



Understanding molecular mechanisms underlying *Mycobacterium tuberculosis* persistence

Thesis submitted in accordance with the requirements of the University of Liverpool
for the degree of Doctor of Philosophy

Victor Sengelwayo Ndhlovu

August 2018

Faculty of Health and Life Sciences
Institute of Infection and Global Health
University of Liverpool

DECLARATION OF AUTHORSHIP

I declare that the material contained in “Understanding Molecular Mechanisms of Mycobacterium tuberculosis persistence” is my own work. Work on culture of *Mycobacterium tuberculosis* isolates was done with the assistance of Mercy Kamdolozi in the College of Medicine/MLW TB laboratory whereas that on Large sequence polymorphism-PCR was done with the assistance of Nadia Kontogianni at the Liverpool School of Tropical Medicine. Work on generating persister in vitro lipid *Mtb* organisms was done under the collaborative guidance Jonathan Decker at the University of Leicester. The material has not been presented, nor is currently being presented, either wholly or in part for a PhD, any other degree or qualification at the University of Liverpool or any other university.

Signature:



Printed Name: Victor Sengelwayo Ndhlovu

Date: 30 August 2018

ABSTRACT

Background: Although *Mtb* strains have been shown to exhibit nucleotide level similarity of >99%, such similarity is rarely replicated in the phenotype. This phenotypic heterogeneity has been demonstrated by a subpopulation of cells called “persisters”. Persisters constitute a minority of the organisms found in clinical specimens but statistical modelling studies suggest that they predominate after the first few days of treatment. The rate of their elimination is likely the principal determinant of the length of therapy and likelihood of relapse. Understanding the complete biology of persistence is among the most important scientific barriers to shortening TB treatment.

Methods: A total of 64 bacteriologically confirmed *Mtb* clinical isolates from Malawian patients recruited for a cohort study of determinants of treatment outcome were genotyped using LSP-PCR to determine the molecular epidemiology of TB in Blantyre, Malawi and to evaluate the possible impact of lineage on outcome. Eighteen of the 64 isolates including two “persister” strains defined by lipid body staining and clinical relapse were further subjected to whole genome sequencing using Single Molecule Real Time (SMRT) sequencing in order to assess differences in the methylome among these strains. Using bioinformatics analysis DNA methylation patterns were determined and factors leading to loss/ absence of methylation could then be elucidated.

Results: The 64 isolates genotyped clustered into 4 main *Mtb* lineages (1-4) with lineage 4 predominating. There was differential occurrence of methylation across the strains. Methylation was found to occur within three different confidently defined sequence motifs (CTGGAG, CACGCAG and GATN₄RTAC) across all the isolates influenced by three different methyltransferases in a lineage specific manner. Three different methylation

disrupting mutations within methyltransferase genes E270A in *mamA*, P360L in *hsdM* the recently characterized S253L mutation in *mamB* were confirmed. For the first time, total loss of methylation in *MamB* was reported due to the S253L mutation. No compelling correlation between DNA methylation pattern and persistence during treatment was observed, at the level of genic or intergenic regions or within specific operons of interest previously associated with persistence *in vitro*.

Conclusions: Four main *Mtb* lineages occur in Blantyre Malawi and DNA methylation in these isolates were lineage specific. Specific mutations are responsible for disrupting methylation within various *Mtb* strains and methylation disrupting mutations could be reliably used to assign *Mtb* lineages for some lineages. Methylation patterns did not differ significantly between persister and non-persister isolates, though current limitations in direct extraction of DNA from clinical samples and numbers of paired isolates from representative longitudinal studies suggest that further work is needed to definitively answer this question.

ACKNOWLEDGEMENTS

This has been a long and at times frustrating journey. There were times I felt like giving up. I am extremely grateful to all the people who supported me through this journey. Firstly, I thank my three supervisors who have been absolutely wonderful. Gerry Davies my principal supervisor, thank you for believing in me and allowing me to take on such an ambitious project. You provided excellent guidance and support. Wilson Mandala, I remember you always used to tell me that “it is possible” and Maxine Caws, I could always count on you whenever I was stuck. I will forever be grateful to the Helse Nord Tuberculosis Initiative (HNTI) for funding this study and to colleagues in the HNTI project office, Mphatso Mwapasa for the administration and logistical support and Marriot Nliwasa and Kruger Kaswaswa for technical and moral support. Thank you all for your encouraging words. This work could not have been possible without the help of two laboratory personnel, Nadia Kontogianni at the Liverpool School of Tropical Medicine and Mercy Kamdolozi at the College of Medicine. I will forever be indebted to you.

