of an inexpensive and sustainable assay to rapidly identify local mutations is vital. The HRM assays designed in this study however failed to differentiate between those circulating SNPs. Therefore, the use of HRM assays for identification and characterisation of drug-resistance conferring mutations limited. Further development and evaluation could be performed, and work is warranted to investigate the possibility assay improvement.

6.8   Conclusion

The results presented in this Chapter suggest that:

1. The synthetic oligo controls could be used as a substitute to the conventional controls such as genomic DNA or plasmids.

2. HRM may provide information that can be reliable and accurate only if this technology is used to examine multiple nucleotide changes. Detecting single nucleotide variations is the limit of this technology.

3. Sample quality highly influences the HRM results.

CHAPTER 7

General discussion

The Balimo and Gogodala region in the Western Province of PNG have an annual TB incidence rate of 727/100,000 people (Diefenbach-Elstob *et al.*, 2019). Multi-drug resistant tuberculosis (MDR-TB) is another serious problem in PNG and is still evolving. Reliable data on the prevalence of MDR-TB is unavailable due to unreliable reporting (Aia *et al.*, 2016a). Several studies have indicated the widespread of MDR-TB in different regions of PNG (Gilpin *et al.*, 2008; Simpson *et al.*, 2011; Ballif *et al.*, 2012; Cross *et al.*, 2014b; Ley *et al.*, 2014a; Diefenbach-Elstob *et al.*, 2019) and dissemination of such drug-resistant strains to healthy people could potentially cause serious issues especially when the available diagnostics and the patient management system are inefficient.

Accurate diagnosis is the main prerequisite to initiate treatment of a suspected TB patient. There are two extremes when it comes to TB diagnosis. Some regions have only clinical diagnosis based on presenting symptoms and smear microscopy available. Whereas access to modern techniques including MGIT™-based culture, molecular diagnostics, radiological techniques including CT and PET scans is more restricted. Hence, the ability to diagnose (or misdiagnose) TB varies greatly across these extremes. When funds, continuous power supply and infrastructure are limited, simple and cost-effective strategies for diagnosing TB remain essential. Although clinical presentation consistent with active pulmonary TB and sputum smear microscopy may aid in diagnosis, clinical diagnosis alone can lead to misdiagnosis and inappropriate treatment. Infection with NTM and indeed other respiratory pathogens can confound accurate TB diagnoses, leading to treatment delays and mortality.

In a large study of respiratory samples from US-affiliated islands of the Western Pacific, Lin *et al.* (2018) reported the rising prevalence of NTM. Of 15,811 respiratory specimens from 5807 patients, 998 patients were AFB positive, 675 were Mtb positive, 323 were NTM positive and five patients were infected with both Mtb and NTM. There was an increasing

pattern of NTM prevalence from 2 cases/1,00,000 in August 2007 to 48 cases/1,00,000 up to December 2011. While the Mtb prevalence remained steady with 43 cases/1,00,000 in 2008 to 58 cases/1,00,000 in 2010 (Lin *et al.*, 2018). A cohort of 29 patients that were NTM positive from May 2010 till November 2010 were analysed by paper chart review. Chronic cough with sputum, reduction in weight and fatigue/malaise were the common recorded signs in this study, mirroring the clinical manifestations of TB. Chronic obstructive pulmonary disease was the frequent concurrent diagnosed condition from the 8 patients and three patients were diagnosed with diabetes mellitus (Lin *et al.*, 2018).

Historically, NTM appears to have a low prevalence in PNG (Guernier *et al.*, 2017), although a recent PNG based study found some NTM from samples tested (Lavu *et al.*, 2019). Hence, reliance on smear microscopy in such a setting may potentially lead to misidentification. Therefore, it is imperative to consider NTM when developing new diagnostics for TB. Furthermore, infection with other bacteria such as *Burkholderia pseudomallei* causes melioidosis ranging from acute, sub-acute to chronic disease. Melioidosis is endemic in the Balimo region of the Western Province of PNG. This infection may mimic the clinical manifestation of TB causing misdiagnosis (Warner and Currie, 2018). Hence diagnostic tests that can readily distinguish Mtb from other organisms causing disease in regions such as PNG, are required.

Molecular based tests are preferred over the gold standards like microscopy and culture. They are relatively rapid with higher specificity and sensitivity, thereby narrowing the chances for misdiagnosis with rapid result reporting (Hinic *et al.*, 2017; Parcell *et al.*, 2017). Additionally, nucleic acid based methods have the potential to differentiate between tuberculous and non-tuberculous species of Mycobacteria (Eichbaum and Rubin, 2002). WHO recommends a range of molecular based diagnostics to identify and characterise Mtb (WHO, 2021), however, their selection, implementation and sustainability depend upon the financial capacity of each nation. Low-income countries often face challenges to remain current with the WHO endorsed commercial diagnostic assay platforms. International support may aid in implementation, but their sustained use becomes difficult when the financial support is discontinued. Therefore, cost-effective substitutes that are reliable,

allow rapid, local reporting of results and require minimal infrastructure, but still have performance comparable to the commercial systems, are required in those settings.

Therefore, this study aimed to develop and evaluate in-house molecular based identification tools to identify and characterise Mtb from genomic DNA samples. The TaqMan-based duplex assay, targeting the IS6110 and *rpoB* genes was designed to identify and assess the quality of DNA templates. High quality templates with *Mycobacterium* DNA, were further analysed for molecular evidence of drug-resistance and susceptibility. Two HRM assays based on the Mtb genes *rpoB* and *katG* were designed to test the utility of HRM for discrimination of drug-resistant and drug-susceptible genotypes. The HRM results were compared with the sequencing data to verify its utility as an in-house mutation-scanning tool.

The TaqMan based duplex designed in this study (Chapter 4) worked well. It was able to identify 24 DNA templates that contained *Mycobacterium* DNA (Section 4.8.8.1). All had sufficient templates to carry out further analysis (Section 4.8.8.2). However, confirmation was required to determine which of these samples contained MTBC versus NTM DNA. The published *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was therefore added to the testing panel to minimise misidentification (Section 4.9). It was assumed that TB causing strain(s) lacking the IS6110 sequence may be present in PNG. The *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was used to detect such strains. However, the current study revealed that the *senX3-regX3* IR gene is not MTBC specific. This gene is present in a diverse range of *Mycobacterium* species and even in some other bacteria (Section 4.9). This finding therefore contradicts Broccolo *et al.* (Broccolo *et al.*, 2003) and the utility of their assay to confirm MTBC is challenged.

The IS6110 assay designed in the current study had higher analytical sensitivity (Section 4.8.6.1) than the published IS6110 assay (Broccolo *et al.*, 2003). The JCU IS6110 assay was even more sensitive than the published assay (Broccolo *et al.*, 2003) when DNA from the supernatant of MGIT™ cultures with a range of different dilutions were used (section 4.8.6.2). The analytical specificity and repeatability of both IS6110 assays was similar (Section 4.8.6.3 and 4.8.6.4). Both IS6110 primer sets, and probes shared sequence identity

with various sequences within NTM genomes. This confirmed previous reports (Broccolo *et al.*, 2003) that the IS6110 sequence is not MTBC specific.

Wang *et al.* recently developed a IS6110 TaqMan-based PCR assay to detect the MTBC in pulmonary and extrapulmonary specimens. In this study, they designed primers and probes from a region of the IS6110 sequence that was identical across the MTBC. The assay was optimised, evaluated and applied on 130 clinical specimens with Roche Cobas TaqMan MTB (CTM) and culture also used to evaluate assay performance (Wang *et al.*, 2019). As part of design of the novel JCU IS6110 assay described in Chapter 4, BLAST searches of target sequences described in Wang *et al.* (2019) were performed. Following exclusion of MTBC from these searches, it was noted that their IS6110 component had sequence identity ranging from 90% to 74.36% in around 12 *Mycobacterium* species (NTM), five different Nocardia species, two different *Xanthomonas species* and even in the *Saccaropolyspora coralli* genome. Hence, using *in silico* analysis, it appeared that their IS6110 assay was not specific to MTBC or even *Mycobacterium* species (Appendix 16).

The *rpoB* TaqMan in the current study was designed to determine the overall quality of the DNA templates for further analysis (i.e., HRM analysis). A high C_t value was indicative of a low DNA titre whereas a low C_t suggested a high DNA titre, which in turn suggested that genomic DNA was sufficient for subsequent analyses. This assay provided information for triage of samples for the subsequent experiments (Section 4.8.8.2). The assay removed seven DNA samples that, although contained *Mycobacterium* DNA, were unlikely to have sufficient DNA for further analysis of single copy genes. This strategy eliminated unnecessary testing of inferior quality samples and thus effective management of resources. (Section 4.9).

HRM assays were designed to analyse the differentiating and mutation scanning potential of this technology. For this, three assays were designed targeting the Mtb 16S rRNA, *rpoB* and *katG* genes. In Chapter 5, it was concluded that the assay based on the *rpoB* and *katG* could be applied for HRM analysis. As a differentiation tool, the 16S rRNA assay when applied on the panel of DNA templates, had low resolution due to overlapping melt curves. This was not completely unexpected as all bacteria harbour this gene and there are many conserved

regions across this gene. The *rpoB* HRM assay was however able to differentiate between the *Mycobacterium* species (Section 5.4.1.2). However, a limitation was the requirement for high quality template DNA sourced from MGIT™ cultures. The *katG* assay proved to be highly specific for MTBC detection. This assay was quite suitable for differentiating bacteria to their species level (Section 5.4.1.3). However, some members of the *Mycobacterium* genus had similar melting curves with all three assays. Other studies have also reported similar findings (Yang *et al.*, 2009; Perng *et al.*, 2012; Khosravi *et al.*, 2017b).

When the *rpoB* and *katG* assays were used from discrimination of wildtype sequence and mutations associated with drug resistance, mixed results were found. The *rpoB* HRM assay was able to accurately assign a wildtype genotype to DNA from MGIT™ tube supernatants, one DNA from a stained sputum smear and one template sourced from decontaminated sputum. Their sequence data was in concordance with the HRM predicted genotype. However, the assay also predicted that ten DNA (extracted from the smeared slides) might have the H445R mutation while the sequencing showed that all had the S450L variation instead (Section 6.5.1.1). The *katG* assay failed to categorise all samples as wildtype or outlier(s), whilst sequence data indicated that they were wildtype. Therefore, it was concluded that using HRM for accurate detection of single nucleotide changes is variable and at the limits of its utility (Section 6.5.1.2).

Taylor (2009) reviews the limitations of HRM: a change in T_m remains undetectable if the base pairs exchange their position moving to the complementary strand (C to G or A to T) while remaining bonded (Taylor, 2009). The 16S rRNA gene HRM melt curves were identical for *M. bovis* and Mtb H37Rv. This was expected, as both are part of the MTBC. Their recorded melting temperatures were 87.98°C and 87.93°C respectively. However, the sequencing data of the *M. bovis* 16S rRNA gene showed numerous nucleotide changes compared to H37Rv (Figure 5.4 and Appendix 12). This result suggests that multiple base changes may not significantly alter the amplicon melting temperature (Figure 5.2).

Previous studies have claimed that HRM could be used as an economical tool to identify mutations associated with DR-TB (Pietzka *et al.*, 2009; Vossen *et al.*, 2009; Ramirez *et al.*, 2010; Wang *et al.*, 2011; Lee *et al.*, 2012; Nagai *et al.*, 2013; Pholwat *et al.*, 2014; Singh *et*

al., 2014; Galarza *et al.*, 2016; Anthwal *et al.*, 2017; Bentaleb *et al.*, 2017; Sharma *et al.*, 2017). Arefzadeh *et. al.* used culture-based DNA templates in HRM analysis to identify the Mtb *rpoB* gene-based mutations. They suggested that the accuracy of HRM assays relies on the sample source, preparation, quality of the DNA isolated, its concentration, amplicon length, GC content, equipment and the dye used (Arefzadeh *et al.*, 2020). The results of the current study did not fully support the finding that HRM could be used as a tool to identify mutations associated with DR-TB. The reason for this discrepancy is likely to relate to issues of sample quality affecting the sensitivity of HRM. All these previous studies used culture/ L J colony(s)-sourced DNA for HRM analysis. This pure and high-quality DNA, free of any inhibitor(s) or contaminant(s) associated with clinical sample types are likely to be superior in these types of assays.

The results of the current study showed that HRM could distinguish wildtype and mutant genotypes when high-quality DNA (e.g., MGIT™ culture) was used as template, which supports the previous studies. Mycobacterial culture cannot be implemented as a routine method in in PNG due to the infrastructure and trained personnel requirements and financial constraints in PNG. Therefore, other types of samples that may be available in PNG were tested in our assays. Using these types of more PNG-relevant samples, HRM did not appear to be able to detect these drug-resistant/susceptible genotypes.

The current study included DNA extracted from sputum samples, archived sputum smears, decontaminated sputum sediments as examples of the types of samples that are available for analysis at the Balimo district hospital (BDH), PNG. They were collected from the patients suspected of having TB or patients being investigated for TB. None of the samples was taken from patients with definitive laboratory evidence of TB. Therefore, it was unknown whether the patient(s) had TB. Furthermore, after sample acquisition, sputum aliquots were temporarily stored and then processed to decontaminate. The decontaminated sputum samples were then stored for subsequent use. The sample quality during collection, the possibility of presence/absence of the Mtb and the number of organisms in each sputum sample was unknown. The sputum samples used in the current study, may had been contaminated with traces of tobacco, betel nut (commonly chewed in PNG), cigarette smoke, alcohol *et cetera*. In addition, the amount of sputum collected

during sampling might not have been of sufficient quantity for detection using these assays. These samples were stored in the freezer at the Balimo district hospital (BDH) for further processing. The hospital doesn't have a continuous (24/7) power supply and a backup power generator. This might further have affected the integrity of the stored samples to some extent. Those stored samples were then imported to Australia and again stored for extended periods prior to DNA extraction where chemicals associated with the extraction method (e.g., phenol) may have further inhibited PCR reactions. Therefore, these suboptimal DNA templates samples may have contributed to the inability of some assays to detect Mtb DNA or different genotypes associated with drug resistance. In the current study, the *rpoB* TaqMan assay was used as a screening tool to determine whether the DNA templates were of quality sufficient for further analysis (i.e., HRM analysis).

The search for a mere single nucleotide base change(s) in this study challenged the detection potential of HRM. Therefore, it can be said that the HRM isn't fit for purpose to be implemented as a routine in-house mutation(s) scanning tool in resource-limited settings. It is unknown whether the change(s) of nucleotide base(s) position to their complementary strand confers mutation(s) or not, and to what extent these types of substitutions lead to changes in the HRM melting curve.

Overall, the assays designed in this study worked and provided the information that they were designed for. The IS6110 TaqMan assay indicated the presence of the *Mycobacterium* DNA, although this and the *senX3-regX3* IR TaqMan assay failed to differentiate the MTBC from NTM (Chapter 4). The *rpoB* TaqMan assay was used to assess overall quantity with those samples with sufficient template further tested in HRM analysis. Similarly, the 16S rRNA HRM assay had utility in checking that the quality of the DNA from the large panel of microorganisms was suitable for use in subsequent HRM analyses. The *rpoB* HRM assay was able to differentiate the *Mycobacterium* species and the *katG* HRM assay was highly specific for the MTBC (Chapter 5). They also confirmed the presence of Mtb DNA in the samples tested. Therefore, collectively these assays were able to provide the answers to the questions: Is it TB, If yes, how much?

Nevertheless, the question: what is the drug susceptibility profile of the infecting Mtb strain? was only partially resolved. The *rpoB* HRM assay was able to accurately assign samples to their respective genotypes i.e., wildtype or mutant when DNA was sourced from MGIT™ culture (Figure 6.5). However, when DNA from other sources was used, variable results were obtained: some were in concordance with the sequencing data whilst the majority were not (Table 6.10). Therefore, the results of this study highlight that samples source and quality are key factors that influence HRM assay performance, particularly when attempting to detect single nucleotide changes in drug resistance gene targets. The *rpoB* S450L mutation is prevalent in Mtb strain(s) in PNG and therefore it appears as though implementation of *rpoB* HRM assays is not suitable in this setting.

Manson *et al.* in their (2017) analysed 5310 Mtb whole genome sequences from many countries and covering all defined lineages. They revealed that there are numerous harbinger mutations that occur before attaining the MDR state. For example, the *katG*S315T mutation represented an epitome that occurred before the acquisition of rifampicin resistance. Other non-*katG* related mutations had the same pattern. These authors suggested that molecular based diagnostics should also target these types of mutations (Manson *et al.*, 2017). The current study addressed the importance of the *katG*S315T mutation by using synthetic oligo controls harbouring this mutation (Table 6.4). However, the novel *katG* HRM assay described here, was unable to accurately identify this or other genotypes in the samples that were analysed (Chapter 6).

The WHO recommends the Xpert that targets only the *rpoB* gene for diagnosis. This gene is highly conserved across many bacteria, and this may contribute to false-positive Mtb diagnoses. Even its newer version, the Xpert Ultra that amplifies three genes (multicopy genes IS6110 and IS1081 and the RRDR of *rpoB*), has improved sensitivity, but its specificity is still uncertain (WHO, 2021). In the current study, numerous different organisms were detected using the *rpoB*-based assay and most had overlapping melting curves suggested that the sequence of these amplicons was similar (Table 5.5).

In addition, its implementation and sustenance cost (cartridge, reagents, calibration, power supply etc.) limits its use in resource constrained settings. (Abebe *et al.*, 2011; Gupta and Anupurba, 2015) It also lacks the positive and negative reactions. The BMS MIC™ qPCR machine could be a cost-effective (A$20,000) alternative to these expensive commercial systems in such settings. The cycler performs real-time PCR, HRM and is portable due to its ideal size. Additionally, a provision of using it on a battery makes it feasible to implement in areas where there is lack of continuous power supply. The device can even be used to set-up pop up labs that can be quite useful in resource-limited settings. Moreover, 48 samples could be analysed in one PCR run. This may reduce the overall cost per test. This may make it more affordable in low- or middle-income countries.

Approximately 10 DNA samples from Balimo had the S450L mutation in their *rpoB* gene. According to their sequencing data their corresponding *katG* gene were all wildtypes. The question arises however, will those Mtb strains with S450L mutation in their *rpoB* genes acquire resistance conferring *katG* mutations as well? (Manson *et al.*, 2017). Longitudinal sampling of Balimo TB patients was not part of the current study, but these questions would be very interesting to answer in future studies. A study by Niehaus *et al.* (Niehaus *et al.*, 2015) analysed 924 African isolates that had DST and LNA probe data available. The study found that amongst MDR-TB isolates, 67% of *katG* mutation(s) only with 15% of those isolates harbouring both *inhA* and *katG* mutation(s). The pre-XDR TB isolates had 53% of *katG* and 43% of *katG* and *inhA* mutation(s). Lastly, the XDR-TB isolates harboured 17% *katG* and 73% of *inhA* and *katG* mutation(s) (Niehaus *et al.*, 2015). This further demonstrated the importance of analysing *katG* and the *inhA* gene for mutations. The current study included an assay that targeted the *katG* gene to conduct the HRM analysis. This assay surprisingly came out to be the most specific for the MTBC (Section 5.4.1.3).

Figure 7.1 is a systematic diagram showing the utility of all the Mtb identification and drug resistance characterisation assays described in the current study. The DNA provided for testing were first subjected to the IS6110 TaqMan assay. The multi-copy IS6110 sequence is specific to the *Mycobacterium* genus. The templates reacting in this assay confirmed the *Mycobacterium* DNA presence in the tested templates. The DNA were then tested for their

overall genomic DNA quality by the *rpoB* TaqMan assay. The templates with low C_t value were selected for further testing. The subsequent testing was done using HRM analysis. Two assays based on the *rpoB* and *katG* were developed and tested. The predictive results from the HRM assays and their respective sequencing data were eventually compared to verify if HRM had the potential to provide reliable data on drug characterisation. The intent was to check whether it could be implemented as a routine in-house mutation-scanning tool in PNG. It is important to note here that the published assay *senX3-regX3* IR assay (Broccolo *et al.*, 2003) was not included in Figure 7.1. The assay was not MTBC specific.
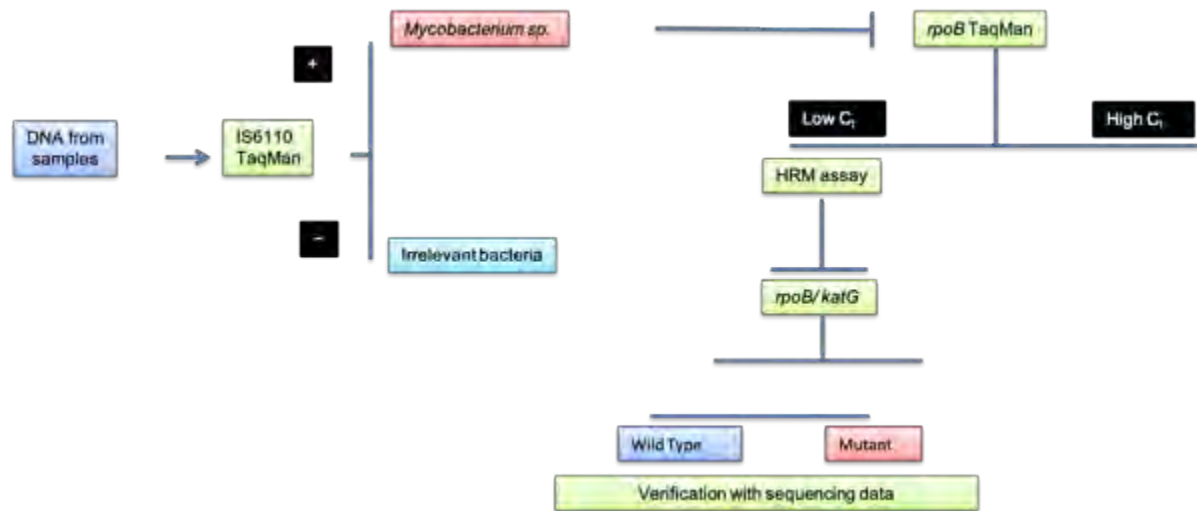
Figure 7.1: A flow chart of systematic testing of the isolates for the identification of Mtb and drug-resistance conferring mutations.

Future perspectives

Mtb detection and its characterisation was the foundation of this project. Successful Mtb identification requires two key things: the presence of the tubercle bacilli in the host lungs and an Mtb specific identification tool(s). Furthermore, the Mtb population size inside the host lungs and the extent of their dissemination across the lungs is important for the detection of the organism. High population size and a widespread of bacterial cells across the lungs probably caused them to appear adequately in the sputum portions while sampling. This possibly made them to have their genomic DNA available to be detected. The assays designed in this study were devised to detect and characterise that Mtb DNA (if present) sourced from the sputum of suspected TB patients.

The TaqMan based duplex (JCU IS6110 and *rpoB*) designed in this study performed well and provided reliable data. However, during specificity testing, cross binding of these primers and probes with other non-relevant members of the *Mycobacterium* genus was unexpected. This instead revealed that the IS6110 sequence is generic to Mtb and NTM. An assay based on this gene can however discriminate the *Mycobacterium* genus from other bacteria, however. The presence of the *rpoB* gene in all bacteria due to its important and conserved function (bacterial DNA replication) is likely to explain why it these assays showed poor specificity.

The designed TaqMan assays worked well on the DNA templates of unknown or potentially low quality. According to Buh Gasparic *et. al.* (2010) the MGB and LNA probes have the potential to even identify a single nucleotide base change in the target sequence (Buh Gasparic *et al.*, 2010). Future assays could incorporate these highly specific probes and such an approach may extend the ability of the Taqman assay(s) to identify the point mutations that are prevalent in PNG. This may serve as a substitute to the commercial expensive systems endorsed by the WHO (WHO, 2021).

A target gene(s) that is highly specific for the MTBC is required for future assays. This specific gene must be interrogated thoroughly for its presence in all *Mycobacterium* species and further extended to taxonomically relevant bacteria such as *Nocardia* species. Results from the current study (Chapter 5) suggest that the *katG* gene could be a MTBC-specific candidate gene. The *katG* assay described here (for HRM analysis) could be re-designed as a TaqMan assay. This revised TaqMan assay should first be tested for its specificity against all possible *Mycobacterium* species.

Incorporation of possibly all the terrestrial as well as marine *Mycobacterium* species to the testing panel would be important. This thorough testing would give rise to a robust and highly MTBC specific assay. Manson *et. al.* (2017) and Niehaus *et. al*. (2015) clearly demonstrated the importance of investigating the *katG* gene that usually acquires mutation earlier than the *rpoB* gene does (Niehaus *et al.*, 2015; Manson *et al.*, 2017).

Guernier *et. al.* (2019) study indicated the presence of the Mtb strains belonging to lineage 2 and 4 in a single human host (Guernier-Cambert *et al.*, 2019). This study provided evidence of a potential possibility of the presence of mixed (genomically different) Mtb strains in one host. Therefore, an assay(s) that could determine the in-host Mtb diversity could also be a useful tool for further studies. This could be achieved by targeting the Mtb lineage specific genes such as *rpsA*, Rv0249c, Rv2971 and Rv2164c which have been identified as lineage markers for the lineage 2, 3, 4 and 6 respectively.

HRM might not be a beneficial tool for detection of point mutation(s). Regardless, it may be an affordable, sustainable, and therefore useful tool to aid in the screening of and triage of Mtb genomes for further sequence analysis to confirm clinically significant mutations.

As a future direction, it would be interesting to take the MIC™ qPCR along with the designed assays in PNG to see:

1. Whether it can work and provide information in a real-world setting.
2. If fresh or high-quality samples/templates are used, will it increase the HRM quality?

APPENDIX 1

BIOCHEMICAL DETAILSOF MTB OF ALL LINEAGES

Appendix 1-Table 1.1: Biochemical properties of some *Mycobacterium* species and all lineage strains (Ngabonziza *et al.*, 2020).

Mycobacterial species/ lineage	Niacin production	Nitrate reduction	Urease hydrolysis	Tween hydrolysis	Catalase production	Aryls ulfat ase
			-	-	-	-
Lineage 1	+	+	+	-	Weak+	-
Lineage 2	+	-	+	-	-	-
Lineage 3	+	+	+	-	-	-
Lineage 4	+	+	+	+	-	-
Lineage 5	+	-	-	+	-	-
Lineage 6	+	-	Weak+	Weak+	-	-
Lineage 7	+	-	Weak+	-	-	-
Lineage 8	+	+	+	Weak +	-	-
M. bovis	-	-	Weak+	-	-	-
M. bovis BCG	-	-	+	-	Weak+	-
M. orygis	+	-	+	-	-	-
M. canettii	-	-	+	+	Weak+	-
M. fortuitum	-	+	+	-	Weak+	+

APPENDIX 2

REAGENT PREPARATION

7% NaCL

7g of NaCL in 100mL sterile distilled or deionised water.

4% NaOH

4g of NaOH in 100mL sterile distilled or deionised water

9% NaCL

9g of NaCL in 100 mL sterile distilled or deionised water.

70% Isopropyl alcohol

70mL of isopropyl alcohol in 30 mL sterile distilled or deionised water.

