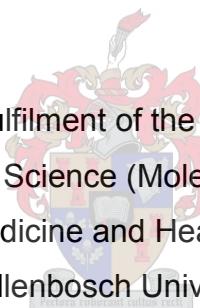


Targeted deep sequencing to detect heterogeneity in *Mycobacterium* *tuberculosis* populations

by

Ncíté Lima Da Camara

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Medical Science (Molecular Biology) in the
Faculty of Medicine and Health Sciences at
Stellenbosch University



Supervisor: Dr Margaretha De Vos

Co-supervisor: Dr Anzaan Dippenaar

Prof Robin Mark Warren

March 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2017

Copyright © 2017 Stellenbosch University

All rights reserved

Abstract

Antibiotic resistance in *Mycobacterium tuberculosis* is a worldwide problem as it drastically affects patient treatment outcome. The development of drug resistance is due to the acquisition of mutations in drug resistance conferring genes. Early detection of drug resistance is vital to improve patient therapy and prevent the transmission of drug resistant strains. It is therefore important to develop a method that is able to accurately detect minority variants conferring drug resistance to prevent treatment failure. The aim of this study was to develop an ultrasensitive method to detect underlying resistance causing variants in specific *M. tuberculosis* fluoroquinolone resistance causing genes (*gyrA* and *gyrB*).

Efficient primer sets were used to amplify the quinolone resistance-determining region of *gyrA* and *gyrB*. Targeted deep sequencing was done using the Ion Torrent Personal Genome Machine (PGM) and Illumina MiSeq platforms and sequencing data were analysed using the appropriate bioinformatics tools for the respective platforms. The method was validated using synthetic heterogeneous mixtures and was subsequently applied to identify underlying variants in patient isolates showing the acquisition of fluoroquinolone resistance.

The Illumina MiSeq platform was shown to be superior to the Ion Torrent PGM platform as it accurately detected the correct proportion of mutant DNA to a minimum frequency of 0.1%. We also showed that targeted deep sequencing is sensitive and able to detect underlying variants emerging and fluctuating during the evolution of fluoroquinolone resistance. These results show great promise for the development of an ultrasensitive diagnostic method for the early detection of fluoroquinolone resistance that could ultimately be used to improve the tuberculosis control program.

Opsomming

Antibiotikum weerstandigheid in *Mycobacterium tuberculosis* is 'n wêreld-wye probleem omdat dit die uitkoms van die behandeling van tuberkulose pasiënte drasties beïnvloed. Die ontwikkeling van middelweerstandigheid is as gevolg van die verkryging van mutasies in middelweerstandige veroorsakende gene. Die vroeë identifisering van middelweerstandigheid is belangrik om die behandeling van pasiënte te bevorder, asook om die transmissie van middelweerstandige stamme te verhoed. Dit is dus belangrik om 'n metode te ontwikkel wat akkuraat is in die identifisering van onderliggende mutasies wat middelweerstandigheid veroorsaak. Die doel van hierdie studie was om 'n uiters sensitiewe metode te ontwikkel vir die identifisering van onderliggende mutasies in die fluoroquinolone weerstandigheids gene (*gyrA* en *gyrB*).

Effektiewe inleiers was gebruik om die quinolone weerstandigheids-bepalende area van *gyrA* en *gyrB* te amplifiseer. Geteikende diep deoksiribonukleïensuur (DNS) volgorde bepaling was gedoen deur gebruik te maak van die Ion Torrent Personal Genome Machine (PGM) en die Illumina MiSeq platform. Die volgorde bepaling data was geanalyseer deur toepaslike bioinformatika sagteware vir die onderskeie platforms. Die metode was getoets deur gebruik te maak van sintetiese heterogeniese DNS mengsels en was verder toegepas om onderliggende mutasies in pasiënt isolate wat die verkryging van fluoroquinolone weerstandigheid wys, te identifiseer.

Die Illumina MiSeq platform was beter as die Ion Torrent PGM platform, siende dat dit met akkuraatheid die korrekte verhouding van mutant DNS by 'n minimum frekwensie van 0.1% kon wys. Die studie het ook gevind dat geteikende diep DNS volgorde bepaling sensitief en instaat is om onderliggende mutasies te identifiseer wat ontluik en wisselend voorkom tydens die evolusie van fluoroquinolone weerstandigheid. Hierdie resultate is belowend vir die ontwikkeling van 'n uiters sensitiewe diagnostiese metode vir die vroeë opsporing van fluoroquinolone weerstandigheid wat teneinde die tuberkulose beheerprogram kan verbeter.

Acknowledgements

I would like to acknowledge the following people for their contribution to the completion of this study.

Above all, I honour God for His amazing grace, protection, God-like peace and guidance that He has bestowed upon me, especially the past year.

I would like to extend a special word of thanks and sincere gratitude to Prof Rob Warren for his guidance, help, patience, trust, encouragement and support, especially during the past year.

Prof Rob Warren and my supervisors: Dr. Margareta De Vos and Dr. Anzaan Dippenaar for their contributions to conception, design, acquisition of data, data analysis and interpretation of data, who were also involved in revising this manuscript critically for intellectual content. In addition, I especially want to thank them for their support and patience during the past year.

Prof Sam Sampson for insightful comments, suggestions, questions and especially her willingness to assist in what was a very difficult time.

Dr. Ruben van der Merwe thank you for insightful comments, suggestions and all your help and advice with lab work and script writing.

Dr. Lizma Streicher thank you for explaining concepts and insightful comments, as well as providing me with sample data.

Dr. Monique Williams for her insightful comments, suggestions and questions.

Prof David Tabb for offering the use of Deepthought workstation which allowed for much faster data processing than the KHAOS server.

Prof Gerhard Tromp and Dr. Ruben van der Merwe for their guidance, support, patience, motivation and encouragement with Ion Torrent data analysis and writing of scripts for data processing.

Dr. Natasha Mothapo, for her friendship, guidance, encouragement, support and critically revising the manuscript.

I am grateful for the financial support provided by our collaborators Dr. Yuri van der Heijden, Dr. Karen Jacobson, and Dr. Charles Horsburgh at Boston University over the past two years, which made this study possible. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

I would like to sincerely thank my mother, Naomi Da Camara, brother, Alphonso Da Camara, and sister, Celine Da Camara, Wynand Glas as well as my family, for all their love, encouragement, and unending support.

Table of Contents

Title	Page
Contents	
Declaration.....	2
Abstract	3
Opsomming	4
Acknowledgements.....	5
Table of Contents.....	7
List of Figures	9
List of Tables	10
List of abbreviations	11
Chapter 1: General introduction	16
Background and motivation of the study.....	16
Hypotheses	18
Aims:	18
Objectives	18
Chapter 2: Literature review: Heteroresistance in <i>Mycobacterium tuberculosis</i> challenging current diagnostic methods	20
Introduction.....	20
Background	23
Clinical relevance of heteroresistance in <i>M. tuberculosis</i>	31
Limitations of different diagnostic methods	32
Outstanding questions.....	33
Conclusion	33
Chapter 3: Materials and Methods	35
Sample selection and DNA extraction	35
PCR optimization.....	41
Primer design.....	41
PCR parameters	42
Electrophoretic evaluation.....	42
Sanger sequencing	43
Targeted deep sequencing of isolates	43
Bioinformatic analysis	44

Ion Torrent data analysis	44
Illumina data analysis	45
FASTQ file format.....	46
Phred quality score.....	46
Quality control	46
Read alignment to the reference genome.....	47
Converting the SAM file to a BAM file.....	47
Variant calling.....	47
SMOR	48
Down sampling of a synthetically created heteroresistant mixture.....	51
Ethics	51
Chapter 4: Results	52
PCR optimisation.....	54
Dilution series - Synthetically made heterogeneous mixes.....	56
Targeted deep sequencing of isolates (Ion Torrent PGM and Illumina MiSeq)	58
Quality assessment of the fastq and alignment files.....	58
Synthetic heterogeneous mixtures	58
Clinical <i>M. tuberculosis</i> isolates of three patients showing evolution of FQ resistance	60
Clinical <i>M. tuberculosis</i> isolates from the OptiQ clinical trial of patients receiving FQ at different concentrations	61
Comparison of variants called by different mappers, i.e. TMAP (Ion Torrent) and Novoalign (Illumina) for:.....	62
Synthetically created heteroresistant mixtures ranging from 10% - 0.1% mutant	62
Comparative analysis between Sanger sequencing and NGS SMOR analyses of targeted deep sequencing isolates.....	63
Synthetically created heteroresistant mixtures ranging from 10% - 0.1% mutant	63
Clinical <i>M. tuberculosis</i> isolates of three patients showing evolution of FQ resistance.....	65
Clinical <i>M. tuberculosis</i> isolates from the OptiQ clinical trial of patients receiving FQ at different concentrations	69
Downsampling.....	71
Chapter 5: Discussion	72
Chapter 6: Conclusion	76
Appendices	77
Appendix A:	77
Appendix B:	82
Appendix C:	83
References	84

List of Figures

Title	Page
Figure 1 Diagnostic sensitivity of current diagnostic assays to detect underlying drug resistance.....	30
Figure 2 Outline of the computational analysis of targeted deep sequencing data.....	45
Figure 3 An example of a read from a fastq file generated by the Illumina sequencing.	46
Figure 4 Flow diagram showing down-sampling of sequenced data (fastq file) to a certain coverage a 100 times.....	51
Figure 5 Flow chart of sample selection and the research methodology.....	53
Figure 6 Visualisation of PCR products of a ten-fold serial dilution	54
Figure 7 Sanger sequencing of synthetic heterogeneous mixture ranging from 100% to 0 % mutant.....	55
Figure 8 Sanger sequencing chromatograms of synthetic heterogeneous mixtures.....	57
Figure 9 Illumina targeted deep sequencing results of the QRDR of <i>gyrA</i> gene from Synthetic heterogeneous mixtures.....	64
Figure 10 A comparison of results of (A) Sanger sequencing with (B and C) Illumina targeted deep sequencing of clinical isolates.....	68

List of Tables

Table	Page
Table 1 A description of the different clinical isolates used in this study.	38
Table 2 Optimal primer sets, for the QRDR region of <i>gyrA</i> and <i>gyrB</i> genes.	41
Table 3 SNP positions in the QRDR of <i>gyrA</i> and <i>gyrB</i> that confer antibiotic resistance.	49
Table 4 Overview of the quantity (read coverage) of raw sequenced data of synthetic heteroresistant mixtures reads generated by the two different platforms, Ion Torrent Personal Genome Machine (PGM) and the Illumina MiSeq platform. Wild type (WT), mutant (MUT).	59
Table 5 Quantity (read coverage) of raw sequenced data of FQ resistant <i>M. tuberculosis</i> clinical isolates prior to bioinformatics analysis.	60
Table 6 Quantity (read coverage) of raw sequenced data of <i>M. tuberculosis</i> clinical isolates from the OptiQ study, prior to bioinformatics analysis.	61
Table 7 Comparison of the variants called by the different mappers (TMAP and Novoalign) for synthetic heteroresistant mixtures, ranging from 10% - 0.1% mutant.	62
Table 8 Targeted deep sequencing results of clinical <i>M. tuberculosis</i> isolates showing acquisition of FQ resistance.	67
Table 9 Targeted deep sequencing results of the clinical <i>M. tuberculosis</i> isolates from the OptiQ study.	70
Table 10 Variants called at various coverages of 0.5% synthetic heterogeneous mix.	71
Table A 1 Primer (Oligonucleotide) sequences of the QRDR of <i>gyrA</i> and - <i>gyrB</i> genes.	77
Table A 2 Different combinations of the forward and reverse primer sequences for the QRDR region of <i>gyrB</i> .	79
Table A 3 Parameters tested to obtain an optimal PCR reaction.	80
Table A 4 PCR parameters of Colman <i>et al.</i> , 2015.	80
Table A 5 Optimised PCR parameters for this study, excluding Betaine solution.	81
Table A 6 PCR thermo-cycling conditions for pure DNA.	81
Table A 7 PCR thermo-cycling conditions for genolysed extracted DNA.	81

List of abbreviations

<u>Abbreviation</u>	<u>Full text</u>
%	Percent
~	Approximately
<	Less than
>	Greater than
≤	Less than or equal to
®	Registered trademark
°C	Degrees Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
µm	Micrometer
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
AMK	Amikacin
BAM	Binary Alignment/Map
bp	Base pairs
cDNA	Complementary DNA
CFZ	Clofazimine
CI	Confidence interval
Conc.	Concentration

CPR	Capreomycin
Cq	Quantification cycle
DH ₂ O	Distilled water
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DRM	Drug resistance mutation
dsDNA	Double-stranded DNA
DST	Drug susceptibility testing
e.g.	Example
EDTA	Ethylenediaminetetraacetate
EMB	Ethambutol
em-PCR	Emulsion PCR
<i>et al.</i>	And others
ETH	Ethionamide
FQ	Fluoroquinolone
FQ-R	Fluoroquinolone resistance
Fwd	Forward
Gb	Gigabases
GS	Genome sequencer
GU	Growth unit
HIV	Human immunodeficiency virus
HIVDR	HIV drug resistance

hrs	Hours
H ₂ O	Water
i.e.	Namely
ID	Identification
INH	Isoniazid
IQR	Interquartile range
kb	Kilobases
KAN	Kanamycin
LJ	Löwenstein-Jensen
M	Molar
Mb	Megabases
MDR-TB	Multidrug-resistant tuberculosis
mg	Milligram
MgCl ₂	Magnesium chloride
min	Minutes
ml	Millilitre
mM	Millimolar
MU	Mutant
MW	Molecular-weight size marker (bp)
N.F.H ₂ O	Nuclease free water
NA	Not available
NC	Negative control
ng	Nanogram

NGS	Next Generation Sequencing
NHLS	National Health Laboratory Service (South Africa)
nM	Nanomolar
nm	Nanometers
NTC	No-template control
OFX	Ofloxacin
PAS	Para-aminosalicylic acid
PCR	Polymerase chain reaction
PGM	Personal genome machine
pH	Potential of hydrogen
PHRED	Phil's read editor
pM	Picomolar
PZA	Pyrazinamide
qPCR	Quantitative PCR
QRDR	Quinolone resistance-determining region
Q-score	Quality score
RC	Reverse complement
Rev	Reverse
RIF	Rifampicin
RNA	Ribonucleic acid
rpm	Reads per run
RT-PCR	Reverse transcription and PCR
SANBI	South African National Bioinformatics Institute
SAM	Sequence Alignment/Map
SBS	Sequencing-by-synthesis

sec	Seconds
SNP	Single-nucleotide polymorphism
SOLiD	Sequencing by oligonucleotide ligation and detection
STR	Streptomycin
SU	Stellenbosch University
TAE	Tris-Acid-EDTA buffer
TB	Tuberculosis
TE	Tris-EDTA buffer
Tm	Melting temperature
™	Trademark
USA	United States of America
UV	Ultraviolet
V	Volts
Vol.	Volume
WHO	World Health Organisation
WT	Wild type
XDR-TB	Extensively drug-resistant TB

Chapter 1: General introduction

Background and motivation of the study

Tuberculosis (TB) control programmes world-wide are faced with the challenge that *Mycobacterium tuberculosis* isolates have evolved to become resistant to the available anti-TB drugs. Of major concern is *M. tuberculosis* that has become resistant to the best available drugs isoniazid and rifampicin, which are defined as multi-drug resistant tuberculosis (MDR-TB). According to the World Health Organization (WHO), MDR-TB is a global problem and an estimated 480 000 people developed MDR-TB in 2015 (WHO, 2015a). Recent MDR-TB prospective cohort studies have demonstrated that approximately 9 to 15% of MDR-TB patients develop resistance to fluoroquinolones (FQs) while on treatment (Cegielski *et al.*, 2014). FQ resistance have been associated with increased mortality (between 50-80%), high risk of progression to XDR-TB (MDR-TB with additional resistance to a second-line injectable as well as a FQ) and the potential spread of preXDR-TB (MDR-TB with additional resistance to either a second-line injectable or a FQ) (Pantel *et al.*, 2012).

FQs are widely used antibiotics with a broad-spectrum antimicrobial activity (Wang *et al.*, 2007). FQs target the enzyme, DNA gyrase, which consist of two alpha and two beta subunits encoded by *gyrA* and *gyrB*, respectively (Miotto P.; Cirillo D.M.; Migliori B., 2015). The function of DNA gyrase is to relieve stress while double-stranded DNA is being unwound by helicase into single stranded DNA which then allows for the initiation of DNA replication by DNA polymerase (Miotto P.; Cirillo D.M.; Migliori B., 2015). The loss of the DNA gyrase function will result in cell death (Miotto P.; Cirillo D.M.; Migliori B., 2015).

Variants in *gyrA* cause the FQ resistance and reduces negative supercoiling which results in large-scale changes to gene expression, altering broad antimicrobial susceptibility as well as fitness and evolutionary adaptability. Amino acid substitutions within a short conserved region within *gyrA* (codon 74 – 113) and *gyrB* (codon 500 – 538) are referred to as the quinolone resistance-determining region (QRDR) (Wang *et al.*, 2007). Mutations within the QRDR regions affect the binding of the antibiotic to the gene protein. Importantly, the majority of known FQ-resistant *M. tuberculosis* strains are associated with mutations in the QRDR of *gyrA* (Al-Mutairi *et al.*, 2011). To date resistance to FQs has been determined by phenotypic culture based methods. However, these methods are slow and require complex infrastructure to ensure biosafety. In 2016 the WHO endorsed the use of the GenoType MTBDRs/ line probe assay which detects mutations in the *gyrA* and *gyrB* gene of *M. tuberculosis*. This method of FQ drug susceptibility testing is envisaged to shorten the turnaround time from 50 days (culture) to <5 days. However the implementation of genetic

drug susceptibility testing has certain limitations which need to be considered when making a diagnosis. Firstly the GenoType MTBDRs/ assay does not detect all of the mutations conferring FQ resistance and secondly the assay is only able to detect subpopulations of FQ resistant *M. tuberculosis* if they are present at >5% (as compared to >1% for culture-based methods) (Folkvardsen, Thomsen, et al., 2013).

We know that FQ resistance in *M. tuberculosis* develops spontaneously through mutation in the *gyrA* and *gyrB* genes (Mayer & Takiff, 2014) and during treatment with an FQ the resistant subpopulation will be selected. The proportion of resistant clones within an individual will be exceptionally low at the point when resistance develops and with time will increase in the presence of FQ to a proportion that is by culture (>1%) and finally until 100% of the population is FQ resistant. Thus the low sensitivity of genetic drug susceptibility tests for detecting underlying populations may be explained by their inability to detect *gyrA* or *gyrB* mutations in some clinical isolates of *M. tuberculosis* which are below 5%. (Chakravorty et al., 2011).

Previous studies have shown the existence of both *gyrA* wildtype and *gyrA* mutant populations in the same *M. tuberculosis* clinical isolate (van Rie et al., 2005; Streicher, Bergval, et al., 2012). The presence of both resistant and susceptible population has been termed heteroresistance. In addition, the above studies also showed that multiple *gyrA* mutations could be present in the same *M. tuberculosis* isolate, and that the frequency of these mutations can vary over time (van Rie et al., 2005; Streicher, Bergval, et al., 2012). The presence of multiple heterogeneous genotypes in an isolate adds to the complexity for the identification of drug resistance as current molecular techniques fail to detect underlying populations below the 5% threshold. Furthermore this study showed that the discordance between genotype and phenotype could be resolved by initially culturing the isolate on FQ containing media and then sequencing the individual colonies. In all instances classical *gyrA* mutations were identified thereby reconfirming the association between *gyrA* mutations and FQ resistance and questioning the role of other gene mutations in FQ resistance.

The sensitivity of currently available genetic drug susceptibility tests to detect underlying resistant subpopulations is greater than 5% for the line probe assays (Engström et al., 2013; Folkvardsen, Thomsen, et al., 2013). Sanger sequencing is of the order of 10% (Folkvardsen, Svensson, et al., 2013), while Xpert MTB/RIF can accurately detect underlying resistance if the proportion of resistant clone is >60% (Engström et al., 2013; Folkvardsen, Thomsen, et al., 2013). Ideally, molecular methods should be able to accurately detect the presence of ≤1% underlying populations in order to identify the emergence of resistant clones during treatment.

We propose to use targeted deep sequencing to identify minority clones harboring FQ resistance conferring mutations in serial patient isolates initiated on MDR-TB treatment. This will allow us to determine whether the resistant clones that emerged on therapy were initially present prior to the start of treatment or were emerged due to antibiotic pressure.

Hypotheses

In this study we hypothesize that targeted deep sequencing can be used to identify underlying drug resistance variants in *M. tuberculosis* clinical isolates.

Aims

To develop an ultrasensitive method to detect resistance-causing variants (mutations) directly from sputum using targeted deep sequencing.

To use targeted deep sequencing to detect underlying mutant variants in synthetically created heterogeneous mixtures.

To use targeted deep sequencing to detect underlying mutant variants in FQ-susceptible clinical isolates.

To use targeted deep sequencing on clinical samples to determine whether underlying resistant clones were initially present, and subsequently selected for under antibiotic pressure or whether drug resistant clones emerge during treatment as a result of a mutation.

Objectives

Therefore, the objectives of this study were to use targeted deep sequencing to:

1. Amplify the quinolone resistance determining region (QRDR) of *gyrA* and *gyrB* genes of *M. tuberculosis*.
2. Create and amplify *in vitro* artificial heterogeneous mixtures (WT:MU) to determine if the targeted deep sequencing method is able to detect approximately known underlying variants (<10%) (e.g. SNP) within the QRDR region of *gyrA* and *gyrB* genes of *M. tuberculosis*.
3. Determine the sensitivity of the method by amplification of diluted artificially created heterogeneous mixtures and Sanger sequencing thereof.
4. Use clinical isolates to determine *M. tuberculosis* population dynamics, i.e. whether the minority population is present and to observe/see whether it becomes dominant and fixed over time (i.e. during treatment period).
5. Optimize a deep sequencing method to detect a 1% minority variant population of *M. tuberculosis* in clinical specimens.