Lastly, I would like to thank my family. My wife Rachel and my kids Martin and Esnart, you put up with my long absences from home. Without your support this would not have been possible.

PUBLICATIONS AND ABSTRACTS RELATED TO THIS WORK

Scientific publications

Victor Ndhlovu, Wilson Mandala, Derek Sloan, Mercy Kamdolozi, Maxine Caws, Geraint Davies Evaluation of the efficacy of two methods for direct extraction of DNA from *Mycobacterium tuberculosis* sputum, *The Journal of Infection in Developing Countries* 2018 (accepted for publication)

Victor Ndhlovu, Anmol Kiran, Derek Sloan, Wilson Mandala, Konstantina Kontogianni, Mercy Kamdolozi, Maxine Caws, Gerry Davies Genetic diversity of *Mycobacterium tuberculosis* clinical isolates in Blantyre, Malawi, *Tuberculosis (Edin)* 2018 (Under review)

Victor Ndhlovu, Anmol Kiran, Derek Sloan, Benjamin Kumwenda, Wilson Mandala, Konstantina Kontogianni, Mercy Kamdolozi, Maxine Caws, Gerry Davies Understanding the methylome of *Mycobacterium tuberculosis* (manuscript in preparation)

Oral and poster presentations

Victor Ndhlovu, Wilson Mandala, Derek Sloan, Mercy Kamdolozi, Maxine Caws, Geraint Davies Evaluation of the efficacy of two methods for direct extraction of DNA from *Mycobacterium tuberculosis* sputum Poster presentation at the 11th Interest Conference, Lilongwe, Malawi, May 2017

Victor Ndhlovu, Anmol Kiran, Derek Sloan, Benjamin Kumwenda, Wilson Mandala, Konstantina Kontogianni, Mercy Kamdolozi, Maxine Caws, Gerry Davies Understanding molecular mechanisms underlying *Mycobacterium tuberculosis* persistence, Oral presentation at the 12th Interest Conference, Kigali Rwanda, May 2018

TABLE OF CONTENTS

DECLARATION OF AUTHORSHIP	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
PUBLICATIONS AND ABSTRACT RELATED TO THIS WORK.....	v
Scientific publications	v
Oral and poster presentations	v
LIST OF FIGURES.....	xiii
LIST OF TABLES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER ONE: INTRODUCTION	1
1. 1 Evolution of Mycobacterium tuberculosis	1
1. 2 History of tuberculosis	3
1. 3 Tuberculosis as a global disease.....	5
1. 4 Tuberculosis in sub-Saharan Africa	7
1. 5 Tuberculosis pathogenesis.....	8
1. 5. 1 The infectious process.....	11
1. 6 The diagnosis of tuberculosis	12
1. 6. 1 Smear microscopy.....	13
1. 6. 2 Culture methods	14

1. 6. 3 Nucleic acid amplification tests	15
1. 6. 4 Other diagnostic methods	19
1. 7 Treatment of tuberculosis	21
1. 7. 1 Drug resistant tuberculosis (DR-TB).....	22
1. 7. 2 Rifampicin resistance.....	23
1. 7. 3 Isoniazid resistance	24
1. 7. 4 Resistance to Fluoroquinolones	26
1. 7. 5 Resistance to Kanamycin, Amikacin, Capreomycin and Viomycin.....	27
1. 8 The Mycobacterium tuberculosis genome.....	28
1. 8. 1 Genotyping of Mycobacterium tuberculosis.....	29
1. 8. 2 Genotyping methods	30
1. 8. 3 IS6110-based restriction fragment length polymorphisms	30
1. 8. 4 Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU/VNTR)	32
1. 8. 5 Spacer oligonucleotide typing (Spoligotyping)	33
1. 8. 6 Other genotyping methods.....	33
1. 9 Bacterial persistence	34
1. 9. 1 Mechanisms generating persister cells.....	36
1. 9. 2 Genetic basis of persistence	36
1. 9. 3 Toxin and Antitoxin (TA) module.....	37