3% HCL

3mL of HCL in 100 mL sterile distilled or deionised water.

95% alcohol

95mL of alcohol in 5mL sterile distilled or deionised water.

0.5% Malachite green

5g of Malachite green in 100 mL sterile distilled or deionised water.

0.4% Sodium sulphite

0.4g of sodium sulphite in 100mL sterile distilled or deionised water.

10% Trigene

10mL of Trigene in 90 mL of sterile distilled or deionised water.

APPENDIX 3

IN-SILICO EVALUATION OF THE IS6110 ASSAYS

In-silico evaluation of the published IS6110 assay (Broccolo *et al.*, 2003)

	Rating	Sequence	Original Position	Position on Alignment	Length bp	Tm ºC	GC %	Hairpin ΔG kcal/mol	Self Dimer ΔG kcal/mol	Cross Dimer ΔG kcal/mol	Run Length bp	GC Clamp	TaOpt ºC	Degenerate Bases
Y17220_Mycobacteriu#000013_22														
Sense TaqMan®	65.4	TGTGGGTAGCAGACCTCACCTATGTGTCGA	799	800	30	77.4	53.3	-1.6	-3.2		3			0
Sense Primer	0.6	GATCGCTGATCCGGCCA	725	726	17	68.4	64.7	-2.1	-4.5		2	0		0
Anti-sense Primer	0.6	AGGCGAACCCTGCCCAG	846	847	17	71.3	70.6	-2.5	-2.5		3	3		0
Product	8.6				122	86.6				-4.3			66.2	

In-silico evaluation of the designed JCU IS6110 assay

	Rating	Sequence	Original Position	Position on Alignment	Length bp	Tm ºC	GC %	Hairpin ΔG kcal/mol	Self Dimer ΔG kcal/mol	Cross Dimer ΔG kcal/mol	Run Length bp	GC Clamp	TaOpt ºC	Degenerate Bases
Y17220_Mycobacteriu#0D0013_22														
Sense TaqMan®	87.9	CATCGCCGCCTCTACCAGTA	1,220	1,222	20	70.1	60	0	-0.5		2			0
Sense Primer	62.6	CATCGAGGATGTCGAGTTG	1,171	1,173	19	64.1	52.6	-3.2	-3.2		2	1		0
Anti-sense Primer	73.6	CGGAGACTCTCTGATCTG	1,335	1,337	18	62	55.6	-1	-2.4		3	1		0
Product	66.4				165	87.4				-2.5			64.9	

APPENDIX 4

PRIMER SETS DETAILS USED FOR SEQUENCING

AND CONTROL PLASMID CONSTRUCTION

The IS6110 primer set

Appendix 4-Table 4.1: Details of primer set used for amplification of the IS6110 sequence for control plasmid construction

Details of primer set for the reference IS6110 plasmid construct:		
Primer set	Sequence detail	Length (bp)
Mtb-IS6110-1045-F	5'-ACC ACG ACC GAA GAA TCC-3'	1045
Mtb-IS6110-1045-R	5'-TGG CGT TGA GCG TAG TAG-3'	1045

The *rpoB* primer set

Appendix 4-Table 4.2: Details of primer set used for *rpoB* gene amplification for control plasmid construction

Details of primer set for the reference *rpoB* plasmid construct:		
Primer name	Sequence detail	Length (bp)
rpoB-PL4-1013-F	5'-ACC AGC GAG CAG ATT GTC-3'	1013
rpoB-PL4-1013-R	5'- ACGTCCATGTAGTCCACC-3'	1013

APPENDIX 5

IN-SILICO SPECIFICITY TESTING OF PRIMER SETS

USED FOR PLASMID CONSTRUCTS

In silico specificity testing

IS6110 sequencing primers

A large number of published sequences with 100% homology to the designed primer sequences were observed while entering the forward and reverse primer sequences one at a time as a query sequence individually on BLAST n suite (Figure 3.1)

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium bovis strain BCG-1 genome	36.2	72.4	100%	2.7	100%	CP011455.1
Mycobacterium bovis BCG strain Russia 368, complete genome	36.2	72.4	100%	2.7	100%	CP009243.1
Mycobacterium bovis strain 1595, complete genome	36.2	36.2	100%	2.7	100%	CP012095.1
Mycobacterium bovis BCG strain 3281, complete genome	36.2	36.2	100%	2.7	100%	CP008744.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F309	36.2	36.2	100%	2.7	100%	LC005482.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F302	36.2	36.2	100%	2.7	100%	LC005481.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F178	36.2	36.2	100%	2.7	100%	LC005473.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F158	36.2	36.2	100%	2.7	100%	LC005472.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F145	36.2	36.2	100%	2.7	100%	LC005471.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F130	36.2	36.2	100%	2.7	100%	LC005469.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F76	36.2	36.2	100%	2.7	100%	LC005463.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F60	36.2	36.2	100%	2.7	100%	LC005462.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F47	36.2	36.2	100%	2.7	100%	LC005460.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F41	36.2	36.2	100%	2.7	100%	LC005459.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F33	36.2	36.2	100%	2.7	100%	LC005458.1
Mycobacterium tuberculosis DNA, insertion sequence:IS6110, VNTR, complete sequence, strain: F31	36.2	36.2	100%	2.7	100%	LC005457.1
Mycobacterium bovis strain ATCC BAA-935, complete genome	36.2	36.2	100%	2.7	100%	CP009449.1
Mycobacterium bovis BCG str. ATCC 35743, complete genome	36.2	36.2	100%	2.7	100%	CP003494.1
Mycobacterium tuberculosis str. Haarlem, complete genome	36.2	325	100%	2.7	100%	CP001664.1
Mycobacterium tuberculosis EAI5, complete genome	36.2	36.2	100%	2.7	100%	CP006578.1
Mycobacterium tuberculosis EAI5/NITR206, complete genome	36.2	36.2	100%	2.7	100%	CP005387.1
Mycobacterium bovis BCG str. Korea 1168P, complete genome	36.2	36.2	100%	2.7	100%	CP003900.2
Mycobacterium tuberculosis 7199-99 complete genome	36.2	325	100%	2.7	100%	HE663067.1
Mycobacterium canettii CIPT 140070008 complete genome	36.2	325	100%	2.7	100%	FO203508.1
Mycobacterium tuberculosis KZN 605, complete genome	36.2	542	100%	2.7	100%	CP001976.1
Mycobacterium tuberculosis strain SCAID 187.0, complete genome	36.2	542	100%	2.7	100%	CP012506.1
Mycobacterium tuberculosis W-148, complete genome	36.2	759	100%	2.7	100%	CP012090.1
Mycobacterium tuberculosis haplotype 20 insertion sequence IS6110 transposase (orfA) gene, complete cds	36.2	36.2	100%	2.7	100%	KP844685.1

(a)

Description	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium bovis strain BCG-1 genome	36.2	72.4	100%	2.7	100%	CP011455.1
Mycobacterium tuberculosis strain SCAID 187.0, complete genome	36.2	542	100%	2.7	100%	CP012506.1
Mycobacterium bovis BCG strain Russia 368, complete genome	36.2	72.4	100%	2.7	100%	CP009243.1
Mycobacterium tuberculosis W-148, complete genome	36.2	759	100%	2.7	100%	CP012090.1
Mycobacterium bovis strain 1595, complete genome	36.2	36.2	100%	2.7	100%	CP012095.1
Mycobacterium bovis BCG strain 3281, complete genome	36.2	36.2	100%	2.7	100%	CP008744.1
Mycobacterium tuberculosis haplotype 36 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844721.1
Mycobacterium tuberculosis haplotype 35 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844720.1
Mycobacterium tuberculosis haplotype 34 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844719.1
Mycobacterium tuberculosis haplotype 33 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844718.1
Mycobacterium tuberculosis haplotype 31 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844716.1
Mycobacterium tuberculosis haplotype 30 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844715.1
Mycobacterium tuberculosis haplotype 29 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844714.1
Mycobacterium tuberculosis haplotype 28 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844713.1
Mycobacterium tuberculosis haplotype 27 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844712.1
Mycobacterium tuberculosis haplotype 26 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844711.1
Mycobacterium tuberculosis haplotype 25 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844710.1
Mycobacterium tuberculosis haplotype 24 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844709.1
Mycobacterium tuberculosis haplotype 23 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844708.1
Mycobacterium tuberculosis haplotype 22 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844707.1
Mycobacterium tuberculosis haplotype 21 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844706.1
Mycobacterium tuberculosis haplotype 20 insertion sequence IS6110 transposase (orfB) gene, partial cds	36.2	36.2	100%	2.7	100%	KP844705.1

(b)

Appendix 5-Figure 5.1: Results from the nucleotide BLAST (BLASTn suite) service showing that both Mtb-IS6110-1045-F (forward) (a) and Mtb-IS6110-1045-R (reverse) (b) primer will produce a PCR product that has 100% homology to the IS6110 sequence.

Both the primer sequences (Mtb-IS6110-1045-F and Mtb-IS6110-1045-R) were 100% specific in identifying the insertion sequence from different sequences present on web. Therefore, it was interpreted that the designed primer pair had the potential for amplifying the target sequence from any relevant source DNA.

Furthermore, upon conducing a homology search using primer sequences against the aligned genomic database in GeneDoc (Figure 5.2), it was found that all the reference sequences from the web had 100% homology with the designed primer pair sequence. Moreover, an evaluation of the designed primers in AlleleID® software indicated that the primer set was superior.
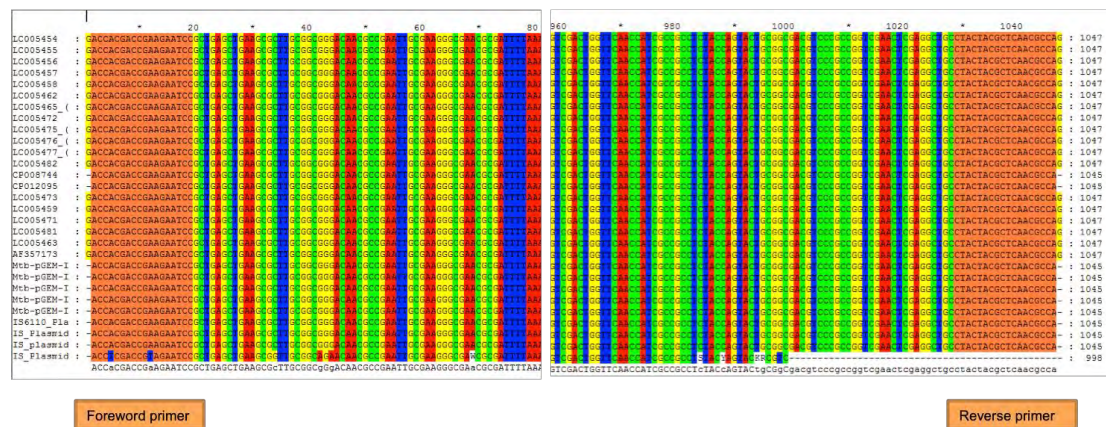
Appendix 5-Figure 5.2: An overall presentation of both primer-binding sites against a genomic database showing the IS6110 sequence between these oligonucleotides sites.

The orange-coloured regions indicate the foreword (Mtb-IS6110 -1045-F) and the reverse primer (Mtb-IS6110 -1045-R) binding sites respectively. The forward primer sequence approximately bound at position 0 to 20bp, while 1030- 1045bp was the position of reverse primer binding site. This 1045bp long sequence was the IS6110 sequence required for the control plasmid construction.

The *rpoB* sequencing primers

Many published sequences with 100% homology to the designed primer sequences were observed while putting the forward and reverse primer sequences as a query sequence individually on BLAST n suite (Figure 5.3).

	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium sp. 3/86Rv chromosome	36.2	36.2	100%	4.1	100%	CP023708.1
Mycobacterium tuberculosis strain TBV4952 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023640.1
Mycobacterium tuberculosis strain TBV4768 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023639.1
Mycobacterium tuberculosis strain TBV4766 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023638.1
Mycobacterium tuberculosis strain TBDM2717 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023637.1
Mycobacterium tuberculosis strain TBDM2699 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023636.1
Mycobacterium tuberculosis strain TBDM2489 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023635.1
Mycobacterium tuberculosis strain TBDM2487 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023634.1
Mycobacterium tuberculosis strain TBDM2444 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023633.1
Mycobacterium tuberculosis strain TBDM2189 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023632.1
Mycobacterium tuberculosis strain TBDM1506 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023631.1
Mycobacterium tuberculosis strain MDRMA2441 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023630.1
Mycobacterium tuberculosis strain MDRMA2260 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023629.1
Mycobacterium tuberculosis strain MDRMA2082 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023628.1
Mycobacterium tuberculosis strain MDRMA2019 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023627.1
Mycobacterium tuberculosis strain MDRMA1565 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023626.1
Mycobacterium tuberculosis strain MDRMA863 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023625.1
Mycobacterium tuberculosis strain MDRMA701 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023624.1
Mycobacterium tuberculosis strain MDRMA203 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023623.1
Mycobacterium tuberculosis strain MDRDM827 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023622.1
Mycobacterium tuberculosis strain LN2900 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023621.1
Mycobacterium tuberculosis strain LN1856 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023620.1
Mycobacterium tuberculosis strain LN1100 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023619.1
Mycobacterium tuberculosis strain LN3695 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023618.1
Mycobacterium tuberculosis strain LN3672 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023617.1
Mycobacterium tuberculosis strain LN3668 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023616.1
Mycobacterium tuberculosis strain LN3589 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023615.1
Mycobacterium tuberculosis strain LN3588 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023614.1
Mycobacterium tuberculosis strain LN3584 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023613.1
Mycobacterium tuberculosis strain LN2978 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023612.1
Mycobacterium tuberculosis strain LN763 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023611.1
Mycobacterium tuberculosis strain LN317 chromosome, complete genome	36.2	36.2	100%	4.1	100%	CP023610.1

(a)

	Max score	Total score	Query cover	E value	Ident	Accession
Mycobacterium sp. 3/86Rv chromosome	36.2	36.2	100%	4.1	100%	CP023708.1
Mycobacterium tuberculosis strain TBV4952 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023640.1
Mycobacterium tuberculosis strain TBV4768 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023639.1
Mycobacterium tuberculosis strain TBV4766 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023638.1
Mycobacterium tuberculosis strain TBDM2717 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023637.1
Mycobacterium tuberculosis strain TBDM2699 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023636.1
Mycobacterium tuberculosis strain TBDM2489 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023635.1
Mycobacterium tuberculosis strain TBDM2487 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023634.1
Mycobacterium tuberculosis strain TBDM2444 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023633.1
Mycobacterium tuberculosis strain TBDM2189 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023632.1
Mycobacterium tuberculosis strain TBDM1506 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023631.1
Mycobacterium tuberculosis strain MDRMA2441 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023630.1
Mycobacterium tuberculosis strain MDRMA2260 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023629.1
Mycobacterium tuberculosis strain MDRMA2082 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023628.1
Mycobacterium tuberculosis strain MDRMA2019 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023627.1
Mycobacterium tuberculosis strain MDRMA1565 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023626.1
Mycobacterium tuberculosis strain MDRMA863 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023625.1
Mycobacterium tuberculosis strain MDRMA701 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023624.1
Mycobacterium tuberculosis strain MDRMA203 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023623.1
Mycobacterium tuberculosis strain MDRDM827 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023622.1
Mycobacterium tuberculosis strain LN2900 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023621.1
Mycobacterium tuberculosis strain LN1856 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023620.1
Mycobacterium tuberculosis strain LN1100 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023619.1
Mycobacterium tuberculosis strain LN3695 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023618.1
Mycobacterium tuberculosis strain LN3672 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023617.1
Mycobacterium tuberculosis strain LN3668 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023616.1
Mycobacterium tuberculosis strain LN3589 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023615.1
Mycobacterium tuberculosis strain LN3588 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023614.1
Mycobacterium tuberculosis strain LN3584 chromosome, complete genome	36.2	578	100%	4.1	100%	CP023613.1

(b)

Appendix 5-Figure 5.3: Results from the nucleotide BLAST (BLASTn suite) search showing that both *rpoB*-PL4-1013-F (a) and *rpoB*-PL4-1013-R (b) primer are predicted to produce a PCR product that has 100% homology to the *rpoB* sequence.

Since none of the published sequences had *rpoB* written in the title, the GenBank data of the very first sequence was obtained (Figure 5.4).

185

```
Sequencing Technology   :: Illumina
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider                   :: NCBI
Annotation Date                       :: 09/28/2017 12:06:12
Annotation Pipeline                   :: NCBI Prokaryotic Genome
                                         Annotation Pipeline
Annotation Method                     :: Best-placed reference protein
                                         set; GeneMarkS+
Annotation Software revision          :: 4.2
Features Annotated                    :: Gene; CDS; rRNA; tRNA; ncRNA;
                                         repeat_region
Genes (total)                         :: 4,156
CDS (total)                           :: 4,105
Genes (coding)                        :: 3,915
CDS (coding)                          :: 3,915
Genes (RNA)                           :: 51
rRNAs                                 :: 1, 1, 1 (5S, 16S, 23S)
complete rRNAs                        :: 1, 1, 1 (5S, 16S, 23S)
tRNAs                                 :: 45
ncRNAs                                :: 3
Pseudo Genes (total)                  :: 190
Pseudo Genes (ambiguous residues)     :: 37 of 190
Pseudo Genes (frameshifted)           :: 123 of 190
Pseudo Genes (incomplete)             :: 52 of 190
Pseudo Genes (internal stop)          :: 20 of 190
Pseudo Genes (multiple problems)      :: 37 of 190
CRISPR Arrays                         :: 2
##Genome-Annotation-Data-END##
FEATURES             Location/Qualifiers
     source          1..18
                     /organism="Mycobacterium sp. 3/86Rv"
                     /mol_type="genomic DNA"
                     /strain="3/86Rv"
                     /isolation_source="Mediastinal lymph node"
                     /host="Cattle"
                     /db_xref="taxon:2041046"
                     /country="India:Uttar Pradesh"
                     /lat_lon="28.36 N 79.43 E"
                     /collection_date="05-Jul-1986"
                     /collected_by="Dr. A.K.Sharma"
     gene            <1..>18
                     /locus_tag="CP498_03715"
     CDS             <1..>18
                     /locus_tag="CP498_03715"
                     /inference="COORDINATES: similar to AA
                     sequence:RefSeq:NP_215181.1"
                     /note="Derived by automated computational analysis using
                     gene prediction method: Protein Homology."
                     /codon_start=1
                     /transl_table=11
                     /product="DNA-directed RNA polymerase subunit beta"
                     /protein_id="ATI69502.1"
                     /translation="MADSRQSKTAASPSPSRPQSSSNNSVPGAPNRVSFAKLREPLEV
                     PGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLEEVLYELSPIEDFSGSMSLSFS
                     DPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTF
                     IINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIPSRGAWLEFDVDKRDTVGV
```

Appendix 5- Figure 5.4: GenBank data of the first appeared sequence after BLAST. On the bottom side, the product was DNA directed RNA polymerase beta subunit.

It was clear from the Figure 3.3 and 3.4 that both primer sequences were 100% specific in identifying the *rpoB* sequence from different published sequences in NCBI Genbank. Therefore, it was interpreted that the designed primer pair could find and amplifying specifically the target sequence from any source DNA.

Homology search using primer sequences against the aligned genomic database in GeneDoc indicated 100% homology. A further evaluation of the designed primers in AlleleID® software indicated that the primer set was very good. Therefore, it was deduced that the chances of being able to amplify the sequence of interest were quite high and it was concluded that the primer set designed to make the reference plasmid was suitable and highly specific.
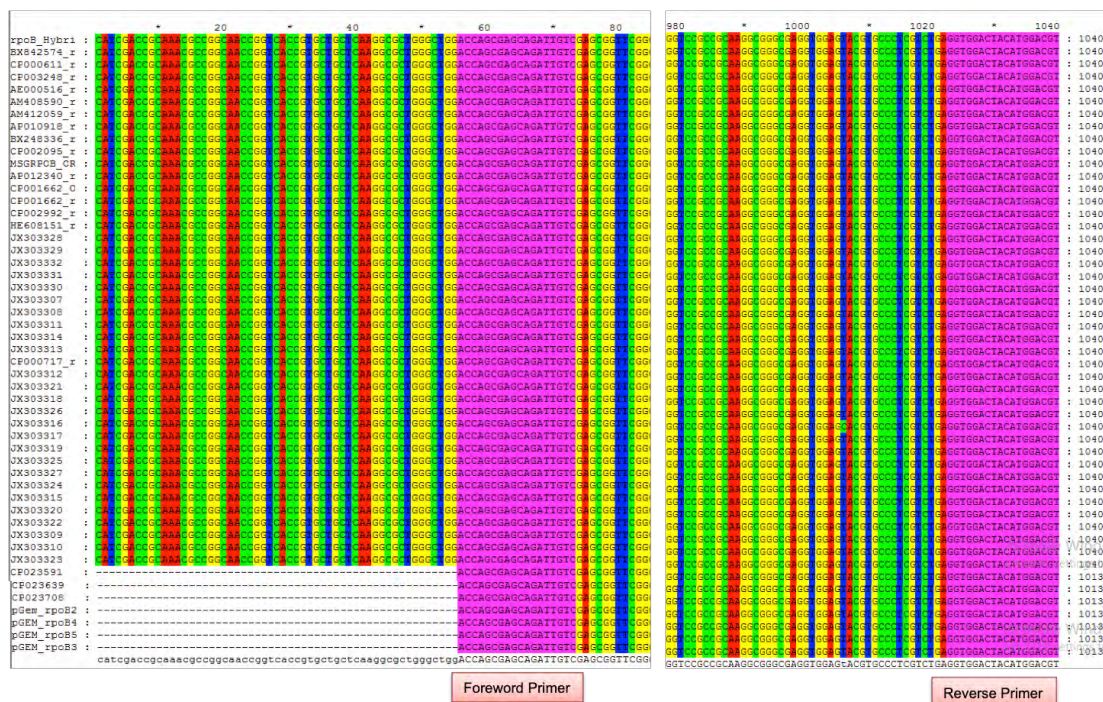
Appendix 5- Figure 5.5: Binding pattern of the *rpoB* sequencing primers against the aligned published *rpoB* sequences. Details of the *rpoB* published sequences are in Appendix 14, Table 14.2.

The pink highlighted (Figure 5.5) areas were the binding positions of the foreword *rpoB*-PL4-1013-F (57 till 73bp approximately) and the reverse primer *rpoB*-PL4-1013-R (1022 till 1040bp) against the aligned published sequences. Both the primers bound to all the aligned sequences. This indicated the high specificity of the primer set towards its target region that was the sequence (1013bp) inside the annealing sites of the primers

APPENDIX 6

DETAILS OF REACTION MIX

USED FOR PLASMID INSERT AMPLIFICATION

Details of reaction mix for target sequence extraction

For the IS6110 sequence:

Appendix 6-Table 6.1: Details of the reaction mix to amplify IS6110 gene for plasmid construct.

Stock Concentration	Reagent	Final Concentration	Reaction (μL)
2X	Master Mix (Promega A#6002)	1X	10
10 μM	Forward Primer Mtb-IS6110-1045-F	0.8 μM	1.6
10 μM	Reverse Primer Mtb-IS6110-1045-R	0.8 μM	1.6
	Nuclease free water		4.8
Template			2
Total Volume			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 10 seconds and 60°C for 45 seconds. The PCR cycling was 40 cycles.

For the *rpoB* gene fragment:

Appendix 6-Table 6.2: Details of the reaction mix to extract the *rpoB* gene segment for plasmid construction

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer *rpoB*- PL4-1013 - F	0.8 µM	1.6
10 µM	Reverse Primer Mtb- *rpoB*- PL4-1013 - R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total volume of one reaction mix			20

The standard cycling protocol was 40 cycles of 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 45 seconds.

APPENDIX 7

REAGENT PREPARATION FOR BLUE-WHITE COLONY SCREENING

Appendix 7-Table 7:1: Table of ingredients and their concentration used for blue-white screening

Stock Concentration	Reagent	Final Concentration	Volume of medium in one LB plate (50mL)
100 mM	IPTG	0.5mM	0.25
50 mg	X-gal	0.08mg	0.08
100mg	Ampicillin	0.1mg	0.05

IPTG (0.1M)

Add 1.2 g in 50ml of DI water. Filters sterilize it with 0.22μm filters.

Dispense 2ml of solution separately in each tube to make 25 tubes.

X-Gal (50mg/ml)

Add 0.5g in 10ml of DMSO (Use glass containers, as it may dissolve plastic). Cover it with aluminium foil and store at -20C. Allocate this reagent in 17 tubes containing 600μl X-Gal in each tube.

Ampicillin (100mg/ml)

Take 1g, add it to 5 ml of DI water and 5ml of absolute ethanol, making 10ml total volume.

Make 25 tubes each containing 400μl of this solution.

Luria Bertani (LB) Broth

Tryptone- 10g, NaCL- 10g, Yeast extract -5g, Distilled water- 1L

Dissolve the contents in 900 mL of distilled water. Adjust pH to 7 with 1N NaOH.

Make the volume to 1L by adding more distilled water. Sterilise it by autoclaving.

Pour the sterilised broth into sterilise petri dishes (25mL) as required. Store at 4°C.

APPENDIX 8

BIOINFORMATIC ANALYSIS OF

CROSS REACTIVITY OF THE IS6110 ASSAYS

In-silico specificity of Broccolo's IS6110 assay against *M. abscessus and M. intracellulare*

(a)

(b)

(c)
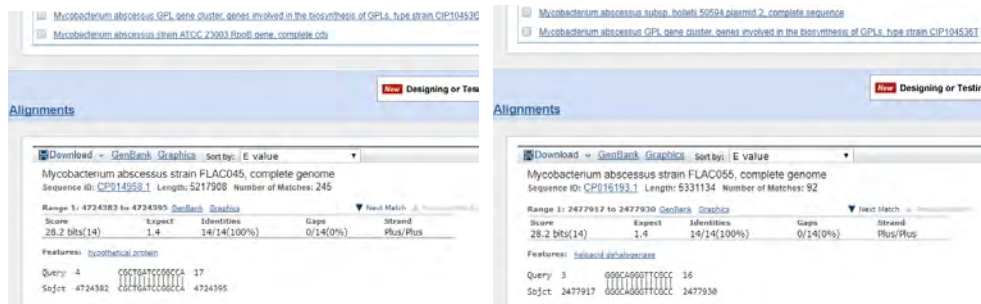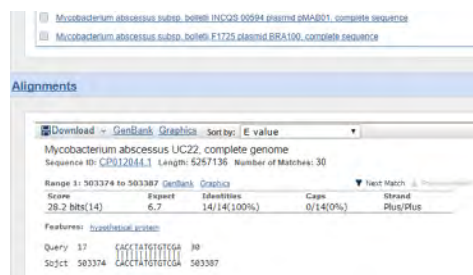
Appendix 8-Figure 8.1: Binding of forward (a), reverse primer (b) and the probe (c) of the published IS6110 assay (Broccolo *et al.*, 2003) with the random DNA segments of the *Mycobacterium abscessus* all with 100% identity of 14 bases of both primers pair and the probe.