6. Determine whether FQ resistant *M. tuberculosis* clones that emerge in MDR-TB patients while on therapy were initially present (minority variants) and were simply selected for under antibiotic pressure or whether they developed as a result of de novo mutations.

Chapter 2: Literature review: Heteroresistance in *Mycobacterium tuberculosis* challenging current diagnostic methods

Introduction

Tuberculosis (TB) infected 9.6 million people globally and caused 1.5 million deaths in 2014 (World Health Organization, 2016). This highly infectious disease is caused by *Mycobacterium tuberculosis* a bacterium that is able to develop drug resistance to the available antibiotics that constitute the five drug anti-TB regimen. Drug resistance to multiple drugs in the first-line anti-TB drug regimen, including at least rifampicin (RIF) and isoniazid (INH), has been termed multi-drug resistant tuberculosis (MDR-TB) (WHO, 2015b). MDR-TB is a global concern as an estimated 480,000 cases of MDR-TB are detected each year (WHO, 2015b). Furthermore, current second-line anti-TB treatment is toxic, poor treatment outcomes are common and inappropriate or interrupted treatment may promote amplification of resistance (Cegielski *et al.*, 2014).

Successful treatment of MDR-TB cases is dependent on prior knowledge of the infecting strain's drug susceptibility pattern. Guidelines for drug susceptibility testing (DST) as well as critical concentrations for first and second-line drugs have been formulated by the World Health Organization (WHO) (WHO, 2008a). These guidelines recommend the proportion method for the identification of subpopulations of drug-susceptible and drug resistant bacilli (Folkvardsen, Svensson, *et al.*, 2013). Using the current guidelines, a drug resistant subpopulation consisting of $\geq 1\%$ of the total population can be detected when *M. tuberculosis* is exposed to fixed concentrations of antimicrobials (e.g. INH, RIF, ethambutol (EMB), and streptomycin (SM)) (Engström *et al.*, 2013). This method is highly sensitive and remains the gold standard for phenotypic DST, despite, has an average four week turnaround time. Alternatively, molecular diagnostic methods (e.g. Xpert MTB/RIF (Cepheid, Sunnyvale, USA) and Genotype MTBDRplus (Hain Lifescience, Nehren, Germany)) have been endorsed by the WHO to identify genomic variations conferring drug resistance in clinical *M. tuberculosis* isolates (World Health Organization, 2007; Schurch, Kremer, Daviena, *et al.*, 2010; Schurch, Kremer, Kiers, *et al.*, 2010). The Xpert MTB/RIF assay can detect mutations in the rifampicin resistance determining region (RRDR) of *rpoB*, while the Genotype MTBDRplus assay can be used as a diagnostic and confirmatory test for MDR-TB as it can also identify mutations in *katG* and the promoter region of *inhA* promoter (Rahman

et al., 2016). WHO-approved molecular diagnostic methods are notably less sensitive than culture-based methods to detect subpopulations of resistant bacilli (Blakemore *et al.*, 2010). The current molecular methods are able to detect drug resistant subpopulations that represent ≥10% of the total population, however, the detection ability varies greatly between molecular methods (Engström *et al.*, 2013; Folkvardsen, Svensson, *et al.*, 2013; Folkvardsen, Thomsen, *et al.*, 2013).

This variability implies that genetic DSTs may fail to identify subpopulations of resistant bacilli, which may in part explain the observed discordance between phenotype and genotype in a limited proportion of cases (Streicher, Muller, *et al.*, 2012; Folkvardsen, Svensson, *et al.*, 2013; Folkvardsen, Thomsen, *et al.*, 2013) where genetic DST classifies the *M. tuberculosis* strain as susceptible, whilst phenotypic DST classifies the strain as resistant. Failure to identify the resistant subpopulations leads to antibiotic selection of this subpopulation with concomitant treatment failure (van Rie *et al.*, 2005). Another limitation of genetic DST is that it does not detect the full repertoire of mutations conferring resistance and thus specimens harbouring rare mutations conferring resistance will not be detected (Campbell *et al.*, 2011). For this reason genetic DST is a “rule in” test and a secondary test may be required if genetic resistance is not detected (Wedajo *et al.*, 2014). This limits the usefulness of genetic tests for detecting the emergence of resistance during treatment and explains why culture based methods remain the gold standard for DST.

Adding to the complexity of *M. tuberculosis* drug resistance and diagnosis thereof is intra-patient co-existence of clonal subpopulations displaying different drug resistance allele combinations, despite having the same IS6110 fingerprint, otherwise known as heteroresistance (Merker *et al.*, 2013). Possible causes of heteroresistance are (I) in host evolution of a single strain as a result of acquisition of drug resistance during treatment (Sun *et al.*, 2012), or (II) infection with more than one strain (mixed infection) as a result of transmission (van Rie *et al.*, 2005; Wang *et al.*, 2015). Currently, the prevalence of heteroresistance in *M. tuberculosis* is unknown due to a paucity of studies in this area (Folkvardsen, Thomsen, *et al.*, 2013). A possible reason is the lack of a standard definition of heteroresistance, which may result in misidentification of heterogeneous strains as heteroresistance, and may hinder proper assessment of the clinical significance of heteroresistance (El-Halfawy & Valvano, 2015). For example, heteroresistance in *M. tuberculosis* was defined as the presence of both susceptible and resistant strains in different proportions in the same patient (Hofmann-Thiel *et al.*, 2009; Folkvardsen, Thomsen, *et al.*, 2013). Subsequently, it was redefined as the coexistence of populations with different variants (nucleotides) at a drug resistance conferring locus within a sample of organisms (Eilertson *et al.*, 2014). For the purpose of this review, we consider heteroresistance to be

the coexistence of drug-resistant and -susceptible bacteria in the same patient sample. This is in accordance with Hofmann-Thiel *et al.* (2009), Folkvardsen *et al.* (2013), and Thomsen, *et al.* (2013).

In this review, we assess the published literature on heteroresistance in *M. tuberculosis* and compare the sensitivities of current diagnostic methods to detect heteroresistance in *M. tuberculosis*. We also critically review the published literature on next generation sequencing (NGS) technology, specifically targeted deep sequencing, as a potential genetic test to diagnose low-frequency drug resistant subpopulations of $\leq 1\%$. Furthermore, we discuss the clinical relevance of heteroresistance and of NGS technology as a diagnostic tool to detect heteroresistance, as well as provide suggestions for future studies on heteroresistance.

Background

The prevalence of heteroresistance in *M. tuberculosis* is currently unknown. However, studies have identified heteroresistance in *M. tuberculosis* against numerous anti-TB drugs, showing that heteroresistance is not restricted to a particular resistance associated gene (Rinder *et al.*, 2001; Zhang *et al.*, 2012). For example, heteroresistance has been documented in *rpoB* for RIF (Hofmann-Thiel *et al.*, 2009), in *katG* and *oxyR-ahpC* for INH (Zhang, Lu, *et al.*, 2014), in quinolone resistant determining (QRDR) regions of both *gyrB* and *gyrA* genes for fluoroquinolone (FQ) (Streicher, Bergval, *et al.*, 2012; Eilertson *et al.*, 2014).

Furthermore, heteroresistance contributes to *M. tuberculosis* fitness, since, it allows for alteration of strain dominance according to the type of external threats that *M. tuberculosis* strains are exposed to (Streicher, Bergval, *et al.*, 2012). Black *et al.* (2015) observed similar results with the emergence and disappearance of variants during the evolution of drug resistance, as a result of anti-TB drugs exerting a selective pressure on the *M. tuberculosis* strains under study (Black *et al.*, 2015). The fluctuating variants resulted in a purifying effect through the loss of heterogeneous variants (Black *et al.*, 2015). This is in agreement with an earlier study that also observed strain dominance being influenced by treatment and adherence (van Rie *et al.*, 2005). Alternatively, treatment with less-effective second-line drugs led to re-emergence of a drug-susceptible strain in a patient with mixed infection, demonstrating that strain dominance in sputum culture is determined by the selection pressure of the antibiotic therapy (van Rie *et al.*, 2005). This in turn allows the existence, persistence and spread of TB disease with various degrees of drug resistance, which affect diagnosis (Meacci *et al.*, 2005).

Diagnosis of heteroresistance in *M. tuberculosis* depends on various factors. Primarily, the source of isolates, since different sources (i.e. original clinical sputum specimen (genolyse samples) vs cultured isolate) may result in different ratios of the wild type and mutant strains within the *M. tuberculosis* population (Rinder *et al.*, 2001; Mekonnen *et al.*, 2015). In particular, culturing in the presence or absence of a drug, may alter the ratio of drug sensitive and drug resistant populations, depending on the fitness cost of the drug resistance conferring mutation (van Rie *et al.*, 2005; Zetola *et al.*, 2014). Furthermore, decontamination methods also influence the detection of the true presence of heteroresistance within a sample as a large proportion of mycobacteria is inadvertently killed by the process (Chatterjee *et al.*, 2013). For example, 4% NaOH, a strong decontaminant, has been observed to kill 60% of tubercle bacilli present in the specimen (Chatterjee *et al.*, 2013). As a result, the true composition of resistant and susceptible populations present within the decontaminated specimens may be altered and in some cases result in heteroresistance

being overlooked (Chatterjee *et al.*, 2013). Furthermore, different growth media result in variation in growth rates, for example faster growth is observed in liquid media compared to Löwenstein-Jensen solid medium and solid Middlebrook 7H11 agar, which would in turn affect phenotypic resistance tests and molecular assays (Rinder *et al.*, 2001). It has also been shown that some drug resistance causing mutations are associated with a fitness cost in the absence of the antibiotic (van Rie *et al.*, 2005) and therefore it is possible that the fitter strain (i.e. susceptible or resistant strain, depending on whether the antibiotic is present or not) will outgrow the weaker strain and as a result alter the true ratio of wild type and mutant strains. The above mentioned factors question the reliability of culture as an accurate reflection of the composition of the true variants in a clinical sample (El-Halfawy & Valvano, 2015). Lastly, the sensitivity of the diagnostic method used will determine whether the low-frequency resistance-causing variants in *M. tuberculosis* will be detected or not, see (**Figure 1**) (Mekonnen *et al.*, 2015).

Current diagnostic methods apply conventional culture-based phenotypic DST on solid or liquid culture media, which is highly sensitive for detecting drug-resistant TB (Folkvardsen, Thomsen, *et al.*, 2013; Hughes *et al.*, 2014). This gold standard phenotypic test is able to classify an *M. tuberculosis* population as resistant or susceptible and to detect underlying resistant bacilli if at least 1% of the bacterial population is resistant (Engström *et al.*, 2013; Zhang, Lu, *et al.*, 2014). Furthermore, studies have found that solid culture (i.e. the broth dilution method) has a greater sensitivity in detecting heteroresistance (e.g. INH and RIF) compared to liquid culture (i.e. Mycobacteria Growth Indicator Tube (MGIT) DST) (Zhang, Lu, *et al.*, 2014; Zhang, Wang, *et al.*, 2014). In particular, the broth dilution method is able to detect INH resistance if 0.5% of resistant bacteria are present, typically harbouring mutations in *katG* and/or *oxyR-ahpC* genes (according to the study's observation), compared to the MGIT system that is able to detect 1% INH resistant *M. tuberculosis* bacteria (Zhang, Lu, *et al.*, 2014). Furthermore, the broth dilution method is able to detect INH heteroresistance in *M. tuberculosis* isolates with mutations in the promoter region of the *inhA* gene as opposed to liquid culture that cannot, owing to low-level INH resistance (Zhang, Lu, *et al.*, 2014). Conversely, the detection ability of MGIT DST for RIF resistance depends on the type of *M. tuberculosis* lineage, for example, MGIT DST that was done in the same laboratory was able to detect 1% RIF resistance for Haarlem strains, as opposed to identifying 5% RIF resistance for the Beijing strains (Folkvardsen, Thomsen, *et al.*, 2013). It was however speculated that the different *rpoB* mutations found in the particular Beijing and Haarlem strains are associated with different fitness costs which may affect the ability of DST methods to detect heteroresistance. Notably, there is a lack of information on the ability to

detect heteroresistance for the other anti-TB antibiotics when using either solid or liquid media DST methods.

Alternatively, heteroresistance can be detected from sputum or cultured samples by making use of traditional polymerase chain reaction (PCR) and gel electrophoresis, followed by restriction fragment length polymorphism analysis of polymerase chain reaction products (PCR-RFLP) (Rinder *et al.*, 2001). The extensive use of PCR and gel electrophoresis as diagnostic techniques in the laboratory has led to advances in PCR based detection technology to improve its diagnostic ability. For example real-time PCR and digital PCR has enabled quantification and improved detection of heteroresistance in *M. tuberculosis* to a level of 10% and 0.1%, respectively, indicating that PCR-based technologies also vary greatly in their sensitivity to detect heteroresistance (Pholwat *et al.*, 2013). Furthermore, the addition of sloppy molecular beacons (SMBs) to real-time PCR methods allows for rapid detection of FQ resistance and emergence of as little as 5% FQ heteroresistance in clinical samples from TB patients (Chakravorty *et al.*, 2011). However, concern over consistency of different PCR runs with the same clinical genotypes has been raised, following the outcome of a study that showed different genotypes from reruns of the same sample (Rinder *et al.*, 2001). The authors suggest that the inconsistency could be due to stochastic events occurring during the initial PCR cycles (Rinder *et al.*, 2001). A possible solution to this problem is to repeat the PCR step a number of times, however this is very time and resource consuming.

An alternative applied diagnostic, namely Sanger sequencing, is an automated method of DNA sequencing built upon the chemistry of PCR. The method results in the termination of the extending DNA fragment, by the addition of fluorescently labelled dideoxy chain-terminating nucleotides by DNA polymerase during *in vitro* DNA replication. Sanger sequencing is used for analysing sequence variants as well as for detecting resistance-causing mutations (Sanger & Coulson 1975, Sanger, Nicklen & Coulson 1992). This genotypic test is also able to detect low frequency variants (heteroresistance) at a sub-population of $\geq 10\%$ (Folkvardsen, Svensson, *et al.*, 2013). A different DNA sequencing method, namely Pyrosequencing, relies on the detection of pyrophosphate release, detected as emitted light, on nucleotide incorporation (Engström *et al.*, 2013). This method is not as sensitive as Sanger sequencing, but has the ability to detect 35-50% low-frequency resistance-associated mutations in *M. tuberculosis* (Engström *et al.*, 2013).

In 2008, the WHO endorsed the line-probe assay (LPA) GenoType MTBDRplus VER1.0 (Hain Lifescience, Nehren, Germany) for rapid diagnosis of MDR-TB (WHO, 2008b; Hughes *et al.*, 2014). This PCR-based genotypic test amplifies genes where mutations associated

with resistance are common (for example mutations in the *rpoB* gene that result in RIF resistance) and then hybridizes the target sequences to a membrane strip (Folkvardsen, Svensson, et al., 2013). Resistance is then represented by the presence of specific mutation bands or the absence of wild type bands on the visualised strip (Folkvardsen, Svensson, et al., 2013). Drug resistance to the first line drugs RIF and INH can be detected directly from smear-positive sputum specimens with the GenoType MTBDRplus (Hughes et al., 2014). Furthermore, the GenoType MTBDRplus is also able to detect heteroresistance when the relative proportion of resistant *M. tuberculosis* organisms is $\geq 5\%$ and harbour the specified known resistance causing mutations (Hofmann-Thiel et al., 2009; Folkvardsen, Svensson, et al., 2013; Folkvardsen, Thomsen, et al., 2013). Folkvardsen et al. (2013) has also showed that the GenoType MTBDRplus is superior to INNO-LiPA in detecting heteroresistance (Folkvardsen, Thomsen, et al., 2013).

Subsequently, in 2010, the WHO endorsed the Xpert MTB/RIF (Cepheid, Sunnyvale, USA) for TB diagnosis (WHO, 2011), and it has subsequently become the routine first-line diagnostic test in South Africa (Hughes et al., 2014). This assay is highly regarded for its high specificity and fast turnaround time (Mayer & Takiff, 2014). However, the Xpert MTB/RIF assay is only able to detect RIF resistance-conferring mutations only if the *rpoB* mutation is present in $\geq 65\%$ of the bacterial population. A possible reason is the principle of the Xpert test that only detects resistance when the signal of the molecular beacon probes is either lost or reduced, attributable to mutations in the *rpoB* gene that inhibit hybridization of one or more molecular beacons specific to *rpoB* gene. An alternative reason is the particular *rpoB* mutation that could influence the level at which heterogeneity can be detected, since common RIF resistance causing mutations result in the complete inhibition of probe hybridization whereas uncommon *rpoB* mutations only results in partial probe hybridization (Blakemore et al., 2010). Studies have also observed high false-positive RIF resistance results within strains with a wild type *rpoB* gene sequence (Boehme et al., 2011; Marlowe et al., 2011; Theron et al., 2011; van Rie et al., 2013).

Recently, an increasing number of studies have reported using highly improved genomic diagnostic techniques such as whole genome sequencing (WGS) and targeted deep sequencing to identify resistance causing variants, establish relapse or re-infection, reveal genomic heterogeneity within *M. tuberculosis*, to name a few (Bryant et al., 2013; Black et al., 2015; Colman et al., 2016). These techniques provide a detailed view of the genomic structure of the pathogen, including alterations (for example single nucleotide polymorphisms (SNPs), short insertions and deletions (in/dels) and regions of difference (RDs)) in the genome which may assist with TB drug resistance diagnostics. Alternatively, these techniques can be used for epidemiological studies, phylogenetic studies, population

studies and research in drug development of anti-TB drugs (Gardy *et al.*, 2011; Ioerger *et al.*, 2013; Walker *et al.*, 2013).

Recent studies have investigated the ability of WGS to identify genetic heterogeneity (Black *et al.* (2015) and showed that a read frequency cut-off of 30% is adequate to identify low frequency sequencing variants in *M. tuberculosis* with high confidence, at an ever decreasing cost and increasing throughput (Black *et al.*, 2015). However, WGS currently cannot be used as a routine clinical diagnostic tool due to cost and computational intensiveness, but despite these limitations, it remains a useful discovery tool in academic settings.

Advancement in NGS technology and computational capacity, more specifically next generation targeted deep sequencing, has made it possible to detect drug resistance causing genetic variants present at a low frequency ($\geq 0.1\%$) prior to or during treatment of *M. tuberculosis* infected patients (Colman *et al.*, 2015). The extremely high sensitivity, allows targeted deep sequencing to out-perform all of the previous genomic diagnostic assays, making it the most sensitive and specific genomic assay presently available.

Targeted deep sequencing refers to sequencing of a defined genomic region multiple times, where sequence depth (depth of coverage) refers to the number of times a given region is covered by individual reads (“Illumina”, 2016). Targeted deep sequencing is a powerful tool to investigate evolving bacterial as well as viral populations, provided that variants are accurately identified during data analysis. Targeted deep sequencing can also be useful for drug resistance diagnostics, as it can be applied to detect an extremely low-level ($\geq 0.1\%$) of resistance conferring genetic variants at a minimum depth of 10 000x coverage (Colman *et al.*, 2015). The latter study showed that the limit of detection of heterogeneity using targeted deep sequencing is inversely proportional to the total depth of coverage (Colman *et al.*, 2015). However, the clinical relevance of $\geq 0.1\%$ subpopulation detection is yet to be confirmed (Colman *et al.*, 2015).

In the study by Colman *et al.* (2015), rare *M. tuberculosis* drug resistant subpopulations were detected using targeted deep sequencing (with an Illumina sequencing platform) together with single molecule-overlapping reads (SMOR) analysis. Initially, the method (SMOR) was validated using standardized mixtures of DNA from susceptible and resistant *M. tuberculosis* isolates followed with DNA directly extracted from clinical sputum samples (Colman *et al.*, 2015). The SMOR analysis allows for tracking and quantification of low frequency subpopulations by using sequence information from overlapping paired-end reads (forward and reverse). A minimum of 2000x depth of coverage (1000x paired reads) is required for the detection of a 1% subpopulation, which is equal to 10 pairs of reads (Colman *et al.*,

2015). In the study, the same DNA molecule was used for the quantification and detection of drug resistance conferring SNPs in six different target gene regions (*rpoB*, *inh* promoter, *katG*, *gyrA*, *rrs* and *eis* promoter) in artificially mixed subpopulations (Colman *et al.*, 2015). These mixed subpopulations consisted of an extensively drug resistant- (XDR) (ranging from 0.025% to 70% of the mixture) and a pan-susceptible *M. tuberculosis* strain (Colman *et al.*, 2015). The results showed that targeted deep sequencing with SMOR analysis had a greater sensitivity than standard molecular approaches (Sanger sequencing, NGS analysis and other WHO endorsed molecular methods) (Colman *et al.*, 2015). In addition, SMOR analysis allows for *in silico* examination of the genetic changes occurring over time in mixed subpopulations of *M. tuberculosis*, which allows rapid characterization and evaluation for clinical significance (Colman *et al.*, 2015). However, this study used the SMOR analysis by only calling variants at specific positions in the target gene regions. This approach may prevent the discovery of novel drug resistance causing mutations and mask the presence of low frequency (0.1-0.4%) false positive variants across the entire targeted area as the error rate of Illumina is around 0.4% (Quail *et al.*, 2012).

Other NGS platforms have also been utilized for targeted deep sequencing, namely the Ion Torrent Personal Genome Machine (PGM) and Ion Proton, the Pacific Bioscience RS and Illumina MiSeq sequencers, however each has its own strengths and limitations, which have been discussed elsewhere in a comparative study of NGS platforms (Quail *et al.*, 2012). The latter study included both Prokaryotic (haploid) and eukaryotic (diploid) organisms which represented a range of genotypic features such as GC-rich regions, AT-rich regions and repetitive regions (Quail *et al.*, 2012). The maximum sequencing yield (referring to the number of reads per run) of each platform ranges from $3.5\text{--}7.5 \times 10^4$ for the PacBio RS II: P6-C4 platform, to 8.2×10^7 for Ion Torrent (PGM), and 8×10^9 (paired) for Illumina HiSeq 2500 (Rhoads & Au, 2015). The sequences generated from these platforms differ in quality which can be used to inform the analysis of the data for certain diagnostic applications. Therefore it is important to consider the following before implementing targeted deep sequencing: (I) the type of platform, (II) research question (hypothesis), (III) available resources, (IV) sequencing and data analysis expertise, (V) finances, (VI) infrastructure, (VII) repeat elements, (VIII) GC-rich -, and (IX) AT-rich content of the organism under study (Quail *et al.*, 2012). The type of sequencing platform may determine the diagnostic applications and the limit of detection.