1. 10	<i>Mycobacterium tuberculosis</i> persistence.....	39
1. 10. 1	The special populations hypothesis and <i>Mtb</i> metabolic activity	40
1. 10. 2	Laboratory investigations into <i>Mycobacterium tuberculosis</i> persistence....	41
1. 10. 3	Mechanisms of persister formation in <i>Mycobacterium tuberculosis</i>	43
1. 10. 4	Persistent <i>Mycobacterium tuberculosis</i> in sputum	50
1. 11	Epigenetics in <i>Mycobacterium tuberculosis</i>	51
1. 11. 1	Role of DNA methylation.....	53
1. 11. 2	Detection of modified bases.....	54
1. 12	Aims of this thesis	58
CHAPTER TWO: MATERIALS AND METHODS.....		60
2. 1	Experimental procedures	60
2. 1. 1	<i>Mycobacterial</i> strains and growth conditions	60
2. 1. 2	Decontamination of sputum samples	66
2. 1. 3	The saline wash procedure.....	66
2. 1. 4	Solid plate preparation and determination of colony counts	67
2. 1. 5	Liquid culture.....	68
2. 1. 6	Solid culture inoculation.....	68
2. 1. 7	Ziehl Neelsen (ZN) staining.....	69
2. 1. 8	Nitric Oxide (NO) induction of H37Rv isolates.....	70
2. 1. 9	Drug susceptibility testing	70

2. 1. 10 DNA extraction -from culture.....	71
2. 1. 11 Phenol/Chloroform/Isoamyl alcohol.....	71
2. 1. 12 The Chloroform/ Isoamyl alcohol method.....	72
2. 1. 13 DNA extraction -from sputum.....	74
2. 1. 14 Sample preparation	75
2. 1. 15 Genotyping of <i>Mycobacterium tuberculosis</i> isolates using large sequence polymorphisms.....	75
2. 1. 16 Agarose gel electrophoresis	77
2. 2 DNA sequencing	77
2. 2. 1 Bioinformatics analyses of SMRT sequence data	78
2. 2. 2 Analysis of DNA methylation patterns using SMRT sequencing	79
2. 2. 3 Prediction of promoter regions	79
2. 2. 4 Comparative analysis of methylated sites within persisters and non-persisters	80
2. 2. 5 SNP calling and identification of mutations causing loss of methylation	80
2. 2. 6 Comparison of Malawian isolates against a global sample	81
2. 2. 7 Ethical approval.....	82
CHAPTER THREE:OPTIMIZATION OF A DNA EXTRACTION METHOD FOR SINGLE MOLECULE REAL TIME (SMRT) SEQUENCING.....	83
3. 1 Introduction	83

3. 2 Aims	84
3. 3 Results	84
3. 3. 1 The Chloroform/Phenol/Isoamyl alcohol method	85
3. 3. 2 The Chloroform/Isoamyl alcohol method.....	88
3. 4 Discussion.....	90
CHAPTER FOUR: COMPARISON OF PRIMEXTRACT DNA EXTRACTION KIT AND AN IN- HOUSE CTAB METHOD.....	94
4. 1 Introduction	94
4. 2 Results	95
4. 3 Discussion.....	102
CHAPTER FIVE: GENETIC DIVERSITY OF MYCOBACTERIUM TUBERCULOSIS ISOLATES IN BLANTYRE, MALAWI.....	106
5. 1 Introduction	106
5. 2 Results	108
5. 2. 1 Large sequence polymorphism (LSP)-PCR.....	108
5. 2. 2 Whole genome sequence analysis.....	111
5. 3 Discussion.....	114
CHAPTER SIX: CHARACTERIZATION OF DNA METHYLATION IN CULTURED <i>MTB</i> ISOLATES USING SMRT SEQUENCING	120
6. 1 Introduction	120