The same pattern was observed in the genome of *M. intracellulare* (Figure8.2).

4.

(b)

(c)

Appendix 8-Figure8.2: Binding pattern of the published assay with the random DNA segments of *Mycobacterium intracellulare.* 13 Bases of the foreword primer (a), reverse primer (b) and the probe (c) bound with 100% identity without any gaps.

(a)

(b)

(c)

Appendix 8-Figure 8.3: 13 bases of JCU IS6110 foreword, reverse and the probe bound to the non-specific DNA segments of *Mycobacterium abscessus* with 100% identity.

As in published assay, the primer pair and probe of JCU IS6110 bound non-specifically to the bits of *M. abscessus* genome with 100% identity having no mismatches (Figure 1.8) and *M. intracellulare* (Figure 1.9).

(a)

(b)

(c)

Appendix 8-Figure 8.4: Binding of JCU IS6110 foreword (a), reverse (b) and the probe (c) to the non-specific DNA segments of *Mycobacterium intracellulare* with 100% identity.

THE PRESENCE OF *SENX3-REGX3* IR GENE

IN NTM AND OTHER BACTERIA
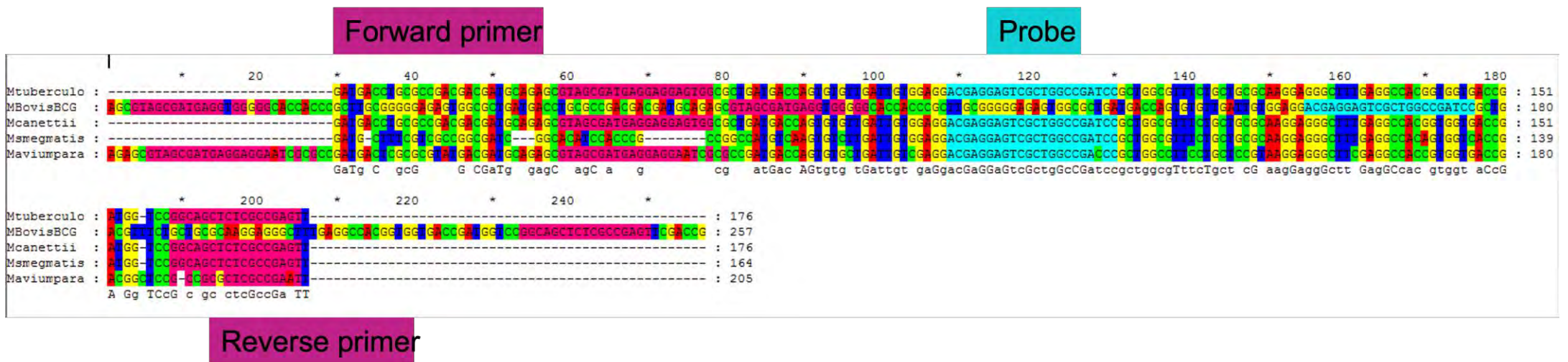
Appendix 9-Figure 9.1: Annealing of the published assay (*senX3-regX3* IR) primers and probe (Broccolo *et al.*, 2003) against the published and aligned *senX3-regX3* IR gene sequences. Alignment and viewing of the sequences were done using the GeneDoc program. The primer set and probe were specific to the sequences sourced from the Mtb and other NTM. No binding of the forward primer occurred in the *M. bovis* sequence. This demonstrated that the *senX3-regX3* IR sequence is not specific to the Mtb. The aligned sequences were from the Mtb, *M. bovis* (BCG), *M. canettii, M. smegmatis, M. avium* paratuberculosis. This assay has been claimed to be Mtb specific.

In-silico specificity testing of the *senX3- regX3* IR sequence

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Uncultured organism clone pLE07 hypothetical protein gene, complete cds	uncultured organism	75.2	136	38%	3e-09	89.83%	2752	HQ156906.1
Pseudonocardia autotrophica NBRC 12743 DNA, complete genome	Pseudonocardia autotro...	92.4	92.4	30%	4e-14	85.90%	7246130	AP018920.1
Pseudonocardia sp. HH130630-07 chromosome, complete genome	Pseudonocardia sp. HH...	92.4	92.4	30%	4e-14	85.90%	6186048	CP013854.1
Pseudonocardia sp. EC080619-01, complete genome	Pseudonocardia sp. EC...	92.4	92.4	30%	4e-14	85.90%	6149338	CP012184.1
Pseudonocardia sp. EC080610-09, complete genome	Pseudonocardia sp. EC...	92.4	92.4	30%	4e-14	85.90%	6138223	CP012181.1
Pseudonocardia sp. EC080625-04, complete genome	Pseudonocardia sp. EC...	92.4	92.4	30%	4e-14	85.90%	6135769	CP010989.1
Pseudonocardia broussonetiae strain Gen 01 chromosome, complete genome	Pseudonocardia brouss...	90.6	90.6	29%	2e-13	85.71%	7177725	CP053564.1
Saccharopolyspora sp. ASAGF58 chromosome, complete genome	Saccharopolyspora sp.	88.7	88.7	29%	5e-13	85.53%	8045912	CP040605.1
Nakamurella multipartita DSM 44233 chromosome, complete genome	Nakamurella multipartita...	88.7	88.7	29%	5e-13	85.53%	6060298	CP001737.1
Saccharothrix algeriensis strain NRRL B-24137 chromosome	Saccharothrix algeriensis	89.7	89.7	31%	2e-13	85.19%	6880068	CP072788.1
Saccharothrix espanaensis DSM 44229 complete genome	Saccharothrix espanaen...	89.7	89.7	31%	2e-13	85.19%	9360653	HE804045.1
Saccharothrix sp. 6-C chromosome, complete genome	Saccharothrix sp. 6-C	89.7	89.7	31%	2e-13	85.19%	9045220	CP064192.1
Kutzneria albida DSM 43870, complete genome	Kutzneria albida DSM 4...	84.2	84.2	29%	7e-12	84.21%	9874926	CP007155.1
Lentzea guizhouensis strain DHS C013 chromosome, complete genome	Lentzea guizhouensis	85.1	85.1	31%	7e-12	83.95%	9997872	CP016793.1
Pseudonocardia sp. AL041005-10, complete genome	Pseudonocardia sp. AL0...	83.3	83.3	30%	2e-11	83.33%	6143341	CP011862.1
Pseudonocardia sp. HH130629-09, complete genome	Pseudonocardia sp. HH...	83.3	83.3	30%	2e-11	83.33%	6058802	CP011868.1
Amycolatopsis aidingensis strain YIM 96748 chromosome, complete genome	Amycolatopsis aidingensis	79.7	79.7	29%	3e-10	82.89%	7657695	CP076538.1
Saccharothrix syringae strain NRRL B-16468 chromosome, complete genome	Saccharothrix syringae	80.6	80.6	31%	8e-11	82.72%	10929570	CP034550.1
Kibdelosporangium sp. MJ126-NF4 complete genome	Kibdelosporangium sp.	79.7	79.7	32%	3e-10	81.93%	12113479	LN877229.1
Pseudonocardia sp. DSM 110487 chromosome, complete genome	Pseudonocardia sp. DS...	77.0	77.0	29%	1e-09	81.82%	10132709	CP080521.1
Actinopolyspora erythraea strain YIM 90600, complete genome	Actinopolyspora erythraea	75.2	75.2	29%	3e-09	81.58%	5243930	CP022752.1
Nakamurella panacisegetis strain P4-7,KCTC 19426,CECT 7604 genome assembly, chromosome: I	Nakamurella panaciseg...	75.2	75.2	29%	3e-09	81.58%	5013970	LT629710.1

(a)

Nakamurella panacisegetis strain P4-7,KCTC 19426,CECT 7604 genome assembly, chromosome: I	Nakamurella panaciseg...	75.2	75.2	29%	3e-09	81.58%	5013970	LT629710.1
Alloactinosynnema sp. L-07 genome assembly Alloactinosynnema sp. L-07, chromosome : I	Alloactinosynnema sp. L...	79.7	79.7	35%	3e-10	79.12%	7405112	LN850107.1
Arthrobacter sp. 7749 genome	Arthrobacter sp. 7749	77.0	77.0	35%	1e-09	78.26%	4125942	CP022462.1
Paeniglutamicibacter sp. Y32M11 chromosome, complete genome	Paeniglutamicibacter sp...	77.0	77.0	35%	1e-09	78.26%	4173408	CP079107.1
Yinghuangia sp. ASG 101 chromosome, complete genome	Yinghuangia sp. ASG 101	82.4	82.4	38%	2e-11	78.22%	8343587	CP088911.1
Saccharothrix sp. AJ9571 chromosome, complete genome	Saccharothrix sp. AJ9571	98.7	98.7	49%	3e-16	78.12%	9915751	CP091382.1
Amycolatopsis sp. CA-230715 chromosome, complete genome	Amycolatopsis sp. CA-2...	95.1	95.1	51%	4e-15	77.44%	10363158	CP059997.1
Actinotalea sp. JY-7876 chromosome, complete genome	Actinotalea sp. JY-7876	78.8	78.8	42%	3e-10	77.19%	3546809	CP059493.1
Saccharomonospora xinjiangensis strain 31sw chromosome, complete genome	Saccharomonospora xin...	89.7	89.7	49%	2e-13	76.56%	4713652	CP038101.1
Prauserella marina strain DSM 45268 chromosome, complete genome	Prauserella marina	89.7	89.7	49%	2e-13	76.56%	6851918	CP016353.1
Amycolatopsis methanolica 239, complete genome	Amycolatopsis methanol...	89.7	89.7	49%	2e-13	76.56%	7237391	CP009110.1
Pseudonocardia sp. CGMCC 4.1532 chromosome, complete genome	Pseudonocardia sp. CG...	91.5	91.5	50%	4e-14	76.52%	6490139	CP060131.1
Rothia kristinae strain FDAARGOS_1001 chromosome, complete genome	Rothia kristinae	77.0	77.0	41%	1e-09	76.42%	2513835	CP066078.1
Amycolatopsis sp. YIM 10 chromosome, complete genome	Amycolatopsis sp. YIM 10	95.1	95.1	49%	4e-15	76.38%	10312959	CP045480.1
Saccharomonospora viridis DSM 43017, complete genome	Saccharomonospora viri...	95.1	95.1	49%	4e-15	76.38%	4308349	CP001683.1
Micromonospora krabiensis strain DSM 45344 genome assembly, chromosome: I	Micromonospora krabie...	82.4	82.4	42%	2e-11	76.36%	7074838	LT598496.1
Micromonospora carbonacea strain africana chromosome, complete genome	Micromonospora carbon...	82.4	82.4	42%	2e-11	76.36%	6786809	CP058905.1
Micromonospora carbonacea strain aurantiaca chromosome, complete genome	Micromonospora carbon...	82.4	82.4	42%	2e-11	76.36%	7528335	CP058322.1
Nakamurella sp. PAMC28650 chromosome, complete genome	Nakamurella sp. PAMC2...	92.4	92.4	49%	4e-14	76.15%	5629850	CP060298.1
Geodermatophilus obscurus DSM 43160, complete genome	Geodermatophilus obsc...	92.4	92.4	49%	4e-14	75.78%	5322497	CP001867.1
Allosaccharopolyspora coralli strain E2A chromosome, complete genome	Allosaccharopolyspora c...	90.6	90.6	49%	2e-13	75.59%	4832009	CP045929.1
Amycolatopsis albispora strain WP1 chromosome, complete genome	Amycolatopsis albispora	90.6	90.6	49%	2e-13	75.59%	9121443	CP015163.1
Saccharopolyspora pogona strain NRRL30141 chromosome, complete genome	Saccharopolyspora pog...	90.6	90.6	49%	2e-13	75.59%	9436548	CP031142.1
Saccharopolyspora spinosa strain CCTCC M206084 chromosome, complete genome	Saccharopolyspora spin...	90.6	90.6	49%	2e-13	75.59%	8876435	CP061007.1

(b)

☑	Saccharopolyspora spinosa strain CCTCC M206084 chromosome, complete genome	Saccharopolyspora spin...	90.6	90.6	49%	2e-13	75.59%	8876435	CP061007.1
☑	Actinoalloteichus hoggarensis strain DSM 45943 chromosome, complete genome	Actinoalloteichus hoggar...	89.7	89.7	49%	2e-13	75.57%	6606717	CP022521.1
☑	Allokutzneria albata strain DSM 44149 genome assembly, chromosome: I	Allokutzneria albata	87.8	87.8	49%	5e-13	75.38%	8565619	LT629701.1
☑	Kutzneria sp. CA-103260 chromosome, complete genome	Kutzneria sp. CA-103260	87.8	87.8	49%	5e-13	75.38%	11609901	CP073318.1
☑	Actinoalloteichus sp. AHMU CJ021 chromosome, complete genome	Actinoalloteichus sp. AH...	87.8	87.8	49%	5e-13	75.00%	6822750	CP025990.1
☑	Pseudonocardia dioxanivorans CB1190, complete genome	Pseudonocardia dioxani...	87.8	87.8	49%	5e-13	75.00%	7096571	CP002593.1
☑	Actinosynnema pretiosum subsp. pretiosum strain ATCC 31280 chromosome, complete genome	Actinosynnema pretiosu...	86.9	86.9	51%	2e-12	75.00%	8340266	CP029607.1
☑	Actinosynnema pretiosum strain X47 chromosome, complete genome	Actinosynnema pretiosum	86.9	86.9	51%	2e-12	75.00%	8131572	CP023445.1
☑	Actinosynnema pretiosum subsp. pretiosum strain ATCC 31280 chromosome, complete genome	Actinosynnema pretiosu...	86.9	86.9	51%	2e-12	75.00%	8340580	CP073249.1
☑	Actinosynnema mirum DSM 43827, complete genome	Actinosynnema mirum D...	86.9	86.9	51%	2e-12	75.00%	8248144	CP001630.1
☑	Arthrobacter sp. NtRootA9 DNA, complete genome	Arthrobacter sp. NtRootA9	86.0	86.0	49%	2e-12	74.80%	4177067	AP024636.1
☑	Micromonospora endophytica (Xie et al. 2001) Li et al. 2019 NBRC 109090 DNA, complete genome	Micromonospora endop...	73.4	73.4	42%	1e-08	74.55%	6586761	AP023358.1
☑	Amycolatopsis acidiphila strain KCTC 39523 chromosome, complete genome	Amycolatopsis acidiphila	76.1	76.1	49%	3e-09	74.22%	8183622	CP090063.1
☑	Arthrobacter sp. PGP41 chromosome, complete genome	Arthrobacter sp. PGP41	81.5	81.5	49%	8e-11	74.02%	4270237	CP026514.1
☑	Saccharopolyspora erythraea strain NRRL 23338 chromosome, complete genome	Saccharopolyspora eryt...	81.5	81.5	49%	8e-11	74.02%	8305652	CP069353.1
☑	Saccharopolyspora erythraea NRRL2338 complete genome	Saccharopolyspora eryt...	81.5	81.5	49%	8e-11	74.02%	8212805	AM420293.1
☑	Kibdelosporangium phytohabitans strain KLBMP1111 chromosome, complete genome	Kibdelosporangium phyt...	77.9	77.9	52%	1e-09	73.91%	11759770	CP012752.1
☑	Streptomyces antibioticus strain DSM 41481 chromosome, complete genome	Streptomyces antibioticus	74.3	74.3	49%	1e-08	73.85%	8473575	CP050692.1
☑	Amycolatopsis sp. FDAARGOS 1241 strain FDAARGOS_1241 chromosome, complete genome	Amycolatopsis sp. FDA...	74.3	74.3	49%	1e-08	73.85%	9796944	CP069526.1
☑	Kocuria palustris strain MU14/1, complete genome	Kocuria palustris	85.1	85.1	56%	7e-12	73.79%	2854447	CP012507.1
☑	Modestobacter marinus str. BC501 chromosome, complete genome	Modestobacter marinus	81.5	81.5	55%	8e-11	73.29%	5575517	FO203431.1
☑	Salinispora tropica CNB-440, complete genome	Salinispora tropica CNB-...	79.7	79.7	50%	3e-10	73.28%	5183331	CP000667.1
☑	Streptomyces sp. JB150 chromosome	Streptomyces sp. JB150	77.0	77.0	49%	1e-09	73.23%	7318937	CP049780.1
☑	Streptomyces tendae strain 139 chromosome, complete genome	Streptomyces tendae	77.0	77.0	49%	1e-09	73.23%	6900068	CP043959.1

(c)

☑	Streptomyces sp. ETH9427 chromosome	Streptomyces sp. ETH9...	77.0	77.0	49%	1e-09	73.23%	7745357	CP029624.1
☑	Streptomyces nigra strain 452 chromosome, complete genome	Streptomyces nigra	77.0	77.0	49%	1e-09	73.23%	7641029	CP029043.1
☑	Streptomyces glaucescens strain GLA.O, complete genome	Streptomyces glaucescens	77.0	77.0	49%	1e-09	73.23%	7453200	CP009438.1
☑	Saccharopolyspora erythraea strain SCSIO 07745 chromosome, complete genome	Saccharopolyspora eryt...	77.0	77.0	49%	1e-09	73.23%	8243897	CP054839.1
☑	Streptomyces ferrugineus strain CCTCC AA2014009 chromosome, complete genome	Streptomyces ferrugineus	77.0	77.0	49%	1e-09	73.23%	9959088	CP063373.1
☑	Catellatospora sp. IY07-71 DNA, complete genome	Catellatospora sp. IY07-71	78.8	78.8	53%	3e-10	72.86%	8595826	AP023360.1
☑	Actinoalloteichus hymeniacidonis strain HPA177(T) (=DSM 45092(T)) chromosome, complete genome	Actinoalloteichus hymen...	74.3	74.3	49%	1e-08	72.66%	6306386	CP014859.1
☑	Streptomyces sp. NA02536 chromosome, complete genome	Streptomyces sp. NA02...	72.5	72.5	49%	4e-08	72.44%	7707218	CP054939.1
☑	Streptomyces sp. Z423-1 chromosome	Streptomyces sp. Z423-1	72.5	72.5	49%	4e-08	72.44%	9460473	CP053109.1
☑	Streptomyces sp. S1A1-8 chromosome, complete genome	Streptomyces sp. S1A1-8	72.5	72.5	49%	4e-08	72.44%	12037094	CP041612.2
☑	Streptomyces sp. S1D4-23 chromosome	Streptomyces sp. S1D4-23	72.5	72.5	49%	4e-08	72.44%	12057750	CP041613.2
☑	Streptomyces sp. S1A1-7 chromosome	Streptomyces sp. S1A1-7	72.5	72.5	49%	4e-08	72.44%	11713216	CP041604.2
☑	Streptomyces sp. RLB3-6 chromosome	Streptomyces sp. RLB3-6	72.5	72.5	49%	4e-08	72.44%	12338170	CP041602.2
☑	Streptomyces sp. RLB3-17 chromosome, complete genome	Streptomyces sp. RLB3-17	72.5	72.5	49%	4e-08	72.44%	12022941	CP041610.2
☑	Streptomyces albogriseolus strain LBX-2 chromosome, complete genome	Streptomyces albogrise...	72.5	72.5	49%	4e-08	72.44%	7210477	CP042594.1
☑	Streptomyces galilaeus strain ATCC 14969 chromosome, complete genome	Streptomyces galilaeus	72.5	72.5	49%	4e-08	72.44%	7756194	CP023703.1
☑	Streptomyces sp. WAC06273 strain WAC6273 substr. delta orf15 pCRISPR-Cas9 chromosome, compl...	Streptomyces sp. WAC0...	72.5	72.5	49%	4e-08	72.44%	7556430	CP042278.1
☑	Streptomyces sp. S1A1-3 chromosome	Streptomyces sp. S1A1-3	72.5	72.5	49%	4e-08	72.44%	12042091	CP041611.1
☑	Streptomyces sp. RLB1-9 chromosome, complete genome	Streptomyces sp. RLB1-9	72.5	72.5	49%	4e-08	72.44%	11940408	CP041654.1
☑	Streptomyces variabilis strain ARRS001 chromosome	Streptomyces variabilis	72.5	72.5	49%	4e-08	72.44%	6101786	CP040941.1
☑	Streptomyces sp. SGAir0924 chromosome, complete genome	Streptomyces sp. SGAir...	72.5	72.5	49%	4e-08	72.44%	7653753	CP027297.1
☑	Streptomyces fungicidicus strain F2 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6740841	CP086178.1
☑	Streptomyces fungicidicus strain F10 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6740826	CP086174.1
☑	Streptomyces fungicidicus strain F9 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6740828	CP086175.1

(d)

☑ Streptomyces fungicidicus strain ATCC 21013 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6740828	CP086173.1
☑ Streptomyces fungicidicus strain F8 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6741257	CP086176.1
☑ Streptomyces fungicidicus strain F1 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6740829	CP086179.1
☑ Streptomyces fungicidicus strain F7 chromosome, complete genome	Streptomyces fungicidicus	72.5	72.5	49%	4e-08	72.44%	6741246	CP086177.1
☑ Streptomyces sp. A144 chromosome, complete genome	Streptomyces sp. A144	72.5	72.5	49%	4e-08	72.44%	8085430	CP083267.1
☑ Streptomyces sp. MG62 chromosome, complete genome	Streptomyces sp. MG62	72.5	72.5	49%	4e-08	72.44%	8599513	CP075896.1
☑ Streptomyces chromofuscus strain DSM 40273 chromosome, complete genome	Streptomyces chromofu...	72.5	72.5	49%	4e-08	72.44%	7483282	CP063374.1
☑ Streptomyces lincolnensis strain B48 chromosome, complete genome	Streptomyces lincolnensis	72.5	72.5	49%	4e-08	72.44%	10008966	CP046024.1

(e)

Appendix 9-Figure 9.2: Figures (a-e) represent the overall specificity of the *senX3-regX3* sequence. Importantly, in this search, the search parameters were customised (*Mycobacterium* genus was excluded and somewhat similar sequences were chosen) to check the presence of the gene in other bacteria. It was present in a wide range of *Mycobacterium*, *Pseudonocardia and Streptomyces* species. It was not specific to the Mtb at all.

198

APPENDIX 10

IN-SILICO SPECIFICITY TESTING

OF HRM ASSAYS

Primer design

16S rRNA assay

Details of published 16S rRNA primers

Appendix 10-Table 10.1: Published universal primers capable of amplifying 1.3kb of 16S

rRNA gene.

Primer name	Sequence detail	Target region (1.3 Kb)	Reference
27F	5'-AGAGTTTGATCMTGGCTCAG-3'	16S rRNA region	(Turner *et al.*, 1999)
1391R	5'-GACGGGCGGTGTGTRCA-3'	16S rRNA region	(Frank *et al.*, 2008)

Designed 16S rRNA assay

In silico specificity of the designed primer sets were checked on the relevant published

sequences. Primers having high specificity were used.
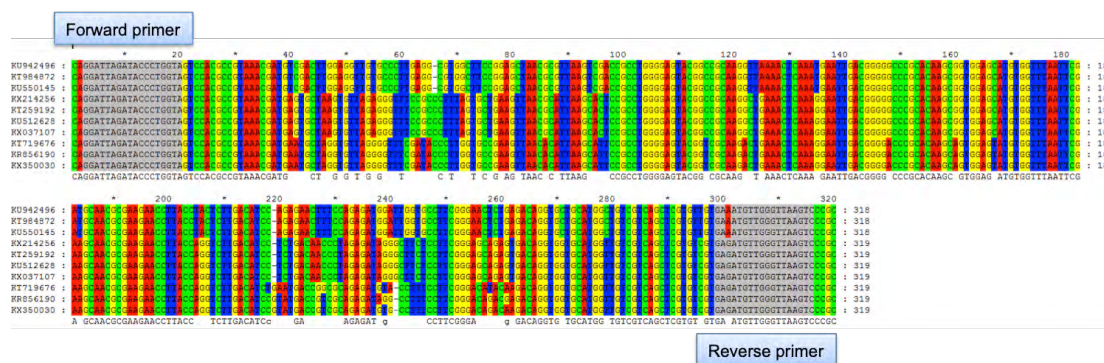
Appendix 10-Figure 10.1: Representation of the binding sites of 16S rRNA primer set for HRM against the aligned published sequences of different organisms' DNA. The sequence of interest was in between the flanking primers set.

The grey coloured regions (Figure 4.1) indicated the foreword (RibRNA-312-F) and the reverse primer (RibRNA-312-R) binding sites respectively. The foreword primer sequence approximately bound at position 0 to 20bp, while 300 till 320bp was the position of reverse primer binding site. The 320bp (approx.) long sequence positioned inside the flanking primer sequences was the location of the 16S rRNA sequence required for the HRM testing to differentiate between organisms to their species level.

The *rpoB* assay
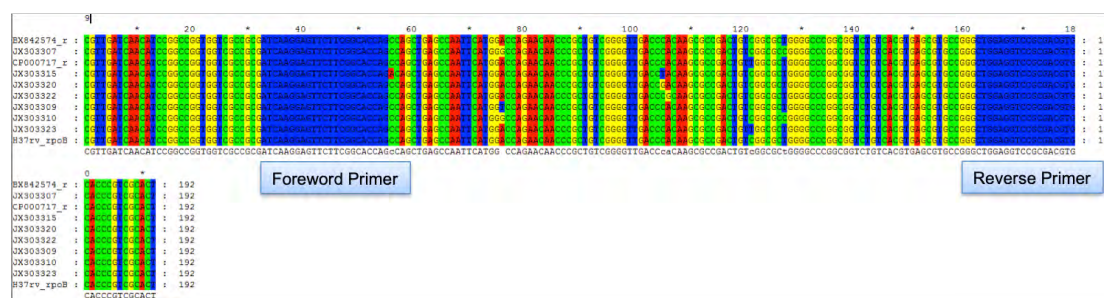
Appendix 10-Figure 10.2: Demonstration of the binding pattern of the *rpoB* primer set for HRM against the published genomic database. The target located in between the flanking primers set. Published sequences detail is in Appendix 14, Table 14.2.

The blue coloured areas (Figure 4.2) indicated the foreword (Mtb-*rpoB*-147-F) and the reverse primer (Mtb-*rpoB*-147-R) binding sites respectively. The foreword primer sequence approximately bound at position 35 to 55bp, while 160 till 180bp was the position of

reverse primer binding site. The 170bp (approx.) long sequence positioned inside the flanking primer sequences was the location of the *rpoB* segment required for the mutation(s) scanning by the HRM assay.

The *katG* assay
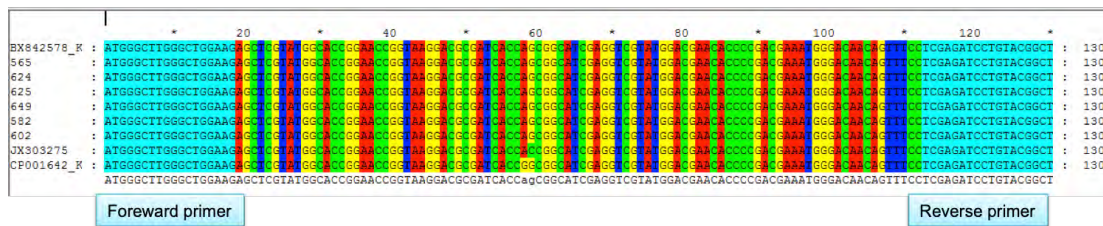
Appendix 10-Figure10.3: Illustration of binding sites of the *katG* primer set against the published *katG* gene sequences. The blue colour area represented the binding sites of the forward and reverse primer. The sequence between the set was the target sequence. The published katG sequences are described in Appendix 14, Table 14.4.