Furthermore, the research question determines the type of platform required for sequencing, software needed to analyse the data and the minimum depth of coverage required, which is also influenced by the available expertise, as currently available data analysis software is

largely specialised. The GC content of the organism can also influence the outcome and performance of the various NGS platforms (Roetzer *et al.*, 2013). Furthermore, the depth of coverage, read length, and error rate of NGS platforms also contribute to the confidence, accuracy and validity of the variants called by a particular variant caller (analysis tool kit).

The major advantage of targeted deep sequencing, as seen above, is that it has the potential to detect low frequency genetic variants in heterogeneous populations. Other advantages of targeted deep sequencing are fast turnaround (large numbers of samples in a single run and target multiplexing), the ability to sequence low levels of degraded DNA, and the ability to provide realistic reconstruction of individual bacterial haplotypes because of an increase in read length (McElroy *et al.*, 2014).

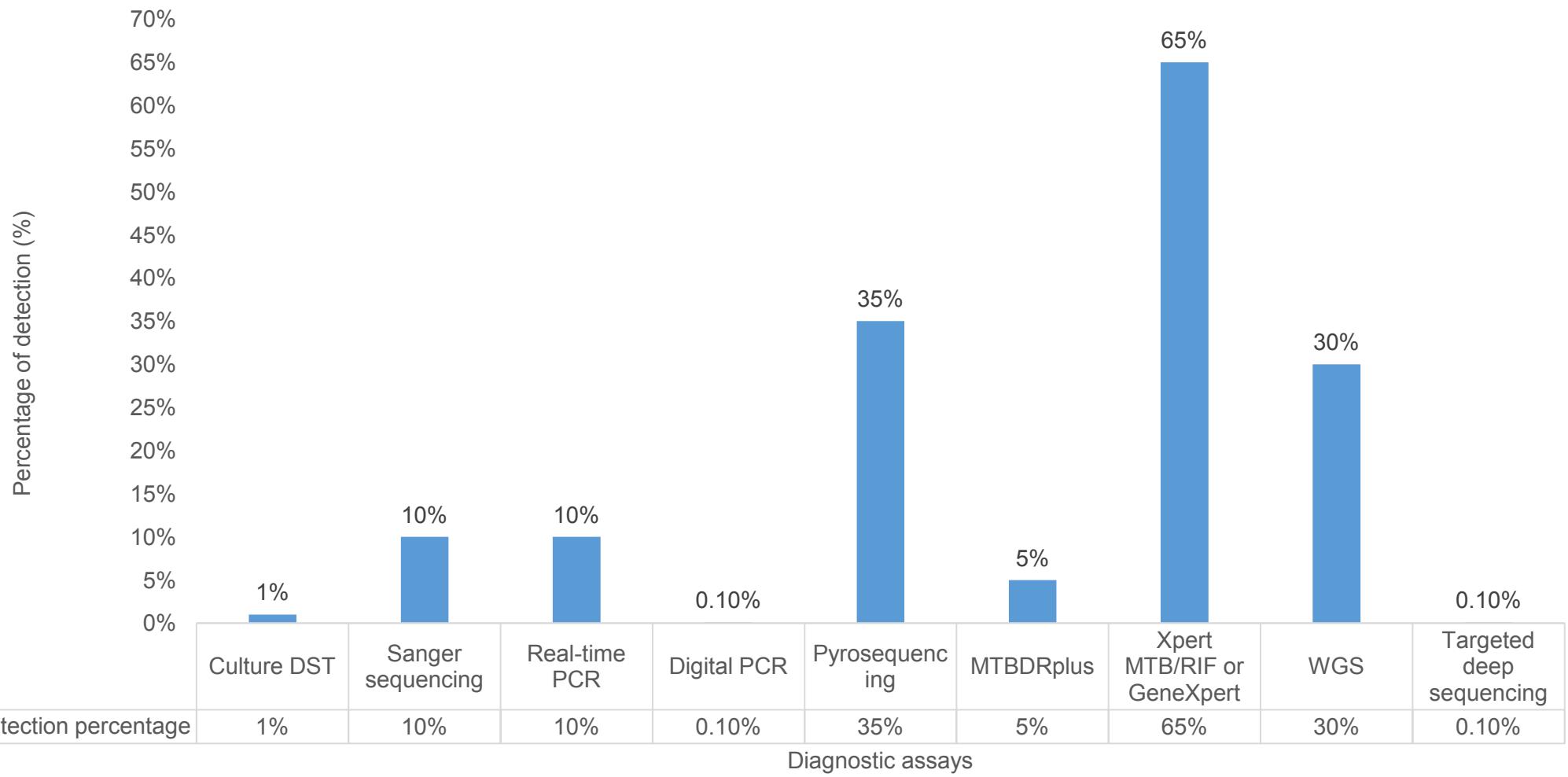


Figure 1 Diagnostic sensitivity of current diagnostic assays to detect underlying drug resistance. DST=drug susceptibility test. PCR=polymerase chain reaction. WGS=whole genome sequencing.

Clinical relevance of heteroresistance in *M. tuberculosis*

The current literature reflects ambiguity surrounding heteroresistance, possibly driven by the lack of a standard definition. Furthermore, an editorial by Behr (2004) postulated that, in a mixed population, the resistant strain would be selected through antibiotic treatment, ultimately leading to dominance of the resistant strain and poor treatment outcome (Behr, 2004). This was confirmed in several subsequent studies where patients were dually infected prior to treatment, but the proportion of the resistant strain was initially too low to be detected by traditional molecular methods. However, after receiving treatment, an increase of the resistant strain was observed due to antibiotic selection (van Rie *et al.*, 2005; Streicher, Bergval, *et al.*, 2012; Eldholm *et al.*, 2014; Mayer & Takiff, 2014). These studies also showed that following the withdrawal of antibiotic pressure, the drug susceptible population reappeared, suggesting that the drug susceptible population was more fit than the drug resistant population (van Rie *et al.*, 2005; Streicher, Bergval, *et al.*, 2012; Eldholm *et al.*, 2014; Mayer & Takiff, 2014). Van Rie *et al.* (2005) used highly sensitive PCR based methods to describe the *M. tuberculosis* strain population structure in serial clinical isolates to determine the mechanisms whereby drug resistance emerges in a patient over time (van Rie *et al.*, 2005). Three mechanisms were identified, namely acquisition of resistance after prolonged therapy, reinfection and mixed infection (van Rie *et al.*, 2005), which was also confirmed by other studies (Hofmann-Thiel *et al.*, 2009; Sun *et al.*, 2012). All of these scenarios may have resulted in heteroresistance, posing a diagnostic challenge if the minority population was drug resistant.

The low sensitivity of current diagnostic methods to accurately predict the presence of resistant subpopulations may lead to diagnostic confusion, especially when serial isolates show different resistance patterns when the susceptible population re-emerges after withdrawal of the therapy, which could subsequently impact therapy and therapeutic response. Information regarding the presence of heteroresistant populations may be vital when considering the pharmacogenetic implications of the different drug resistance causing mutations as well as transmission prevention strategies. Alternatively, diagnostic delay, e.g. a 4 to 8 weeks turn-around time for phenotypic DST, causes a clinical dilemma of which regimen to start the patient on, limiting future treatment options if patients are incorrectly treated. For instance if a patient starts treatment with first-line drugs while awaiting diagnostic outcome, ineffective treatment (due to the inability to detect underlying resistance to either INH or RIF) could pose a risk of amplification of resistance to pyrazinamide and ethambutol (van Rie *et al.*, 2013). Under these conditions the patient may still be infectious, which could subsequently lead to transmission of the resistant or susceptible TB strain and/or poor treatment outcomes (van Rie *et al.*, 2005, 2013). On the other hand, not starting treatment with any drug poses infection control issues and increased risk of death (van Rie *et al.*, 2005, 2013). It is for these reasons important to note that incorrect TB treatment of

undetected heteroresistance can lead to limited treatment options and propagation of undetected resistant or susceptible strains.

A foreseeable solution will be the implementation of higher-resolution sequencing methods that can be applied directly to sputum samples to overcome some of the above mentioned confounders that play a role in heteroresistance detection. In addition, high resolution sequencing diagnostic tests can provide insight into within host complexity of infection and the dynamic response of the organisms under the selective pressure of treatment (Cohen *et al.*, 2016).

Limitations of different diagnostic methods

Diagnosing heteroresistance is challenging, with many factors playing a role in detection of heteroresistance. The primary reason is failure to harvest all existing variants within a patient's sputum sample. This may occur through loss of rare variants during sample transport or decontamination (Rinder, 2001), during the DNA extraction procedure. Alternatively, the ratio of heteroresistant variants may be altered during culturing when subpopulations may outcompete each other (Cohen *et al.*, 2016) or may be lost during culturing in the presence of an antimicrobial agent (Streicher, Bergval, *et al.*, 2012; Zhang, Lu, *et al.*, 2014; Colman *et al.*, 2015). Additional challenges that influence the detection of heteroresistance in *M. tuberculosis* are the different mutations conferring resistance to a single anti-TB drug, necessitating sequencing of more than one genomic region (Colman *et al.*, 2015). Furthermore, the detection of heteroresistance also depends on the sensitivity of the molecular technique used in a study, since the current molecular techniques have variable detection limits for detection of low frequency underlying genomic mutations conferring drug resistance. Alternatively, culture-based phenotypic DST is a more accurate and sensitive diagnostic technique, especially when the mechanism of action is unknown and is currently considered as the gold standard for DST (WHO, 2014). However, The limitations of culture-based phenotypic DST are that it is very time consuming (4 to 6 weeks from receipt of clinical samples), and requires a very good infrastructure and trained personnel, which is costly (Tolani *et al.*, 2012).

NGS-based diagnostic methods are also costly and require highly advanced technology and expertise. Furthermore, unknown resistance-conferring mutations may be overlooked, targeted deep sequencing platforms possess intrinsic sequencing error rates, possibly obscuring true variants, and stochastic PCR effects will alter resistant allele frequencies (Rinder, 2001; Colman *et al.*, 2015). The latter, refers to the initial DNA amplification step, a confounder that affects all culture-independent molecular resistance prediction tests, which all consequent steps rely on (Rinder, 2001).

In addition, the available methods to detect heteroresistance are non-standardized methods that are costly, vary in their detection abilities and are resource dependant and labour intensive,

indicating that these factors are preventing the evaluation of the severity and clinical magnitude of this phenomenon, and prevents implementation of proper guidelines and therapeutic interventions (El-Halfawy & Valvano, 2015). Furthermore, the clinical relevance of variants in less than 1% of the population still needs to be determine.

Outstanding questions

Significant progress has been made in TB research, more specifically discovering heteroresistance, determining some of the causes of heteroresistance, and establishing which methods are appropriate to diagnose heteroresistance. However, more research is required as many more questions have been raised, some of which can form the basis of future studies. For example, to determine which of the following occur more frequently, (I) acquisition or (II) infection with mixed strains. Alternatively, future studies can show the clinical relevance of heteroresistance in a carefully designed experiment or determine at what point the drug regimen needs to change. Furthermore, in a comparative study of diagnostic techniques, a particular diagnostic technique could be highlighted that is sensitive and accurate enough to detect heteroresistance in all of the *M. tuberculosis* lineages. Alternative studies could indicate how the decontamination step carried out before starting the culture affects the heteroresistant populations and show the percentage of organisms that gets lost during the decontamination step. It will also be very interesting to prove that underlying variants (e.g. 0.1%) can result in clinical drug resistance. Novel methodologies are needed to allow targeted deep sequencing to be done directly from sputum specimens. Currently the effect of the decontamination step carried out before culturing of hetero-resistant populations is unknown (Rinder, 2001). Finding answers and solutions to the pressing challenge of heteroresistance might significantly impact TB research and assist in prevention and management of the disease.

Conclusion

The clinical outcome of MDR-TB cases remains poor, with an estimated 50% treatment success (completion or cure) globally (WHO, 2015a). A contributing factor is heteroresistance, which exacerbates poor clinical outcomes (El-Halfawy & Valvano, 2015). In addition, current genetic DSTs are only able to detect underlying resistance if the proportion of resistant clones (strains) are $\geq 5\%$ (Engström *et al.*, 2013; Folkvardsen, Thomsen, *et al.*, 2013) resulting in an underestimation of the true prevalence of heteroresistance. The low sensitivity of these and other molecular methods compared to culture, cause it to be used as “rule in” assays and should be supplemented with phenotypic testing to identify the emergence of resistant clones during treatment.

Ideally, molecular methods should be able to accurately detect the presence of $\leq 1\%$ underlying populations prior to advising treatment or in order to identify the emergence of resistant clones

during treatment. However, as discussed in this review, the current diagnostic methods have variable *M. tuberculosis* detection abilities and are limited to detect a certain percentage of resistant *M. tuberculosis* strains. Targeted NGS is currently the most advanced and powerful tool to detect underlying populations (Faino & Thomma, 2014). Targeted deep sequencing has been used extensively in the field of human immunodeficiency virus (HIV) and has the potential to become the next gold standard molecular diagnostic tool for the detection of underlying populations (Li & Kuritzkes, 2013) as its detection limit is similar to that achieved by culture based methods. However, currently diagnostic methods, based on NGS technology are being streamlined to shorten the turnaround time in order to obtain results instantly.

The prospect of implementing targeted deep sequencing as a routine clinical diagnostic could potentially provide early detection of antibiotic resistant subpopulations (heteroresistance). Targeted deep sequencing could thus be used to detect resistance-associated variants at low frequencies prior to implementation of treatment or during treatment, allowing the prescription of tailored treatment regimens. In turn, this could potentially prevent the selection of resistant populations, improve treatment outcome and prevent the consequent transmission of resistant strains.

Chapter 3: Materials and Methods

Sample selection and DNA extraction

Three different types of *Mycobacterium tuberculosis* DNA samples were used in this study, namely (I) purified DNA extracted from a *gyrA*-wild type (H37Rv) and -mutant (BO1) isolate (which was also used to make synthetic heterogeneous mixtures), (II) crude DNA extracted from heat inactivated cultured *M. tuberculosis* isolates, and (III) crude DNA extracted from decontaminated sputum specimens using the GenoLyse kit (Hain Lifescience, Germany) as well as crude DNA extracted from heat inactivated MGIT cultured *M. tuberculosis* isolates.

(I) The purified DNA used in this study was extracted from cultures of two strains housed at the Department of Biomedical Sciences, Stellenbosch University, namely the *M. tuberculosis* H37Rv laboratory isolate and an *in vitro* selected ofloxacin-mono resistant *M. tuberculosis* isolate (BO1). BO1 harbours the following polymorphisms in *gyrA*: GAG21CAG, GAC94GGC, AGC95ACC, GGC668GAC, and no polymorphisms in *gyrB*. The *gyrA* GAC94GGC mutation is known to confer resistance to fluoroquinolones (Chakravorty *et al.*, 2011). These isolates were previously characterised with standard DST and whole genome sequencing (McGrath *et al.* 2016). DNA was extracted using the phenol chloroform method as previously described (Warren *et al.*, 2006). The *M. tuberculosis* cultures were heat inactivated by incubating the culture plates in a pre-heated oven at 80°C for 2 hours. The bacteria were then harvested by scraping the bacteria off from the media using a sterile disposable 10 µl plastic loop and placing the bacteria directly into a 50 ml polypropylene centrifuge tube containing 20 x 5 mm glass beads and 6 ml extraction buffer (5 % sodium glutamate, 50 mM Tris-HCl (pH7.4) and 25mM EDTA). The bacterial colonies were dispersed by vigorous vortex mixing for 2 min. Five hundred microliter (500 µl) of lysozyme (25mg, Roche, Germany) at a concentration of 100 mg/ml was then added to each bacterial suspension and incubated at 37°C for 2 hours to degrade the cell wall and to digest RNA. Subsequently, 600 µl of 10X Proteinase K buffer (5% sodium dodecyl sulphate, 100nM Tris-HCl (pH 7.8), 50mM EDTA) together with 1.5 mg proteinase K (150 µl of a 10mg/ml stock solution) was added and the suspension was incubated at 45°C for 16 hours to digest all bacterial proteins. Thereafter, 5 ml phenol/chloroform/isoamylalcohol (24:23:1) was added and gently inverted every 30 minutes (min) for 2 hours at room temperature. Each tube was then centrifuged at 3000 revolutions per min (rpm) for 20 min at room temperature to ensure complete phase separation. The aqueous phase was carefully transferred to a new 50 ml polypropylene centrifuge tube which contained 5 ml of chloroform/isoamyl alcohol (24:1), followed by gentle mixing by inversion and centrifugation at 3000 rpm for 20 min at room temperature. The aqueous phase was carefully aspirated and transferred to a new 50 ml polypropylene centrifuge tube containing 600 µl 3M sodium-acetate pH 5.2. To precipitate the DNA, 7 ml of ice-cold isopropanol

was added and the tube was gently inverted. The precipitated DNA were collected immediately on a thin glass rod and incubated for 10 min in a 1.5 ml tube containing 1 ml 70 % ethanol. The glass rod with precipitated DNA was then transferred to a new 1.5 ml tube and allowed to dry at room temperature. Subsequently, the DNA was rehydrated by adding 300 µl Tris-EDTA (TE) buffer (10mM Tris-HCl (pH 8.0), 1mM EDTA), followed by gentle mixing. The DNA was incubated at 65°C for 2 hours to allow for the DNA to re-dissolve. The DNA concentration was determined using the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

The synthetic heterogeneous mixtures (with the ratios ranging from 100% mutant to 0% mutant) (**Table 1**) were prepared by adding a constant volume of purified BO1 mutant DNA (2.5 µl) to different volumes of purified H37Rv wild type DNA. The purified mutant DNA concentration was used to determine the volume of purified wild type DNA needed to make synthetic heterogeneous mixtures at an expected percentage (%), e.g. a 10 % mutant DNA sample (i.e. 90:10 ratio). A nanodrop reading was taken of each mixture and the mixture was diluted to a final DNA concentration of 25 ng/µl. Subsequently a 10-fold serial dilution of the heterogeneous mixtures was made using TE buffer (ranging from 25 ng/µl to 25×10^{-6} ng/µl DNA concentration). A DNA concentration of 25×10^{-6} ng/µl is approximately equivalent to 5 copies of DNA template (see **Appendix B**). This was done to test the sensitivity of the PCR amplification method used in this study.

(II) A total of eleven serial isolates from three patients (see **Table 1**) showing the acquisition of fluoroquinolone (FQ) resistance were selected from a longitudinal *M. tuberculosis* strain bank housed at the Department of Biomedical Sciences (Study: “Fluoroquinolone resistance in patients with multidrug-resistant tuberculosis study” (N14/08/106)). Crude DNA was extracted from the selected isolates by incubating an aliquot (200 µl) of the cultured isolate at 100 °C for 30min. The boiled isolates were then centrifuged and the supernatant was aspirated into a clean eppendorf tube.

(III) A total of eleven serial culture confirmed isolates (base line and follow up) from five patients collected during the OptiQ Phase 2 clinical trial (Efficacy and Safety of Levofloxacin for the Treatment of Multi-Drug Resistant Tuberculosis, Opti-Q; NIAID U01AI100805-02) were selected. Patients that participated in this trial received various dosages of levofloxacin in combination with the standard MDR regimen. This collection included baseline FQ susceptible MDR-TB isolates as well as follow up isolates collected between week 8 and 16 (see **Table 1**). Crude DNA was extracted from the baseline sputum specimens using the GenoLyse kit (Hain Lifescience, Germany) according to manufacturer’s instructions. In addition, crude DNA was extracted from the 8 to 16 week follow-up MGIT cultured isolates using heat inactivation as described above.

The latter method was used to determine whether the resistant clones that emerge on therapy were initially present and was simply selected for under antibiotic pressure or whether they emerged as the result of a mutation while on treatment.

A clinical isolate R2658 was used as a positive control to validate the NGS data analysis pipeline. Isolate R2658 was chosen because it has been characterized by targeted deep sequencing (on the Illumina MiSeq 2x300bp platform) in a previous study. Therefore the results generated in this study can be compared to what was found in the previous study. Targeted deep sequencing of the R2658 clinical isolate (work that was done previously) revealed that it had multi-drug resistance conferring mutations in *katG*315 (AGC to GGT) and *rpoB*531 (TCG to TTG), as well as FQ resistant conferring mutations in *gyrA*94 (GAC to GGC) at a variant frequency of 44.5% for GGC (the minor component).

Table 1 A description of the different clinical isolates used in this study.