6. 2 Results	122
6. 2. 1 Comparative bioinformatics analysis of “persister” and “non-persister” <i>Mtb</i> isolates.....	123
6. 2. 2 Occurrence of methylated motifs within persister and non-persister isolates.....	125
6. 2. 3 “Persister” gene and promoter methylation	132
6. 2. 4 Specific location of differentially methylated gene promoters in persisters	133
6. 2. 5 Differential methylation within persister related genes	136
6. 2. 6 Stability of methylation patterns in <i>Mycobacterium tuberculosis</i> genome under different growth conditions.....	138
6. 2. 7 Comparison of MGIT and LJ cultured isolates.....	139
6. 2. 8 Lineage specific methylation patterns across 18 <i>Mtb</i> strains.....	141
6. 2. 9 Comparative analysis of Malawian isolates against a global sample	150
6. 3 Discussion.....	153
6. 3. 1 Detection of methylated motifs in <i>Mycobacterium tuberculosis</i> strains	153
6. 3. 2 Methylation within promoters and genes of persister <i>Mtb</i> strains	156
6. 3. 3 Lineage specific methylation in <i>Mtb</i> strains from Malawi	158
CHAPTER SEVEN: FINAL DISCUSSIONS AND CONCLUSIONS	164
7. 1 Introduction	164
7. 2 Isolation of persisters and DNA extraction	165

7. 3 Lineage assignment and SMRT sequencing.....	167
7. 4 Key findings of the study	168
7. 5 Limitations of the study.....	170
7. 6 Future direction	171
8. REFERENCES.....	173

LIST OF FIGURES

Figure 1. 1 Global geographical distribution of <i>Mycobacterium tuberculosis</i> lineages	2
Figure 1. 2 Top 10 leading causes of death in globally.....	6
Figure 1. 3 Top 10 countries with new TB cases per 100,000 populations	8
Figure 1. 4 Granuloma formation.....	11
Figure 1. 5 The Xpert-MTB/RIF assay	18
Figure 1. 6 Chromosome of <i>Mycobacterium tuberculosis</i> strain.....	31
Figure 1. 7 Chemical structure of some methylated bases.....	53
Figure 1. 8 Detection of methylated DNA bases using Single Molecule Real Time Sequencing	57
Figure 3. 1 Comparison of bead beating duration on DNA concentration	86
Figure 3.2 Gel electrophoresis image of genomic DNA obtained using the bead beating method.....	87
Figure 3. 3 Comparison of DNA concentration by extraction method.	88
Figure 3. 4 Comparison of DNA yields from LJ slants and MGIT culture	89
Figure 4. 1 Concentrations of DNA extracted using two different methods.....	97
Figure 4. 2 Yields of DNA extracted using two different methods.	98
Figure 4. 3 Purity of DNA extracted using two different methods.....	99
Figure 4. 4 Comparison of DNA concentrations obtained by CTAB method with and without a saline wash step.....	101
Figure 4. 5 Comparison of DNA purity obtained by CTAB method with and without a saline wash step.....	101
Figure 5. 1 Gel electrophoresis of RD-105 deletion PCR.....	110

Figure 5. 2 Gel electrophoresis of RD-239 deletion PCR.....	110
Figure 5. 3 Diversity of <i>Mycobacterium tuberculosis</i> in Blantyre, Malawi	111
Figure 5. 4 Phylogenetic analysis of the 18 sequenced <i>Mycobacterium tuberculosis</i> strains from Blantyre Malawi.	113
Figure 5. 5 Presence of a 7 bp deletion in the polyketide synthase (pks15/1) gene.....	114
Figure 6. 1 Pathways differentially methylated in persisters	135
Figure 6. 2 Occurrence of mutations within the gene Rv3263 encoding the methyltransferase <i>MamA</i> and targeting the motif CTGGAG leading to partial and complete loss of methylation in three isolates	145
Figure 6. 3 Occurrence of mutations within the gene Rv2024c encoding the methyltransferase <i>MamB</i> and targeting the motif CACGCAG leading to partial and complete loss of methylation	147
Figure 6. 4 Occurrence of mutations within the gene Rv2756c encoding the methyltransferase <i>HsdM</i> which is known to methylate the GATN4RTAC motif leading to loss of methyltransferase activity.....	149