The light blue coloured areas (Figure 4.3) indicated the foreword (Mtb-*katG*-130-F) and the reverse primer (Mtb-*katG*-130-R) binding sites respectively. The foreword primer sequence bound at position 0 to 20bp, while 112 till 130bp was the position of reverse primer binding site. The 130bp long sequence positioned inside the flanking primer sequences was the location of the *katG* segment. It was required to scan any mutation(s) from the isolates by the HRM assay.

Therefore, it was concluded that the designed primer sets were highly specific towards their target regions. Therefore, it was concluded that the designed primer sets were fit to apply on the panel for their differentiation testing.

Appendix 10-Table 10.2: The C_t values recorded during the reproducibility data of the published IS6110 assay. Three different H37Rv genomic DNA dilutions were run in duplicates three times and the C_t value was noticed and compared.

DNA name and dilution	C_t value on 1st attempt	C_t value on 2nd attempt	C_t value on 3rd attempt
H37Rv genomic DNA 10^{-1}	23.01	24.23	24.27
H37Rv genomic DNA 10^{-1}	22.76	24.08	24.24
H37Rv genomic DNA 10^{-2}	29.82	31.44	31.18
H37Rv genomic DNA 10^{-2}	30.18	31.08	31.29
H37Rv genomic DNA 10^{-3}	33.38	35.45	35.37
H37Rv genomic DNA 10^{-3}	33.71	35.27	35.18

Appendix 10-Table 10.3: The C_t values recorded during the reproducibility data of the JCU IS6110 assay. Three different IS6110 plasmid dilutions were run in duplicates three times and the C_t value was noticed and compared.

Plasmid name and dilution	C_t value on 1st attempt	C_t value on 2nd attempt	C_t value on 3rd attempt
IS6110 plasmid 10^{-1}	17.02	17.20	16.18
IS6110 plasmid 10^{-1}	17.05	17.30	15.90
IS6110 plasmid 10^{-1}	16.93	17.16	16.01
IS6110 plasmid 10^{-2}	19.77	20.28	19.12
IS6110 plasmid 10^{-2}	20.02	20.19	19.35
IS6110 plasmid 10^{-2}	19.81	20.25	19.02
IS6110 plasmid 10^{-3}	22.05	22.27	20.88
IS6110 plasmid 10^{-3}	22.10	22.28	20.90
IS6110 plasmid 10^{-3}	22.01	22.36	21.06

Appendix 10-Table 10.4: The C_t values recorded during the reproducibility data of the JCU *rpoB* assay. Three different H37Rv genomic DNA dilutions were run in duplicates three times and the C_t value was noticed and compared.

DNA name and dilution	C_t value on 1st attempt	C_t value on 2nd attempt	C_t value on 3rd attempt
H37Rv genomic DNA 10^{-1}	27.18	26.35	27.03
H37Rv genomic DNA 10^{-1}	27.09	26.36	27.25
H37Rv genomic DNA 10^{-2}	34.57	32.55	33.38
H37Rv genomic DNA 10^{-2}	34.04	33.17	33.29
H37Rv genomic DNA 10^{-3}	39.17	36.81	37.27
H37Rv genomic DNA 10^{-3}	38.50	36.74	37.72

APPENDIX 11

DETAILS OF REACTION MIX OF HRM ASSAYS

16S assay

Appendix 11-Table 11.1: Details of reaction mix of 16S rRNA assay for HRM

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	qPCR Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer RibRNA-312-F	0.8 µM	1.6
10 µM	Reverse Primer RibRNA-312-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total volume of one reaction mix			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 59°C for 15 seconds and 72°C for 20 seconds for 35 cycles with green channel. Melt condition: 72-95°C, 0.3°C/sec.

The *rpoB* assay

Appendix 11-Table 11.2 Details of constituents of the *rpoB* assay reaction mix for HRM

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	qPCR Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer Mtb-*rpoB*-147-F	0.8 µM	1.6
10 µM	Reverse Primer Mtb-*rpoB*-147-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total volume of one reaction mix			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds for 40 cycles with green channel. Melt condition: 72-97°C, 0.3°C/sec.

The *katG* assay

Appendix 11-Table 11.3: The *katG* assay reaction mix description for melt analysis

Stock Concentration	Reagent	Final Concentration	Reaction (µL)
2X	qPCR Master Mix (Promega A#6002)	1X	10
10 µM	Forward Primer Mtb-*katG*-130-F	0.8 µM	1.6
10 µM	Reverse Primer Mtb-*katG*-130-R	0.8 µM	1.6
	Nuclease free water		4.8
Template			2
Total volume of one reaction mix			20

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 60°C for 15 seconds for 35 cycles with green channel. Melt condition: 72-95°C, 0.3°C/sec.

APPENDIX 12

ENLARGED IMAGES OF

SEQUENCING DATA FOR COMPARISON

```
                 *        20        *        40        *        60        *        80        *       100        *       120        *       140        *       160        *       180
MG995565  : -GATTAGATACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATCCGTGCCGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTC : 179
M.intracel : G................................................................................................................................................................................. : 180
M.fortuitu : G................................................................................................................................................................................. : 180
M.bovis312 : AATCC.C..G.T...CGG.C..C..T.AATTC.TT..A..TT...T..C..CCG.A..C..CA.------.C.GG...CT...T..G...GC...GG.A..AA..A..G.AAC..A..C--...GT..C..CC.TTTAC.G..T..A.TAC.---------.G.T..C.AATCCTG... : 163
M.avium-31 : G....................T.........T....C.......GGA.A....AC.TTT....C................G................................................................................C..... : 180
M.abscessu : G...........................................................................................................T.....A........................................................... : 180
M.intracel : G................................................................................................................................................................................. : 180
M.peregrin : G................................................................................................................................................................................. : 180


                 *       200        *       220        *       240        *       260        *       280        *       300        *
MG995565  : GATGCAACGCGAAGAACCTTACCTGGGTTTGACATGCACAGGACGCGTCTAGAGATAGGCGTTCCCTTGTGGCCTGTGTGCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTC- : 311
M.intracel : ..........................................................................................................................- : 312
M.fortuitu : ..........................................................................................................................- : 312
M.bovis312 : .C.C.----------..ACG.T.TC.C.CCT..GCGT...TTACT.C.C.....CCC..C..G.CACC.......A.CTGC.CAT.C..CC...A-.AC....CA..C.-----------------C.- : 265
M.avium-31 : ..................AA..C.......A...C...TCGT..C......G.TTC......T..G....A...........T..........................................- : 312
M.abscessu : ..........................................TAC..........TA..............................G...........................-........- : 311
M.intracel : ...............................................CAG........TTG..............................................................- : 312
M.peregrin : .............................................GCGGC.....GTCGT...............................................................- : 312
```

Figure 5.4: The sequencing data of 16S rRNA gene of different members of the *Mycobacterium* genus compared to the reference strain of the Mtb (MG995565).

```
           99
                   *        20        *        40        *        60        *        80        *       100        *       120        *       140
BX842574_r : ATCAAGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTG : 147
BC-30      : ..........................................................G.................................................................................... : 147
M68        : ..................................................................................T............................................................. : 147
JQ411539   : .......................G...............C..TC.............C..C..........C........C.........C........C.....T.....- : 146
CP029332   : .....................TC...G................C.C..........C.........C.G..T........C.G.....G............................................ : 147
AY544893   : .............................----------........................................................................................................... : 140
```

Figure 5.7: The sequencing data of the *rpoB* gene of different members of the *Mycobacterium* genus compared to the Mtb reference strain DNA (BX842574).

APPENDIX 13

DETAILS OF THE PRIMER SET

FOR OLIGO EXTENSION

Appendix 13-Table 13.1: Primer pair for *rpoB* oligo sequence extension

Primer name	Sequence	Target gene	Length of the amplicon (bp)
Mtb-*rpoB*-147H-F	5'-ATCAAGGAGTTCTTCGGCACCAGC-3'	*rpoB*	147
Mtb-*rpoB*-147H-R	5'-CGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTG-3'	*rpoB*	147

APPENDIX 14

DETAILS OF PUBLISHED SEQUENCES USED

IN SEQUENCING DATA FOR COMPARISON

Appendix 14-Table 14.1: Details of published sequences used to compare with sequencing data.

Accession Number	Organism	Gene	Isolation source	Country	Year
LC005454	*M.tuberculosis strain F9*	IS6110	*Homo sapiens*	Japan	2016
LC005455	*M.tuberculosis strain F16*	IS6110	*Homo sapiens*	Japan	2016
LC005456	*M.tuberculosis strain F20*	IS6110	*Homo sapiens*	Japan	2016
LC005457	*M.tuberculosis strain F31*	IS6110	*Homo sapiens*	Japan	2016
LC005458	*M.tuberculosis strain F33*	IS6110	*Homo sapiens*	Japan	2016
LC005462	*M.tuberculosis strain F60*	IS6110	*Homo sapiens*	Japan	2016
LC005465	*M.tuberculosis strain F97*	IS6110	*Homo sapiens*	Japan	2016
LC005472	*M.tuberculosis strain F158*	IS6110	*Homo sapiens*	Japan	2016
LC005475	*M.tuberculosis strain F194*	IS6110	*Homo sapiens*	Japan	2016
LC005476	*M.tuberculosis strain F200*	IS6110	*Homo sapiens*	Japan	2016
LC005477	*M.tuberculosis strain F208*	IS6110	*Homo sapiens*	Japan	2016

LC005482	*M.tuberculosis strain F309*	IS6110	*Homo sapiens*	Japan	2016
CP008744	*M. tuberculosis variant bovis BCG strain 3281*	Whole genome	*Homo sapiens*	China	2015
CP012095	*M. bovis strain 1595*	Whole genome	Larynooharyngeal lymph node of animal (Bos Taurus coreanae) with TB	Korea	2015
LC005473	*M. tuberculosis strain F178*	IS6110	*Homo sapiens*	Japan	2016
LC005459	*M. tuberculosis strain F41*	IS6110	*Homo sapiens*	Japan	2016
LC005471	*M. tuberculosis strain F145*	IS6110	*Homo sapiens*	Japan	2016
LC005481	*M.tuberculosis strain F302*	IS6110	*Homo sapiens*	Japan	2016
LC005463	*M. tuberculosis strain F76*	IS6110	*Homo sapiens*	Japan	2016
AF357173	*M.tuberculosis isolate DS71*	IS6110		USA	2001

Appendix 14-Table 14.2: Details of published sequences used to compare with sequencing data.

Accession Number	Organism	Gene	Isolation source	Country	Year
BX842574	*M. tuberculosis H37Rv*	Whole genome		UK	1998
CP000611	*M. tuberculosis H37Ra ATCC 25177*	Whole genome	ATCC	China	2008
CP003248	*M. tuberculosis H37Rv*	Whole genome		USA	2014
AE000516	*M. tuberculosis CDC1551*	Whole genome	CDC	USA	2002
AM412059	*M. tuberculosis variant bovis BCG str. Moreau RDJ*	Whole genome	SL21 FAP RJ Passage B8S2 vaccine culture	Brazil	2012
AP010918	*M. bovis* BCG str. Tokyo 172 DNA	Whole genome		Japan	2009
BX248336	*M. bovis* subsp. bovis AF2122/97	Whole gonome		UK	2003
CP002095	*M. tuberculosis* variant bovis BCG str. Mexico 1931	Whole genome	Laboratorios de Biológicos y Reactivos de Mexico	Mexico	2011

MSGRPOB					
AP012340	*M. tuberculosis* str. Erdman (ATCC 35801)	Whole genome	ATCC	Japan	2012
CP001662	*M. tuberculosis* KZN 4207	Whole genome	Homo sapiens	South Africa	2009
CP002992	*M. tuberculosis* CTRI-2	Whole genome	Central region of the European Russia	Russia	2013
HE608151	*M. tuberculosis* UT205	Whole genome	TB patient	Colombia	2012
JX303328	*M. tuberculosis* MTB001	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303329	*M. tuberculosis* MTB003	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303332	*M. tuberculosis* MTB007	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303331	*M. tuberculosis* MTB006	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303330	*M. tuberculosis* MTB005	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012

JX303307	*M. tuberculosis* strain 14641	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303308	*M. tuberculosis* strain 14655	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303311	*M. tuberculosis* strain 14666	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303314	*M. tuberculosis* strain 14949	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303313	*M. tuberculosis* strain 14671	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
CP000717	*M. tuberculosis* F11				2013
JX303312	*M. tuberculosis* strain 14669	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303321	*M. tuberculosis* strain 3100029	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303318	*M. tuberculosis* strain 2961678	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303326	*M. tuberculosis* strain 3275126	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012

JX303316	*M. tuberculosis* strain 2919984	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303317	*M. tuberculosis* strain 2937646	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303319	*M. tuberculosis* strain 2997164	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303325	*M. tuberculosis* strain 3264812	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303327	*M. tuberculosis* strain 3290183	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303324	*M. tuberculosis* strain 3247469	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303315	*M. tuberculosis* strain 2875299	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303320	*M. tuberculosis* strain 3050732	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303322	*M. tuberculosis* strain 3150565	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303309	*M. tuberculosis* strain 14657	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012

JX303310	*M. tuberculosis* strain 14665	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303323	*M. tuberculosis* strain 3174991	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
CP023591	*M. tuberculosis* strain TBV5365	Whole genome	Homo sapiens (sputum)	Peru	2008
CP023639	*M. tuberculosis* strain TBV4768	Whole genome	Homo sapiens (sputum)	Peru	2008
CP023708	*Mycobacterium* sp. 3/86Rv	Whole genome	Cattle (mediastinal lymph node)	India	2013

Appendix 14-Table 14.3 Details of published sequences used to compare with sequencing data.

Accession Number	Organism	Gene	Isolation source	Country	Year
H37Rv_refr (BX842574)	*M.tuberculosis* H37Rv	Whole genome		UK	1998
JX303307	*M.tuberculosis* H37Rv strain 14641	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
CP000717	*M.tuberculosis* H37Rv strain F11				2013
JX303315	*M.tuberculosis* H37Rv strain 2875299	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303320	*M. tuberculosis* strain 3050732	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303322	*M. tuberculosis* strain 3150565	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303309	*M. tuberculosis* strain 14657	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303310	*M. tuberculosis* strain 14665	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012
JX303323	*M. tuberculosis* strain 3174991	*rpoB* gene	Homo sapiens (sputum)	South Africa	2012

Accession Number	Organism	Gene	Isolation source	Country	Year
BX842578	*M.tuberculosis* H37Rv	Whole genome		UK	1998
JX303275	*M. tuberculosis* strain 3290183	katG gene	Homo sapiens (sputum)	South Africa	2012
CP001642	*M. tuberculosis* CCDC 5180	Whole genome	Patient (Homo Sapiens) with TB (sputum)	China	2011

Appendix-Table 14.4 Details of published *katG* sequences used to compare with sequencing data.

APPENDIX 15

CHARACTERISATION OF THE PANEL ORGANISMS

15.1 Introduction

Saah and Hoover defined analytical specificity as the ability of an assay to measure one particular organism or a substance, rather than others, in a sample (Saah and Hoover, 1997). This signifies the requirement of the irrelevant or the other organisms so that the assay(s) to be tested for its specificity can specifically identify the target organism. This project aimed at the development and validation of in-house assays to identify and characterise the Mtb in terms of the strain drug susceptibility. For assays validation, for the specificity determination, testing the assay with and without the relevant organisms was required. This would reveal any chance(s) of cross-reaction in the organisms' DNA samples having no target sequence in their genome. Keeping this in view, a set of around 40 cultured grown organisms were sorted.

The genomic DNA from each of those grown bacteria was extracted and stored at -80°C. This set was comprised of taxonomically and clinically relevant organisms. The 16S rRNA gene is present in all the bacteria. Therefore, a broad range PCR using a set of published universal primers targeting the 16S rRNA gene for the identification of all the bacterial isolates to their species level was performed. Published primer selection was done from the following website:

http://lutzonilab.org/16s-ribosomal-dna/

Further, the amplified 16S rRNA sequence from each isolate was subjected to the Sanger sequencing. This panel of DNA after their verification by sequencing was used for the determination of specificity of the developed assays (mentioned in the Chapter 4).

15.2 Selection and acquisition of the panel organisms

Since *Mycobacterium tuberculosis* is facultative intracellular bacteria (Sharma *et al.*, 2009). It initiates the infection after getting a successful entry via water droplets

into the lung alveolus (Orme, 2014). Acquisition of similar pathogens that infect the lower respiratory tract of humans was ideal for the specificity analysis. In addition, the second manifestation of TB, the EPTB was also kept in the mind and the pathogens infecting other sites such as skin were also included.

Around 40 different organisms (Table15.1) were acquired from different sources based on their ease of availability. They were categorised as a set of clinically and taxonomically relevant organisms. The former set was comprised of 33 organisms in total. Of 33, all were gained as cultured organisms except the Ureaplasma plasmid suspension and the DNA 12, an unidentified genomic DNA sample. The seven taxonomically relevant organisms were comprised of bacteria of *Mycobacterium* genus with varying species. The genomic DNA of all seven was availed instead of the live organisms due to their infectivity risk.

15.3 Identification of the acquired panel organisms

The 31 live organisms were obtained in the cultured form in separate petri plates. These plates were provided by A/Prof. Jeffrey Warner. Initial morphological analysis of all the organisms was done by microscopy. Further, the biochemical testing using specific API kits was conducted to phenotypically confirm the growth and isolation of all the desired organisms. Gram staining for all the clinically relevant organisms and Zeihl Neelsen (ZN) staining for the taxonomically relevant organisms was carried out to observe the characteristic morphology of the organisms. Different API kits suggested specifically for organisms were used to confirm the right organisms based on their biochemical reactions. All the phenotypic observations were recorded on an excel sheet to verify the microscopy and biochemical test results with the sequencing. The whole data has been mentioned in the Appendix 18.

15.4 Extraction of the genomic DNA from all the bacterial isolates

Two bacterial colonies (grown for 24 hours) from each of the 31 plates were aseptically transferred to a microfuge tube containing 200 μL of sterile distilled water. A new disposable loop was used every time transfer was done. This was carried out under a Biosafety hood. All the 40-microfuge tubes were then transferred to a dry heating block to perform the cell lysis at 63°C for 10 minutes. 100μL of the cell lysate from each of all 31 heated microfuge tubes was separately collected and transferred into the new sterilised microfuge tubes to initiate genomic DNA extraction.

High pure PCR template preparation kit (Product No. 11796828001) from Roche was used to extract the genomic DNA of 31 all bacterial isolates. Each DNA isolate (40) was tested for its concentration and purity ($A_{260/280}$) using Nano spectrometer. All the DNA isolates were stored at -80°C freezer. The protocols were performed strictly according to the manufacturer's instructions. The section 3.10 and 3.11 of Chapter 3 described the whole procedure.

Melt analysis in Bio molecular systems (BMS) MIC™ real-time PCR was also carried out. This was to confirm the purity of the amplified 16 S sequence from each of the 40 DNA. Based on the acquired isolates specific melt curves, comparative analysis for differences was further done. The details of the primers used, and the cycling has been mentioned in 15.5.1 and 15.5.2 section respectively.

Table 15.1: List of organisms and their respective source of acquisition.

Organisms	Source
Clinically associated organisms	
Acinetobacter baumannii	JCU Microbiology Culture Collection
Aeromonas hydrophila	JCU Microbiology Culture Collection
Bacillus cereus	JCU Microbiology Culture Collection
Bacillus mycoides	JCU Microbiology Culture Collection
Brachybacterium sp.	JCU Microbiology Culture Collection
Enterobacter cloacae	JCU Microbiology Culture Collection
Enterococcus faecalis (ATCC)	JCU Microbiology Culture Collection
Erysipelothrix rhusiopathiae	JCU Microbiology Culture Collection
Escherichia coli (ATCC)	JCU Microbiology Culture Collection
Alcaligenes sp.	JCU Microbiology Culture Collection
Klebsiella pneumoniae	JCU Microbiology Culture Collection
Listeria monocytogenes	JCU Microbiology Culture Collection
Micrococcus luteus	JCU Microbiology Culture Collection
Moraxella catarrhalis	JCU Microbiology Culture Collection
Neisseria gonorrhoeae	JCU Microbiology Culture Collection
Neisseria meningitidis	JCU Microbiology Culture Collection
Nocardia sp.	JCU AITHM PC3 lab
Paenibacillus polymyxa	JCU Microbiology Culture Collection
Proteus mirabilis	JCU Microbiology Culture Collection
Pseudomonas aeruginosa (ATCC)	JCU Microbiology Culture Collection
Salmonella typhimurium (ATCC)	Dr. Jenny Elliman, JCU
Staphylococcus aureus (ATCC)	JCU Microbiology Culture Collection
Staphylococcus epidermidis (ATCC)	JCU Microbiology Culture Collection
Staphylococcus saprophyticus	JCU Microbiology Culture Collection
Streptococcus agalactiae	JCU Microbiology Culture Collection
Streptococcus dysgalactiae	JCU Microbiology Culture Collection
Streptococcus pneumoniae (ATCC)	JCU Microbiology Culture Collection
Streptococcus pyogenes	JCU Microbiology Culture Collection

Streptococcus viridians	JCU Microbiology Culture Collection
Ureaplasma plasmid	DNA 2.0
Corynebacterium diphtheriae	JCU Microbiology Culture Collection
Bacillus cereus	JCU Microbiology Culture Collection
Taxonomically associated organisms	
M. abscessus	JCU AITHM PC3 lab
M. avium	Dr. Jackie Picard, JCU
M. bovis	JCU AITHM PC3 lab
M. intracellulare	Dr. Jackie Picard, JCU
M. fortuitum	Dr. Jackie Picard, JCU
M. peregrinum	Dr. Jackie Picard, JCU
DNA 12 (M. intracellulare)	Dr. Jackie Picard, JCU
M. tuberculosis H37Rv	JCU AITHM PC3 lab

15.5 Primer set for the ribosomal sequences

15.5.1 Primer set designed at JCU

The primer set (RibRNA -312-F and RibRNA -312-R) to amplify the universal 16S rRNA gene
was designed using the AlleleID® v 7.7 (Premier Biosoft International, Palo Alto, CA). Around
24 published sequences were acquired from the National Centre for Biotechnology
Information (NCBI) nucleotide database, GenBank®. These sequences were then aligned
using Geneious software. Muscle option was opted to carry out to multiply align those
downloaded sequences.

The size of the amplicon produced by these primers was 312bp according to the *in-silico*
evaluation of the primer set in the AlleleID® program. Genedoc program was further used to
view and test the designed primer strings on the aligned published sequences. All the
guidelines to make good primer set were taken into consideration while designing the
primer pair.

Primer details are:

Primer name	Sequence details	Target region (312Kb)
RibRNA-312-F	5'-GGATTAGATACCCTGGTA-3'	16S rRNA region
RibRNA -312-R	5'-GACTTAACCCAACATCTC-3'	16S rRNA region

15.5.2 Published primer pair

Another published sequencing primers (sourced from http://lutzonilab.org/16s-ribosomal-dna/) set was used to amplify the entire 1.39kb of ribosomal sequence. This set was also a universal target group. Primer details are:

Primer name	Sequence detail	Target region (1.3 Kb)	Reference
27F	5'-AGAGTTTGATCMTGGCTCAG-3'	16S rRNA region	(Turner *et al.*, 1999)
1391R	5'-GACGGGCGGTGTGTRCA-3'	16S rRNA region	(Frank *et al.*, 2008)

Lyophilized vials of above-mentioned primer sets were vigorously centrifuged and reconstituted with nuclease free water to produce stock solution of 100µM concentration followed by the working stock, achieved by further diluting the stock solution to the concentration of 10µM prior to use for PCR. The work was done inside Biosafety hood. All the vials were stored at -80°C.

15.5.3 16S analysis on the panel of 40 organisms DNA

The amplification of target sequence was carried out in the Bio molecular systems (BMS) MIC real time PCR. Each reaction was carried out in a dedicated 48 well PCR plate, each containing a 20μL reaction mixture with the components as follows:

For primer set designed at JCU to amplify 300bp region of 16S panel organisms

10μL qPCR master Mix (Promega A#6002),

1.6μL of 10μM RibRNA -312-F (foreword primer),

1.6μL of 10μM RibRNA -312-R (reverse primer),

4.8μL of nuclease free water and

2μL of the template DNA.

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 59°C for 15 seconds and 72°C for 20 seconds.

For the published primer set to amplify 1.39Kb region of 16S panel organisms

10μL qPCR master Mix (Promega A#6002),

1.6μL of 10μM 27F (foreword primer),

1.6μL of 10μM 1391R (reverse primer),

4.8μL of nuclease free water and

2μL of the template DNA.

The reaction mixture was subjected to 95°C for 2 minutes, 95°C for 5 seconds and 59°C for 15 seconds and 72°C for 45 seconds.

15.6 Sequencing of the DNA isolates extracted from the panel organisms

Aliquots (20μL) of all the 40 DNA isolates were send off to Macrogen for Sanger sequencing. Two primer sets (RibRNA -312-F and R and 27F and 1391R) both of 10μM concentration were used. The purpose of using two sets of primers both targeting the 16S ribosomal region of the 40 panel organisms was to ensure the good quality of the DNA extracted from them. In addition, the PCR products gained after the run could be analysed individually by sequencing to verify the right amplified sequence from the known source organism.

The primer set designed at JCU was able to amplify 312bp 16S ribosomal region, while the published set had the capability to amplify the amplicon of 1.3Kb segment of the same ribosomal sequence. An amplicon of 1.3kb length was enough for the organisms' species determination and to generate the phylogenetic tree. The intent of using the shorter primer (312bp) set using the PCR product of 1.3kb primer set as a template was to further assure the presence of target sequence.

A C_t value in the range of 15 to 25 was indicative of good quality DNA that could be sequenced straightaway. All the isolates were expected to react in the PCR reaction due to the universal nature of both primer sets. Sanger sequencing using both primer sets was carried out to sequence the amplified 16S rRNA region. The PCR product of 27F and 1391 R primer set was used as a template for the primer set (RibRNA -312-F and R). The sequencing result of ribosomal region verified the DNA quality and the source organism.

Macrogen provided the sequencing data as ab1 files. For each isolate, there were four ab1 files (one pair from the 312bp primer set and the other from 1.3kb set). These files were processed in Geneious software. Around 40 folders each for one organism was created and named in Geneious. Isolate specific files were then transported into the organism specific folder and *de novo* assembly of acquired four files was carried out. Sequence editing was performed wherever needed. The resultant contig was then copied and pasted as a query sequence in nucleotide blast on NCBI website to carry out the homology search of the

identical published sequences available in the sequence database. This was necessary to identify the organism to its species level.