Study	Origin	Description	Date	Sanger Sequencing	Deep Sequencing (Platform)
Synthetic H37Rv & BO1 (mutant) heteroresistant mixtures	BO1 (mutant)	Ratio WT:MU (0:100)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (10:90)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (20:80)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (30:70)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (40:60)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (50:50)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (60:40)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (70:30)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (80:20)	NA	Yes	Not sequenced
	H37Rv & BO1 (mutant)	Ratio WT:MU (90:10)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (92:8)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (94:6)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (96:4)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (99:1)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.2:0.8)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.4:0.6)	NA	Yes	Ion Torrent PGM & Illumina MiSeq

	H37Rv & BO1 (mutant)	Ratio WT:MU (99.5:0.5)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.8:0.2)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.9:0.1)	NA	Yes	Ion Torrent PGM & Illumina MiSeq
	H37Rv	Ratio WT:MU (100:0)	NA	Yes	No
Evolution of FQ resistance	Patient 1	1st clinical isolate	05/01/2010	Yes	Illumina MiSeq
		2nd clinical isolate	23/03/2010	Yes	Illumina MiSeq
	Patient 2	1st clinical isolate	28/08/2009	Yes	Illumina MiSeq
		2nd clinical isolate	02/10/2009	Yes	Illumina MiSeq
		3rd clinical isolate	27/10/2009	Yes	Illumina MiSeq
		4th clinical isolate	27/10/2009	Yes	Illumina MiSeq
	Patient 3	1st clinical isolate	29/07/2010	Yes	Illumina MiSeq
		2nd clinical isolate	23/09/2010	Yes	Illumina MiSeq
		3rd clinical isolate	25/10/2010	Yes	Illumina MiSeq
OptiQ study	Patient 1	Baseline clinical isolate	27/01/2016	Yes	Illumina MiSeq
		Week 8 clinical isolate	23/04/2016	Yes	Illumina MiSeq
	Patient 2	Baseline clinical isolate	01/06/2015	Yes	Illumina MiSeq
		Week 8 clinical isolate	22/08/2015	Yes	Illumina MiSeq
		Week 10 clinical isolate	06/09/2015	Yes	Illumina MiSeq
	Patient 3	Baseline clinical isolate	08/03/2015	Yes	Illumina MiSeq
		Week 16 clinical isolate	01/07/2015	Yes	Illumina MiSeq

	Patient 4	Baseline clinical isolate	25/05/2016	Yes	Illumina MiSeq
		Week 8 clinical isolate	13/08/2016	Yes	Illumina MiSeq
	Patient 5	Baseline clinical isolate	24/02/2015	Yes	Illumina MiSeq
		Week 8 clinical isolate	14/05/2015	Yes	Illumina MiSeq

Fluoroquinolone (FQ). Wild type (WT):Mutant (MU).

PCR optimization

Primer design

Primer3Plus (Untergasser *et al.*, 2007) was used to design primers for the amplification of the QRDR region of *gyrA* and *gyrB*, with product sizes not greater than 300 base pairs in length (Appendix **Table A 1**). The length of the amplicon plays an important role in determining the amount of sequencing coverage that can be generated per sample as the sequencing technology used in this study was limited to 300 bp reads. In addition, we evaluated the amplification efficiency of previously designed primers (Chakravorty *et al.*, 2011 and Colman *et al.*, 2015) (Appendix **Table A 1**). Each primer sequence was then compared to the reference sequence making use of the Basic Local Alignment Search Tool (BLAST) on the NCBI database, to determine whether the designed primers had multiple binding sites elsewhere in the genome of *M. tuberculosis* (NCBI, 1997). Primer sets selected for this study are shown in **Table 2**, all of which provided reliable, specific and efficient amplification of the target regions. The primer sets were either used in a mono-plex PCR reaction when only one of the primer sets were present in the reaction (during testing for the efficiency of amplification), or used in a duplex PCR reaction where both primers sets were present in the reaction (during testing for the efficiency of amplification as well as during amplification for Sanger sequencing and Targeted deep sequencing).

Table 2 Optimal primer sets, for the QRDR region of *gyrA* and *gyrB* genes.

Name of primer set	Gene	Forward primer	Reverse Primer	Size (bp)
*Primer set A	<i>gyrA</i>	5'-gggtgcttatgcaatgttcgat-3'	5'-gggcttcggtgtacccatc-3'	237
Primer set B	<i>gyrB</i>	5'-accgacatcggtggattg-3'	5'-gatatcgaactcgtcgatggat-3'	261

*Primer set A described by Colman *et al.*, 2015

PCR parameters

The optimized PCR conditions were as follows: The PCR protocol was according to Colman *et al.* (2015), with the exclusion of 5 µl Betaine solution (Appendix **Table A 5**). The 25 µl PCR reaction consisted of the following reagents: 9 µl Nuclease-Free water (Ambion), 12.5 µl Q5 Hot Start High-Fidelity 2X Master Mix (The Q5 Hot Start High-Fidelity 2X Master Mix contains: dNTPs, 2 mM Mg⁺⁺, proprietary broad-used buffer, and hot start DNA polymerase with 3' – 5' exonuclease activity (New England Biolabs Inc.), 10 µM forward primer (1.25 µl), 10 µM reverse primer (1.25 µl), and 1 µl pure DNA (at a concentration of 25 ng/µl). A volume of 2 µl was used for DNA samples extracted using the GenoLyse kit. To reduce primer dimer formation, the primer concentration was decreased to 0.4 µM forward primer (0.05 µl) and 0.4 µM reverse primer (0.05 µl), these concentrations were also used for the duplex PCR. To compensate for the decreased primer volumes we also increased the volume of Nuclease-free water (Ambion) to 11.5 µl (Appendix **Table A 1**). The thermocycling conditions for cultured sample were as follows: an initial denaturation at 98°C for 1 min, followed by 35 cycles of denaturing at 98°C for 10 sec, annealing at 60°C for 15 sec, and extension at 72°C for 20 sec and a final extension at 72°C for 2 min (Protocol for Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs Inc.)). To improve the efficiency of amplification, the number of cycles was increased to 45 cycles with the annealing temperature increased to 62 °C (for 15 sec) when using crude DNA (extracted using the GenoLyse kit) as template (Appendix **Table A 2** and - **Table A 3**). Each procedure was conducted in a separate room to minimise the risk of cross-contamination. RNase-free water was used as a negative control to regulate for reagent contamination.

Electrophoretic evaluation

A 1.5% agarose solution was prepared by dissolving 1.5 g of agarose in 100 ml of 1x tris base acetic acid ethylene diamine tetra acetic acid (TAE) buffer (pH 8.0) and heating the solution in a microwave oven until fully dissolved. Once the mixture has cooled, 5 µl ethidiumbromide (10 mg/ml) was added to the solution casted into a gel tray and allowed to cool to room temperature. A 5 µl aliquot of the PCR products (25 µl) were mixed with 1 µl loading dye (0.25% Xylene Cyanol, 30% glycerol) and was loaded onto the gel. The size of the amplified DNA samples were determined by a 100 bp Plus DNA ladder (GeneRuler, Thermo Scientific, USA) that was loaded into a well adjacent to the samples. The gel was run at 100 V for approximately 3-4 hours in 1x TAE buffer and visualized under ultra violet light using the Experion automated electrophoresis system. To avoid any form of amplicon contamination, all PCR reactions were done in duplicate. The opened PCR tube (containing the remaining 20 µl PCR product) was send for Sanger sequencing and only PCR products

from closed PCR tubes were sent to Central Analytical Facility (CAF) or Agricultural Research Council (ARC) for next generation sequencing.

Sanger sequencing

The PCR products were subjected to post PCR manual bead clean-up and sequenced with only the forward primer from primer set A using Sanger Sequencing at the Central Analytical Facility (CAF), Stellenbosch University. The reason is that variants are found to occur in *gyrA* more often and seldom in *gyrB*. The sequencing files received from CAF were aligned to the *M. tuberculosis* H37Rv laboratory reference sequence using DNAMAN (Lynnon, Quebec, Canada), while Chromas version 2.4 (Technelysium, Australia) was used to visually inspect the chromatograms of the sequences for underlying variants which are represented by double peaks on the chromatogram file.

Targeted deep sequencing of isolates

Ion Torrent and Illumina were used as the two sequencing platforms to investigate the presence of underlying fluoroquinolone resistant populations in the DNA isolated from the *M. tuberculosis* isolates selected for the present study, which were subjected to targeted deep sequencing. Unpurified PCR products were submitted to the relevant sequencing facilities, namely CAF (for sequencing on the Ion Torrent platform) and ARC (for sequencing on the Illumina platform) for library preparation. During the library preparation, indexing adapters were ligated to the DNA fragments to allow for samples to be pooled on a sequencing run. The depth of the coverage for all isolates was estimated to be at least 60000 times to ensure a high level of confidence for identifying low frequency variation in the genes of interest.

Ion Torrent sequencing was done using the Ion 318 semiconductor chip on the Ion Torrent Personal Genome Machine (PGM) System (Life technologies, California, USA). Two target regions (237bp QRDR of *gyrA* and 261bp QRDR of *gyrB*) of the 10 synthetic heterogeneous mixes, ranging from 0% to 10% mutant (**Table 1**). An amplicon library was prepared prior to sequencing using the Ion Amplicon Library Preparation (Fusion Method) protocol at CAF.

The Illumina MiSeq platform using the 2x300bp version 3 sequencing chemistry (Illumina, California, USA) was used for targeted deep sequencing on (I) the 10 synthetic heterogeneous mixtures (ranging from 0% to 10% mutant), (II) crude DNA extracted from serial patient isolates displaying the acquisition of fluoroquinolone resistance and (III) crude DNA extracted from serial patient isolates receiving levofloxacin at different dosages in a clinical trial (**Table 1**). TruSeq Nano DNA Library Prep guide was used for the library preparation.

Bioinformatic analysis

Ion Torrent data analysis

The BAM files generated by the Ion Torrent PGM platform was analysed by the Torrent Suite 4.0 on the Torrent Server (<http://github.com/iontorrent/TS>) at CAF. DNA sequences were aligned to the reference genome *M. tuberculosis* H37Rv. The generated BAM data files of the targeted deep sequenced synthetic heterogeneous mixtures of *M. tuberculosis* isolates were provided by CAF and were shared via a secure file transfer protocol (FTP). Subsequently, variants were called using the stand alone Torrent Variant Caller Plugin (ThermoFisherScientific, 2013). The variants called by the Ion Torrent Variant Calling Algorithms are based on the consensus accuracy which is independent of the variants identified in raw reads. The specific Ion Torrent Variant Calling parameters were adjusted as follows: (1) the duration of time ("wall time") the torrent server spends on a plugin was increased from 96 hours to 960 hours to ensure completion of the analyses and (2) the coverage ("downsample_to_coverage") was increased from 2000 times coverage (the default) to 60000 times coverage. The resulting files were provided in a variant caller format (VCF). The VCF file, is a tab-separated file that contains information regarding the position, alternative allele observations, read depth at the locus (coverage) and quality score in phred scale of each variant.

Illumina data analysis

The fastq targeted deep sequence data files of the sequenced *M. tuberculosis* isolates were provided by the ARC and were shared via a secure FTP. A computational program, md5sum was used to verify if the raw sequenced data files (fastq files) were completely downloaded and similar to the files on the FTP site. Computational analysis was done using the protocol described by Colman *et al.* (2015) (Colman *et al.*, 2015).

The workflow for data analysis of targeted deep sequences on the Illumina platform is summarised in **Figure 2**.

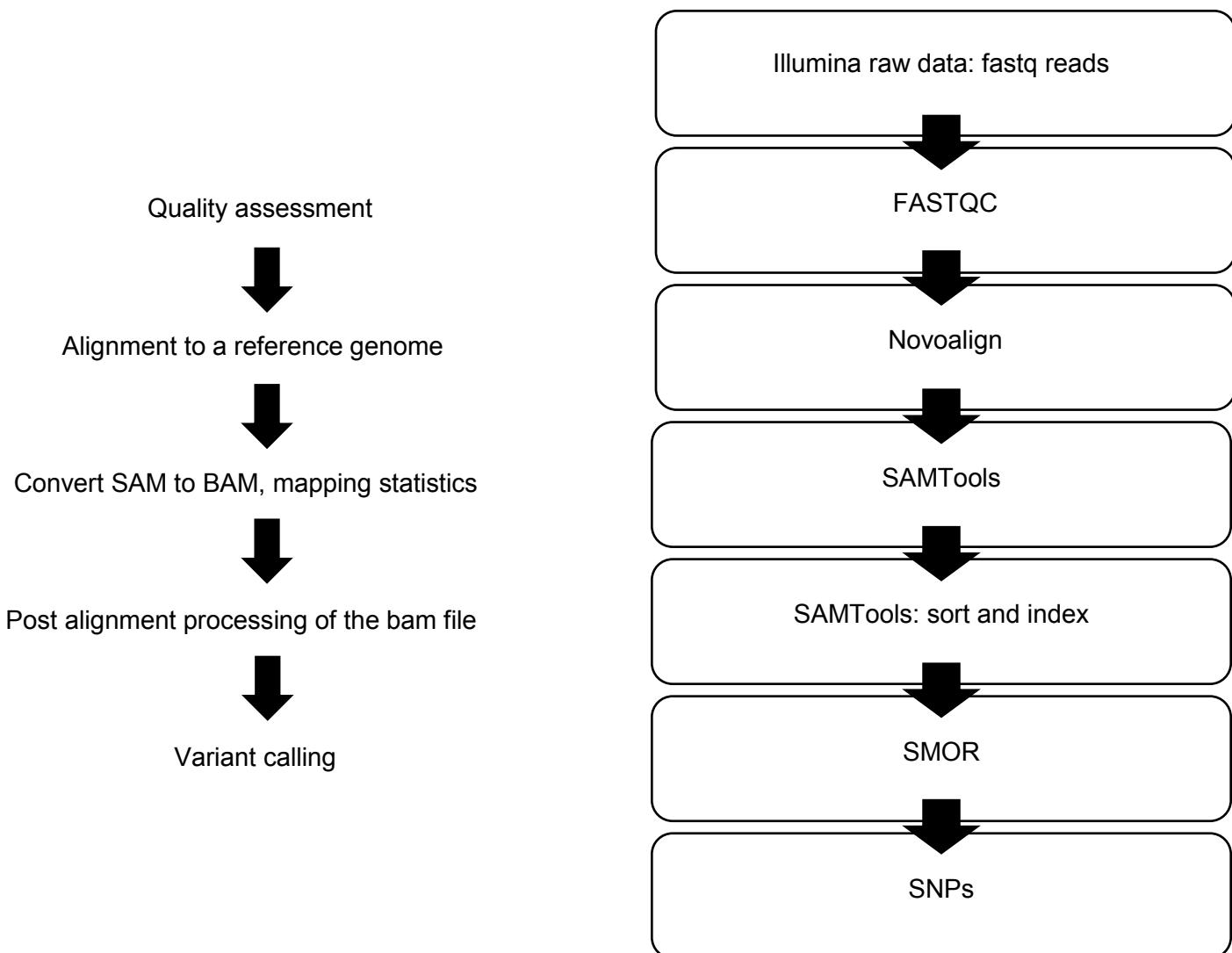


Figure 2 Outline of the computational analysis of targeted deep sequencing data. SAM: Sequence Alignment/Map format, BAM: Binary alignment format. SMOR: Single molecule-overlapping reads analysis, SNPs: Single-nucleotide polymorphism.

FASTQ file format

Biological sequences and the corresponding quality information recorded from the sequencer are stored in a text-based format (i.e. fastq format) comprising of four lines. The first line begins with a "@" character and is followed by a sequence identifier. The second line gives the raw sequence letters. The third, optional line starts with a "+" character and is optionally followed by the same sequence identifier. The fourth line encodes the phred quality value of the sequence, with one character pertaining to each base in the second line of a read entry. (Cock *et al.*, 2010) (**Figure 3**).

Line 1	@MISEQ_M01232:13:000000000-AMWWWD:1:1101:10838:1362 1:N:0:GTTTCG
Line 2	GGGCTTCGGTGTACCTCATGCCGCCGGTGGGTATTGCCTGGCGAGCCGA
Line 3	+
Line 4	CCCCCGGGGGGGGGFGGFGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG

Figure 3 An example of a read from a fastq file generated by the Illumina sequencing.

Phred quality score

Sanger and Illumina phred quality scores (Q) are depicted as American standard code for information change (ASCII) characters with an offset of +33. Each character in Line 4 of a fastq file is correlated with a number. For example, the character "C" (Line 4 in **Figure 3**) represents a phred quality score (Q) of 34 that correlates to an error probability of 0.000040, which means that the quality of the base scored was high.

NGS sequencing technologies generate phred scores that are assigned to each nucleotide base called in an automated sequencing run (Cock *et al.*, 2010). A phred score is logarithmically linked to the error probability of a base called, e.g. $Q = -10 \log_{10} P$, where: Q is equivalent to phred score and P is equivalent to error probability.

A high base accuracy is represented by a phred score of >10. If a base has a phred score of 10 it means that the probability of an incorrect base call is 1 in 10, resulting in a base call accuracy of 90%.

Quality control

The quality of the raw sequenced data for each isolate analysed was assessed with FastQC software (Babraham Institute, UK) using the generated fastq files. FastQC is a java-based program that evaluates various aspects of the raw data and produces the results in HTML (HyperText Markup Language) format (Consortium, 1999). The FastQC analysis is a 7 step module which includes: I) general statistics together with a description of the platform used,

the type of input file, the amount of reads processed, read length, and percentage GC content, II) calculates per read quality score by making use of the mean scores of all the reads, III) calculates the per base sequence content to determine the distribution of the four nucleotides throughout the reads, IV) calculates the GC content throughout the reads and comparing it to a theoretical normal distribution and mean value, V) calculates the possibility of contamination in the reads, VI) calculates the number of uncalled bases throughout the reads (number of "N"s), and VII) calculates the amount of duplicate sequences. A detailed FastQC report is generated from the results which provides a warning if quality scores are below that of the pre-set parameters.

Read alignment to the reference genome

Novoalign (Needleman-Wunsch algorithm) (Novocraft Technologies <http://www.novocraft.com>) mapping software were used to accurately map short sequencing reads (in the fastq format) to an indexed reference genome, *M. tuberculosis* H37Rv. Indexing of the reference genome was done by the "novoindex" command, using a k-mer indexing size of 13 and an indexing step size of 1. The fastq reads were used as input and the alignment was computed using the Needle-Wunsch algorithm. Novoalign software does global gapped alignment and for this analysis the default value of 6 was used, thus allowing six mismatches per alignment, producing an output in the Sequence Alignment/Map (SAM) format. The SAM format is compact in size, can be indexed (allowing for the efficient and fast retrieval of all the reads at a specific chromosomal locus. This format also allows most operations on the alignment to work on a stream without loading the entire alignment into the memory.

Converting the SAM file to a BAM file

SAMTools (<http://samtools.sourceforge.net>) provides a set of utilities for manipulating and post-processing of SAM files (Li *et al.*, 2009). The software package includes tools to sort, merge, and index SAM files. The SAMTools commands "view" and "sort" was used to convert the SAM alignment files into a sorted binary alignment (BAM) format. Likewise, SAMTools was used to assess the percentage mapped reads using the "flagstat" function.

Variant calling

An independent variant caller was used to identify SNPs with regards to the reference genome used for the alignment. The mapping file obtained from the previous step was analysed with a SNP caller, single molecule-overlapping reads analysis (SMOR) (DePristo *et al.*, 2011; Colman *et al.*, 2015). The resulting variants are given in the variant call format (vcf) file format. The variant caller was used to detect underlying variants present in deep sequenced data.

SMOR

The SMOR analysis script (<https://github.com/TGenNorth/SMOR>) (Colman *et al.*, 2015) was used to call the frequency of each nucleotide present at each position in the sequence. The SMOR script (paired_haplotype_smor.pl found at <https://github.com/TGenNorth/SMOR>) is preloaded with a set of amplicon identifiers and gene positions that are known to confer resistance to anti-TB drugs. Additional gene targets or gene positions can be added. Alternative gene targets or gene positions can also be removed from the SMOR script, which will allow for an unbiased variant calling approach. The script accepts a BAM file as input, and searches the file for complete overlap of the forward and reverse paired-end reads from the same DNA molecule, calculating the total frequency at which each nucleotide appears at a certain position in single- and paired reads. The information is presented in a table of counts for each position of interest, e.g. the output text file contains the position of interest, the number of times a nucleotide is called by a pair of reads, the total coverage, as well as the total number of times a nucleotide is called by a single read.

For analysis purposes the QRDR regions of *gyrA* and *gyrB* were targeted in *M. tuberculosis* which contains drug resistance conferring SNPs. The entire QRDR of both *gyrA* and *gyrB* (unbiased approach), as well as forty-five previously published SNP loci conferring resistance to ofloxacin (OFX) (**Table 3**) were examined during the SMOR analysis.

Table 3 SNP positions in the QRDR of *gyrA* and *gyrB* that confer antibiotic resistance.