LIST OF TABLES

Table 1. 1 Genes implicated in <i>Mtb</i> persistence.....	46
Table 2. 1 Patient characteristics from whom bacterial isolates used in the study were obtained.....	61
Table 3. 1 General Summary of DNA extraction methods and yield/purity.....	90
Table 4. 1 Summary of results of DNA extraction using PrimeXtract and CTAB method.....	99
Table 6. 1 Motifs and Modifications within “persisters” and “non-persisters”	124
Table 6. 2 General genome methylation for 4 <i>Mtb</i> clinical isolates	127
Table 6. 3 Methylation within genic regions in “persisters” and “non persisters” by strand.....	128
Table 6. 4 Methylation within intergenic regions in persisters and non persisters by strand.....	129
Table 6. 5 Methylation within promoter regions by strand for persisters and non-persisters compared to the rest of intergenic region.....	131
Table 6. 6 Methylation within promoter regions of selected persister related genes....	135
Table 6. 7 Sites differentially methylated within selected persister related genes.....	137
Table 6. 8 Comparison of methylation patterns between MGIT and LJ cultured isolates.....	141
Table 6. 9 Genome-wide methylation patterns among the 18 sequenced <i>Mtb</i> isolates .	143

LIST OF ABBREVIATIONS

Abbreviations have been defined at the time of first use. The following list may also be useful:

(p)ppgp guanosine tetra-phosphate

AFB Acid Fast Bacilli

BCG Bacillus Calmette-Guerin

CFP-10 Cultures Filtered Protein-10

CFU Colony Forming Unit

CGR Centre for Genomic Research

CTAB cetyltrimethylammonium

DC Dendritic Cells

dosR dormancy survival Regulator

DOT Directly Observed Therapy

ELISA enzyme linked immunoabsorbent assay

ESAT-6 early secretory antigenic target-6

ETH Ethambutol

FDA Food and Drugs Administration

FM Fluorescent Microscopy

FQs Fluoroquinolones

GFF General Features Format

GPI glycosylphostidylinositol

HBC High Burden Countries

HGAP Hierarchical Genome Assembly

Hip High persister

ICL isocitrate lyase

IGRA Interferon Gamma Release Assay

INF- interferon- γ

INH Isoniazid

IPD Inter-pulse duration

IS6110 Insertion Sequence 6110

LAM Lipoarabinomannan

LAMP Loop-mediated isothermal amplification

LD Lipid Droplets

LED Light Emitting Diode

LJ Lowenstein-Jensen

LSP Large Sequence Polymorphism

m4C 4-methylCytosine

m6A 6-methylAdenine

MAFFT Multiple sequence Alignment using Fast Fourier Transform

MamA Mycobacterium adenine methyltransferase A

MamB Mycobacterium adenine methyltransferase B

MDR Multi Drug Resistant

MGIT Microbacterial Growth Indicator Tube

MIC Minimum Inhibitory Concentration

MIRU-VNTR Mycobacterial Interspersed Repetitive Unit –Variable Number Tandem Repeat