Five published sequences with 100% or at least 99% identity were selected, downloaded, and finally imported to Geneious. Further processing was done in the Geneious, where the contig was multiply aligned with these downloaded database sequences. The resultant alignment was further imported to a viewing utility tool, the GeneDoc to view the sequence alignment. Same protocol was done to all the 40 isolates.

All the 40 contigs were processed separately. Each contig with at least 5 identical published sequences (reference sequences) was separately aligned and exported to GeneDoc tool. Editing was performed as needed. All the Uracil (U) bases were converted to Thymine (T) bases and file was saved in fasta format. The resultant file was imported to Mega 7 program to generate a phylogenetic tree. Since the tree was quite big, therefore it was not included in this thesis. The intent was to ensure that all 40 DNA were different belonging to their respective organism.

This made a ground of the fact that all the 40 samples that were claimed to be the genomic DNA of 40 different organisms really contained DNA in them, which in turn indicated that the genomic extraction process from all the organisms was fruitful. This would potentially knock out any suspicion that may emerge during the specificity testing of the designed TaqMan assays (IS6110 and *rpoB*, discussed later) especially during the non-reaction (failure in the amplification of IS6110 and *rpoB* genes). A reaction/non-reaction in any isolate(s) would purely reflect the specificity of the assay(s).

15.7 Results

15.7.1 Identification of acquired organisms

Referring to the **Table 2,** culturing on specific media and microscopy analysis gave suggestive names of the 31 assorted organisms. They were further confirmed by carrying out a series of biochemical tests using organism specific API kits. Further genomic analysis using 16S rRNA sequence as a target; definitive organism identification to their species level was achieved. Except few (six) successful sequencing results of all the isolates were received. On carrying out the homology search of all 34 acquired contigs (obtained by the *de novo* assembly of sequencing ab1 files on Geneious program) on NCBI revealed 99-100% identity having no gap with the published sequences present on the database. This provided the surety of the organism selected.

15.7.2 Extraction of the genomic DNA from all the bacterial isolates

Melt analysis on the panel organisms' DNA was carried out to verify the purity of the extracted genomes (**Figure3.1**). Appearance of a sharp clear peak indicated the good quality of the amplified gene. All isolates had their respective peaks with different melting temperatures. Nearly all the isolates had sharp peaks. This in turn verified that the genomic DNA of the isolates was quality DNA and successful amplification of the 16S rRNA sequence of around 312bp length from the entire assorted panel isolates occurred. This further led to the conclusion that the designed primer set (RibRNA -312-F and RibRNA -312-R) could be used on a wide range of bacteria and had applicability in carrying out the melt analysis along with the target gene amplification.

The interpretation of the characteristic peaks of all the isolates at similar position made the comparative analysis bit difficult. Therefore, specific isolates of interest were then selected and further compared.
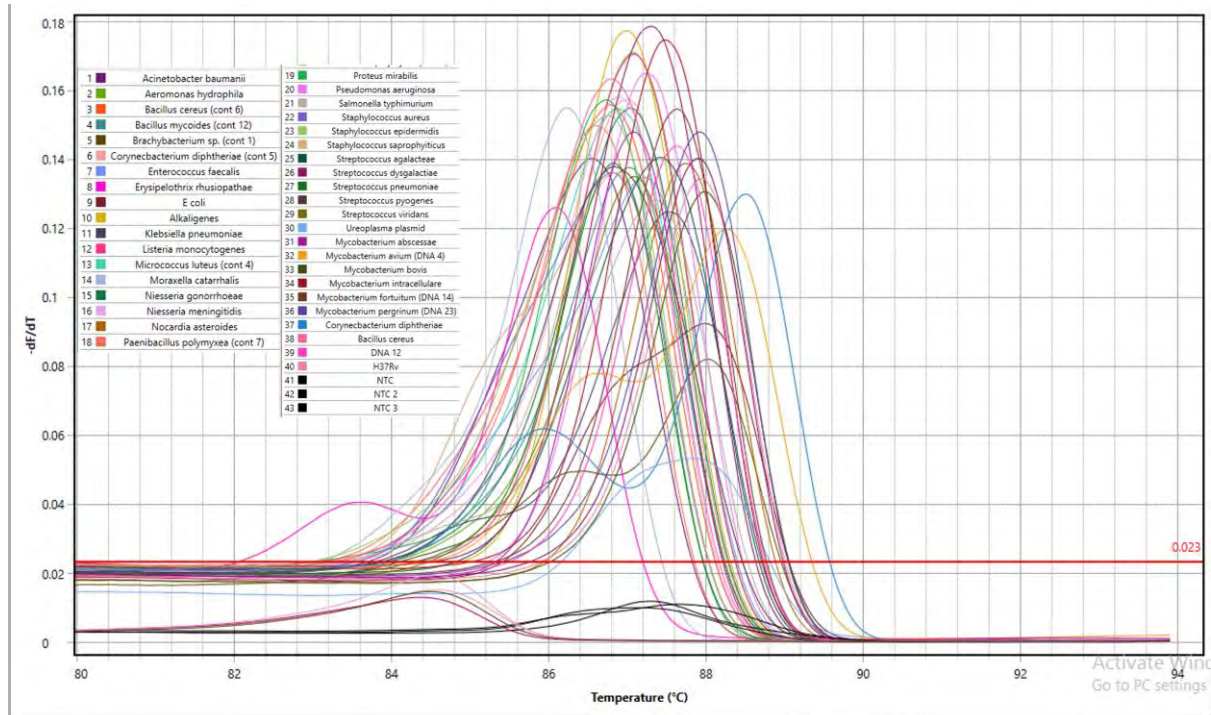
Figure 15.1: Melt analysis of the panel organisms' DNA for purity verification. Every single isolate had its own characteristic peak reflecting good quality of the amplicon (312bp long 16S rRNA gene).

Figure 15.1a indicate the melting behaviour of the amplified 16S rRNA sequence from the H37Rv and *Mycobacterium bovis* genomic DNA. Both look almost identical with same melting temperature of 88°C. These two were considered as reference in order to carry out the comparison with other *Mycobacterium* species.
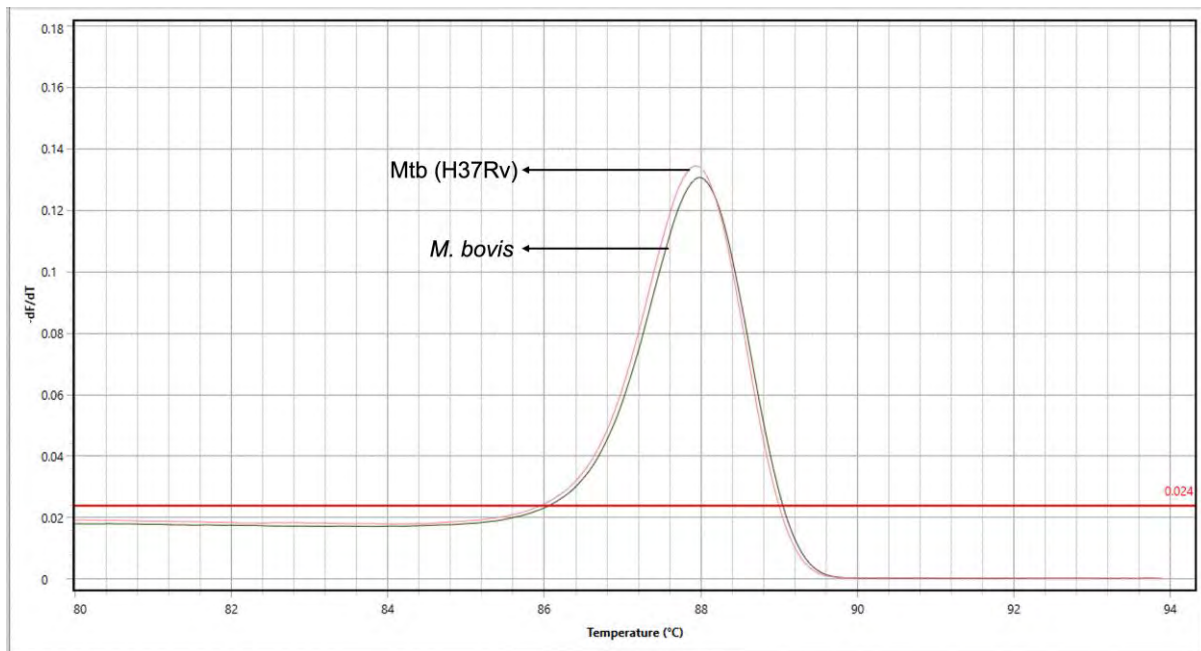
Figure 15.1a: Melt analysis of 16S rRNA gene segment from the H37Rv and *Mycobacterium bovis.* Both melts represented the high similarity between the two organisms.

It is apparent from the Figure 15.1b that the representative peaks of the *Mycobacterium* species other than the H37Rv and *M. bovis* were clearly distinguishable. This was due to the variation in the 16S rRNA sequence among different *Mycobacterium* species leading to different melting temperatures. This led to conclude that the designed primer set was able to differentiate between the taxonomically relevant organisms on considering the H37Rv and *M. bovis* as the reference genotype.
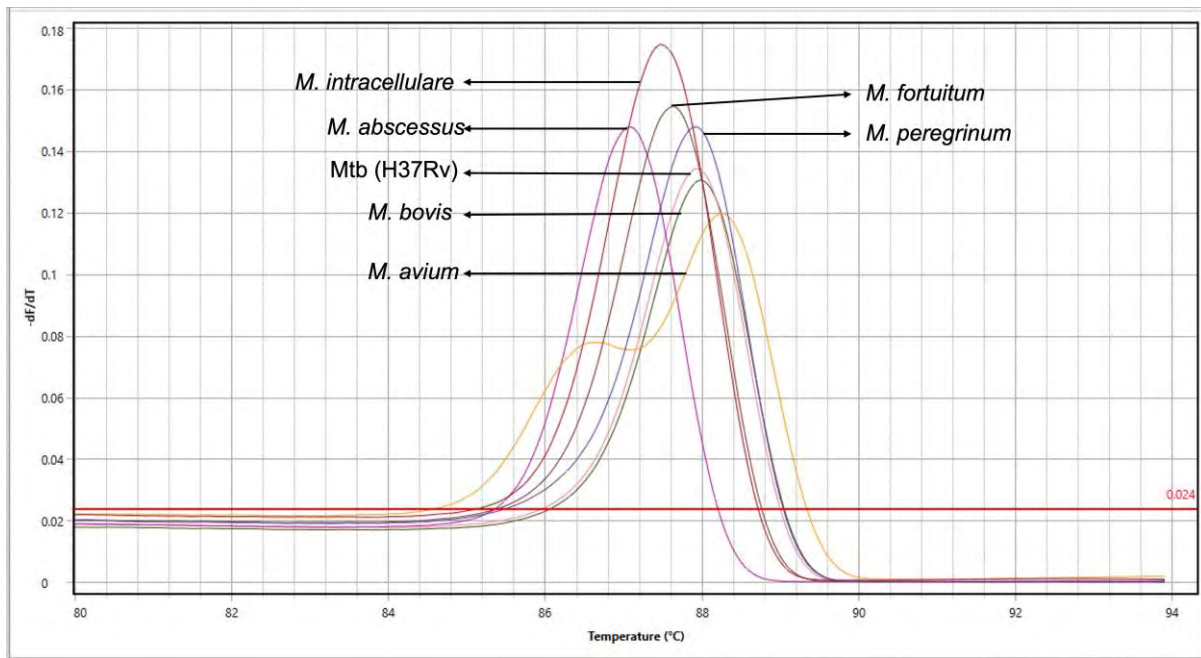
Figure15.1b: Comparison of melt curves of *Mycobacterium* species with H37Rv and *M. bovis.* Significant variation of the amplified gene among different species could be observed.

Further, Figure 15.1c depicts the variation in the melting temperature of some isolates that tend to infect the human lower respiratory tract. This overall gave the variation in the peaks of the isolates. Comparing all the isolates to the reference H37Rv and *M. bovis*, there was a drastic shift to the left-hand side that indicated high variation in the amplified target sequence. This verified that the assay was also able to differentiate the selected members of the clinically relevant organisms as well. Intentionally, *Salmonella* typhimurium was also selected for comparison and a significant difference in melting temperature was noticed.
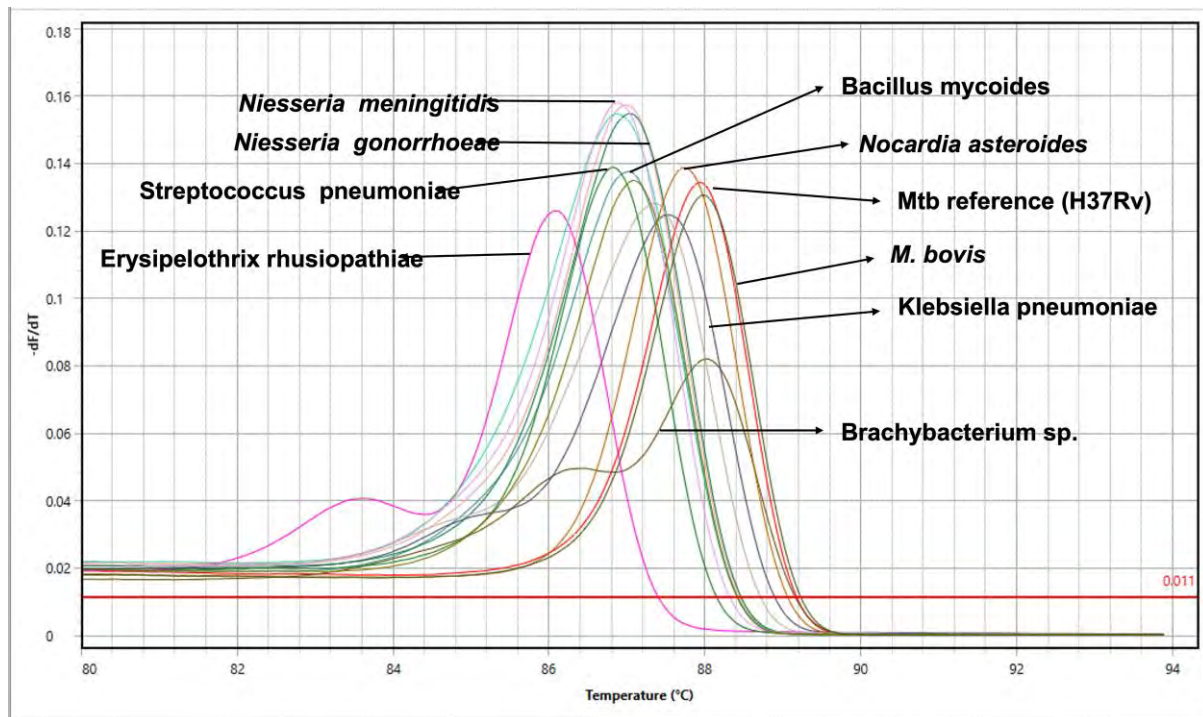
Figure 15.2: Melt of all *Mycobacterium* members

Comparative analysis keeping the H37Rv and *M. bovis* as reference genotypes helped to deduce that the assay could be incorporated into the panel of assays to identify and characterise the Mtb. Since it was clearly able to differentiate between the clinically and even the taxonomically relevant organisms. Its utility while being economic, simple, and easy to perform with short turnover time could be ideal to implement it as a routine identification tool for testing the ZN smear positive cases.

15.7.3 Primer set for the ribosomal sequence

A homology search using primer sequences of both sets against the aligned genomic database in GeneDoc (**Figure 1 and 2**) was carried out. It was found that all the reference sequences from the web had nearly 100% homology with the designed and published primer pair sequences. Therefore, it was deduced that the chances of getting the sequence of interest were quite high.

15.7.4 Binding analysis of primer set designed at JCU

The regions highlighted with grey colour in Figure 15.3 indicated the approximate position of 18 bases long forward primer sequence (RibRNA-312-F) and a corresponding domain of 18 bases of reverse primer sequence (RibRNA-312-R) binding site. A mismatch of one nucleotide was observed (red vertical bar) in case of reverse primer with some of the top most aligned sequences. It can be deduced that the region in between the foreword and reverse primer was the location of 312bp long 16S rRNA segment.

Figure 15.3: An overall presentation of both primer-binding sites against published sequences giving a hint of 312bp length of 16S rRNA sequence which is in between these oligonucleotides' sites.

Therefore, it was concluded that both primer sequences were specific and the observed difference in isolated sequence imparted no difference in the acquired results.

15.7.5 Binding pattern of published primer set

Similarly, Figure 15.4 indicated the binding pattern of the published primers against the reference sequences. It was observed that the published primers were also able to bind and amplify the target sequence. However, 27F primer didn't overlay the 12 bp position in all the downloaded and aligned sequences. Similar findings were observed with reverse primer at 3bp position, where the isolated sequence differed by one nucleotide (vertical red bar just in middle of 27F primer and the vertical blue bar near the starting of 1391R binding site). It was deduced that the annealing of both primers did not the cover the region of noticed difference.

27F

1391R
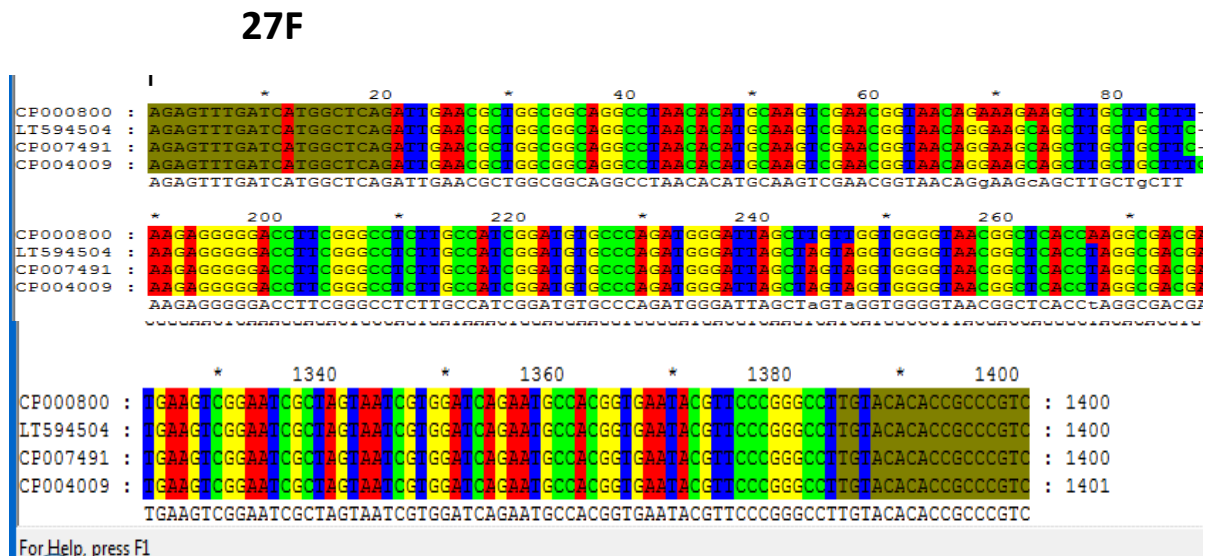
Figure 15.4: Presentation of both primer-binding sites against published sequences indicating 1391bp length of 16S rRNA sequence that is in between these oligonucleotides' sites.

Therefore, it was concluded that both primer sequences were specific and the observed difference in isolated sequence imparted no difference in the acquired results.

15.7.6 16S Analysis on the panel of 40 organisms DNA

15.7.6.1 Analysis with primer set targeting 312kb region of 16S rRNA

The primer set RibRNA -312-F and RibRNA -312-R was used to amplify the 312bp long ribosomal 16S region. All of the tested organisms DNA reacted. Sigmoid curves were observed in all isolates. This was indicative of successful amplification. No amplification was observed in the NTC replicates. This was expected to happen. Most of the samples C_t value lied in the range of 9 to 23 except the Ureaplasma plasmid suspension.
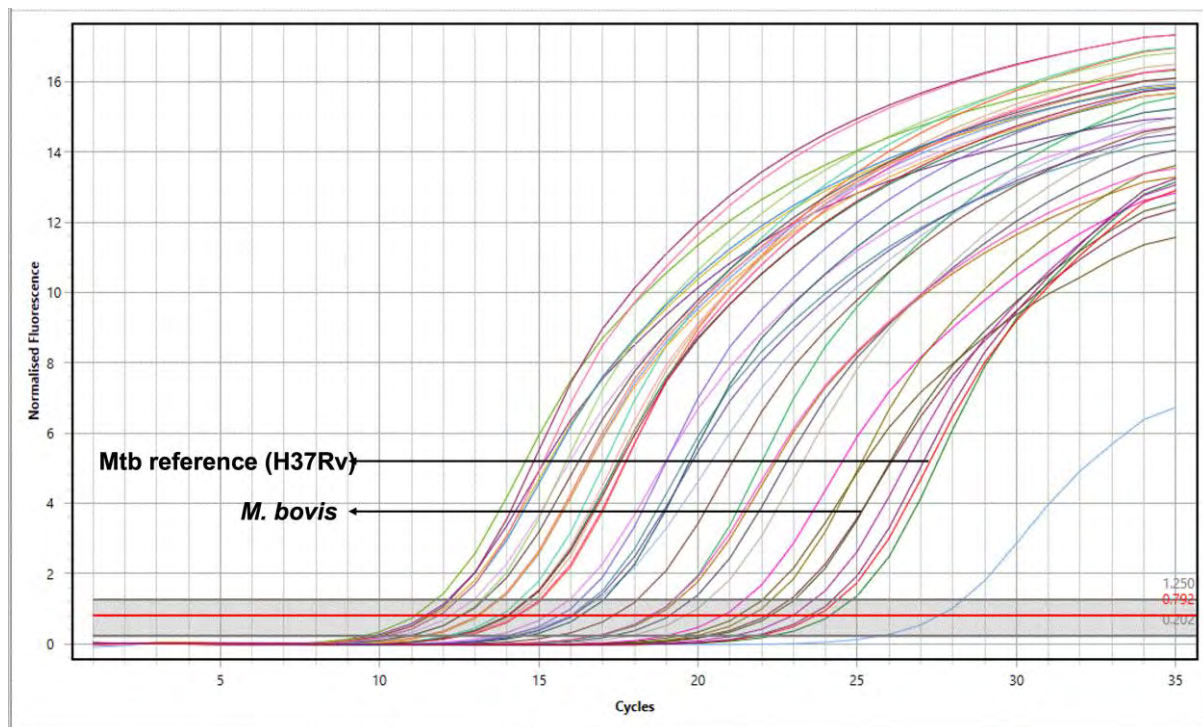
Figure 15.5: Application of designed primer set to isolate 312bp of 16S ribosomal region from a panel of 40 assorted organisms' genomic DNA. As expected, all isolates reacted. All isolates had C_t value between 15 and 25.

This indicated that genomic extraction of all 40 assorted organisms remained successful. The designed primer set was able to amplify the target region from all the isolates. Based on the C_t value of all the isolates, it was therefore concluded that the genomic DNA of all the 40 organisms was of high quality that can be sequenced straight away. Deductions were that the primer set was very specific in nature and able to amplify the required 16S ribosomal region of all assorted organisms.

15.7.6.2 Analysis with primer set targeting 1.39kb region of 16S rRNA

The published primer set 27F and 1391R was used to amplify the whole 1.3kb ribosomal 16S region. As expected, all the isolates reacted. Typical sigmoid curves in most isolates were observed. All the isolates had C_t value in range of 7 to 25 that could be sequenced straight away. This was indicative of high-quality genomic DNA. It was bit surprising that H37Rv in triplicate lately reacted and had C_t value of around 30.A tenfold dilution of H37Rv genomic DNA was used.
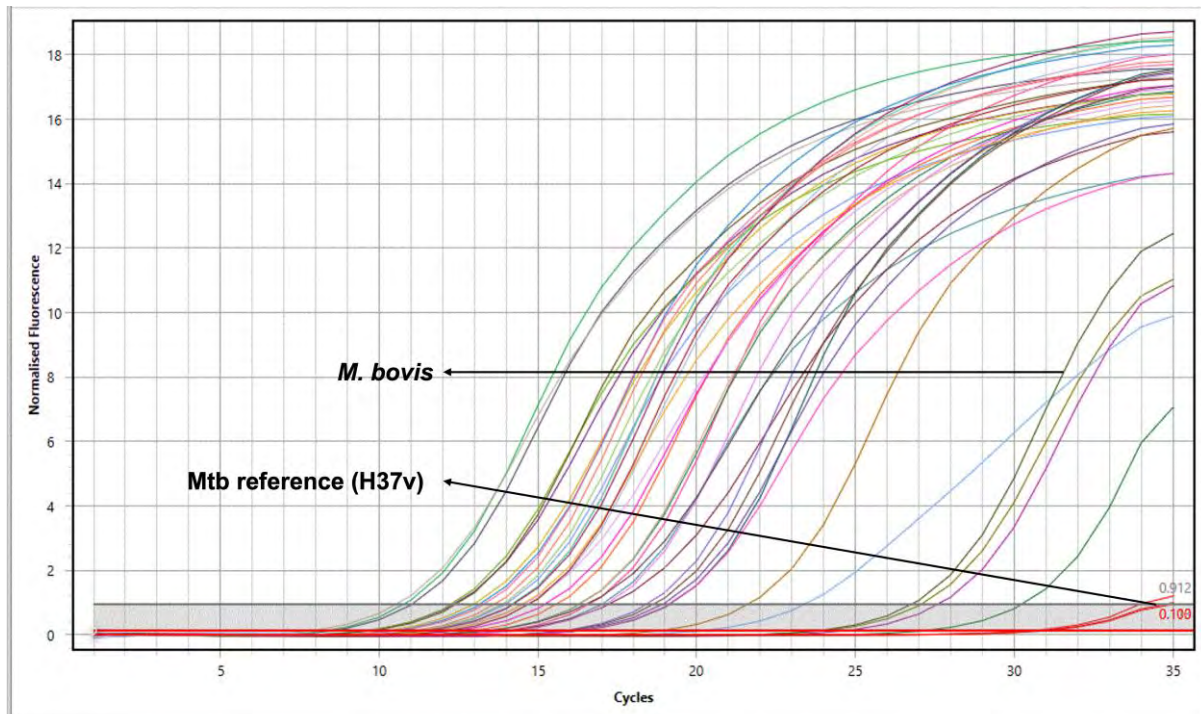
Figure 15.6: Application published primer set to isolate 1.3kb segment of 16S ribosomal region from a panel of 40 assorted organisms' genomic DNA. As expected, all isolates reacted. All most all the isolates had C_t value between 15 and 25.

234

Therefore, the above data interpreted the potential of the published primer set to successfully amplify the whole 16S region. The assay had good specificity and sensitivity. This primer set was perfect to use it as a sequencing primer pair. However, the main intent of designing another primer set to amplify a small (312bp) region was to acquire more concrete results after applying this primer pair on the PCR product of all the isolates amplified by the published primer set during the sequencing event.