Gene	codon/ nucleotide	Mutation	codon position in H37Rv	SNP location in H37Rv	Wild type	Mutant
<i>gyrA</i>	88	GGC-TGC	7563-7565	7563	G	T
<i>gyrA</i>	88	GGC-GCC	7563-7565	7564	G	C
<i>gyrA</i>	89	GAC-AAC	7566-7568	7566	G	A
<i>gyrA</i>	90	GCG-G(A/G/T)G	7569-7571	7570	C	(A/G/T)
<i>gyrA</i>	91	TCG-(C/G)CG	7572-7574	7572	T	(C/G)
<i>gyrA</i>	92	ATC-ATG	7575-7577	7577	C	G
<i>gyrA</i>	94 a	GAC-(A/C/T)AC	7581-7583	7581	G	A/C/T
<i>gyrA</i>	94 b	GAC-G(T/C/G)C	7581-7583	7582	A	T/C/G
<i>gyrA</i>	95**	AGC-ACC	7584-7586	7585	G	C
<i>gyrA</i>	102	CCC-CAC	7605-7607	7606	C	A
<i>gyrA</i>	109	CTG-GTG	7626-7628	7626	C	G
<i>gyrA</i>	112**	GGC-GGG	7635-7637	7637	C	G
<i>gyrA</i>	120**	GGC-GGG	7659-7661	7661	C	G
<i>gyrA</i>	125**	GCG-GCA	7674-7676	7676	G	A
<i>gyrB</i>	90	GCC-GGC	5507-5509	5508	C	G
<i>gyrB</i>	446	CGT-TGT	6575-6577	6575	C	T
<i>gyrB</i>	446	CGT-C(A/T)T	6575-6577	6576	C	T
<i>gyrB</i>	447	TCC-TTC	6578-6580	6579	C	T
<i>gyrB</i>	461	GAC-(A/C)AC	6620-6622	6620	G	(A/C)
<i>gyrB</i>	461	GAC-GCC	6620-6622	6621	A	C
<i>gyrB</i>	470	GGT-TGT	6647-6649	6647	G	T
<i>gyrB</i>	470	GGT-GCT	6647-6649	6648	G	C
<i>gyrB</i>	486	ATC-CTC	6695-6697	6695	A	C
<i>gyrB</i>	494	GAC-GCC	6719-6721	6720	A	C
<i>gyrB</i>	499	AAC-(T/G)AC	6734-6736	6734	A	(T/G)
<i>gyrB</i>	499	AAC-TAC	6734-6736	6735	A	C
<i>gyrB</i>	499	AAC-TAC	6734-6736	6736	C	(A/G)
<i>gyrB</i>	500	ACC-CCC	6737-6739	6737	A	C
<i>gyrB</i>	500	ACC-C(A/T)C	6737-6739	6738	C	(A/T)
<i>gyrB</i>	501	GAA-GTA	6740-6742	6741	A	T
<i>gyrB</i>	501	GAA-GA(T/C)	6740-6742	6742	A	(T/C)

<i>gyrB</i>	504	GCG-ACG	6749-6751	6749	G	A
<i>gyrB</i>	504	GCG-GTG	6749-6751	6750	C	T
<i>gyrB</i>	507	ACG-ATG	6758-6760	6759	C	T
<i>gyrB</i>	538	CAA-CA(T/C)	6851-6853	6853	A	(T/C)
<i>gyrB</i>	670	GTT-TTT	7247-7249	7247	G	T

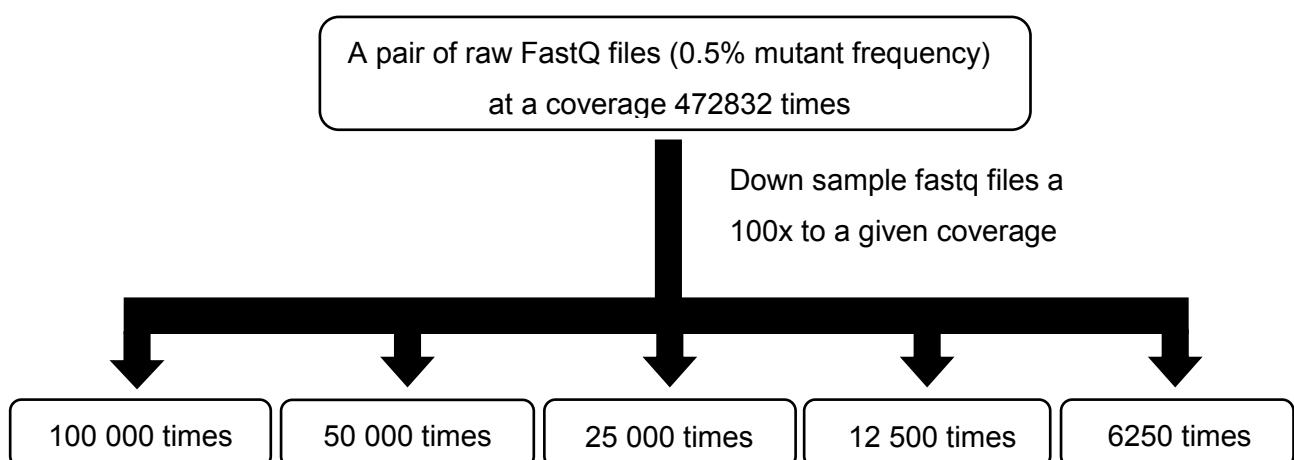
** SNP does not confer resistance, possible position of interest or phylogenetic informative.

See (Coll *et al.*, 2015; Colman *et al.*, 2015) for the SNP positions that confer antibiotic resistance.

Down sampling of a synthetically created heteroresistant mixture

To determine the minimum coverage needed to confidently detect extremely low frequency (<1%) *gyrA* variants, the total depth of coverage of a fastq file with a variant frequency of 0.5% and depth of coverage of 472832 times were down sampled in a step wise manner to 6250 times (**Figure 4**). Each down sampling step was repeated 100 times with the use of an automated in-house bash script (N.L. Da Camara, Department of Biomedical Sciences, Stellenbosch University) that allowed for the use of different seed numbers (i.e. prime numbers ranging from 300 to 5000) in order to generate "random" samples. The automated script made use of the seqtk tool (<https://github.com/lh3/seqtk>) for processing sequences in the FASTA and FASTQ format, in combination with an in house bash script (N.L. Da Camara, Department of Biomedical Sciences, Stellenbosch University) (see **Appendix C**). This resulted in the random generation of a 100 different fastq files at the relevant depth of coverage (e.g. 100000 times coverage).

Figure 4 Flow diagram showing down-sampling of sequenced data (fastq file) to a certain coverage a 100 times.



Ethics

The work described in the present study forms part of a large on-going project which received ethical approval from the Stellenbosch University Health Research Ethics Committee under the title: —Method development to detect minority fluoroquinolone resistant variants for predicting TB treatment failure: N16/04/049.

Chapter 4: Results

In order to determine the accuracy of the Illumina MiSeq and Ion torrent platforms identify underlying minority frequency variants in *gyrA* and *gyrB*, targeted deep sequencing was done using synthetic heterogeneous mixtures of a characterized wildtype (H37Rv) and ofloxacin resistant mutant (BO1). Subsequently, targeted deep sequencing (using the Illumina MiSeq platform) was used to investigate the presence of underlying variants in *gyrA* and *gyrB* in two groups of serial patient isolates: 1. Serial patient isolates collected that showed the acquisition of fluoroquinolone resistance, 2. Serial patient isolates collected from fluoroquinolone susceptible MDR-TB patients receiving different dosages of levofloxacin during a clinical trial. Lastly the study also investigated the minimum depth of coverage that is needed to detect underlying minority frequency variants. An outline of the sample selection and methodologies used in this study is showed in Figure 5.

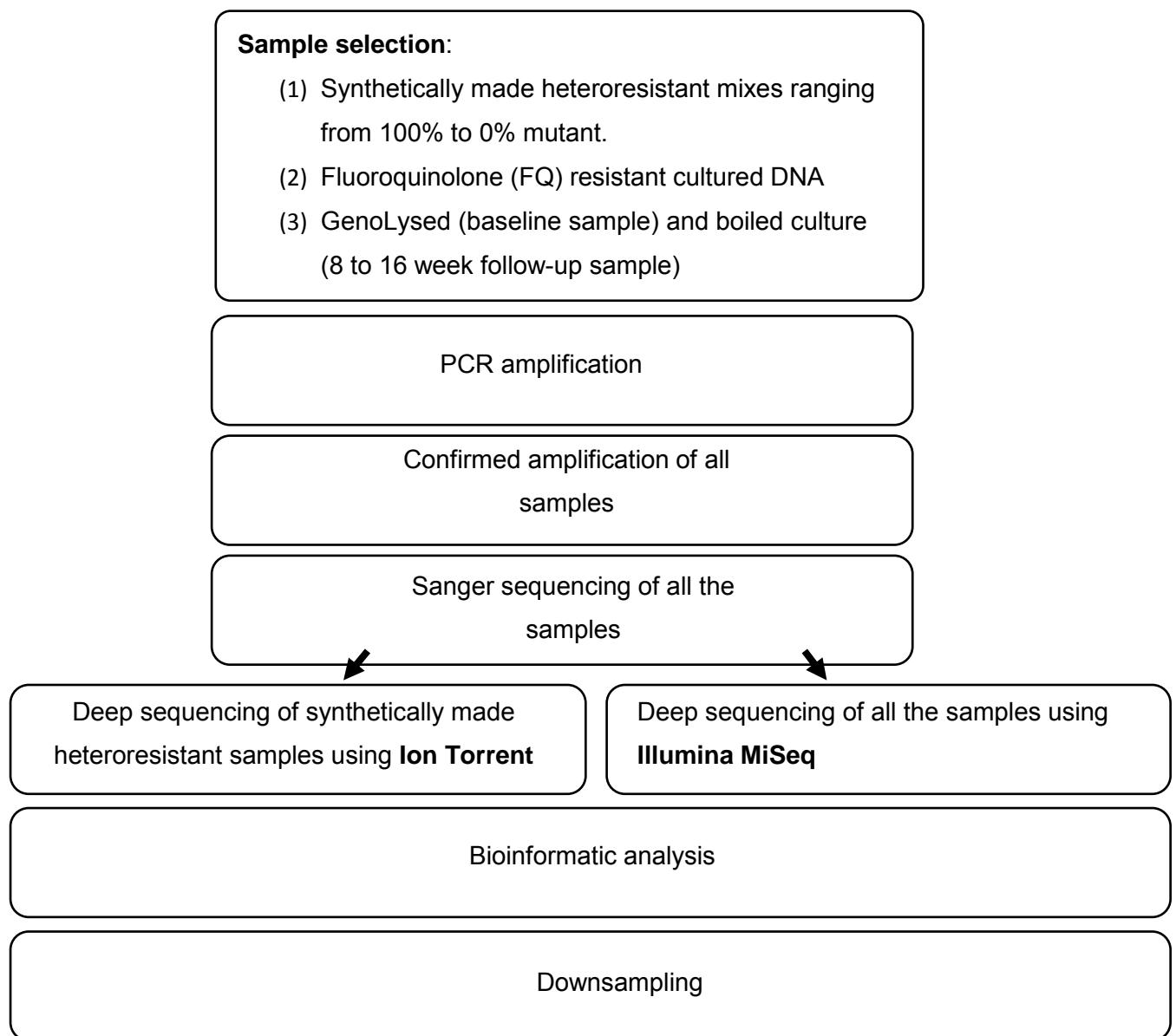


Figure 5 Flow chart of sample selection and the research methodology.

PCR optimisation

We first optimized the amplification of the QRDR of *gyrA* and *gyrB* in monoplex PCR reactions. Subsequently the amplification of both targets was optimized in a duplex PCR reaction, where the annealing temperature was decreased from 62°C to 60°C. In addition the number of reaction cycles was adjusted depending on the type of input DNA. It was found that 35 reaction cycles was efficient when using pure DNA or crude DNA extracted using the heat inactivation method and 45 reaction cycles were required using DNA extracted with the GenoLyse kit. A PCR efficiency test using quantitative Real-Time PCR and a ten-fold dilution series of purified H37Rv DNA showed that the efficiency of the amplification is equivalent for both primer sets used.

We then tested the sensitivity of the duplex PCR reaction using a ten-fold serial dilution (from 25 ng/ μ l to 25×10^{-6} ng/ μ l) of a synthetic heterogeneous mixture with a mutant allele frequency of 10% as DNA template. As shown in Figure 6 the PCR product (with expected product sizes at 261bp for the QRDR of *gyrB* and 237bp for the QRDR of *gyrA*) using a starting template concentration of 25×10^{-6} ng/ μ l were detected by gel electrophoresis and UV visualisation. In addition, underlying peaks of the resistant allele were visible on the chromatograms obtained from Sanger sequencing, shown in Figure 7, when using a starting template concentration of 25×10^{-5} ng/ μ l, which is equivalent to 52 copies of template DNA.

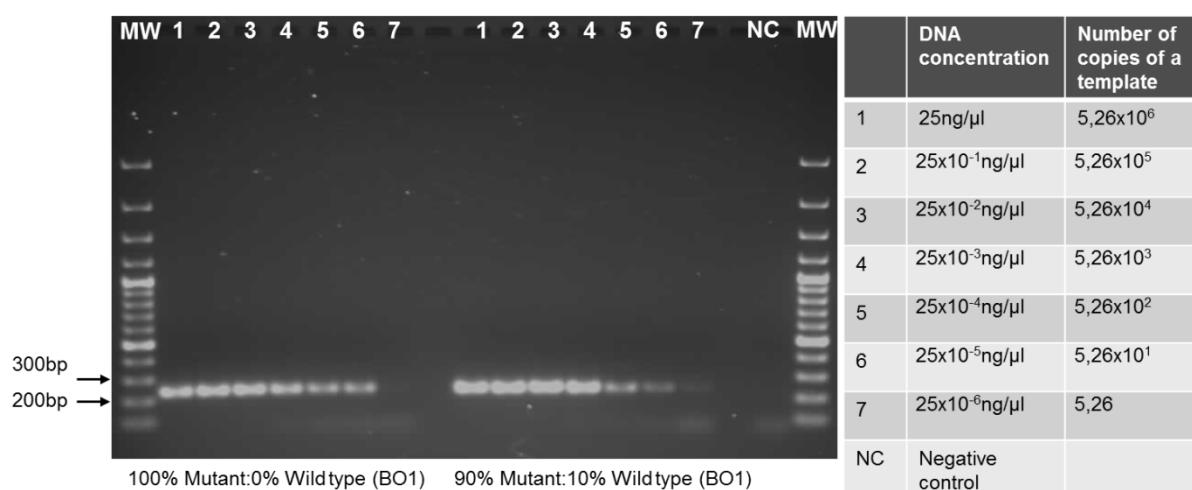


Figure 6 Visualisation of PCR products of a ten-fold serial dilution of using wildtype H37Rv DNA as well as a synthetic heterogeneous mixture with a mutant allele frequency of 10% as DNA template. The expected product sizes of the PCR products were 261bp for the QRDR of *gyrB* and 237bp for the QRDR of *gyrA*. Lanes 1-7 shows the range (25 ng/ μ l to 25×10^{-5} ng/ μ l) of the dilution series that was used as DNA template. Lane NC shows the negative control. Lane MW shows the GeneRuler 100 bp Plus DNA Ladder (GeneRuler, Thermo

Scientific, USA) that was used as a molecular weight marker representing DNA fragments of known size.

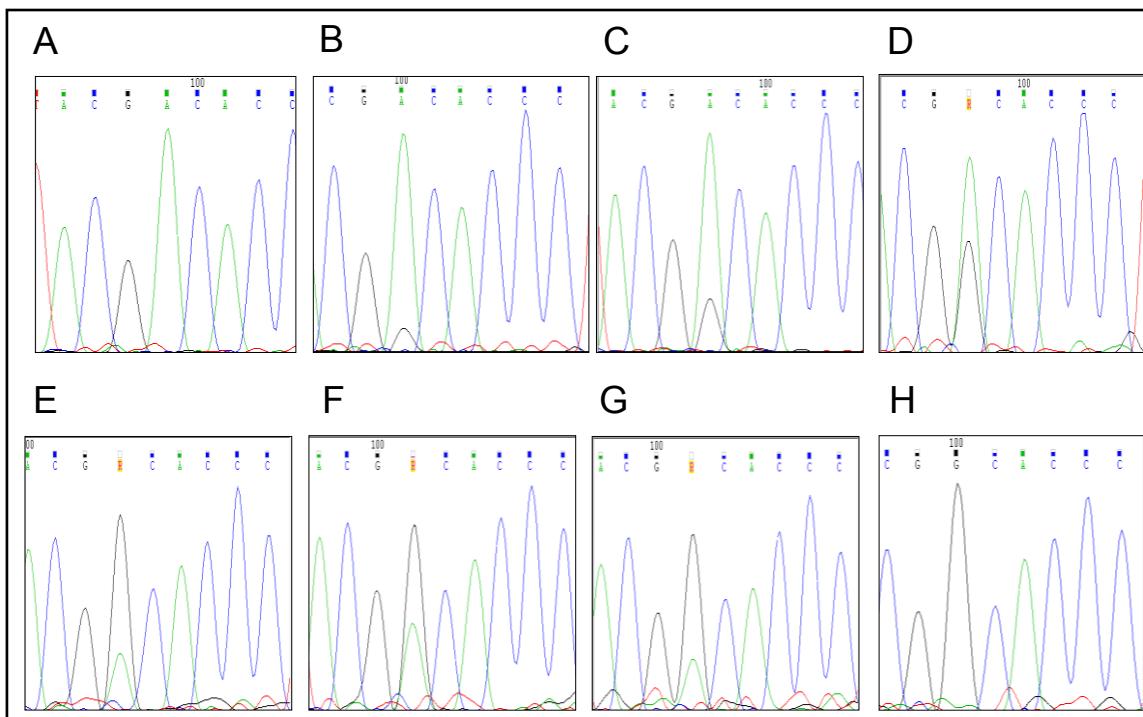


Figure 7 Sanger sequencing of synthetic heterogeneous mixture ranging from 100% to 0 % mutant, and vice versa) consisting of a DNA concentration of 10^{-5} ng/ μ l, which is equivalent to 52 number of DNA copies. The DNA underwent PCR amplification and purified PCR products were Sanger sequenced. For example A) consist of 100%:0% denotes wild type (*M. tuberculosis* H37Rv) to mutant (FQ monoresistant strain (BO1)). B) 90%:10%. C) 80%:20%. D) 70%:30%. E) 30%:70%. F) 20%:80%. G) 10%:90%. H) 0%:100%

Dilution series - Synthetically made heterogeneous mixes

To test the sensitivity of the duplex PCR reaction and Sanger sequencing to detect underlying variants in the amplified targets of *gyrA* and *gyrB*, synthetic heterogeneous mixtures (ranging from 100% to 0 % mutant, and *vice versa*) were used as DNA template. Thereafter the Sanger sequencing chromatograms were visually inspected for the presence of underlying peaks (Figure 8). Underlying variants (i.e. heterogeneous) were identified by the presence of more than one peak at a single base position on the sequencing chromatogram. The limit of detection for Sanger sequencing was shown to be 6%, as no underlying peaks were detecting in samples with less than 6% mutant DNA (Figure 8). This is in agreement with previous studies (Folkvardsen, Svensson, *et al.*, 2013; Pholwat *et al.*, 2013).

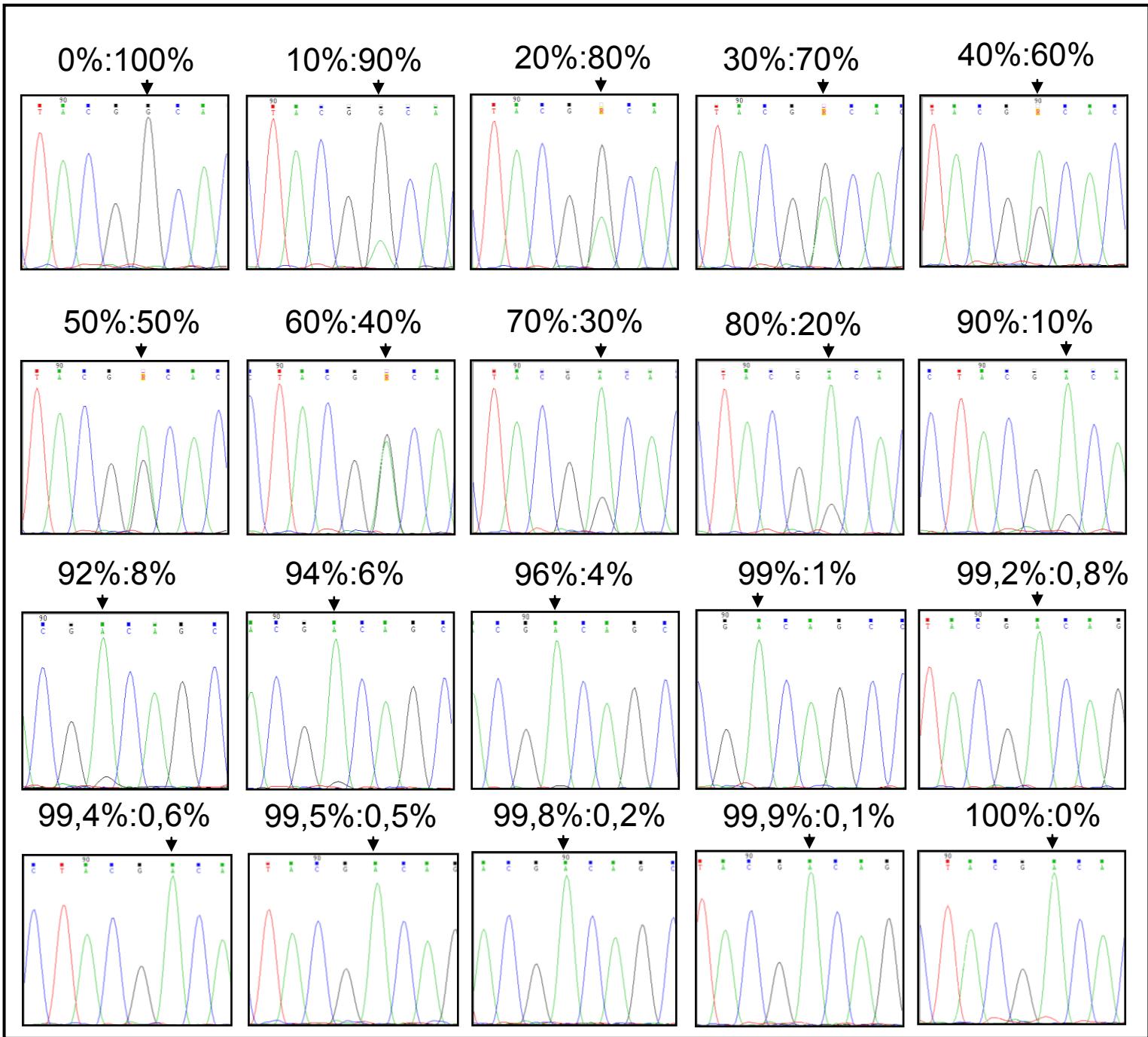


Figure 8 Sanger sequencing chromatograms of synthetic heterogeneous mixtures consisting of ratios of *M. tuberculosis* H37Rv:FQ mono-resistant strain (BO1), ranging from 100% mutant (top left corner) to 0% mutant (bottom right corner) and from 0% wild type (top left corner) to 100% wild type (bottom right corner). Heteroresistance was identified by the double peaks at a single base position seen in the chromatograms.

Targeted deep sequencing of isolates (Ion Torrent PGM and Illumina MiSeq)

Quality assessment of the fastq and alignment files.

Synthetic heterogeneous mixtures

The sequence quality per base of the forward read of all the mixtures were very good (phred score of 25 for 280bp of the read) with a decline in quality at the last 20bp of the forward read. However, the reverse read sequence quality was lower for the last 40bp of the read.

Table 4 shows that an average of 404441 reads generated by the Ion Torrent platform and 509765 reads generated by the Illumina MiSeq platform mapped to the *M. tuberculosis* reference genome, H37Rv when using TMAP and Novoalign as aligners respectively. Only reads with a quality phred score of greater than 20 (good) were allowed to map to the reference genome. The average read length of the reads was 209bp (using Ion Torrent) and 300bp (using Illumina MiSeq). The average coverage at which variants were called was 53219 times (using Ion Torrent) and 185465 times (using Illumina MiSeq).

Table 4 Overview of the quantity (read coverage) of raw sequenced data of synthetic heteroresistant mixtures reads generated by the two different platforms, Ion Torrent Personal Genome Machine (PGM) and the Illumina MiSeq platform. Wild type (WT), mutant (MUT).

Study	Origin	Ratio WT:MUT	Mapped reads TMAP (Ion Torrent PGM)	Mapped reads Novoalign (Illumina)	Coverage of reads with a mapped quality $\geq Q20$ TMAP (Ion Torrent PGM)	Coverage of reads with a mapped quality $\geq Q20$ Novoalign (Illumina)	Read length TMAP (Ion Torrent PGM)	Read length Novoalign (Illumina)
Synthetic H37Rv & BO1 (mutant) heteroresistant mixtures	H37Rv & BO1 (mutant)	90:10	417406	432330	59913	157071	212	300
	H37Rv & BO1 (mutant)	92:8	294978	519192	18687	177551	187	300
	H37Rv & BO1 (mutant)	94:6	245319	544213	59921	205120	202	300
	H37Rv & BO1 (mutant)	96:4	543601	471204	59619	167255	213	300
	H37Rv & BO1 (mutant)	99:1	511031	503189	59773	186021	212	300
	H37Rv & BO1 (mutant)	99.2:0.8	378399	568886	57857	183776	212	300
	H37Rv & BO1 (mutant)	99.4:0.6	566221	461636	59915	170933	215	300
	H37Rv & BO1 (mutant)	99.5:0.5	347993	472831	51828	176022	213	300
	H37Rv & BO1 (mutant)	99.8:0.2	401252	516809	59915	205992	211	300
	H37Rv & BO1 (mutant)	99.9:0.1	338215	607364	44767	224909	216	300

Clinical *M. tuberculosis* isolates of three patients showing evolution of FQ resistance

The average number of Illumina MiSeq reads that mapped to the *M. tuberculosis* reference genome, H37Rv, using Novoalign was 331574 (**Table 5**). The quality of the reads were extremely good with an average phred score of greater than 34 per read. Variants were called at an average coverage of 114462 times using reads with a read length of 301bp.

Table 5 Quantity (read coverage) of raw sequenced data of FQ resistant *M. tuberculosis* clinical isolates prior to bioinformatics analysis.

Study	Patient	Isolate designation	Mapped reads Novoalign (Illumina)	Coverage of reads Novoalign (Illumina)	Read length Novoalign (Illumina)
Evolution of FQ resistance	Patient 1	1 st clinical isolate	382130	117016	301
		2 nd clinical isolate	258314	114914	301
	Patient 2	1 st clinical isolate	387236	162655	301
		2 nd clinical isolate	345390	136528	301
		3 rd clinical isolate	345418	81089	301
		4 th clinical isolate	297100	87286	301
	Patient 3	1 st clinical isolate	280482	99420	301
		2 nd clinical isolate	409253	134035	301
		3 rd clinical isolate	278849	97222	301

Clinical *M. tuberculosis* isolates from the OptiQ clinical trial of patients receiving FQ at different concentrations

The average number of Illumina MiSeq reads that mapped to the *M. tuberculosis* reference genome, H37Rv, using Novoalign was 1892628 (**Table 6**). The quality of the Illumina MiSeq reads had an average phred score of greater than 34 per read indicative of high quality reads. Variants were called at an average coverage of 121450 times using reads with a read length of 301bp.

Table 6 Quantity (read coverage) of raw sequenced data of *M. tuberculosis* clinical isolates from the OptiQ study, prior to bioinformatics analysis.

Study	Patient	Isolate designation	Mapped reads Novoalign (Illumina)	Coverage of reads Novoalign (Illumina)	Read length Novoalign (Illumina)
OptiQ study	Patient 1	Baseline clinical isolate	729108	38193	301
		Week 8 clinical isolate	755884	58606	301
	Patient 2	Baseline clinical isolate	848516	53517	301
		Week 8 clinical isolate	796732	58945	301
		Week 10 clinical isolate	1035072	78944	301
	Patient 3	Baseline clinical isolate	823716	33735	301
		Week 16 clinical isolate	1215724	105641	301
	Patient 4	Baseline clinical isolate	10947148	630140	301
		Week 8 clinical isolate	758700	64225	301
	Patient 5	Baseline clinical isolate	980816	69883	301
		Week 8 clinical isolate	763968	60864	301

Comparison of variants called by different mappers, i.e. TMAP (Ion Torrent) and Novoalign (Illumina) for:

Synthetically created heteroresistant mixtures ranging from 10% - 0.1% mutant

Different frequencies of variants were detected in the TMAP and Novoalign pipelines, which can be attributed to the different technologies used by the two platforms and mapping algorithms used by each mapper. **Table 7** shows the frequencies of variants detected for each synthetic heteroresistant mixture (ranging from 10% - 0.1% mutant) at an approximate coverage of 60000 time for Ion Torrent PGM deep sequenced data and at an approximate coverage of 185000 time for Illumina MiSeq deep sequenced data. Using this analysis method we were able to accurately detect subpopulations at 0.1% and above using the Illumina MiSeq platform. However, when using 60000 times coverage for the data generated by the Ion Torrent platform, only underlying variants with a frequency of greater than 2% could be detected. Ion torrent also did not detect the expected variant frequency at the input proportion. In contrast the Illumina MiSeq was able to detect variants at a similar proportion to the input.

Table 7 Comparison of the variants called by the different mappers (TMAP and Novoalign) for synthetic heteroresistant mixtures, ranging from 10% - 0.1% mutant.

			Variant frequency:	
Study	Origin	Isolate designation	TMAP (Ion Torrent)	Novoalign (Illumina)
Synthetic heteroresistant mixtures	H37Rv & BO1 (mutant)	Ratio WT:MU (90:10)	5%	9%
	H37Rv & BO1 (mutant)	Ratio WT:MU (92:8)	5%	7%
	H37Rv & BO1 (mutant)	Ratio WT:MU (94:6)	3%	5%
	H37Rv & BO1 (mutant)	Ratio WT:MU (96:4)	2%	3%
	H37Rv & BO1 (mutant)	Ratio WT:MU (99:1)		0.91%
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.2:0.8)		0.75%
	H37Rv & BO1 (mutant)	Ratio WT:MU (99.4:0.6)		0.57%
	H37Rv & BO1 (mutant)	Ratio WT:MU		0.41%

	(99.5:0.5)		
H37Rv & BO1 (mutant)	Ratio WT:MU (99.8:0.2)		0.17%
H37Rv & BO1 (mutant)	Ratio WT:MU (99.9:0.1)		0.09%

Comparative analysis between Sanger sequencing and NGS SMOR analyses of targeted deep sequencing isolates

Synthetically created heteroresistant mixtures ranging from 10% - 0.1% mutant

Sanger sequencing was only able to detect underlying resistant subpopulations above 6% (**Figure 9**). Alternatively, SMOR analyses was able to detect a known mutant allele frequency of greater than 0.1% with and without filtering of known FQ resistant loci (**Figure 9**). **Figure 9** shows SMOR analyses at a certain heteroresistant ratio without (**Figure 9**, A1 to A3) and with filtering (**Figure 9**, B1 to B3) known FQ resistant loci.

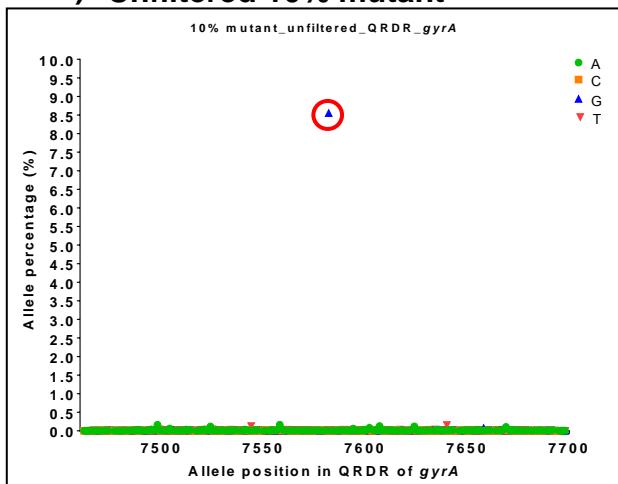
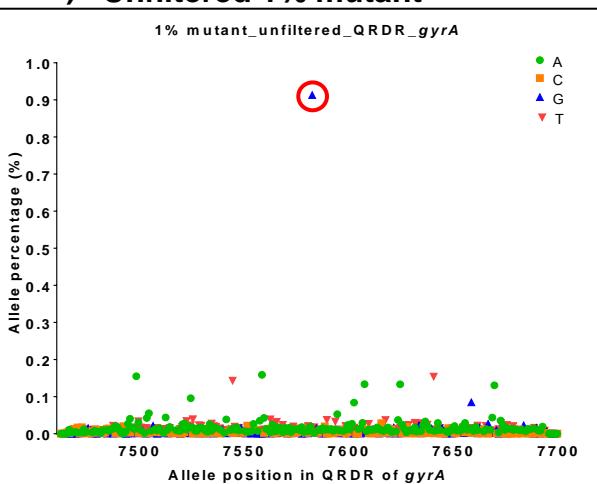
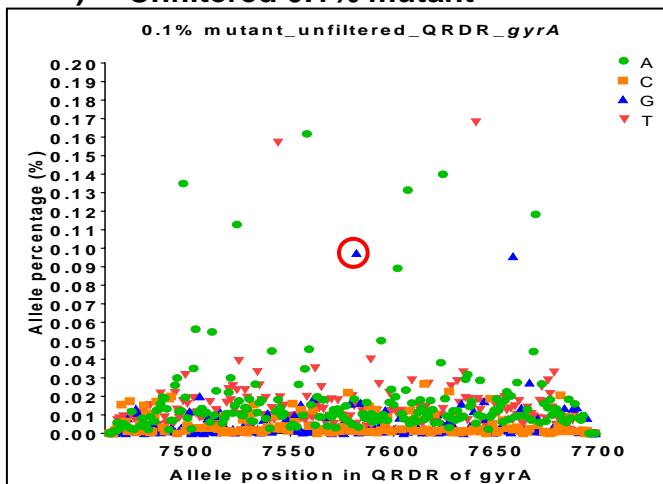
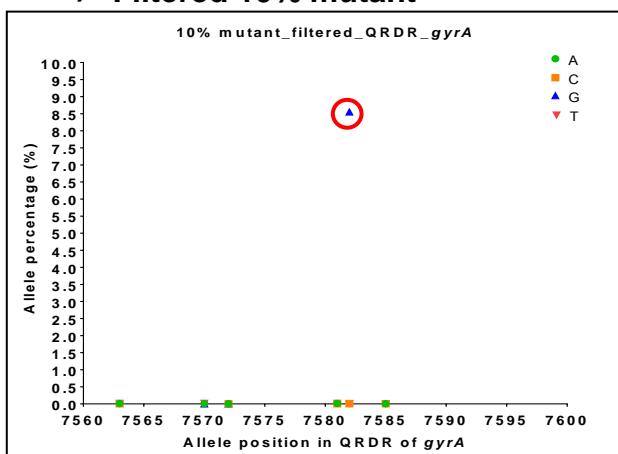
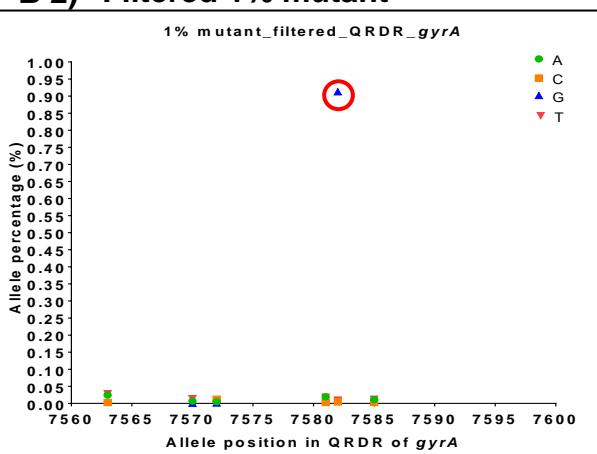
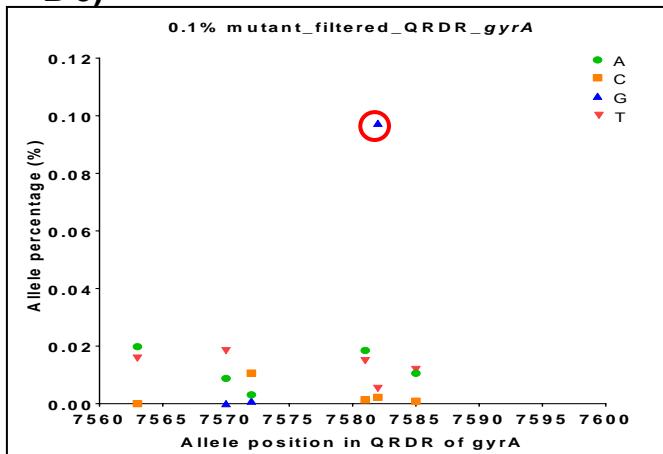
A 1) Unfiltered 10% mutant**A 2) Unfiltered 1% mutant****A 3) Unfiltered 0.1% mutant****B 1) Filtered 10% mutant****B 2) Filtered 1% mutant****B 3) Filtered 0.1% mutant**

Figure 9 Illumina targeted deep sequencing results of the QRDR of *gyrA* gene from Synthetic heterogeneous mixtures, showing SMOR analyses at a certain heteroresistant ratio with and without filtering known FQ resistant loci. Allele position in the QRDR of *gyrA* (x-axis) vs allele percentage (%) (y-axis). A 1) SMOR unfiltered targeted deep sequencing results for the QRDR of *gyrA* showing the presence of underlying mutant variants at A 1) 10%, A 2) 1% and A 3) 0.1%, B) SMOR filtered targeted deep sequencing results for the QRDR of *gyrA*, showing the mutant variants at A 1) 10%, A 2) 1% and A 3) 0.1%.

Clinical *M. tuberculosis* isolates of three patients showing evolution of FQ resistance

Targeted deep sequencing was used to investigate serial patient isolates of three patients that showed the acquisition of FQ resistance by routine DST. As shown in **Table 8**, the first isolate for all three patients showed the absence of FQ resistance according to phenotypic DST, while the follow-up isolates showed the progression to pre-XDR TB.

In **Table 8** the following was observed:

For Patient 1, two isolates were subjected to targeted deep sequencing. The first isolate (R8394) showed no variants above a frequency of 0.02% ("T" allele at codon 90 position 7570 and "G" allele at codon 94 position 7582 were present at a frequency of 0.02%), suggesting that this isolate is wild type (**Table 8**). The second isolate (R9172) showed the acquisition of three underlying mutations (GCG90GTG, GAC94GGC in the QRDR of *gyrA* and GAC461AAC in the QRDR of *gyrB*) with variant frequencies of 3.35% (GCG90GTG), 17.09% (GAC94GGC) in the QRDR of *gyrA*, and 0.16% (GAC461AAC) in the QRDR of *gyrB* (**Table 8**).

Four isolates from patient 2 were analysed using Illumina-based targeted deep sequencing. The first isolate (R6747) showed the presence of an underlying variant (GCG90GTG) at 0.17%, at a total depth of 162655 times coverage (**Table 8** and **Figure 10**). The second isolate (R6971) showed an increase of the mutant allele (GCG90GTG) to 68.18% at a total depth of coverage of 136528 times. The frequency of the mutant allele (GCG90GTG) increases in the third isolate (R7515) to 86.14%, and to 98.63% in the fourth isolate with total depth of 81089 times and 87286 times coverage, respectively. A comparison of results of Sanger sequencing with Illumina targeted deep sequencing of clinical isolates from patient 2 is shown in **Figure 10**.

Four samples of patient 3 underwent Illumina MiSeq targeted deep sequencing. No variants were observed within the first isolate. In the second isolate, a mutant allele (GCG90GTG) was present at 99.91% at a total depth of 99420 times coverage. In the third isolate, the mutant allele frequency decreases to 98.60% at a total depth of 134035 times coverage. Interestingly, a second underlying variant is observed at a frequency of 0.07% in codon GAC94GCC with a total depth of coverage of 135026 times (**Table 8**). In the fourth isolate, six underlying variants are observed. The mutation GCG90GTG decreased to 41.59% with a coverage of 96862 times (**Table 8**). The mutation GAC94GCC increased from 0.07 to 1.33% at 96490 times coverage (**Table 8**). Furthermore, the additional underlying variants were observed in codons 88, 89, 91 and an additional underlying variant was observed in codon 94. At codon 88, the mutant "T" allele (GGC88TGC) is present at 0.54% with a depth of

coverage of 97909 times. Codon 89 has a depth of coverage of 97710 times with an underlying "A" mutant allele (GAC89AAC) at 0.96% (**Table 8**). Furthermore, a 9.07% underlying variant was observed in codon 91 (TCG91GCG) at 96674 times depth of coverage (**Table 8**). Lastly, the lowest underlying variant seen in this sample (R11603) was 0.27%, a "C" mutant allele in codon 94 (GAC94CAC) (position 7581) present at a depth of coverage of 97688 times (**Table 8**).

To determine the reproducibility of the Illumina MiSeq deep sequencing method we compared our results from the above clinical isolates to the results of a previous analysis on isolate R2658 done independently by Translational Genomics Research Institute (TGen). Illumina targeted deep sequenced confirmed the two mutant alleles in codon 94, the "C" (GAC94GCC) mutant allele that is present at 77.41% and the "G" (GAC94GGC) mutant allele that is present at 22.43% with a total depth of coverage of 78728 times (data not shown). The frequencies between the previous targeted deep sequenced and of the current study was very similar. Previously, 74.63% for GAC94GCC and 24.86% for GAC94GGC were obtained, compared to 77.41% for GAC94GGC and 22.43% for GAC94GCC.

Table 8 Targeted deep sequencing results of clinical *M. tuberculosis* isolates showing acquisition of FQ resistance.

Patient	Clinical isolate	Date	Ofloxacin DST	Gene name	Codon change	Mutant allele frequency (%)	Depth of coverage (x) of base called	
1	1 st (R8394)	05/01/2010	FQ susceptible	gyrA	Wild type			
	2 nd (R9172)	23/03/2010	FQ resistant	gyrA	GCG90GTG	3.35	117016	
					GAC94GGC	17.09	114914	
				gyrB	GAC461AAC	0.16	126187	
2	1 st (R6747)	28/08/2009	FQ susceptible	gyrA	GCG90GTG	0,17	162655	
	2 nd (R6971)	02/10/2009	FQ resistant	gyrA	GCG90GTG	68,18	136528	
	3 rd (R7515)	27/10/2009	FQ resistant	gyrA	GCG90GTG	86,14	81089	
	4 th (R7764)	27/10/2009	FQ susceptible	gyrA	GCG90GTG	98,63	87286	
3	1 st (R8569)	20/01/2010		Not targeted deep sequenced				
	2 nd (R10499)	29/07/2010	FQ susceptible	gyrA	GCG90GTG	99.91	99420	
	3 rd (R11193)	23/09/2010	FQ resistant	gyrA	GCG90GTG	98,6	134035	
				gyrA	GAC94GCC	0,07	135026	
	4 th (R11603)	25/10/2010	FQ susceptible	gyrA	GGC88TGC	0,54	97909	
				gyrA	GAC89AAC	0,96	97710	
				gyrA	GCG90GTG	41.59	96862	
				gyrA	TCG91CCG	9.07	96674	
				gyrA	GAC94CAC	0,27	97688	
				gyrA	GAC94GCC	1.33	96490	

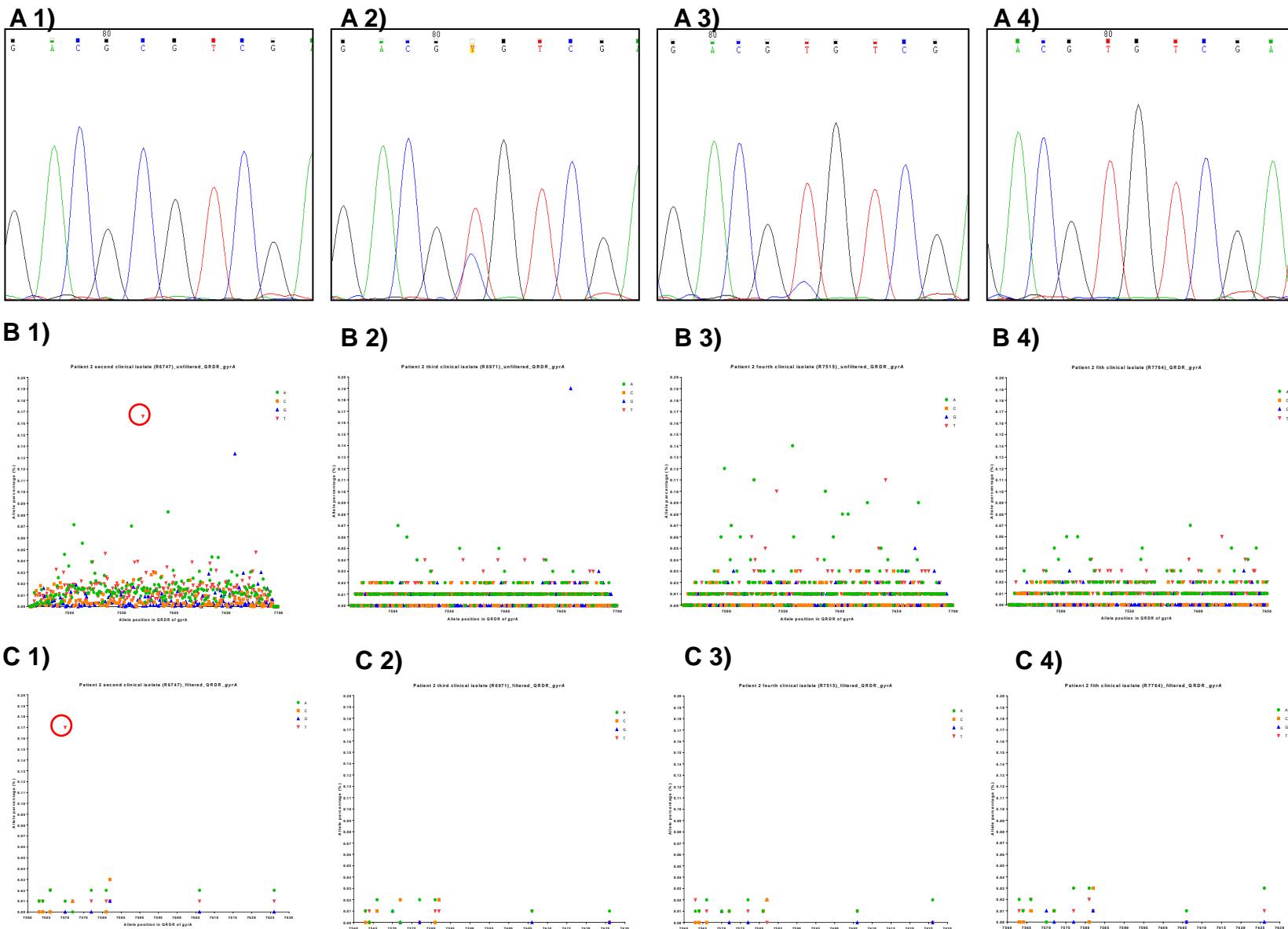


Figure 10 A comparison of results of (A) Sanger sequencing with (B and C) Illumina targeted deep sequencing of clinical isolates from Patient 2. A) Emergence of resistance (GCG90GTG) observed by the presence of underlying mutant peaks within the serial isolates (A 3). B1 to B4) Unfiltered targeted deep sequencing results for the QRDR of *gyrA* showing the presence of underlying variants, as well as the sequencing errors. C1 to C4) Filtered targeted deep sequencing results for the QRDR of *gyrA* that only shows the regions, which are known to contain resistant conferring SNPs. An underlying GCG90GTG variant can be observed at 17.09% (red circle). X-axis allele position in the QRDR of *gyrA*. Y-axis allele percentage in the QRDR of *gyrA*.