15.8 Sequencing analysis

All the isolates were tested by phenotypic tests (microscopy and biochemical testing) and were further verified by sequencing results for the species identification. Referring to the Table 2, except 6 isolates, all other were successfully identified. All had sequence length of 1.3Kb as expected. None of all 40 had any gap(s) or mismatch(s) that made them 100% identical with the published sequences. This concluded that the published primer set (27F and 1391 R) was very specific being able to amplify 1.3Kb 16S rRNA sequence from nearly all the isolates. The 100% identity with presence of no mismatch(s) further gave the surety of the relativeness of the query sequences with the possible published sequences. This led to the deduction that the suggested name of the organism based on the sequence similarity was reliable that could further be compared with the phenotypic observations.

No doubt, the microscopic observation and the biochemical tests can provide significant amount of information useful to identify an organism. Genotypic identification of the bacterial isolates has become an alternative approach to this time consuming and labourious phenotypic identification methods. This involves usage of a broad range set of primers to target a conversed genetic region harbouring no mutation in it at all. The conserved region of the 16S rRNA can be addressed as an example. (Tang *et al.*, 1998; Edwards *et al.*, 2012; Akram *et al.*, 2017). Bacterial identification to its species level (Xiao *et al.*, 2010), ease to perform with fast turnover period makes it a method of choice (Akram *et al.*, 2017).

Woese and his colleagues in 1977 even hypothesised that the Tree of life comprised of three domains: *Archea, Bacteria and Eukarya.* It was based on the phylogenetic analysis of the16S ribosomal sequence (16S rRNA) (Woese and Fox, 1977; Woese *et al.*, 1990). The Bacterial rRNA operons consist of 16S rRNA, 23S rRNA, 5S rRNA and tRNA genes separated by the intergenic spacer regions (Anda *et al.*, 2015). Of these, 16S rRNA sequence being present in nearly all bacteria is generally given more preference for bacterial identification to its species level (Hakovirta *et al.*, 2016).

Successful sequencing of most of the DNA isolates demonstrated the quality of DNA that could be used as a template for PCR and further sequencing could be done on the corresponding PCR product. The primer set **(**RibRNA-312-F and RibRNA-312-R) designed at JCU targeting the shorter sequence (312bp) of 16S rRNA from the PCR product of the published primer set (27F and 1391R) gave further proof of the right sequence and DNA quality.

In addition, it was interesting to note that the assay was clearly able to demonstrate the variation in the melt curves of *Mycobacterium* like organisms and all the other panel organisms. Moreover, it was also capable of differentiating between the closely related *Mycobacterium* non-tuberculous organisms from the reference TB organisms. This implies

that this assay can differentiate the infections in a patient caused by non-tuberculous *Mycobacterium.* This in turn suggests that it is worth adding this assay in a panel of assays that can be applied to firstly identify the infections with acid-fast organisms.

APPENDIX 16

THE IS6110 SEQUENCE IS NOT MTBC SPECIFIC

Appendix 16-Figure 16.1: The IS6110 sequence (in the red box) that Wang *et al.* (2019) used as a target site to develop an assay to identify the Mtb (Wang *et al.*, 2019).

THE TARGET SEQUENCE
CTCGACCTGAAAGACGTTATCCACCACACGGATAGGGGATCTCAGTACACATCGATCCGGTTCAGCGAGCGGCTCGCCGAG

BLAST » blastn suite

Home | Recent Results | Saved Strategies | Help

Standard Nucleotide BLAST

blastn | blastp | blastx | tblastn | tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Reset page | Bookmark

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear

CTCGACCTGAAAGACGTTATCCACCACACGGATAGGGGATCTCAGTACAC
ATCGATCCGGTTCAGCGAGCGGCTCGCCGAG

Query subrange

From

To

New columns added to the **Description Table**
Click 'Select Columns' or 'Manage Columns'.

Or, upload file | Choose File no file selected

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

Nucleotide collection (nr/nt)

Organism
Optional
Mycobacterium tuberculosis complex (taxid:77643) ☑ exclude | Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude
Optional
☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to
Optional
☐ Sequences from type material

Entrez Query
Optional

Enter an Entrez query to limit search

You Tube Create custom database

Program Selection

Optimize for
○ Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
● Somewhat similar sequences (blastn)

Choose a BLAST algorithm

(a)

Sequences producing significant alignments

Download | New Select columns | Show 100 ✓

☑ select all 97 sequences selected

GenBank | Graphics | Distance tree of results | New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Mycobacterium mantenii JCM 18113 DNA, complete genome	Mycobacterium mantenii	97.8	171	100%	1e-16	86.42%	6185541	AP022590.1
Mycobacterium intracellulare subsp. chimaera strain FLAC0070 chromosome, complete ge...	Mycobacterium intracellulare subsp. chi...	97.8	97.8	100%	1e-16	86.42%	5408332	CP023151.1
Mycobacterium rutilum strain DSM 45405 genome assembly, chromosome: I	Mycolicibacterium rutilum	94.2	94.2	91%	2e-15	87.84%	5987931	LT629971.1
Mycobacterium sp. JLS, complete genome	Mycobacterium sp. JLS	86.9	86.9	98%	2e-13	83.75%	6048425	CP000580.1
Mycobacterium heckeshornense strain JCM 15655	Mycobacterium heckeshornense	86.0	244	95%	2e-13	84.42%	4957188	AP024237.1
Mycolicibacterium duvalii JCM 6396 DNA, complete genome	Mycolicibacterium duvalii	84.2	252	100%	8e-13	82.72%	5657388	AP022563.1
Nocardia cyriacigeorgica GUH-2 chromosome complete genome	Nocardia cyriacigeorgica GUH-2	81.5	230	96%	1e-11	83.12%	6194645	FO082843.1
Mycolicibacterium poriferae JCM 12603 DNA, complete genome	Mycolicibacterium poriferae	80.6	80.6	91%	1e-11	83.78%	5712830	AP022570.1
Mycobacterium sp. THAF192 chromosome, complete genome	Mycobacterium sp. THAF192	80.6	80.6	91%	1e-11	83.78%	5780554	CP045325.1
Nocardia sp. WCH-YHL-001 chromosome, complete genome	Nocardia sp. WCH-YHL-001	80.6	322	91%	1e-11	83.78%	8339910	CP059399.1
Nocardia cyriacigeorgica strain MDA3349 chromosome, complete genome	Nocardia cyriacigeorgica	74.3	297	96%	2e-09	80.77%	6320090	CP026746.1
Mycobacterium conspicuum JCM 14738 DNA, nearly complete genome	Mycobacterium conspicuum	69.8	69.8	97%	2e-08	81.01%	6237139	AP022613.1
Mycobacterium chimaera strain ZUERICH-2 plasmid unnamed 1, complete sequence	Mycobacterium intracellulare subsp. chi...	67.1	67.1	91%	2e-07	79.73%	175972	CP015268.1
Azoarcus sp. CIB, complete genome	Azoarcus sp. CIB	61.7	61.7	87%	1e-05	78.87%	5257030	CP011072.1
Nocardia brasiliensis ATCC 700358, complete genome	Nocardia brasiliensis ATCC 700358	60.8	60.8	96%	1e-05	76.92%	9436348	CP003876.1
Mycobacterium parmense JCM 14742 DNA, nearly complete genome	Mycobacterium parmense	59.9	59.9	85%	3e-05	79.71%	5952912	AP022614.1
Nocardia terpenica strain AUSMDU00012715 chromosome, complete genome	Nocardia terpenica	59.0	59.0	95%	3e-05	76.62%	9306871	CP046173.1
Nocardia wallacei FMUON74 DNA, complete genome	Nocardia wallacei	58.1	174	91%	1e-04	77.03%	7832428	AP023396.1
Xanthomonas campestris pv. badrii strain NEB122 chromosome, complete genome	Xanthomonas campestris pv. badrii	56.3	211	77%	4e-04	79.37%	4955824	CP051651.1
Mycolicibacterium phocaicum JCM 15301 DNA, nearly complete genome	Mycolicibacterium phocaicum	55.4	55.4	55%	4e-04	86.67%	5853197	AP022616.1
Mycobacterium avium subsp. hominissuis strain MAC109 chromosome, complete genome	Mycobacterium avium subsp. hominissuis	55.4	108	88%	4e-04	90.00%	5188883	CP029332.1

(b)

☑ Mycobacterium avium subsp. hominissuis strain MAC109 chromosome, complete genome	Mycobacterium avium subsp. hominissuis	55.4	108	88%	4e-04	90.00%	5188883	CP029332.1
☑ Xanthomonas oryzae pv. oryzae strain K2 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4984295	CP050113.1
☑ Xanthomonas oryzae pv. oryzae strain K3 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4945802	CP050114.1
☑ Xanthomonas oryzae pv. oryzae strain K3a chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4880106	CP050115.1
☑ Xanthomonas oryzae pv. oryzae strain IXO704 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4994377	CP040604.1
☑ Xanthomonas oryzae pv. oryzae strain IXO1088 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5093052	CP040687.1
☑ Xanthomonas oryzae pv. oryzae strain K1 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4981423	CP049205.1
☑ Xanthomonas oryzae pv. oryzae strain ITCCBB0002 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4731568	CP046148.1
☑ Xanthomonas oryzae pv. oryzae strain LN18 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5024861	CP045238.1
☑ Xanthomonas oryzae pv. oryzae strain BXO1 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4991257	CP033201.1
☑ Xanthomonas oryzae pv. oryzae strain JW11089 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5010511	CP033193.2
☑ Xanthomonas oryzae pv. oryzae strain PXO61 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4968141	CP033187.3
☑ Xanthomonas oryzae pv. oryzae strain PXO513 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4915674	CP033188.1
☑ Xanthomonas oryzae pv. oryzae strain PXO421 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4909962	CP033189.1
☑ Xanthomonas oryzae pv. oryzae strain PXO404 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4914700	CP033190.1
☑ Xanthomonas oryzae pv. oryzae strain PXO364 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4904940	CP033191.1
☑ Xanthomonas oryzae pv. oryzae strain NX0260 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5050391	CP033192.1
☑ Xanthomonas oryzae pv. oryzae strain KXO85 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4975259	CP033197.1
☑ Xanthomonas oryzae pv. oryzae strain CIAT chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5035285	CP033194.1
☑ Xanthomonas oryzae pv. oryzae strain AUST2013 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4958184	CP033196.1
☑ Xanthomonas oryzae pv. oryzae strain YN24 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	5028971	CP018089.1
☑ Xanthomonas oryzae pv. oryzae strain ScYc-b chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4865653	CP018087.1
☑ Xanthomonas oryzae pv. oryzae strain XM9 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4920509	CP020334.1
☑ Xanthomonas oryzae pv. oryzae SK2-3 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4934446	CP019515.1
☑ Xanthomonas oryzae pv. oryzae strain IX-280 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4963594	CP019226.1
☑ Xanthomonas oryzae pv. oryzae strain PXO142 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4982118	CP031698.1

(C)

☑ Xanthomonas oryzae pv. oryzae strain PXO142 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4982118	CP031698.1
☑ Xanthomonas oryzae pv. oryzae strain ICMP3125 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4990672	CP031697.1
☑ Xanthomonas oryzae pv. oryzae strain ScYc-b chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4865660	CP031469.1
☑ Xanthomonas oryzae pv. oryzae strain JL25 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4900521	CP031457.1
☑ Xanthomonas oryzae pv. oryzae strain PX086 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5015823	CP031463.1
☑ Xanthomonas oryzae pv. oryzae strain PX079 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5026592	CP031462.1
☑ Xanthomonas oryzae pv. oryzae strain YC11 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4867200	CP031464.1
☑ Xanthomonas oryzae pv. oryzae strain JL33 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4898917	CP031459.1
☑ Xanthomonas oryzae pv. oryzae strain JL28 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4697691	CP031458.1
☑ Xanthomonas oryzae pv. oryzae strain JP01 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4948537	CP031460.1
☑ Xanthomonas oryzae pv. oryzae strain OS198 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4908920	CP031461.1
☑ Xanthomonas oryzae pv. oryzae strain HuN37 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4915452	CP031456.1
☑ Xanthomonas oryzae pv. oryzae strain PXO61 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4994045	CP021789.1
☑ Xanthomonas oryzae pv. oryzae strain PXO61 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4980838	CP021788.1
☑ Xanthomonas oryzae pv. oryzae strain PXO61 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	156	81%	0.005	77.27%	4999360	CP020942.1
☑ Cellulomonas sp. PSBB021, complete genome	Cellulomonas sp. PSBB021	52.7	52.7	75%	0.005	78.69%	3794193	CP021430.1
☑ Xanthomonas oryzae pv. oryzae strain XF89b, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4966744	CP011532.1
☑ Xanthomonas oryzae pv. oryzae strain PXO602, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	4951791	CP013679.1
☑ Xanthomonas oryzae pv. oryzae strain PXO563, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4936308	CP013678.1
☑ Xanthomonas oryzae pv. oryzae strain PXO524, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4954304	CP013677.1
☑ Xanthomonas oryzae pv. oryzae strain PXO282, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4961996	CP013676.1
☑ Xanthomonas oryzae pv. oryzae strain PXO236, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4968717	CP013675.1
☑ Xanthomonas oryzae pv. oryzae strain PXO211, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5033346	CP013674.1
☑ Xanthomonas oryzae pv. oryzae strain PXO145, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5039763	CP013961.1
☑ Xanthomonas oryzae pv. oryzae strain PXO71, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4906999	CP013670.1
☑ Xanthomonas oryzae pv. oryzae strain PXO83, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5025428	CP012947.1

(d)

(e)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Xanthomonas oryzae pv. oryzae strain PXO83, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5025428	CP012947.1
Xanthomonas oryzae pv. oryzae PXO99A, complete genome	Xanthomonas oryzae pv. oryzae PXO99A	52.7	105	81%	0.005	77.27%	5238555	CP000967.2
Xanthomonas oryzae pv. oryzae PXO86, complete genome	Xanthomonas oryzae pv. oryzae PXO86	52.7	104	81%	0.005	77.27%	5016623	CP007166.1
Xanthomonas oryzae pv. oryzae strain T7133 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4978475	CP071891.1
Xanthomonas oryzae pv. oryzae strain LN4 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5012583	CP045452.1
Xanthomonas oryzae pv. oryzae strain BXO512 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4934882	CP065228.1
Mycobacterium gilvum Spyr1, complete genome	Mycolicibacterium gilvum Spyr1	52.7	52.7	69%	0.005	80.36%	5547747	CP002385.1
Xanthomonas oryzae pv. oryzae KACC 10331, complete genome	Xanthomonas oryzae pv. oryzae KACC 1…	52.7	105	81%	0.005	77.27%	4941439	AE013598.1
Xanthomonas oryzae pv. oryzicola strain 0-9 chromosome, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4830088	CP045912.1
Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4793207	CP043403.1
Gordonia sp. YC-JH1 chromosome, complete genome	Gordonia sp. YC-JH1	51.8	51.8	96%	0.005	74.36%	4101557	CP025435.1
Xanthomonas oryzae pv. oryzicola strain CFBP2286, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4969419	CP011962.1
Xanthomonas oryzae pv. oryzicola strain RS105, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4779952	CP011961.1
Xanthomonas oryzae pv. oryzicola strain L8, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4796527	CP011960.1
Xanthomonas oryzae pv. oryzicola strain CFBP7341, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	258	77%	0.005	77.78%	5017766	CP011959.1
Xanthomonas oryzae pv. oryzicola strain CFBP7331, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	258	77%	0.005	77.78%	5008292	CP011958.1
Xanthomonas oryzae pv. oryzicola strain BXOR1, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4692590	CP011957.1
Xanthomonas oryzae pv. oryzicola strain BLS279, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4790622	CP011956.1
Xanthomonas oryzae pv. oryzicola strain B8-12, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4794316	CP011955.1
Xanthomonas oryzae pv. oryzicola BLS256, complete genome	Xanthomonas oryzae pv. oryzicola BLS256	51.8	362	77%	0.005	77.78%	4831746	CP003057.2
Xanthomonas oryzae pv. oryzicola strain CFBP7342, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	5080102	CP007221.1
Xanthomonas oryzae pv. oryzae MAFF 311018 DNA, complete genome	Xanthomonas oryzae pv. oryzae MAFF 3…	51.8	103	77%	0.005	77.78%	4940217	AP008229.1
Saccharopolyspora coralli strain E2A chromosome, complete genome	Saccharopolyspora coralli	50.9	50.9	55%	0.018	84.44%	4832009	CP045929.1
Variovorax sp. HW608 genome assembly, chromosome: I	Variovorax sp. HW608	50.9	50.9	74%	0.018	78.33%	7733665	LT607803.1
Mycobacterium sp. PYR15 chromosome, complete genome	Mycobacterium sp. PYR15	50.0	149	70%	0.018	78.95%	6037017	CP023435.1
Mycobacterium sp. WY10, complete genome	Mycobacterium sp. WY10	50.0	50.0	70%	0.018	78.95%	6041408	CP018043.1

(e)

(f)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Xanthomonas oryzae pv. oryzae strain PXO83, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	5025428	CP012947.1
Xanthomonas oryzae pv. oryzae PXO99A, complete genome	Xanthomonas oryzae pv. oryzae PXO99A	52.7	105	81%	0.005	77.27%	5238555	CP000967.2
Xanthomonas oryzae pv. oryzae PXO86, complete genome	Xanthomonas oryzae pv. oryzae PXO86	52.7	104	81%	0.005	77.27%	5016623	CP007166.1
Xanthomonas oryzae pv. oryzae strain T7133 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	104	81%	0.005	77.27%	4978475	CP071891.1
Xanthomonas oryzae pv. oryzae strain LN4 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	105	81%	0.005	77.27%	5012583	CP045452.1
Xanthomonas oryzae pv. oryzae strain BXO512 chromosome, complete genome	Xanthomonas oryzae pv. oryzae	52.7	52.7	81%	0.005	77.27%	4934882	CP065228.1
Mycobacterium gilvum Spyr1, complete genome	Mycolicibacterium gilvum Spyr1	52.7	52.7	69%	0.005	80.36%	5547747	CP002385.1
Xanthomonas oryzae pv. oryzae KACC 10331, complete genome	Xanthomonas oryzae pv. oryzae KACC 1…	52.7	105	81%	0.005	77.27%	4941439	AE013598.1
Xanthomonas oryzae pv. oryzicola strain 0-9 chromosome, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4830088	CP045912.1
Xanthomonas oryzae pv. oryzicola strain GX01 chromosome, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4793207	CP043403.1
Gordonia sp. YC-JH1 chromosome, complete genome	Gordonia sp. YC-JH1	51.8	51.8	96%	0.005	74.36%	4101557	CP025435.1
Xanthomonas oryzae pv. oryzicola strain CFBP2286, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4969419	CP011962.1
Xanthomonas oryzae pv. oryzicola strain RS105, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4779952	CP011961.1
Xanthomonas oryzae pv. oryzicola strain L8, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4796527	CP011960.1
Xanthomonas oryzae pv. oryzicola strain CFBP7341, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	258	77%	0.005	77.78%	5017766	CP011959.1
Xanthomonas oryzae pv. oryzicola strain CFBP7331, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	258	77%	0.005	77.78%	5008292	CP011958.1
Xanthomonas oryzae pv. oryzicola strain BXOR1, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4692590	CP011957.1
Xanthomonas oryzae pv. oryzicola strain BLS279, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4790622	CP011956.1
Xanthomonas oryzae pv. oryzicola strain B8-12, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	4794316	CP011955.1
Xanthomonas oryzae pv. oryzicola BLS256, complete genome	Xanthomonas oryzae pv. oryzicola BLS256	51.8	362	77%	0.005	77.78%	4831746	CP003057.2
Xanthomonas oryzae pv. oryzicola strain CFBP7342, complete genome	Xanthomonas oryzae pv. oryzicola	51.8	310	77%	0.005	77.78%	5080102	CP007221.1
Xanthomonas oryzae pv. oryzae MAFF 311018 DNA, complete genome	Xanthomonas oryzae pv. oryzae MAFF 3…	51.8	103	77%	0.005	77.78%	4940217	AP008229.1
Saccharopolyspora coralli strain E2A chromosome, complete genome	Saccharopolyspora coralli	50.9	50.9	55%	0.018	84.44%	4832009	CP045929.1
Variovorax sp. HW608 genome assembly, chromosome: I	Variovorax sp. HW608	50.9	50.9	74%	0.018	78.33%	7733665	LT607803.1
Mycobacterium sp. PYR15 chromosome, complete genome	Mycobacterium sp. PYR15	50.0	149	70%	0.018	78.95%	6037017	CP023435.1
Mycobacterium sp. WY10, complete genome	Mycobacterium sp. WY10	50.0	50.0	70%	0.018	78.95%	6041408	CP018043.1
Mycobacterium heckeshornense JMUB5695 DNA, complete genome	Mycobacterium heckeshornense	50.0	50.0	70%	0.018	78.95%	4865109	AP024310.1

(f)

Appendix 16-Figure 16.2 (a-f): The presence of IS6110 in a wide range of organism

APPENDIX 17

GRAPH CALCULATIONS

Published IS6110 assay reproducibility data

DNA name and dilution	C_t on 1st attempt	Mean value	C_t on 2nd attempt	Mean value	C_t on 3rd attempt	Mean value
H37Rv genomic DNA 10^{-1}	23.01	22.885	24.23	24.155	24.27	24.255
H37Rv genomic DNA 10^{-1}	22.76		24.08		24.24	
H37Rv genomic DNA 10^{-2}	29.82	30	31.44	31.26	31.18	31.235
H37Rv genomic DNA 10^{-2}	30.18		31.08		31.29	
H37Rv genomic DNA 10^{-3}	33.38	33.545	35.45	35.36	35.37	35.275
H37Rv genomic DNA 10^{-3}	33.71		35.27		35.18	

IS6110 Plasmid dilution	Mean C_t: 1st attempt	Mean C_t: 2nd attempt	Mean C_t: 3rd attempt	Mean	Standard Deviation	CV%
Sample 1	22.885	24.155	24.255	23.765	0.763740794	3.213720992
Sample 2	30	31.26	31.235	30.83166667	0.720352923	2.336406042
Sample 3	33.545	35.36	35.275	34.72666667	1.024235487	2.949420677

JCU IS6110 assay reproducibility data

Plasmid name and dilution	C_t on 1st attempt	Mean value	C_t on 2nd attempt	Mean value	C_t on 3rd attempt	Mean value
IS6110 plasmid 10^{-1}	17.02	17	17.2	17.22	16.18	16.03
IS6110 plasmid 10^{-1}	17.05		17.3		15.9	
IS6110 plasmid 10^{-1}	16.93		17.16		16.01	
IS6110 plasmid 10^{-2}	19.77	19.86666667	20.28	20.24	19.12	19.16333333
IS6110 plasmid 10^{-2}	20.02		20.19		19.35	
IS6110 plasmid 10^{-2}	19.81		20.25		19.02	
IS6110 plasmid 10^{-3}	22.05	22.05333333	22.27	22.3033333	20.88	20.94666667
IS6110 plasmid 10^{-3}	22.1		22.28		20.9	
IS6110 plasmid 10^{-3}	22.01		22.36		21.06	

IS6110 Plasmid dilution	Mean C_t: 1st attempt	Mean C_t: 2nd attempt	Mean C_t: 3rd attempt	Mean	Standard Deviation	CV%
Sample 1	17	17.22	16.03	16.75	0.633166645	3.780099372
Sample 2	19.866	20.24	19.163	19.75633333	0.546811058	2.767776029
Sample 3	22.053	22.303	20.946	21.76733333	0.722195495	3.317794992

JCU *rpoB* assay reproducibility data

DNA name and dilution	C_t value on 1st attempt	Mean value	C_t value on 2nd attempt	Mean value	C_t value on 3rd attempt	Mean value
H37Rv genomic DNA 10^{-1}	27.18	27.135	26.53	26.445	27.03	27.14
H37Rv genomic DNA 10^{-1}	27.09		26.36		27.25	
H37Rv genomic DNA 10^{-2}	34.57	34.305	32.55	32.86	33.38	33.335
H37Rv genomic DNA 10^{-2}	34.04		33.17		33.29	
H37Rv genomic DNA 10^{-3}	39.17	38.835	36.81	36.775	37.27	37.495
H37Rv genomic DNA 10^{-3}	38.5		36.74		37.72	

Mean C_t value on 1st attempt	Mean C_t value on 2nd attempt	Mean C_t value on 3rd attempt	Mean	Standard Deviation	CV%
27.135	26.355	27.14	26.87666667	0.451783503	1.680950649
34.305	32.86	33.335	33.5	0.736495078	2.19849277
38.835	36.775	37.495	37.70166667	1.045434519	2.772913272

APPENDIX 18

MICROBIAL TESTING DATA

Organism	Source	Organism ID	Microscopy	ZN Staining	Confirmed biochemistry	Sequencing						TaqMan assays		HRM analysis	
						Query seq length	Accession number	Identity (%)	Gaps (%)	Associated organism	16s	IS6110	*rpoB*	*rpoB*	*katG*
Clinically associated organisms															
Acinetobacter baumannii	JCU MCC		Gram negative coccobacillus		API 20 NE	1366						Non-Reactor	Reactor	Reactor	Non-Reactor
Aeromonas hydrophila	JCU MCC		Gram negative rods		API 20 NE	1331						Non-Reactor	Reactor	Reactor	Non-Reactor
Bacillus cereus	JCU MCC	Contaminant 6	Gram positive rods		API 50 CHB						B. cereus	Non-Reactor	Reactor	Reactor	Non-Reactor
Bacillus mycoides	JCU MCC	Contaminant 12	Gram positive rods		API 50 CHB						Bacillus sp	Non-Reactor	Reactor	Reactor	Non-Reactor
Brachybacterium spp.	JCU MCC	Contaminant 1	Gram positive irregular rods		API Corynae						Brachybacterium sp	Non-Reactor	Non-Reactor	Reactor	Non-Reactor
Enterobacter cloacae	JCU MCC	Contaminant 5	Gram negative rods		API 20 E						Enterobacter sp	Non-Reactor	Reactor	Reactor	Non-Reactor
Enterococcus faecalis (ATCC)	JCU MCC	8213 NCTC	Gram positive cocci		API 20 S	1338	LC279528.1	1333/1338(99%)	0/1295(0%)	Enterococcus faecalis		Non-Reactor	Reactor	Reactor	Non-Reactor
Erysipelothrix rhusiopathiae	JCU MCC		Gram negative rods		API Corynae	1298	CP017116.1	1295/1295(100%)	0/1295(0%)	Erysipelothrix rhusiopathiae		Non-Reactor	Reactor	Reactor	Non-Reactor
Escherichia coli (ATCC)	JCU MCC	11775 ATCC	Gram negative rods		API 20 E	1058	KT260217.1	1056/1058(99%)	0/1058(0%)	Escherichia coli strain		Non-Reactor	Reactor	Reactor	Non-Reactor
Alkaligenes sp.	JCU MCC		Gram negative rods		API 20 NE	1306	CP023667.1	1297/1306(99%)	0/1306(0%)	Alcaligenes faecalis		Non-Reactor	Reactor	Reactor	Non-Reactor
Klebsiella pneumoniae	JCU MCC	15380 ATCC	Gram negative rods		API 20 E	1335	KJ958509.1	1329/1335(99%)	2/1335(0%)	Klebsiella sp.		Non-Reactor	Reactor	Reactor	Non-Reactor