Clinical *M. tuberculosis* isolates from the OptiQ clinical trial of patients receiving FQ at different concentrations

We have applied the established Illumina MiSeq deep sequencing methodology to clinical isolates of patients receiving FQ at different concentrations in a clinical trial (OptiQ study) to determine whether resistance was present at baseline or develops during treatment. To achieve this, the first available (i.e. base line) isolate and the last available isolate (8 or 16 weeks or 24 weeks) were selected for targeted deep sequencing. We were able to detect underlying variants GAC94GCC in the QRDR of *gyrA* within all the isolates (base line and follow-up) at 0.1% with an error rate of (0.1%) (**Table 9**), however no increase in frequency was observed for this variant in the follow-up isolates. No additional variants were observed in the follow-up isolates.

Table 9 Targeted deep sequencing results of the clinical *M. tuberculosis* isolates from the OptiQ study.

Study	Patient	Isolate designation	Date	Gene name	Codon change	Mutant allele frequency (%)	Depth of coverage (x) of base called
OptiQ study	Patient 1	Baseline	27/01/2016	<i>gyrA</i>	GAC94GCC	0.1%	38193
		Week 8	03/02/2016	<i>gyrA</i>	GAC94GCC	0.09%	58606
	Patient 2	Baseline	01/06/2015	<i>gyrA</i>	GAC94GCC	0.06%	53517
		Week 8	22/08/2015	<i>gyrA</i>	GAC94GCC	0.06%	58945
		Week 10	06/09/2015	<i>gyrA</i>	GAC94GCC	0.08%	78944
	Patient 3	Baseline	08/03/2015	<i>gyrA</i>	GAC94GCC	0.07%	33735
		Week 16	01/07/2015	<i>gyrA</i>	GAC94GCC	0.1%	105641
	Patient 4	Baseline	25/05/2016	<i>gyrA</i>	GAC94GCC	0.1%	630140
		Week 8	13/08/2016	<i>gyrA</i>	GAC94GCC	0.07%	64225
	Patient 5	Baseline	24/02/2015	<i>gyrA</i>	GAC94GCC	0.09%	69883
		Week 8	14/05/2015	<i>gyrA</i>	GAC94GCC	0.09%	60864

Downsampling

Downsampling was done to determine the minimum coverage required per sample to confidently detect 0.5% underlying mutant subpopulations. This was done to see if it is possible to make targeted deep sequencing a more economical approach by downsampling the depth of coverage while still ensuring detection of underlying mutant subpopulations. Detection of the underlying variant (GAC94GGC) was possible at all the various read coverages examined (from 1891328 to 6250) with an average error rate of 0.41% (**Table 10**).

Table 10 Variants called at various coverages of 0.5% synthetic heterogeneous mix.

Study	Origin	Description	Number of reads	Variant called (GAC94GGC)	Coverage	Variant frequency
Synthetic heterogeneous mix	Synthetically created heteroresistant mixture (0.5% mutant)	Raw read	472832	720	176022	0.41%
Downsampling		100000 times coverage	100000	153	37231	0.410%
		50000 times coverage	50000	78	18616	0.421%
		25000 times coverage	25000	38	9314	0.410%
		12500 times coverage	12500	18	4659	0.396%
		6250 times coverage	6250	11	2324	0.387%

Chapter 5: Discussion

In this study we explored the feasibility of using next generation targeted deep sequencing to identify underlying variants conferring FQ resistance. We found that targeted deep sequencing together with SMOR analyses is a more sensitive method to detect heteroresistance compared to the traditional diagnostic methods. Previous studies have shown that the GenoType MTBDR*plus* assay could only detect heteroresistant subpopulations if more than 5% of the resistant bacteria are present (Folkvardsen, Svensson, *et al.*, 2013; Folkvardsen, Thomsen, *et al.*, 2013), while Xpert MTB/RIF could only detect rifampicin resistance if more than 60% of the resistant bacteria are present (Blakemore *et al.*, 2010; Engström *et al.*, 2013; Folkvardsen, Thomsen, *et al.*, 2013). Multiple studies have also showed that the detection limit of Sanger sequencing to identify of underlying variants is 10% (Folkvardsen, Svensson, *et al.*, 2013; Eilertson *et al.*, 2014; Colman *et al.*, 2015). The current study showed that when sequencing at a very high (approximately 100000 times) depth of coverage, underlying *M. tuberculosis* drug resistance conferring variants could be detected at frequencies as low as 0.5%. This approach could be applied in a clinical setting for early detection and monitoring of antibiotic resistant subpopulations, which could then potentially prevent selection of the resistant population and/or treatment failure that may occur if otherwise left undetected. However, when the list of variants produced by SMOR is filtered for only the known FQ resistance causing SNPs, we were able to identify variants at an underlying frequency of 0.1%. We were able to reproduce this sensitivity irrespective of the purity of the DNA sample, as this variant frequency was identified in purified DNA samples, crude DNA samples extracted from heat inactivated MGIT cultures as well as crude DNA extracted from sputum specimens. These findings confirmed our hypothesis that targeted deep sequencing can be used to identify underlying drug resistance causing variants in *M. tuberculosis* from clinical isolates.

The highly optimised amplification-based approach using highly efficient primer sets for specific target regions made it feasible to detect underlying variants in samples with a low number of *M. tuberculosis* DNA copies, which is often the case in clinical samples (Colman *et al.*, 2015). It is very important for the targets of the assay to be specific when used in a clinical setting, especially when using DNA directly extracted from sputum specimens (e.g. genolysed samples) which can also contain human DNA as well as DNA from other organisms. The primers used in the assay were found to be specific across a database of sequences and has a high PCR efficiency to detect the presence of *M. tuberculosis*.

Underlying variants present at a variant frequency of more than 6% in the synthetic heterogeneous mixtures as well as in some of the clinical isolates used in this study could be detected by Sanger sequencing. This finding is in agreement with previous studies (Folkvardsen, Svensson, *et al.*, 2013; Zhang *et al.*, 2014).

When comparing the sensitivity of the Ion Torrent and Illumina MiSeq platforms for the detection of underlying variants in *gyrA* and *gyrB* using synthetic heterogeneous mixtures, we found that the Illumina MiSeq targeted deep sequencing was able to detect variants below a frequency of 1%. In contrast, targeted deep sequencing done by the Ion Torrent platform could only detect variants with a frequency of more than 2%. This is in agreement with a previous study that showed that Illumina is more suitable for detecting very low frequencies of variants than other platforms (McElroy *et al.*, 2014). This may be due to the high sequencing error of Ion Torrent which limited our ability to identify underlying low frequency variants with confidence (Quail *et al.*, 2012). In addition, the analysis of sequencing data produced by Ion torrent was problematic as one is confined to only use the Ion Torrent Suite which also has a slow processing time compared to the open source tools that are available for the analysis of Illumina sequencing data.

Our analysis of the serial patient isolates progressing from MDR-TB to pre-XDR-TB showed that two different *gyrA* mutations were acquired in patient one after two months compared to the initial FQ susceptible isolate. The follow up isolate also showed the presence of a *gyrB* GAC461AAC mutation at 0.16% that is known to confer FQ drug resistance (Coll *et al.*, 2015). This short time frame suggests that the bacteria are able to rapidly evolve when under a selective pressure (antibiotic treatment). An alternative reason may be that isolate 2 was sampled from an alternative lesion.

Furthermore, we were able to show the presence of a low frequency FQ resistance conferring mutation (0.17%) in the first isolate of patient 2. The presence of the underlying FQ resistant population was however missed by standard DST due to the detection limit of the diagnostic tool. The frequency of this underlying mutant variant was found to substantially increase over time in the follow up isolates (**Figure 10** or **Table 8**), which provided insight in the evolution of drug resistance among serial clinical *M. tuberculosis* isolates.

Conversely, we were able to show in patient 3 the decrease of a high frequency mutation over time, with the acquisition of additional FQ resistance conferring mutations in the follow up isolates. This is in agreement with a previous study that found that heteroresistance allows for alteration of strain dominance according to the type of external threats that *M. tuberculosis* strains are exposed to (Streicher, Bergval, *et al.*, 2012). Current TB treatment rely on treating TB infected patients with multiple drugs. This may efficiently suppress low level resistant causing variants present within a patient over time. Alternatively, combination therapy can compensate for each other when one drug is lost as a result of acquisition of resistance to it, or combination therapy can kill off clones containing the resistant mutant, thereby managing the TB burden and possibly prevent transmission of resistant strains.

Discrepancies between targeted deep sequencing and phenotypic DST were found for the fourth isolate of patient 2 as well as in the second - and fourth isolate of patient 3 (**Table 8**). This may be

attributed to incorrect phenotypic DST possibly due to the low level resistance that the mutation(s) cause with a minimum inhibitory concentration (MIC) very close to the critical concentration, as targeted deep sequencing shows the presence of the variants at very high frequencies (98.63% and 99.91%, respectively).

The clinical *M. tuberculosis* isolates from the OptiQ study remained FQ susceptible, however shows underlying variants in codon 94 of *gyrA* (GAC94GGC) for all the clinical isolates (collected over a period of 8 weeks or 10 weeks or 16 weeks). This may be indicative of underlying FQ resistance in persisting *M. tuberculosis* strains that may slowly evolve to FQ resistant *M. tuberculosis* isolates as mutations found in codon 94 of *gyrA* gene is known to confer FQ resistance (Farhat *et al.*, 2013).

To compensate for the high cost of targeted deep sequencing, down sampling of the high coverage of the raw reads of 0.5% synthetic heterogeneous mutant were done to determine a more economical approach for detection of low frequency variants. We found that a minimum coverage of 6250 times is adequate to detect 0.5% underlying frequency variant. This is in agreement with Colman *et al.* (2015) that suggest a minimum of 1000x of paired reads (2000x total coverage) is needed in order to detect a 1% subpopulation (10 pairs of reads representing the minor resistant subpopulation) (Colman *et al.*, 2015).

The benefit of next generation targeted deep sequencing is that it is potentially a much faster diagnostic assay, once the workflow has been established, compared to culture diagnostic method that requires a minimum of 4 to 6 weeks (Tolani *et al.*, 2012). Furthermore, next generation targeted deep sequencing provides a resistance profile as well as quantitative analysis of variants present within *M. tuberculosis* populations. The ratio of resistant to susceptible alleles allows for intensive examination of the emergence of low-level genetic resistance during the course of treatment. The Illumina platform was also found to be cost effective and more reliable to use than the Ion torrent platform for targeted deep sequencing as it has a lower error rate (e.g. 0.2% compared to Ion torrent (e.g. 0.5%) which needs to be considered especially when one wants to detect low frequency underlying variants with confidence.

Several limitations were identified in this study. Firstly, not all the mutations conferring FQ resistance are known. Secondly, different mutations confers resistance to a single anti-TB drug (e.g. FQs), necessitating sequencing of more than one genomic region. Thirdly, rare variants can become lost or altered during the DNA extraction procedure, sample transport, sample decontamination procedure or may be altered during culturing. Fourthly, the detection of heteroresistance also depends on the sensitivity of the molecular technique used in a study, since the current molecular techniques have variable detection limits for detection of low frequency underlying genomic mutations conferring drug resistance. Furthermore, NGS-based diagnostic methods are also costly, resource dependant and labour intensive and require highly advanced

technology and expertise. Lastly, targeted deep sequencing platforms possess intrinsic sequencing error rates, possibly obscuring true variants.

In future studies, we want to expand the study using a larger sample set and investigate more drug resistance conferring genes, which will provide us with more comprehensive information on the prevalence of heteroresistance. Competition assays using different FQ resistant mutants at different ratios can be done *in vitro* to determine the fitness cost in relation to the initial inoculum of both mutant isolates. Furthermore, the growth rate of minority variants can be determined by studying the growth of underlying variants on different growth media, as different growth media result in variation in growth rates which would affect phenotypic resistance tests (Rinder *et al.*, 2001). Alternatively, multiple FQ susceptible isolates can be sequenced to determine the sequencing error rate at each position of the target to increase our ability to distinguish between low frequency variants (<0.2%) and sequencing error.

In summary, next generation targeted deep sequencing is a potential emerging diagnostic method that allows for sample and target multiplexing. Thus, allowing for simultaneous detection and characterisation of numerous genetic targets within a short time frame (approximately 4 weeks, when considering all the steps from DNA isolation, PCR amplification, library preparation, deep sequencing and data analysis), compared to culture (i.e. 6 to 8 weeks). Furthermore, next generation targeted deep sequencing which involves amplicon sequencing can be applied to low levels of DNA concentration (e.g. serial dilution) and is able to detect very low frequencies of genetic variant subpopulations at very high coverage (Colman *et al.*, 2015).

Chapter 6: Conclusion

Targeted deep sequencing was used to detect underlying variants within the *M. tuberculosis* fluoroquinolone resistance causing genes (*gyrA* and *gyrB*). Using synthetic heterogeneous mixtures we were able to establish a targeted deep sequencing method that could accurately detect the correct proportion of mutant DNA. The Illumina MiSeq platform were shown to be superior to the Ion Torrent PGM platform as it accurately reflected the known proportion of mutant DNA to a minimum frequency of 0.1% when using the analysis pipeline as described by Colman *et al.* (2015) (Colman *et al.*, 2015). We also found that a minimum coverage of 6250 time is adequate to detect low frequency underlying variant, which were in agreement with previous findings (Colman *et al.*, 2015). From this study we have seen that NGS targeted deep sequencing can confidently diagnose underlying variants present at very low frequencies in patient samples that were deemed FQ sensitive by standard DST. These minority variants were consequently found to be fixed in the follow up patient isolates, suggesting the evolution and the subsequent selection of the FQ resistant population during the course of treatment. Having this knowledge may guide prescription of adequate treatment to patients and prevent the amplification of drug resistance. In contrast FQ susceptible (as determined by standard DST) patient isolates from a clinical trial showed constant low frequency variants (<0.1%) in the baseline as well as follow up isolates. This could possibly be due to the predicted error rate of the Illumina MiSeq sequencing platform or effect of combination treatment that effectively suppresses the minority FQ resistant population. This method also allowed for the identification of variants in *gyrA* with fluctuating frequencies during the evolution of FQ resistance, indicating the dynamic mechanisms for the development of FQ resistance. In summary the aims of this study were successfully achieved as an ultra-sensitive method to detect underlying resistance causing mutations using a targeted deep sequencing approach was developed.

Appendices

Appendix A:

Optimisation protocol

Primer sequences were tested to find optimal primer sets for both, the QRDR of *gyrA* and - *gyrB* genes. Two primer sets were used in this study (methods **Table 2**).

Table A 1 Primer (Oligonucleotide) sequences of the QRDR of *gyrA* and - *gyrB* genes.

Label	Gene	Position	Primer sequences (5' – 3')	Size (bp)	Tm (° C)	GC	Product size (bp)	Reference
<i>gyrA_Option1_F</i>	<i>gyrA</i>	Forward	ccggtcgggtgccgagacc	19	70.6	73.7	107	(Chakravorty <i>et al.</i> , 2011)
<i>gyrA_Option1_R</i>		Reverse	ccagcgggtagcgcagcgaccag	23	61.8	60.0		
<i>gyrA_Option2_F</i>	<i>gyrA</i>	Forward	ggtgcttatgcaatgttcg	20	59.3	50.0	236	
<i>gyrA_Option2_R</i>		Reverse	gggcttcggtgtacacctcat	19	59.9	57.9		
<i>gyrA_Colman_F</i>	<i>gyrA</i>	Forward	gggtgcttatgcaatgttcgat	23	64.5	47.8	237	(Colman <i>et al.</i> , 2015)
<i>gyrA_Colman_R</i>		Reverse	gggcttcggtgtacacctcatc	20	61.8	60.0		
<i>gyrA_Option3_F</i>	<i>gyrA</i>	Forward	ggtgcttatgcaatgttcg	20	59.3	50.0	272	
<i>gyrA_Option3_R</i>		Reverse	tttccctcagcatctccatc	20	60.2	50.0		
<i>gyrA_Option4_F</i>	<i>gyrA</i>	Forward	tgacatcgagcaggagatgc	20	62	55	344	
<i>gyrA_Option4_R</i>		Reverse	gggcttcggtgtacacctcatc	20	62	60		
<i>gyrB_Option1_F</i>	<i>gyrB</i>	Forward	accgacatcggtggattg	18	60.3	55.6	284	
<i>gyrB_Option1_R</i>		Reverse	atcttgtggtagcgcagctt	20	60.0	50.0		
<i>gyrB_Option2_F</i>	<i>gyrB</i>	Forward	cgttaaggcacgagagtgg	20	60.3	55.0	320	
<i>gyrB_Option2_R</i>		Reverse	atcttgtggtagcgcagctt	20	60.0	50.0		

<i>gyrB</i> _Option3_F	<i>gyrB</i>	Forward	accgacatcggtggattg	18	60.3	55.6	261	
<i>gyrB</i> _Option3_R		Reverse	gatatcgaactcgtcgtggat	21	59.0	47.6		
<i>gyrB</i> _Option4_F	<i>gyrB</i>	Forward	aagagcgccaccgacatc	18	62.18	61.11	300	
<i>gyrB</i> _Option4_R		Reverse	cagcacgatcttgtggtagc	20	62.45	55		

Additional optimisation of different combinations of the forward and reverse primer sequences for the QRDR region of *gyrB* were tested, to determine the most specific and sensitive primer set (**Appendix Table A 2**).

Table A 2 Different combinations of the forward and reverse primer sequences for the QRDR region of *gyrB*.

Name of primer set	Label	Position	Primer sequence	Average Tm (° C)	Start	product size
1	<i>gyrB</i> _Option1_F	Forward	accgacatcggtggattg	59,9	1414	291
	<i>gyrB</i> _Option4_R	Reverse	cagcacgatcttgtggtagc		1704	
2	<i>gyrB</i> _Option1_F	Forward	accgacatcggtggattg	59,65	1414	261
	<i>gyrB</i> _Option3_R	Reverse	gatatcgaactcgtcgtggat		1674	
3	<i>gyrB</i> _Option2_F	Forward	cgtaaggcacgagagttggt	59,9	1378	327
	<i>gyrB</i> _Option4_R	Reverse	cagcacgatcttgtggtagc		1704	
4	<i>gyrB</i> _Option2_F	Forward	cgtaaggcacgagagttggt	59,65	1378	299
	<i>gyrB</i> _Option3_R	Reverse	gatatcgaactcgtcgtggat		1674	
5	<i>gyrB</i> _Option4_F	Forward	aagagcgcaccgacatc	60,95	1405	300
	<i>gyrB</i> _Option4_R	Reverse	cagcacgatcttgtggtagc		1704	
6	<i>gyrB</i> _Option4_F	Forward	aagagcgcaccgacatc	61,09	1405	295
	<i>gyrB</i> _Option1_R	Reverse	atcttgtggtagcgcagctt		1697	
7	<i>gyrB</i> _Option4_F	Forward	aagagcgcaccgacatc	60,7	1405	270
	<i>gyrB</i> _Option3_R	Reverse	gatatcgaactcgtcgtggat		1674	

PCR primers (Table 1 and Table 2) were first optimized in monoplex PCR reactions. A PCR protocol adapted from Colman *et al.*, 2015 (Table 4) were used in the combination of different PCR parameters (Table 3). This included testing different magnesium chloride concentrations as well in the presence or absence of Betaine and DMSO. Different PCR conditions including different melting temperatures, cycle numbers and elongation times were also tested (Table 3). The following parameters were selected finally after several trials to find the most optimum method to amplify the regions of interest (Table 5).

Table A 3 Parameters tested to obtain an optimal PCR reaction.

	Parameters
Melting temperature (°C)	55 – 65
Magnesium chloride (MgCl ₂)	1.5 mM, 2 mM, 3 mM
Betaine solution	With and without
Dimethyl sulfoxide (DMSO)	With and without
Number of cycles	25 – 45
Elongation time (sec)	10 - 40

Table A 4 PCR parameters of Colman *et al.*, 2015.

Q5 High-Fidelity 2X Master Mix with Betaine solution	
Components	25 µl Reaction
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µl
10 µM Forward Primer	1.25 µl
10 µM Reverse Primer	1.25 µl
5 M Betaine solution (Sigma-Aldrich)	5 µl
Template DNA	2 µl
Nuclease-Free Water	to 25 µl

Table A 5 Optimised PCR parameters for this study, excluding Betaine solution.

Q5 High-Fidelity 2X Master Mix without Betaine solution	
Components	25 µl Reaction
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µl
10 µM Forward Primer	0.05 µl
10 µM Reverse Primer	0.05 µl
Template DNA	1 µl
Nuclease-Free Water	11.4 (to 25 µl)

Table A 6 PCR thermo-cycling conditions for pure DNA.

Thermo-cycling conditions for PCR	
Initial denaturation	98°C for 1 min
Denaturation	98°C for 10 sec
Annealing	60°C for 15 sec
Extension	72°C for 2 min
Number of cycles	35

Table A 7 PCR thermo-cycling conditions for genolysed extracted DNA.

Thermo-cycling conditions for PCR	
Initial denaturation	98°C for 1 min
Denaturation	98°C for 10 sec
Annealing	62°C for 15 sec
Extension	72°C for 2 min
Number of cycles	45

Appendix B:

In vitro mixtures

The isolated DNA for both H37Rv and BO1 was pre-heated at 65 °C for 10 minutes. DNA concentrations were determined using the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). A DNA concentration of 2051.033 ng/ul was determined for the H37Rv while, a DNA concentration of 1529.567 ug/ml was determined for BO1. Due to the high concentration of the H37Rv DNA, the DNA was firstly diluted to a lower concentration (683.677 ug/ml) by mixing 600ul of H37Rv DNA with 1200 ul TE buffer. The actual concentration of this mixture was then confirmed with the NanoDrop spectrophotometer to be 701.3 ng/ul. For BO1, a 2:1 dilution in TE buffer was first made, resulting in a concentration as determined by the Nanodrop of 774 ng/ul. Subsequently, this concentration was further diluted to a final concentration of 200 ug/ml. A dilution series ranging from 100% (5 ul mutant DNA in the background of 0 ul wildtype DNA) and 0% (0 ul mutant DNA in the background of 5 ul wildtype DNA) was then made using the diluted DNA samples of H37Rv and BO1. Certain pre-calculated volumes, dependent on the DNA concentration, of each DNA sample was used to create the different mutant to wild type ratios. For this purpose a constant volume of BO1 DNA was added to different volume of H37Rv DNA. A nanodrop reading was taken of each mixture and diluted to 25 ug/ml needed for the PCR reactions.

Appendix C:

Script used to down sample coverage of sequenced data (fastq file) is available on GitHub (https://github.com/Ncite/Down_sampiling/blob/master/downsample_script.sh).

References

- Al-Mutairi, N.M., Ahmad, S. & 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.
- Behr, M.A. 2004. Tuberculosis due to multiple strains: a concern for tuberculosis control? *American Journal of Respiratory and Critical Care Medicine*. 169(5):553–554.
- Black, P.A., Vos, M. De, Louw, G.E., Merwe, R.G. Van Der, Dippenaar, A., Streicher, E.M., Abdallah, A.M. & Sampson, S.L. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genomics*. 1–14.
- Blakemore, R., Story, E., Helb, D., Kop, J., Banada, P., Owens, M.R., Chakravorty, S., Jones, M., et al. 2010. Evaluation of the Analytical Performance of the Xpert MTB/RIF Assay. *Journal of Clinical Microbiology*. 48(7):2495–2501.
- Boehme, C.C., Nicol, M.P., Nabeta, P., Michael, J.S., Gotuzzo, E., Tahirli, R., Gler, M.T., Blakemore, R., et al. 2011. Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: A multicentre implementation study. *The Lancet*. 377(9776):1495–1505.
- Bryant, J.M., Harris, S.R., Parkhill, J., Dawson, R., Diacon, A.H., van Helden, P., Pym, A., Mahayiddin, A.A., et al. 2013. Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *The Lancet. Respiratory medicine*. 1(10):786–792.
- Campbell, P.J., Morlock, G.P., Sikes, R.D., Dalton, T.L., Metchock, B., Starks, A.M., Hooks, D.P., Cowan, L.S., et al. 2011. Molecular Detection of Mutations Associated with First- and Second-Line Drug Resistance Compared with Conventional Drug Susceptibility Testing of *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*. 55(5):2032–2041.
- Cegielski, J.P., Dalton, T., Yagui, M., Wattanaamornkiet, W., Volchenkov, G. V., Via, L.E., Van Der Walt, M., Tupasi, T., et al. 2014. Extensive Drug Resistance Acquired During Treatment of Multidrug-Resistant Tuberculosis. *Clinical Infectious Diseases*. 59(8):1049–1063.
- Chakravorty, S., Aladegbami, B., Thoms, K., Lee, J.S., Lee, E.G., Rajan, V., Cho, E.J., Kim, H., et al. 2011. Rapid detection of fluoroquinolone-resistant and heteroresistant *Mycobacterium tuberculosis* by use of sloppy molecular beacons and dual melting-temperature codes in a real-time PCR assay. *Journal of Clinical Microbiology*. 49(3):932–940.
- Chatterjee, M., Bhattacharya, S., Karak, K. & Dastidar, S.G. 2013. Effects of different methods of decontamination for successful cultivation of *Mycobacterium tuberculosis*. *Indian Journal of Medical Research*. 138(OCT):541–548.

Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L. & Rice, P.M. 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*. 38(6):1767–1771.

Cohen, T., Chindelevitch, L., Misra, R., Kempner, M.E., Galea, J., Moodley, P. & Wilson, D. 2016. Within-Host Heterogeneity of *Mycobacterium tuberculosis* Infection Is Associated With Poor Early Treatment Response A Prospective Cohort Study. *The Journal of infectious diseases*. 213:2005–2011.

Coll, F., McNerney, R., Preston, M.D., Guerra-Assunção, J.A., Warry, A., Hill-Cawthorne, G., Mallard, K., Nair, M., et al. 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome medicine*. 7(1):51.

Colman, R.E., Schupp, J.M., Hicks, N.D., Smith, D.E., Buchhagen, L., Valafar, F., Crudu, V. & Romancenco, E. 2015. Detection of Low-Level Mixed-Population Drug Resistance in *Mycobacterium tuberculosis* Using High Fidelity Amplicon Sequencing. 1–18.

Colman, R.E., Anderson, J., Lemmer, D., Lehmkuhl, E., Georghiou, S.B., Heaton, H., Wiggins, K., Gillece, J.D., et al. 2016. Rapid Drug Susceptibility Testing of Drug Resistant *Mycobacterium tuberculosis* Directly from Clinical Samples using Amplicon Sequencing: A Proof of Concept Study. *Journal of Clinical Microbiology*. (May):JCM.00535–16.

Consortium, W.W.W. 1999. HTML 4.01 specification.

DePristo, M.A., Banks, E., Poplin, R.E., Garimella, K. V, Maguire, J.R., Hartl, C., Rivas, M.A., Hanna, M., et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. 43(5):491–498.

Eilertson, B., Maruri, F., Blackman, A., Herrera, M., Samuels, D.C. & Sterling, T.R. 2014. High proportion of heteroresistance in gyra and gyrb in fluoroquinolone-resistant *mycobacterium tuberculosis* clinical isolates. *Antimicrobial Agents and Chemotherapy*. 58(6):3270–3275.

Eldholm, V., Norheim, G., von der Lippe, B., Kinander, W., Dahle, U.R., Caugant, D. a, Mannsåker, T., Mengshoel, A.T., et al. 2014. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biology*. 15(11):490.

El-Halfawy, O.M. & Valvano, M.A. 2015. Antimicrobial heteroresistance: An emerging field in need of clarity. *Clinical Microbiology Reviews*. 28(1):191–207.

Engström, A., Hoffner, S. & Juréen, P. 2013. Detection of heteroresistant *Mycobacterium tuberculosis* by pyrosequencing. *Journal of Clinical Microbiology*. 51(12):4210–4212.

Farhat, M.R., Shapiro, B.J., Kieser, K.J., Sultana, R., Jacobson, K.R., Victor, T.C., Warren, R.M., Streicher, E.M., et al. 2013. Genomic analysis identifies targets of convergent positive selection in

drug-resistant *Mycobacterium tuberculosis*. *Nature genetics*. 45(10):1183–1189.

Folkvardsen, D.B., Thomsen, V.O., Rigouts, L., Rasmussen, E.M., Bang, D., Bernaerts, G., Werngren, J., Toro, J.C., et al. 2013. Rifampin heteroresistance in *Mycobacterium tuberculosis* Cultures as Detected by phenotypic and genotypic drug susceptibility test methods. *Journal of Clinical Microbiology*. 51(12):4220–4222.

Folkvardsen, D.B., Svensson, E., Thomsen, V., Rasmussen, E.M., Bang, D., Werngren, J., Hoffner, S., Hillemann, D., et al. 2013. Can molecular methods detect 1% isoniazid resistance in *mycobacterium tuberculosis*? *Journal of Clinical Microbiology*. 51(5):1596–1599.

Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodkin, E., Rempel, S., Moore, R., et al. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *The New England journal of medicine*. 364(8):730–739.

Hofmann-Thiel, S., Van Ingen, J., Feldmann, K., Turaev, L., Uzakova, G.T., Murmusaeva, G., Van Soolingen, D. & Hoffmann, H. 2009. Mechanisms of heteroresistance to isoniazid and rifampin of *Mycobacterium tuberculosis* in Tashkent, Uzbekistan. *European Respiratory Journal*. 33(2):368–374.

Hughes, J., Cardiff, M.B.B., Osman, M., Chb, M.B., Dip, P.G. & Management, H. 2014. Diagnosis and management of drug-resistant tuberculosis in South African adults. 104(12).

Illumina. 2016. [Online], Available: <http://www.illumina.com/technology/next-generation-sequencing/deep-sequencing.html> [2016, January 15].

Ioerger, T.R., O'Malley, T., Liao, R., Guinn, K.M., Hickey, M.J., Mohaideen, N., Murphy, K.C., Boshoff, H.I., et al. 2013. Identification of new drug targets and resistance mechanisms in *Mycobacterium tuberculosis*. *PLoS one*. 8(9):e75245.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. 25(16):2078–2079.

Marlowe, E.M., Novak-Weekley, S.M., Cumpio, J., Sharp, S.E., Momeny, M.A., Babst, A., Carlson, J.S., Kawamura, M., et al. 2011. Evaluation of the cepheid xpert MTB/RIF assay for direct detection of *mycobacterium tuberculosis* complex in respiratory specimens. *Journal of Clinical Microbiology*. 49(4):1621–1623.

Mayer, C. & Takiff, H. 2014. The Molecular Genetics of Fluoroquinolone Resistance in *Mycobacterium tuberculosis*. *Microbiology Spectrum*. 2(4):1–22.

McElroy, K., Thomas, T. & Luciani, F. 2014. Deep sequencing of evolving pathogen populations: applications, errors, and bioinformatic solutions. *Microbial informatics and experimentation*. 4(1):1.

McGrath, M; Warren, RM; van Pittius, CG; Sampson, S. 2016. Evolution of XDR-TB and the associated proteome. Stellenbosch University.

Meacci, F., Orrù, G., Iona, E., Piersimoni, C., Pozzi, G., Oggioni, M.R., Orru, G. & Giannoni, F. 2005. Drug Resistance Evolution of a *Mycobacterium tuberculosis* Strain from a Noncompliant Patient Drug Resistance Evolution of a *Mycobacterium tuberculosis* Strain from a Noncompliant Patient. 43(7):3114–3120.

Mekonnen, D., Admassu, A., Mulu, W., Amor, A., Benito, A., Gelaye, W., Biadglegne, F. & Abera, B. 2015. Multidrug-resistant and heteroresistant *Mycobacterium tuberculosis* and associated gene mutations in Ethiopia. *International Journal of Infectious Diseases*. 39:34–38.

Merker, M., Kohl, T.A., Roetzer, A., Truebe, L., Richter, E., Rusch-Gerdes, S., Fattorini, L., Oggioni, M.R., et al. 2013. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PloS one*. 8(12):e82551.

Miotto P.; Cirillo D.M.; Migliori B; 2015. Drug Resistance in *Mycobacterium tuberculosis*. *Antimicrobial Resistance in Developing Countries*. 117.

NCBI. 1997. Standard Nucleotide BLAST. [Online], Available: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch [2015, May 26].

New England Biolabs Inc. n.d. Q5® Hot Start High-Fidelity 2X Master Mix. [Online], Available: <https://www.neb.com/products/m0494-q5-hot-start-high-fidelity-2x-master-mix> [2004, May 20].

Pantel, A., Petrella, S., Veziris, N., Brossier, F., Bastian, S., Jarlier, V., Mayer, C. & Aubry, A. 2012. Extending the definition of the GyrB quinolone resistance-determining region in *Mycobacterium tuberculosis* DNA gyrase for assessing fluoroquinolone resistance in *M. tuberculosis*. *Antimicrobial Agents and Chemotherapy*. 56(4):1990–1996.

Pholwat, S., Stroup, S., Foongladda, S. & Houpt, E. 2013. Digital PCR to detect and quantify heteroresistance in drug resistant *Mycobacterium tuberculosis*. *PloS one*. 8(2):e57238.

Quail, M.A., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., et al. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics*. 13(1):341.

Rahman, A., Sahrin, M., Afrin, S., Earley, K. & Ahmed, S. 2016. Comparison of Xpert MTB / RIF Assay and GenoType MTBDR plus DNA Probes for Detection of Mutations Associated with Rifampicin Resistance in *Mycobacterium tuberculosis*. 1–11.

Rhoads, A. & Au, K.F. 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*. 13(5):278–289.

van Rie, A., Victor, T.C., Richardson, M., Johnson, R., van der Spuy, G.D., Murray, E.J., Beyers,

- N., van Pittius, N.C.G., et al. 2005. Reinfection and Mixed Infection Cause Changing *Mycobacterium tuberculosis* Drug-Resistance Patterns. *American Journal of Respiratory and Critical Care Medicine.* 172(5):636–642.
- van Rie, A., Mellet, K., John, M., Scott, L., Dansey, H. & Victor, T. 2013. False-positive rifampicin resistance on Xpert® MTB / RIF : case report and clinical implications. *Int J Tuberc Lung Dis.* 16(2):206–208.
- Rinder, H. 2001. Hetero-resistance: An under-recognised confounder in diagnosis and therapy? *Journal of Medical Microbiology.* 50(12):1018–1020.
- Rinder, H., Mieskes, K.T. & Löscher, T. 2001. Heteroresistance in *Mycobacterium tuberculosis*. *International Journal of Tuberculosis and Lung Disease.* 5(4):339–345.
- Roetzer, A., Diel, R., Kohl, T.A., Ruckert, C., Nubel, U., Blom, J., Wirth, T., Jaenicke, S., et al. 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS medicine.* 10(2):e1001387.
- Schurh, A.C., Kremer, K., Daviena, O., Kiers, A., Boeree, M.J., Siezen, R.J. & van Soolingen, D. 2010. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster. *Journal of clinical microbiology.* 48(9):3403–3406.
- Schurh, A.C., Kremer, K., Kiers, A., Daviena, O., Boeree, M.J., Siezen, R.J., Smith, N.H. & van Soolingen, D. 2010. The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases.* 10(1):108–114.
- Streicher, E.M., Bergval, I., Dheda, K., Bottger, E.C., Gey van Pittius, N.C., Bosman, M., Coetzee, G., Anthony, R.M., et al. 2012. *Mycobacterium tuberculosis* population structure determines the outcome of genetics-based second-line drug resistance testing. *Antimicrobial Agents and Chemotherapy.* 56(5):2420–2427.
- Streicher, E.M., Muller, B., Chihota, V., Mlambo, C., Tait, M., Pillay, M., Trollip, A., Hoek, K.G., et al. 2012. Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases.* 12(4):686–694.
- Sun, G., Luo, T., Yang, C., Dong, X., Li, J., Zhu, Y., Zheng, H., Tian, W., et al. 2012. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *The Journal of infectious diseases.* 206(11):1724–1733.
- ThermoFisherScientific. 2013. *Torrent Suite Software 4.0 User Guide.* [Online], Available: <https://ioncommunity.thermofisher.com/docs/DOC-9589> [2016, February 12].

- Theron, G., Peter, J., Van Zyl-Smit, R., Mishra, H., Streicher, E., Murray, S., Dawson, R., Whitelaw, A., et al. 2011. Evaluation of the Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in a high HIV prevalence setting. *American Journal of Respiratory and Critical Care Medicine*. 184(1):132–140.
- Tolani, M.P., D'Souza D, T. & Mistry, N.F. 2012. Drug resistance mutations and heteroresistance detected using the GenoType MTBDRplus assay and their implication for treatment outcomes in patients from Mumbai, India. *BMC Infect Dis*. 12(1):9.
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J.A. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research*. 35(Web Server issue):W71–4.
- Walker, T.M., Ip, C.L., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., et al. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet infectious diseases*. 13(2):137–146.
- Wang, J.Y., Lee, L.N., Lai, H.C., Wang, S.K., Jan, I.S., Yu, C.J., Hsueh, P.R. & Yang, P.C. 2007. Fluoroquinolone resistance in *Mycobacterium tuberculosis* isolates: associated genetic mutations and relationship to antimicrobial exposure. *The Journal of antimicrobial chemotherapy*. 59(5):860–865.
- Wang, X., Liu, H., Wei, J., Wu, X., Yu, Q., Zhao, X., Lyu, J., Lou, Y., et al. 2015. An investigation on the population structure of mixed infections of *Mycobacterium tuberculosis* in Inner Mongolia, China. *Tuberculosis (Edinburgh, Scotland)*. 95(6):695–700.
- Warren, R., Kock, M. De, Engelke, E., Myburgh, R., Pittius, N.G. Van, Victor, T. & Helden, P. Van. 2006. Safe *Mycobacterium tuberculosis* DNA Extraction Method That Does Not Compromise Integrity. 44(1):254–256.
- Warren, R.M., Victor, T.C., Streicher, E.M., Richardson, M., Beyers, N., Gey van Pittius, N.C. & van Helden, P.D. 2004. Patients with active tuberculosis often have different strains in the same sputum specimen. *American journal of respiratory and critical care medicine*. 169(5):610–614.
- Wedajo, W., Schön, T., Bedru, A., Kiros, T., Hailu, E., Mebrahtu, T., Yamuah, L., Ängeby, K., et al. 2014. A 24-well plate assay for simultaneous testing of first and second line drugs against *Mycobacterium tuberculosis* in a high endemic setting. *BMC research notes*. 7:512.
- WHO. 2008a. Policy guidance on drug-susceptibility testing (DST) of second-line antituberculosis drugs World Health Organization. *World Health*. 1–20.
- WHO. 2008b. Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB). *Policy statement*. (June):1–9. [Online], Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:MOLECULAR+LINE+PROBE+AS>

SAYS+FOR+RAPID+SCREENING+OF+PATIENTS+AT+RISK+OF+MULTIDRUG-RESISTANT+TUBERCULOSIS+POLICY+STATEMENT#0.

WHO. 2011. AUTOMATED REAL-TIME NUCLEIC ACID AMPLIFICATION TECHNOLOGY FOR RAPID AND SIMULTANEOUS DETECTION OF TUBERCULOSIS AND RIFAMPICIN RESISTANCE : Xpert MTB / RIF SYSTEM POLICY STATEMENT. *World Health*. 1–35.

WHO. 2014. Xpert MTB/RIF implementation manual Technical and operational “how-to”: practical considerations. *WHO Manual*.

WHO. 2015a. Global tuberculosis report 2015. *Igarss 2014*. (1):1–5.

WHO. 2015b. Tuberculosis mortality nearly halved since 1990. WHO. [Online], Available: <http://www.who.int/mediacentre/news/releases/2015/tuberculosis-mortality/en/>.

World Health Organization. 2007. *TB diagnostics and laboratory strengthening - WHO policy. Definition of a new sputum smear-positive TB case, 2007*. Geneva, Switzerland: WHO, 2007. [Online], Available: http://www.who.int/tb/laboratory/policy_sputum_smearpositive_tb_case/en/ [2015, November 11].

World Health Organization. 2016. *Tuberculosis*. [Online], Available: <http://www.who.int/mediacentre/factsheets/fs104/en/> [2016, April 13].

Zetola, N.M., Shin, S.S., Tumedi, K.A., Moeti, K., Ncube, R., Nicol, M., Collman, R.G., Klausner, J.D., et al. 2014. Mixed Mycobacterium tuberculosis complex infections and false-negative results for rifampin resistance by genexpert MTB/RIF are associated with poor clinical outcomes. *Journal of Clinical Microbiology*. 52(7):2422–2429.

Zhang, X., Zhao, B., Liu, L., Zhu, Y., Zhao, Y. & Jin, Q. 2012. Subpopulation analysis of heteroresistance to fluoroquinolone in Mycobacterium tuberculosis isolates from Beijing, China. *Journal of clinical microbiology*. 50(4):1471–4.

Zhang, Z., Lu, J., Wang, Y., Pang, Y. & Zhao, Y. 2014. Automated liquid culture system misses isoniazid heteroresistance in Mycobacterium tuberculosis isolates with mutations in the promoter region of the inhA gene. *European Journal of Clinical Microbiology & Infectious Diseases*. 34(3):555–560.

Zhang, Z., Wang, Y., Pang, Y. & Liu, C. 2014. Comparison of different drug susceptibility test methods to detect the rifampin heteroresistance in Mycobacterium tuberculosis. *Antimicrobial agents and chemotherapy*. 58(155):5632–5635.