Species	Source	Strain	Gram		API	bp	Accession	Match	Mismatch	BLAST ID					
Listeria monocytogenes	JCU MCC		Gram positive rods		API Listeria	1345	CP025443.1	1342/1342(100%)	0/1342(0%)	Listeria monocytogenes		Non-Reactor	Reactor	Reactor	Non-Reactor
Micrococcus luteus	JCU MCC	Contaminant 4	Gram positive cocci		API Staph							Non-Reactor	Non-Reactor	Reactor	Non-Reactor
Moraxella catarrhalis	JCU MCC		Gram positive diplococci		API NH	1324	JF150763.1	1314/1321(99%)	1/1321(0%)	Uncultured bacterium clone		Non-Reactor	Reactor	Reactor	Non-Reactor
Neisseria gonorrhoeae	JCU MCC	49226 ATCC	Gram positive diplococci		API NH	1324	LT906472.1	1323/1324(99%)	1/1324(0%)	Neisseria gonorrhoeae		Non-Reactor	Reactor	Reactor	Non-Reactor
Neisseria meningitidis	JCU MCC		Gram positive diplococci		API NH	1334	CP020422.1	1331/1331(100%)	0/1331(0%)	Neisseria meningitidis		Non-Reactor	Reactor	Reactor	Non-Reactor
Nocardia sp.	PC3		Gram positive rods		API Corynae	1313	LC055494.1	1310/1310(100%)	0/1310(0%)	Nocardia sp.		Non-Reactor	Reactor	Reactor	Non-Reactor
Paenibacillus polymyxa	PC3	Contaminant 7	Gram positive rods		API 50 CHB	1343	MG430181.1	1343/1343(100%)	0/1343(0%)	Bacillus sp		Non-Reactor	Reactor	Reactor	Non-Reactor
Proteus mirabilis	JCU MCC	13315 ATCC	Gram negative rods		API 20 E	1329	MF541400.1	1328/1329(99%)	1/1329(0%)	Proteus mirabilis		Non-Reactor	Reactor	Reactor	Non-Reactor
Pseudomonas aeruginosa (ATCC)	JCU MCC	7700 ATCC	Gram negative rods		API 20 E	1329	KY458572.1	1329/1329(100%)	0/1329(0%)	Pseudomonas aeruginosa		Non-Reactor	Reactor	Reactor	Non-Reactor
Salmonella typhimurium (ATCC)	JENNY		Gram negative rods		API Staph	1337	KF828874.1	1330/1335(99%)	2/1335(0%)	Salmonella sp		Non-Reactor	Non-Reactor	Reactor	Non-Reactor
Staphylococcus aureus (ATCC)	JCU MCC	12600 ATCC	Gram positive cocci		API Staph	1324	KY218901.1	1324/1324(100%)	0/1324(0%)	Staphylococcus aureus subsp. anaerobius strain		Non-Reactor	Reactor	Reactor	Non-Reactor
Staphylococcus epidermidis (ATCC)	JCU MCC		Gram positive cocci		API Staph	1344	MG759527.1	1344/1344(100%)	0/1344(0%)	Staphylococcus sp. strain S14		Non-Reactor	Reactor	Reactor	Non-Reactor
Staphylococcus saprophyticus	JCU MCC		Gram positive cocci		API Strep	1344	MF776606.1	1344/1344(100%)	0/1344(0%)	Staphylococcus saprophyticus strain MBL_B9		Non-Reactor	Reactor	Reactor	Non-Reactor
Streptococcus agalactiae	JCU MCC	13813 ATCC	Gram positive cocci		API Strep	1341	HQ658089.1	1341/1341(100%)	0/1341(0%)	Streptococcus agalactiae strain GY105		Non-Reactor	Reactor	Reactor	Non-Reactor
Streptococcus dysgalactiae	JCU MCC	4669 NCTC	Gram positive cocci		API Strep	1341	KF828882.1	1338/1340(99%)	0/1340(0%)	Streptococcus sp. XJ150-1212-NJR1		Non-Reactor	Reactor	Reactor	Non-Reactor
Streptococcus pneumoniae (ATCC)	JCU MCC	10015 ATCC	Gram positive cocci		API Strep							Non-Reactor	Reactor	Reactor	Non-Reactor
Streptococcus pyogenes	JCU MCC		Gram positive cocci		API Strep	1336	KU058410.1	1335/1337(99%)	1/1337(0%)	Streptococcus sp. strain HMU71		Non-Reactor	Reactor	Reactor	Non-Reactor
Streptococcus viridians	JCU MCC		Gram positive cocci		API Strep	1339	LC006301.1	1336/1336(100%)	0/1336(0%)	Uncultured alpha proteobacterium gene		Non-Reactor	Reactor	Reactor	Non-Reactor

Ureaplasma plasmid	DNA 2.0					1337	KX349997.1	1334/1334(100%)	0/1334(0%)	Escherichia coli strain F3-1-9		Non-Reactor	Reactor	Reactor	Non-Reactor
Taxonomically associated organisms															
M. abscessus	PC3			Gram positive acid fast rods		1311	AP018436.1	1305/1311(99%)	3/1311(0%)	Mycobacterium abscessus subsp. bolletii BD DNA		Reactor	Non-Reactor	Reactor	Reactor
M. avium	JACKIE	Sample 4		Gram positive acid fast rods		1325	MF429135.1	1323/1323(100%)	0/1323(0%)	Kocuria kristinae strain CAU4519		Non-Reactor	Reactor	Reactor	Non-Reactor
M. bovis	ABDUL			Gram positive acid fast rods		1323	CP025607.1	1320/1322(99%)	2/1322(0%)	Mycobacterium tuberculosis strain 186-10		Reactor	Reactor	Reactor	Reactor
M intracellulare	JACKIE			Gram positive acid fast rods		1308	GU142933.1	1301/1307(99%)	2/1307(0%)	Mycobacterium fortuitum strain M213		Non-Reactor	Reactor	Reactor	Reactor
M. fortuitum	JACKIE	Sample 14		Gram positive acid fast rods		1313	JF729474.1	1300/1300(100%)	0/1300(0%)	Mycobacterium sp. G2Z43		Non-Reactor	Reactor	Reactor	Non-Reactor
M. peregrinum	JACKIE	Sample 23		Gram positive acid fast rods		1311	KJ873240.2	1303/1311(99%)	3/1311(0%	Mycobacterium sp. GPK 1020		Non-Reactor	Reactor	Reactor	Non-Reactor
Corynebactrium diptheriae	JCU MCC		Gram positive rods		API Corynae	1281	LT960556.1	1281/1281(100%)	0/1281(0%)	Corynebacterium diphtheriae		Non-Reactor	Reactor	Reactor	Non-Reactor
Bacillus cereus	JCU MCC		Gram positive rods			1343	MG738303.1	1343/1344(99%)	1/1344(0%)	Bacillus cereus strain HU2		Non-Reactor	Reactor	Reactor	Non-Reactor
DNA 12 (M. intracellulare)	JACKIE			Gram positive acid fast rods		1312	CP023149.1	1310/1310(100%)	0/1310(0%)	Mycobacterium intracellulare strain FLAC0181		Non-Reactor	Reactor	Reactor	Non-Reactor
H37Rv	PC3			Gram positive acid fast rods								Reactor	Reactor	Reactor	Reactor

REFERENCES

Abebe, G., Paasch, F., Apers, L., Rigouts, L. and Colebunders, R. (2011) Tuberculosis drug resistance testing by molecular methods: opportunities and challenges in resource limited settings. *Journal of Microbiology Methods* 84: 155-60

Abraham, P.R., Chauhan, A., Gangane, R., Sharma, V.D. and Shivannavar, C. (2013) Prevalence of Mycobacterium tuberculosis with multiple copies of IS6110 elements in Gulbarga, South India. *International Journal of Mycobacteriology* 2: 237-239

Adikaram, C.P., Perera, J. and Wijesundera, S.S. (2012) Geographical profile of rpoB gene mutations in rifampicin resistant Mycobacterium tuberculosis isolates in Sri Lanka. *Microbial Drug Resistance* 18: 525-30

Aia, P., Kal, M., Lavu, E., John, L.N., Johnson, K., Coulter, C., Ershova, J., Tosas, O., Zignol, M., Ahmadova, S. and Islam, T. (2016a) The Burden of Drug-Resistant Tuberculosis in Papua New Guinea: Results of a Large Population-Based Survey. *PLoS One* 11: e0149806

Aia, P., Kal, M., Lavu, E., John, L.N., Johnson, K., Coulter, C., Ershova, J., Tosas, O., Zignol, M., Ahmadova, S. and Islam, T. (2016b) The Burden of Drug-Resistant Tuberculosis in Papua New Guinea: Results of a Large Population-Based Survey. *Public Library of Science One* 11: e0149806

Ajbani, K., Lin, S.Y., Rodrigues, C., Nguyen, D., Arroyo, F., Kaping, J., Jackson, L., Garfein, R.S., Catanzaro, D., Eisenach, K., Victor, T.C., Crudu, V., Gler, M.T., Ismail, N., Desmond, E., Catanzaro, A. and Rodwell, T.C. (2015) Evaluation of pyrosequencing for detecting extensively drug-resistant Mycobacterium tuberculosis among clinical isolates from four high-burden countries. *Antimicrobial Agents and Chemotherapy* 59: 414-20

Akobeng, A.K. (2007) Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatrica* 96: 338-41

Akram, A., Maley, M., Gosbell, I., Nguyen, T. and Chavada, R. (2017) Utility of 16S rRNA PCR performed on clinical specimens in patient management. *International Journal of Infectious Diseases* 57: 144-149

Al-Mutairi, N.M., Ahmad, S. and Mokaddas, E. (2011) First report of molecular detection of fluoroquinolone resistance-associated gyrA mutations in multidrug-resistant clinical Mycobacterium tuberculosis isolates in Kuwait. *BMC Research Notes* 4: 123

Alifano, P., Palumbo, C., Pasanisi, D. and Tala, A. (2015) Rifampicin-resistance, rpoB polymorphism and RNA polymerase genetic engineering. *Journal of Biotechnology* 202: 60-77

Alonso, H., Samper, S., Martin, C. and Otal, I. (2013) Mapping IS6110 in high-copy number Mycobacterium tuberculosis strains shows specific insertion points in the Beijing genotype. *BioMed Central Genomics* 14: 422

Amin, I., Idrees, M., Awan, Z., Shahid, M., Afzal, S. and Hussain, A. (2011) PCR could be a method of choice for identification of both pulmonary and extra-pulmonary tuberculosis. *BioMed Central Research Notes* 4: 332

Anda, M., Ohtsubo, Y., Okubo, T., Sugawara, M., Nagata, Y., Tsuda, M., Minamisawa, K. and Mitsui, H. (2015) Bacterial clade with the ribosomal RNA operon on a small

plasmid rather than the chromosome. *Proceedings of the National Academy of Sciences of the United States of America* 112: 14343-7

Andersson, D.I. and Hughes, D. (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology* 8: 260-71

Andre, E., Goeminne, L., Cabibbe, A., Beckert, P., Kabamba Mukadi, B., Mathys, V., Gagneux, S., Niemann, S., Van Ingen, J. and Cambau, E. (2017) Consensus numbering system for the rifampicin resistance-associated rpoB gene mutations in pathogenic mycobacteria. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 23: 167-172

Anthwal, D., Gupta, R.K., Bhalla, M., Bhatnagar, S., Tyagi, J.S. and Haldar, S. (2017) Direct Detection of Rifampin and Isoniazid Resistance in Sputum Samples from Tuberculosis Patients by High-Resolution Melt Curve Analysis. *J Clin Microbiol* 55: 1755-1766

Ao, W., Aldous, S., Woodruff, E., Hicke, B., Rea, L., Kreiswirth, B. and Jenison, R. (2012) Rapid detection of rpoB gene mutations conferring rifampin resistance in Mycobacterium tuberculosis. *Journal of Clinical Microbiology* 50: 2433-40

Arefzadeh, S., Azimi, T., Nasiri, M.J., Nikpor, Z., Dabiri, H., Doustdar, F., Goudarzi, H. and Allahyartorkaman, M. (2020) High-resolution melt curve analysis for rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a single-centre study in Iran. *New microbes and new infections* 35: 100665

Armand, S., Vanhuls, P., Delcroix, G., Courcol, R. and Lemaitre, N. (2011) Comparison of the Xpert MTB/RIF test with an IS6110-TaqMan real-time PCR assay for direct detection of Mycobacterium tuberculosis in respiratory and nonrespiratory specimens. *Journal of Clinical Microbiology* 49: 1772-6

Arthofer, W., Steiner, F.M. and Schlick-Steiner, B.C. (2011) Rapid and cost-effective screening of newly identified microsatellite loci by high-resolution melting analysis. *Molecular genetics and genomics* 286: 225-35

Aye, K.S., Nakajima, C., Yamaguchi, T., Win, M.M., Shwe, M.M., Win, A.A., Lwin, T., Nyunt, W.W., Ti, T. and Suzuki, Y. (2016) Genotypic characterization of multi-drug-resistant Mycobacterium tuberculosis isolates in Myanmar. *Journal of Infection and Chemotherapy* 22: 174-9

Bachmann, L., Daubl, B., Lindqvist, C., Kruckenhauser, L., Teschler-Nicola, M. and Haring, E. (2008) PCR diagnostics of Mycobacterium tuberculosis in historic human long bone remains from 18th century burials in Kaiserebersdorf, Austria. *BioMed Central Research Notes* 1: 83

Bainomugisa, A., Duarte, T., Lavu, E., Pandey, S., Coulter, C., Marais, B.J. and 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. *Microbial genomics* 4

Ballif, M., Harino, P., Ley, S., Carter, R., Coulter, C., Niemann, S., Borrell, S., Fenner, L., Siba, P., Phuanukoonnon, S., Gagneux, S. and Beck, H.P. (2012) Genetic diversity of Mycobacterium tuberculosis in Madang, Papua New Guinea. *The International Journal of Tuberculosis and Lung Disease* 16: 1100-7

Barken, K.B., Haagensen, J.A. and Tolker-Nielsen, T. (2007) Advances in nucleic acid-based diagnostics of bacterial infections. *Clinica Chimica Acta* 384: 1-11

Ben-Selma, W., Ben-Kahla, I., Marzouk, M., Ferjeni, A., Ghezal, S., Ben-Said, M. and Boukadida, J. (2009) Rapid detection of Mycobacterium tuberculosis in sputum by Patho-TB kit in comparison with direct microscopy and culture. *Diagnostic Microbiology and Infectious Disease* 65: 232-5

Bengtson, H.N., Homolka, S., Niemann, S., Reis, A.J., da Silva, P.E., Gerasimova, Y.V., Kolpashchikov, D.M. and Rohde, K.H. (2017) Multiplex detection of extensively drug resistant tuberculosis using binary deoxyribozyme sensors. *Biosensors & bioelectronics* 94: 176-183

Bentaleb, E.M., El Messaoudi, M.D., Abid, M., Messaoudi, M., Yetisen, A.K., Sefrioui, H., Amzazi, S. and Ait Benhassou, H. (2017) Plasmid-based high-resolution melting analysis for accurate detection of rpoB mutations in Mycobacterium tuberculosis isolates from Moroccan patients. *BMC Infectious Diseases* 17: 548

Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., Milovic, A., Jones, M., O'Brien, S.M., Persing, D.H., Ruesch-Gerdes, S., Gotuzzo, E., Rodrigues, C., Alland, D. and Perkins, M.D. (2010) Rapid molecular detection of tuberculosis and rifampin resistance. *The New England Journal of Medicine* 363: 1005-15

Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., Zoppo, M., Orgeur, M., Khanna, V., Hardy, D., Mangenot, S., Barbe, V., Medigue, C., Ma, L., Bouchier, C., Tavanti, A., Larrouy-Maumus, G. and Brosch, R. (2020) TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis lineages. *Nature Communication* 11: 684

Brandis, G., Wrande, M., Liljas, L. and Hughes, D. (2012) Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology* 85: 142-51

Breslauer, K.J., Frank, R., Blocker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences (U S A)* 83: 3746-50

Brites, D. and Gagneux, S. (2012) Old and new selective pressures on Mycobacterium tuberculosis. *Infection, Genetics and Evolution* 12: 678-85

Brites, D. and Gagneux, S. (2015) Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunological reviews* 264: 6-24

Broccolo, F., Scarpellini, P., Locatelli, G., Zingale, A., Brambilla, A.M., Cichero, P., Sechi, L.A., Lazzarin, A., Lusso, P. and Malnati, M.S. (2003) Rapid diagnosis of mycobacterial infections and quantitation of Mycobacterium tuberculosis load by two real-time calibrated PCR assays. *Journal of Clinical Microbiology* 41: 4565-72

Buh Gasparic, M., Cankar, K., Zel, J. and Gruden, K. (2008) Comparison of different real-time PCR chemistries and their suitability for detection and quantification of genetically modified organisms. *BioMed Central Biotechnology* 8: 26

Buh Gasparic, M., Tengs, T., La Paz, J.L., Holst-Jensen, A., Pla, M., Esteve, T., Zel, J. and Gruden, K. (2010) Comparison of nine different real-time PCR chemistries for qualitative and quantitative applications in GMO detection. *Analytical and Bioanalytical Chemistry* 396: 2023-9

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* 55: 611-22

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S.R., Ignatyeva, O., Kontsevaya, I., Corander, J., Bryant, J., Parkhill, J., Nejentsev, S., Horstmann, R.D., Brown, T. and Drobniewski, F. (2014) Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature genetics* 46: 279-86

Castro, R.A.D., Borrell, S. and Gagneux, S. (2020) The within-host evolution of antimicrobial resistance in Mycobacterium tuberculosis. *FEMS Microbiology Reviews*

Chadha, V.K., Vaidyanathan, P.S., Jagannatha, P.S., Unnikrishnan, K.P. and Mini, P.A. (2003) Annual risk of tuberculous infection in the northern zone of India. *Bulletin of the World Health Organization* 81: 573-80

Chakravorty, S., Simmons, A.M., Rowneki, M., Parmar, H., Cao, Y., Ryan, J., Banada, P.P., Deshpande, S., Shenai, S., Gall, A., Glass, J., Krieswirth, B., Schumacher, S.G., Nabeta, P., Tukvadze, N., Rodrigues, C., Skrahina, A., Tagliani, E., Cirillo, D.M., Davidow, A., Denkinger, C.M., Persing, D., Kwiatkowski, R., Jones, M. and Alland, D. (2017) The New Xpert MTB/RIF Ultra: Improving Detection of Mycobacterium tuberculosis and Resistance to Rifampin in an Assay Suitable for Point-of-Care Testing. *mBio* 8

Cohen, K.A., Abeel, T., McGuire, A.M., Desjardins, C.A., Munsamy, V., Shea, T.P., Walker, B.J., Bantubani, N., Almeida, D.V., Alvarado, L., Chapman, S.B., Mvelase, N.R., Duffy, E.Y., Fitzgerald, M.G., Govender, P., Gujja, S., Hamilton, S., Howarth, C., Larimer, J.D., Maharaj, K., Pearson, M.D., Priest, M.E., Zeng, Q.D., Padayatchi, N., Grosset, J., Young, S.K., Wortman, J., Mlisana, K.P., O'Donnell, M.R., Birren, B.W., Bishai, W.R., Pym, A.S. and Earl, A.M. (2015) Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal. *PloS One Medicine* 12: 22

Coll, F., Preston, M., Guerra-Assuncao, J.A., Hill-Cawthorn, G., Harris, D., Perdigao, J., Viveiros, M., Portugal, I., Drobniewski, F., Gagneux, S., Glynn, J.R., Pain, A., Parkhill, J., McNerney, R., Martin, N. and Clark, T.G. (2014) PolyTB: a genomic variation map for Mycobacterium tuberculosis. *Tuberculosis (Edinburgh)* 94: 346-54

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S. and Gagneux, S. (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nature genetics* 44: 106-10

Coscolla, M. and Gagneux, S. (2014) Consequences of genomic diversity in Mycobacterium tuberculosis. *Seminars in immunology* 26: 431-44

Cross, G.B., Coles, K., Nikpour, M., Moore, O.A., Denholm, J., McBryde, E.S., Eisen, D.P., Warigi, B., Carter, R., Pandey, S., Harino, P., Siba, P., Coulter, C., Mueller, I., Phuanukoonnon, S. and Pellegrini, M. (2014a) TB incidence and characteristics in the remote gulf province of Papua New Guinea: a prospective study. *BioMed Central Infectious Diseases* 14: 93

Cross, G.B., Coles, K., Nikpour, M., Moore, O.A., Denholm, J., McBryde, E.S., Eisen, D.P., Warigi, B., Carter, R., Pandey, S., Harino, P., Siba, P., Coulter, C., Mueller, I., Phuanukoonnon, S. and Pellegrini, M. (2014b) TB incidence and characteristics in the remote gulf province of Papua New Guinea: a prospective study. *BMC Infectious Diseases* 14: 93

Daniel, T.M. (2006) The history of tuberculosis. *Respiratory Medicine* 100: 1862-70

Daum, L.T., Fourie, P.B., Bhattacharyya, S., Ismail, N.A., Gradus, S., Maningi, N.E., Omar, S.V. and Fischer, G.W. (2014) Next-generation sequencing for identifying pyrazinamide resistance in Mycobacterium tuberculosis. *Clincal Infectious Diseases* 58: 903-4

de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M., Jackson-Sillah, D.J., Fox, A., DeRiemer, K., Gagneux, S., Borgdorff, M.W., McAdam, K., Corrah, T., Small, P.M. and Adegbola, R.A. (2008) Progression to active tuberculosis, but not transmission, varies by M. tuberculosis lineage in The Gambia. *The Journal of infectious diseases* 198: 1037-43

DeJesus, M.A., Sacchettini, J.C. and Ioerger, T.R. (2013) Reannotation of translational start sites in the genome of Mycobacterium tuberculosis. *Tuberculosis (Edinburgh)* 93: 18-25

Deshpande, A., Gans, J., Graves, S.W., Green, L., Taylor, L., Kim, H.B., Kunde, Y.A., Leonard, P.M., Li, P.E., Mark, J., Song, J., Vuyisich, M. and White, P.S. (2010) A rapid multiplex assay for nucleic acid-based diagnostics. *Journal of Microbiology Methods* 80: 155-63

Diefenbach-Elstob, T., Guernier, V., Burgess, G., Pelowa, D., Dowi, R., Gula, B., Puri, M., Pomat, W., McBryde, E., Plummer, D., Rush, C. and Warner, J. (2019) Molecular Evidence of Drug-Resistant Tuberculosis in the Balimo Region of Papua New Guinea. *Tropical medicine and infectious disease* 4

Donkor, E.S. (2013) Sequencing of bacterial genomes: principles and insights into pathogenesis and development of antibiotics. *Genes (Basel)* 4: 556-72

Dorman, S.E. and Chaisson, R.E. (2007) From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nature Medicine* 13: 295-8

Dorman, S.E., Schumacher, S.G., Alland, D., Nabeta, P., Armstrong, D.T., King, B., Hall, S.L., Chakravorty, S., Cirillo, D.M., Tukvadze, N., Bablishvili, N., Stevens, W., Scott, L., Rodrigues, C., Kazi, M.I., Joloba, M., Nakiyingi, L., Nicol, M.P., Ghebrekristos, Y., Anyango, I., Murithi, W., Dietze, R., Lyrio Peres, R., Skrahina, A., Auchynka, V., Chopra, K.K., Hanif, M., Liu, X., Yuan, X., Boehme, C.C., Ellner, J.J. and Denkinger, C.M. (2018) Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *The Lancet Infectious diseases* 18: 76-84

Drobniewski, F., Nikolayevskyy, V., Maxeiner, H., Balabanova, Y., Casali, N., Kontsevaya, I. and Ignatyeva, O. (2013) Rapid diagnostics of tuberculosis and drug resistance in the industrialized world: clinical and public health benefits and barriers to implementation. *BioMed Central Medicine* 11: 190

Druml, B. and Cichna-Markl, M. (2014) High resolution melting (HRM) analysis of DNA--its role and potential in food analysis. *Food chemistry* 158: 245-54

Dye, C. (2009) Doomsday postponed? Preventing and reversing epidemics of drug-resistant tuberculosis. *Nature Reviews Microbiology* 7: 81-7

Edwards, K.J., Logan, J.M., Langham, S., Swift, C. and Gharbia, S.E. (2012) Utility of real-time amplification of selected 16S rRNA gene sequences as a tool for detection and identification of microbial signatures directly from clinical samples. *Journal of Medical Microbiology* 61: 645-52

Eichbaum, Q. and Rubin, E.J. (2002) Tuberculosis. Advances in laboratory diagnosis and drug susceptibility testing. *American Journal of Clinical Pathology* 118 Suppl: S3-17

Elbir, H. and Ibrahim, N.Y. (2014) Frequency of mutations in the rpoB gene of multidrug-resistant Mycobacterium tuberculosis clinical isolates from Sudan. *Journal of Infection in Developing Countries* 8: 796-798

Fogel, N. (2015) Tuberculosis: A disease without boundaries. *Tuberculosis (Edinburgh)* 95: 527-31

Fraga, D., Meulia, T. and Fenster, S. (2008) Real-Time PCR, Current Protocols Essential Laboratory Techniques. John Wiley & Sons, Inc.

Frank, J.A., Reich, C.I., Sharma, S., Weisbaum, J.S., Wilson, B.A. and Olsen, G.J. (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and Environmental Microbiology* 74: 2461-70

Friedrich, S.O., Rachow, A., Saathoff, E., Singh, K., Mangu, C.D., Dawson, R., Phillips, P.P., Venter, A., Bateson, A., Boehme, C.C., Heinrich, N., Hunt, R.D., Boeree, M.J., Zumla, A., McHugh, T.D., Gillespie, S.H., Diacon, A.H. and Hoelscher, M. (2013) Assessment of the sensitivity and specificity of Xpert MTB/RIF assay as an early sputum biomarker of response to tuberculosis treatment. *The Lancet Respiratory Medicine* 1: 462-70

Gagneux, S. (2012) Host-pathogen coevolution in human tuberculosis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367: 850-9

Galarza, M., Fasabi, M., Levano, K.S., Castillo, E., Barreda, N., Rodriguez, M. and Guio, H. (2016) High-resolution melting analysis for molecular detection of multidrug resistance tuberculosis in Peruvian isolates. *BMC Infectious Diseases* 16: 260

Gilpin, C.M., Simpson, G., Vincent, S., O'Brien, T.P., Knight, T.A., Globan, M., Coulter, C. and Konstantinos, A. (2008) Evidence of primary transmission of multidrug-resistant tuberculosis in the Western Province of Papua New Guinea. *Medical Journal of Australia* 188: 148-52

Guernier, V., Diefenbach-Elstob, T., Pelowa, D., Pollard, S., Burgess, G., McBryde, E.S. and Warner, J. (2017) Molecular diagnosis of suspected tuberculosis from archived smear slides from the Balimo region, Papua New Guinea. *International Journal of Infectious Diseases* 67: 75-81

Guernier, V., Diefenbach-Elstob, T., Pelowa, D., Pollard, S., Burgess, G., McBryde, E.S. and Warner, J. (2018) Molecular diagnosis of suspected tuberculosis from archived smear slides from the Balimo region, Papua New Guinea. *International Journal of Infectious Diseases* 67: 75-81

Guernier-Cambert, V., Diefenbach-Elstob, T., Klotoe, B.J., Burgess, G., Pelowa, D., Dowi, R., Gula, B., McBryde, E.S., Refrégier, G., Rush, C., Sola, C. and Warner, J. (2019) Diversity of Mycobacterium tuberculosis in the Middle Fly District of Western Province, Papua New Guinea: microbead-based spoligotyping using DNA from Ziehl-Neelsen-stained microscopy preparations. *Scientific reports* 9: 15549

Gupta, A. and Anupurba, S. (2015) Detection of drug resistance in Mycobacterium tuberculosis: Methods, principles and applications. *Indian Journal of Tuberculosis* 62: 13-22

Hakovirta, J.R., Prezioso, S., Hodge, D., Pillai, S.P. and Weigel, L.M. (2016) Identification and Analysis of Informative Single Nucleotide Polymorphisms in 16S rRNA Gene Sequences of the Bacillus cereus Group. *Journal of Clinical Microbiology* 54: 2749-2756

Harries, A.D., Lin, Y., Satyanarayana, S., Lonnroth, K., Li, L., Wilson, N., Chauhan, L.S., Zachariah, R., Baker, M.A., Jeon, C.Y., Murray, M.B., Maher, D., Bygbjerg, I.C.,

Enarson, D.A., Billo, N.E. and Kapur, A. (2011) The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. *International Journal of Tuberculosis and Lung Disease* 15: 1436-44, i

Hatherell, H.A., Colijn, C., Stagg, H.R., Jackson, C., Winter, J.R. and Abubakar, I. (2016) Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *Bmc Medicine* 14: 13

Heather, J.M. and Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107: 1-8

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T. and Voelkerding, K.V. (2006) Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. *Clinical Chemistry* 52: 494-503

Heyckendorf, J., Andres, S., Koser, C.U., Olaru, I.D., Schon, T., Sturegard, E., Beckert, P., Schleusener, V., Kohl, T.A., Hillemann, D., Moradigaravand, D., Parkhill, J., Peacock, S.J., Niemann, S., Lange, C. and Merker, M. (2018) What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi- and Extensively Drug-Resistant Tuberculosis. *Antimicrobial Agents and Chemotherapy* 62

Hinic, V., Feuz, K., Turan, S., Berini, A., Frei, R., Pfeifer, K. and Goldenberger, D. (2017) Clinical evaluation of the Abbott RealTime MTB Assay for direct detection of Mycobacterium tuberculosis-complex from respiratory and non-respiratory samples. *Tuberculosis (Edinburgh, Scotland)* 104: 65-69

Homolka, S., Post, E., Oberhauser, B., George, A.G., Westman, L., Dafae, F., Rusch-Gerdes, S. and Niemann, S. (2008) High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. *BioMed Central Microbiology* 8: 103

Hughes, D. and Andersson, D.I. (2015) Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nature reviews, Genetics* 16: 459-71

Huyen, M.N., Tiemersma, E.W., Kremer, K., de Haas, P., Lan, N.T., Buu, T.N., Sola, C., Cobelens, F.G. and van Soolingen, D. (2013) Characterisation of Mycobacterium tuberculosis isolates lacking IS6110 in Viet Nam. *International Journal of Tuberculosis and Lung Disease* 17: 1479-85

Ismail, N.A., Ismail, M.F., Noor, S.S. and Camalxaman, S.N. (2016) Genotypic Detection of rpoB and katG Gene Mutations Associated with Rifampicin and Isoniazid Resistance in Mycobacterium Tuberculosis Isolates: A Local Scenario (Kelantan). *The Malaysian Journal of Medical Sciences* 23: 22-6

Josefsen, M.H., Lofstrom, C., Sommer, H.M. and Hoorfar, J. (2009) Diagnostic PCR: comparative sensitivity of four probe chemistries. *Molecular and Cellular Probes* 23: 201-3

Keeler, E., Perkins, M.D., Small, P., Hanson, C., Reed, S., Cunningham, J., Aledort, J.E., Hillborne, L., Rafael, M.E., Girosi, F. and Dye, C. (2006) Reducing the global burden of tuberculosis: the contribution of improved diagnostics. *Nature* 444 Suppl 1: 49-57

Khosravi, A.D., Alami, A., Meghdadi, H. and Hosseini, A.A. (2017a) Identification of Mycobacterium tuberculosis in Clinical Specimens of Patients Suspected of Having Extrapulmonary Tuberculosis by Application of Nested PCR on Five Different Genes. *Frontiers in cellular and infection microbiology* 7: 3

Khosravi, A.D., Hashemzadeh, M., Hashemi Shahraki, A. and Teimoori, A. (2017b) Differential Identification of Mycobacterial Species Using High-Resolution Melting Analysis. *Frontiers in Microbiology* 8: 2045

Knez, K., Spasic, D., Janssen, K.P. and Lammertyn, J. (2014) Emerging technologies for hybridization based single nucleotide polymorphism detection. *The Analyst* 139: 353-70

Koch, A., Mizrahi, V. and Warner, D.F. (2014) The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? *Emerg Microbes Infect* 3: e17

Kulesh, D.A., Loveless, B.M., Norwood, D., Garrison, J., Whitehouse, C.A., Hartmann, C., Mucker, E., Miller, D., Wasieloski, L.P., Jr., Huggins, J., Huhn, G., Miser, L.L., Imig, C., Martinez, M., Larsen, T., Rossi, C.A. and Ludwig, G.V. (2004) Monkeypox virus detection in rodents using real-time 3'-minor groove binder TaqMan assays on the Roche LightCycler. *Laboratory Investigation* 84: 1200-8

Lavu, E.K., Johnson, K., Banamu, J., Pandey, S., Carter, R., Coulter, C., Aia, P., Majumdar, S.S., Marais, B.J., Graham, S.M. and Vince, J. (2019) Drug-resistant tuberculosis diagnosis since Xpert(®) MTB/RIF introduction in Papua New Guinea, 2012-2017. *Public Health Action* 9: S12-s18

Lee, A.S., Ong, D.C., Wong, J.C., Siu, G.K. and Yam, W.C. (2012) High-resolution melting analysis for the rapid detection of fluoroquinolone and streptomycin resistance in Mycobacterium tuberculosis. *PLoS One* 7: e31934

Lee, J.Y. (2015) Diagnosis and treatment of extrapulmonary tuberculosis. *Tuberculosis and Respiratory Diseases (Seoul)* 78: 47-55

Lemos, A.C.M. and Matos, E.D. (2013) Multidrug-resistant tuberculosis. *Brazilian Journal of Infectious Diseases* 17: 239-246

Ley, S.D., Harino, P., Vanuga, K., Kamus, R., Carter, R., Coulter, C., Pandey, S., Feldmann, J., Ballif, M., Siba, P.M., Phuanukoonnon, S., Gagneux, S. and Beck, H.P. (2014a) Diversity of Mycobacterium tuberculosis and drug resistance in different provinces of Papua New Guinea. *BioMed Central Microbiology* 14: 307

Ley, S.D., Riley, I. and Beck, H.P. (2014b) Tuberculosis in Papua New Guinea: from yesterday until today. *Microbes Infect* 16: 607-14

Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E. and Wittwer, C. (2004) Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry* 50: 1156-64

Lin, C., Russell, C., Soll, B., Chow, D., Bamrah, S., Brostrom, R., Kim, W., Scott, J. and Bankowski, M.J. (2018) Increasing Prevalence of Nontuberculous Mycobacteria in Respiratory Specimens from US-Affiliated Pacific Island Jurisdictions(1). *Emerging infectious diseases* 24: 485-491

Lin, P.L. and Flynn, J.L. (2018) The End of the Binary Era: Revisiting the Spectrum of Tuberculosis. *Journal of immunology (Baltimore, Md. : 1950)* 201: 2541-2548

Mabey, D., Peeling, R.W., Ustianowski, A. and Perkins, M.D. (2004) Diagnostics for the developing world. *Nature Reviews Microbiology* 2: 231-240

Mackay, I.M., Arden, K.E. and Nitsche, A. (2002) Real-time PCR in virology. *Nucleic Acids Research* 30: 1292-305

Manson, A.L., Cohen, K.A., Abeel, T., Desjardins, C.A., Armstrong, D.T., Barry, C.E., 3rd, Brand, J., Chapman, S.B., Cho, S.N., Gabrielian, A., Gomez, J., Jodals, A.M., Joloba, M., Jureen, P., Lee, J.S., Malinga, L., Maiga, M., Nordenberg, D., Noroc, E.,

Romancenco, E., Salazar, A., Ssengooba, W., Velayati, A.A., Winglee, K., Zalutskaya, A., Via, L.E., Cassell, G.H., Dorman, S.E., Ellner, J., Farnia, P., Galagan, J.E., Rosenthal, A., Crudu, V., Homorodean, D., Hsueh, P.R., Narayanan, S., Pym, A.S., Skrahina, A., Swaminathan, S., Van der Walt, M., Alland, D., Bishai, W.R., Cohen, T., Hoffner, S., Birren, B.W. and Earl, A.M. (2017) Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. *Nature genetics* 49: 395-402

McGinn, S. and Gut, I.G. (2013) DNA sequencing - spanning the generations. *New biotechnology* 30: 366-72

McNerney, R., Cunningham, J., Hepple, P. and Zumla, A. (2015) New tuberculosis diagnostics and rollout. *International Journal of Infectious Diseases* 32: 81-6

McNerney, R., Maeurer, M., Abubakar, I., Marais, B., McHugh, T.D., Ford, N., Weyer, K., Lawn, S., Grobusch, M.P., Memish, Z., Squire, S.B., Pantaleo, G., Chakaya, J., Casenghi, M., Migliori, G.B., Mwaba, P., Zijenah, L., Hoelscher, M., Cox, H., Swaminathan, S., Kim, P.S., Schito, M., Harari, A., Bates, M., Schwank, S., O'Grady, J., Pletschette, M., Ditui, L., Atun, R. and Zumla, A. (2012) Tuberculosis diagnostics and biomarkers: needs, challenges, recent advances, and opportunities. *Journal of Infectious Diseases* 205 Suppl 2: S147-58

Meghdadi, H., Khosravi, A.D., Ghadiri, A.A., Sina, A.H. and Alami, A. (2015) Detection of Mycobacterium tuberculosis in extrapulmonary biopsy samples using PCR targeting IS6110, rpoB, and nested-rpoB PCR Cloning. *Frontiers in Microbiology* 6: 675

Morey, M., Fernandez-Marmiesse, A., Castineiras, D., Fraga, J.M., Couce, M.L. and Cocho, J.A. (2013) A glimpse into past, present, and future DNA sequencing. *Molecular genetics and metabolism* 110: 3-24

Muller, Z., Stelzl, E., Bozic, M., Haas, J., Marth, E. and Kessler, H.H. (2004) Evaluation of automated sample preparation and quantitative PCR LCx assay for determination of human immunodeficiency virus type 1 RNA. *Journal of Clinical Microbiology* 42: 1439-43

Nagai, Y., Iwade, Y., Hayakawa, E., Nakano, M., Sakai, T., Mitarai, S., Katayama, M., Nosaka, T. and Yamaguchi, T. (2013) High resolution melting curve assay for rapid detection of drug-resistant Mycobacterium tuberculosis. *Journal of Infection and Chemotherapy* 19: 1116-1125

Naveen, G. and Peerapur, B.V. (2012) Comparison of the Lowenstein-Jensen Medium, the Middlebrook 7H10 Medium and MB/BacT for the Isolation of Mycobacterium Tuberculosis (MTB) from Clinical Specimens. *Journal of Clinical and Diagnostic Research* 6: 1704-9

Nebenzahl-Guimaraes, H., Jacobson, K.R., Farhat, M.R. and Murray, M.B. (2014) Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in Mycobacterium tuberculosis. *The Journal of antimicrobial chemotherapy* 69: 331-42

Ngabonziza, J.C.S., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfadia, O., Antoine, R., Niyigena, E.B., Mulders, W., Fissette, K., Diels, M., Gaudin, C., Duthoy, S., Ssengooba, W., André, E., Kaswa, M.K., Habimana, Y.M., Brites, D., Affolabi, D., Mazarati, J.B., de Jong, B.C., Rigouts, L., Gagneux, S., Meehan, C.J. and Supply, P. (2020) A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nature communications* 11: 2917

Niehaus, A.J., Mlisana, K., Gandhi, N.R., Mathema, B. and Brust, J.C.M. (2015) High Prevalence of inhA Promoter Mutations among Patients with Drug-Resistant Tuberculosis in KwaZulu-Natal, South Africa. *PLOS ONE* 10: e0135003

Nurwidya, F., Handayani, D., Burhan, E. and Yunus, F. (2018) Molecular Diagnosis of Tuberculosis. *Chonnam Medical Journal* 54: 1-9

Orme, I.M. (2014) A new unifying theory of the pathogenesis of tuberculosis. *Tuberculosis (Edinburgh)* 94: 8-14

Palomino, J.C. (2005) Nonconventional and new methods in the diagnosis of tuberculosis: feasibility and applicability in the field. *European Respiratory Journal* 26: 339-50

Parcell, B.J., Jarchow-MacDonald, A.A., Seagar, A.L., Laurenson, I.F., Prescott, G.J. and Lockhart, M. (2017) Three year evaluation of Xpert MTB/RIF in a low prevalence tuberculosis setting: A Scottish perspective. *The Journal of infection* 74: 466-472

Park, S.H., Kim, C.K., Jeong, H.R., Son, H., Kim, S.H. and Park, M.S. (2014) Evaluation and comparison of molecular and conventional diagnostic tests for detecting tuberculosis in Korea, 2013. *Osong Public Health and Research Perspectives* 5: S3-7

Peng, J., Yu, X., Cui, Z., Xue, W., Luo, Z., Wen, Z., Liu, M., Jiang, D., Zheng, H., Wu, H., Zhang, S. and Li, Y. (2016) Multi-Fluorescence Real-Time PCR Assay for Detection of RIF and INH Resistance of M. tuberculosis. *Frontiers in Microbiology* 7: 618

Perng, C.L., Chen, H.Y., Chiueh, T.S., Wang, W.Y., Huang, C.T. and Sun, J.R. (2012) Identification of non-tuberculous mycobacteria by real-time PCR coupled with a high-resolution melting system. *The Journal of Medical Microbiology* 61: 944-51

Pettersson, E., Lundeberg, J. and Ahmadian, A. (2009) Generations of sequencing technologies. *Genomics* 93: 105-11

Pholwat, S., Stroup, S., Gratz, J., Trangan, V., Foongladda, S., Kumburu, H., Juma, S.P., Kibiki, G. and Houpt, E. (2014) Pyrazinamide susceptibility testing of Mycobacterium tuberculosis by high resolution melt analysis. *Tuberculosis (Edinb)* 94: 20-5

Pietzka, A.T., Indra, A., Stoger, A., Zeinzinger, J., Konrad, M., Hasenberger, P., Allerberger, F. and Ruppitsch, W. (2009) Rapid identification of multidrug-resistant Mycobacterium tuberculosis isolates by rpoB gene scanning using high-resolution melting curve PCR analysis. *The Journal of antimicrobial chemotherapy* 63: 1121-7

Rachow, A., Zumla, A., Heinrich, N., Rojas-Ponce, G., Mtafya, B., Reither, K., Ntinginya, E.N., O'Grady, J., Huggett, J., Dheda, K., Boehme, C., Perkins, M., Saathoff, E. and Hoelscher, M. (2011) Rapid and accurate detection of Mycobacterium tuberculosis in sputum samples by Cepheid Xpert MTB/RIF assay--a clinical validation study. *Proceedings of the National Academy of Sciences (U S A)* 6: e20458

Ramirez, M.V., Cowart, K.C., Campbell, P.J., Morlock, G.P., Sikes, D., Winchell, J.M. and Posey, J.E. (2010) Rapid detection of multidrug-resistant Mycobacterium tuberculosis by use of real-time PCR and high-resolution melt analysis. *J Clin Microbiol* 48: 4003-9

Reddington, K., O'Grady, J., Dorai-Raj, S., Maher, M., van Soolingen, D. and Barry, T. (2011) Novel multiplex real-time PCR diagnostic assay for identification and differentiation of Mycobacterium tuberculosis, Mycobacterium canettii, and Mycobacterium tuberculosis complex strains. *Journal of Clinical Microbiology* 49: 651-7

Reichmuth, M.L., Hömke, R., Zürcher, K., Sander, P., Avihingsanon, A., Collantes, J., Loiseau, C., Borrell, S., Reinhard, M., Wilkinson, R.J., Yotebieng, M., Fenner, L.,

Böttger, E.C., Gagneux, S., Egger, M. and Keller, P.M. (2020) Natural Polymorphisms in Mycobacterium tuberculosis Conferring Resistance to Delamanid in Drug-Naive Patients. *Antimicrobial agents and chemotherapy* 64

Riccardi, G. and Pasca, M.R. (2014) Trends in discovery of new drugs for tuberculosis therapy. *The Journal of Antibiotics* 67: 655-659

Roychowdhury, T., Mandal, S. and Bhattacharya, A. (2015) Analysis of IS6110 insertion sites provide a glimpse into genome evolution of Mycobacterium tuberculosis. *Sci Rep* 5: 12567

Saah, A.J. and Hoover, D.R. (1997) "Sensitivity" and "specificity" reconsidered: the meaning of these terms in analytical and diagnostic settings. *Annals of Internal Medicine* 126: 91-4

Sacchettini, J.C., Rubin, E.J. and Freundlich, J.S. (2008) Drugs versus bugs: in pursuit of the persistent predator Mycobacterium tuberculosis. *Nature Reviews Microbiology* 6: 41-52

Schofield, D.A., Sharp, N.J. and Westwater, C. (2012) Phage-based platforms for the clinical detection of human bacterial pathogens. *Bacteriophage* 2: 105-283

Seifert, M., Catanzaro, D., Catanzaro, A. and Rodwell, T.C. (2015) Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. *PLoS One* 10: e0119628

Sharma, K., Modi, M., Kaur, H., Sharma, A., Ray, P. and Varma, S. (2015) rpoB gene high-resolution melt curve analysis: a rapid approach for diagnosis and screening of drug resistance in tuberculous meningitis. *Diagnostic Microbiology and Infectious Disease* 83: 144-9

Sharma, K., Sharma, M., Singh, S., Modi, M., Sharma, A., Ray, P. and Varma, S. (2017) Real-time PCR followed by high-resolution melting curve analysis: A rapid and pragmatic approach for screening of multidrug-resistant extrapulmonary tuberculosis. *Tuberculosis (Edinburgh, Scotland)* 106: 56-61

Sharma, S., Sharma, M. and Bose, M. (2009) Mycobacterium tuberculosis infection of human monocyte-derived macrophages leads to apoptosis of T cells. *Immunology and Cell Biology* 87: 226-34

Shendure, J. and Lieberman Aiden, E. (2012) The expanding scope of DNA sequencing. *Nature Biotechnology* 30: 1084-94

Simpson, G., Coulter, C., Weston, J., Knight, T., Carter, R., Vincent, S., Robertus, L. and Konstantinos, A. (2011) Resistance patterns of multidrug-resistant tuberculosis in Western Province, Papua New Guinea. *The International Journal of Tuberculosis and Lung Disease* 15: 551-2

Singh, M., Gupta, V.H., Amarapurkar, D.N., Joshi, J.M., Baijal, R., Ramegowda, P.H., Amarapurkar, A.D. and Wangikar, P.P. (2014) Association of genetic variants with anti-tuberculosis drug induced hepatotoxicity: a high resolution melting analysis. *Infect Genet Evol* 23: 42-8

Skoura, E., Zumla, A. and Bomanji, J. (2015) Imaging in tuberculosis. *International Journal of Infectious Diseases* 32: 87-93

Slomka, M., Sobalska-Kwapis, M., Wachulec, M., Bartosz, G. and Strapagiel, D. (2017) High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies. *International Journal of Molecular Sciences* 18

Sohn, H., Aero, A.D., Menzies, D., Behr, M., Schwartzman, K., Alvarez, G.G., Dan, A., McIntosh, F., Pai, M. and Denkinger, C.M. (2014) Xpert MTB/RIF testing in a low

tuberculosis incidence, high-resource setting: limitations in accuracy and clinical impact. *Clinical Infectious Diseases* 58: 970-6

Stucki, D., Malla, B., Hostettler, S., Huna, T., Feldmann, J., Yeboah-Manu, D., Borrell, S., Fenner, L., Comas, I., Coscolla, M. and Gagneux, S. (2012) Two new rapid SNP-typing methods for classifying Mycobacterium tuberculosis complex into the main phylogenetic lineages. *Public Library of Science One* 7: e41253

Tang, Y.W., Ellis, N.M., Hopkins, M.K., Smith, D.H., Dodge, D.E. and Persing, D.H. (1998) Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. *Journal of Clinical Microbiology* 36: 3674-9

Taune, M., Ustero, P., Hiashiri, S., Huang, K., Aia, P., Morris, L., Main, S., Chan, G., du Cros, P. and Majumdar, S.S. (2019) Successful implementation of bedaquiline for multidrug-resistant TB treatment in remote Papua New Guinea. *Public Health Action* 9: S73-s79

Taylor, C.F. (2009) Mutation scanning using high-resolution melting. *Biochemical Society transactions* 37: 433-7

Tong, S.Y. and Giffard, P.M. (2012) Microbiological applications of high-resolution melting analysis. *Journal of Clinical Microbiology* 50: 3418-21

Toosky, M. and Javid, B. (2014) Novel diagnostics and therapeutics for drug-resistant tuberculosis. *British Medical Bulletin* 110: 129-40

Tortoli, E., Urbano, P., Marcelli, F., Simonetti, T.M. and Cirillo, D.M. (2012) Is real-time PCR better than conventional PCR for Mycobacterium tuberculosis complex detection in clinical samples? *Journal of Clinical Microbiology* 50: 2810-3

Trauner, A., Borrell, S., Reither, K. and Gagneux, S. (2014) Evolution of drug resistance in tuberculosis: recent progress and implications for diagnosis and therapy. *Drugs* 74: 1063-72

Turner, S., Pryer, K.M., Miao, V.P. and Palmer, J.D. (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *The Journal of Eukaryotic Microbiology* 46: 327-38

Verettas, D., Kazakos, C., Tilkeridis, C., Dermon, A., Petrou, H. and Galanis, V. (2003) Polymerase chain reaction for the detection of Mycobacterium tuberculosis in synovial fluid, tissue samples, bone marrow aspirate and peripheral blood. *Acta Orthopaedica Belgica* 69: 396-9

Versalovic, J. and Lupski, J.R. (2002) Molecular detection and genotyping of pathogens: more accurate and rapid answers. *Trends in Microbiology* 10: S15-21

Vossen, R.H., Aten, E., Roos, A. and den Dunnen, J.T. (2009) High-resolution melting analysis (HRMA): more than just sequence variant screening. *Human mutation* 30: 860-6

Wada, T., Maeda, S., Tamaru, A., Imai, S., Hase, A. and Kobayashi, K. (2004) Dual-probe assay for rapid detection of drug-resistant Mycobacterium tuberculosis by real-time PCR. *Journal of Clinical Microbiology* 42: 5277-85

Walzl, G., McNerney, R., du Plessis, N., Bates, M., McHugh, T.D., Chegou, N.N. and Zumla, A. (2018) Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *The Lancet Infectious Diseases* 18: e199-e210

Walzl, G., Ronacher, K., Hanekom, W., Scriba, T.J. and Zumla, A. (2011) Immunological biomarkers of tuberculosis. *Nature Reviews Immunology* 11: 343-54

Wang, F., Shen, H., Guan, M., Wang, Y., Feng, Y., Weng, X., Wang, H. and Zhang, W. (2011) High-resolution melting facilitates mutation screening of rpsL gene associated with streptomycin resistance in Mycobacterium tuberculosis. *Microbiol Res* 166: 121-8

Wang, F.F., Shao, L.Y., Fan, X.P., Shen, Y.J., Diao, N., Jin, J.L., Sun, F., Wu, J., Chen, J.Z., Weng, X.H., Cheng, X.J., Zhang, Y. and Zhang, W.H. (2015) Evolution and Transmission Patterns of Extensively Drug-Resistant Tuberculosis in China. *Antimicrobial Agents and Chemotherapy* 59: 818-825

Wang, H.Y., Lu, J.J., Chang, C.Y., Chou, W.P., Hsieh, J.C., Lin, C.R. and Wu, M.H. (2019) Development of a high sensitivity TaqMan-based PCR assay for the specific detection of Mycobacterium tuberculosis complex in both pulmonary and extrapulmonary specimens. *Scientific reports* 9: 113

Warner, J.M. and Currie, B.J. (2018) Melioidosis in Papua New Guinea and Oceania. *Tropical medicine and infectious disease* 3

Weyer, K., Mirzayev, F., Migliori, G.B., Van Gemert, W., D'Ambrosio, L., Zignol, M., Floyd, K., Centis, R., Cirillo, D.M., Tortoli, E., Gilpin, C., de Dieu Iragena, J., Falzon, D. and Raviglione, M. (2013) Rapid molecular TB diagnosis: evidence, policy making and global implementation of Xpert MTB/RIF. *European Respiratory Journal* 42: 252-71

WHO. (2021) Global Tuberculosis report.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America* 74: 5088-90

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4576-9

Xiao, M., Kong, F., Sorrell, T.C., Cao, Y., Lee, O.C., Liu, Y., Sintchenko, V. and Chen, S.C. (2010) Identification of pathogenic Nocardia species by reverse line blot hybridization targeting the 16S rRNA and 16S-23S rRNA gene spacer regions. *Journal of Clinical Microbiology* 48: 503-11

Yang, Q. and Rui, Y. (2016) Two Multiplex Real-Time PCR Assays to Detect and Differentiate Acinetobacter baumannii and Non- baumannii Acinetobacter spp. Carrying blaNDM, blaOXA-23-Like, blaOXA-40-Like, blaOXA-51-Like, and blaOXA-58-Like Genes. *PLoS One* 11: e0158958

Yang, S., Ramachandran, P., Rothman, R., Hsieh, Y.H., Hardick, A., Won, H., Kecojevic, A., Jackman, J. and Gaydos, C. (2009) Rapid identification of biothreat and other clinically relevant bacterial species by use of universal PCR coupled with high-resolution melting analysis. *The Journal of Cinical Microbiology* 47: 2252-5

Yu, C.X., Zhao, Z.Y., Lv, J.X. and Zhu, L. (2013) A dual molecular beacon approach for fast detection of Mycobacterium tuberculosis. *Molecular Biology Reports* 40: 1883-92

Zumla, A., Nahid, P. and Cole, S.T. (2013) Advances in the development of new tuberculosis drugs and treatment regimens. *Nature Review Drug Discovery* 12: 388-404