Mtase methyltransferase

Mtb Mycobacterium tuberculosis

NAAT Nucleic acid amplification test

NRP Non-replicating persistence

PANTA Polymyxin B amphotericin B trimethoprim and azlocilin

PBS Phosphate Buffered Saline

PCR Polymerase Chain Reaction

PE/PPE Proline-glutamic acid and proline-proline-glutamic acid

PGRS Guanine-cytosine repetitive sequences

PW Pulse width

PZA Pyrazinamide

RAxML Random accelerated maximum likelihood

RD Region of Difference

RE Restriction Endonuclease

RFLP Restriction Fragment Length Polymorphism

RIF Rifampicin

RM Restriction Modification

rpoB RNA polymerase beta subunit

RRDR Rifampicin Resistance Determining Region

SAM S-adenosyl-methionine

SDS Sodium Dodecyl Sulfate

SMRT Single Molecule Real Time Sequencing

SNP Single Nucleotide Polymorphism

SP Surfactant Protein

TA Toxin-Antitoxin

TAG triacylglycerol

tgs Triacylglycerol synthase

TNF- α Tumor Necrosis Factor- α

TSS Transcription start site

WGS Whole Genome Sequencing

XDR Extensively Drug Resistant

ZMW Zero Mode Wave guide

ZN Ziehl-Neelsen

CHAPTER ONE: INTRODUCTION

1. 1 Evolution of *Mycobacterium tuberculosis*

Tuberculosis (TB) is a debilitating human illness which mainly targets the lungs (pulmonary TB) but may also target other body organs (extra-pulmonary). The bacillus *Mycobacterium tuberculosis* (*Mtb*), which is the causative agent of TB is inhaled via aerosols and infects and activates macrophages and dendritic cells (DCs) in the lungs. After this primary infection, only 5% of the infected population will ultimately develop active TB while the rest carry a lifelong latent infection (Dreher, Nicod 2002). It has been hypothesized that the genus *Mycobacterium* evolved more than 150 million years ago (Hayman 1984, Daniel, T. M. 2006) ultimately forming complex species. It is generally believed that all members of the *Mtb* complex shared a common African ancestor some 35000-15000 years ago (Kapur, Whittan et al. 1994, Brosch, Gordon et al. 2002). This early ancestor of *Mtb* complex could have been present in East Africa some 3 million years ago (Gutierrez, Brisse et al. 2005). The *M. tuberculosis* complex is composed of several members including *M. bovis*, *M. africanum*, *M. canetti* and *M. microti*, *M. pinniped* and *M. caprae*, each one capable of causing disease in different hosts (Stead, Eisenach et al. 1995, Müller, Roberts et al. 2014). They are all genetically similar and only differ in particular nucleotide pairs. These differences may include single nucleotide polymorphisms (SNPs), large sequence polymorphisms (LSPs), minisatellites and microsatellites (Davies, P. D. O., Barnes et al. 2008). Previously the accepted hypothesis was that the ancestral descendant of *M. tuberculosis* was *M. bovis* which mainly causes disease in cattle. This hypothesis has been seriously challenged with the advent of DNA sequencing and other related methods (Smith 2003). Subsequently genome sequencing

projects have revealed that *M.bovis*, while maintaining 99.95% sequence identity with *Mtb* has several significant deletions (Brosch, Gordon et al. 2002). When 875 strains of *Mtb* from different parts of the world were analyzed using large sequence polymorphism (LSP) PCR and sequencing, six main lineages were identified (Figure 1.1).

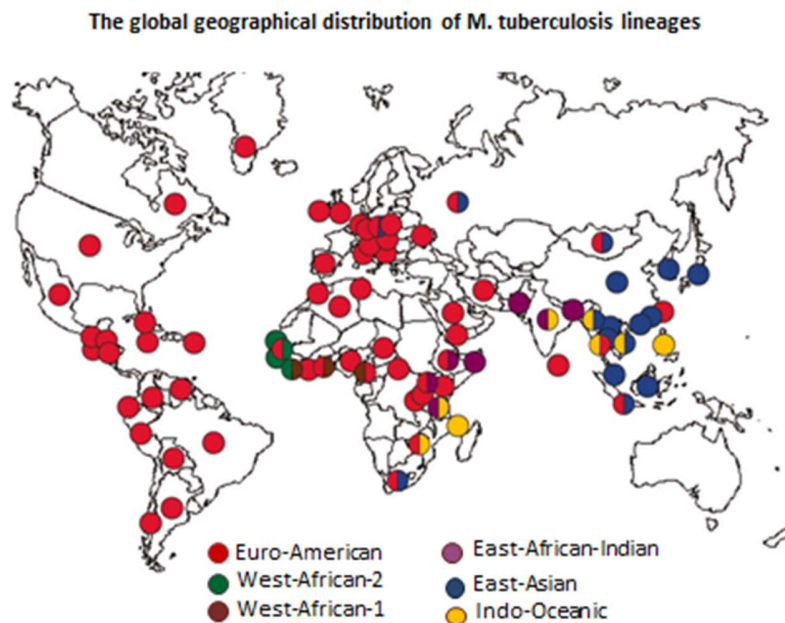


Figure 1. 1 Global geographical distribution of *Mycobacterium tuberculosis* lineages
The Euro-American lineage is dominant in parts of Europe and America; the East African lineage and two West-African lineages are dominant in Africa; the East African-Indian lineage is dominant in the Eastern region of Africa and the Indo-Oceanic region; the East-Asian lineage is dominant in the Indian sub-continent; the Indo-Oceanic lineage occurs in and around the Indian Ocean. (Figure adapted from Gagneux, DeRiemer et al. 2006).

These lineages have been associated with several factors including virulence, ability to cause epidemic spread and retain virulence on mutation to multidrug resistance (Davies, P. D. O., Barnes et al. 2008). The Indo-Oceanic lineage occurs predominantly around the Indian Ocean (Gagneux, DeRiemer et al. 2006) and includes a group of strains referred to as “ancestral”. This group has a conserved TbD1 genomic region which is lacking in “modern” strains (Brosch, Gordon et al. 2002, Gagneux, DeRiemer et al. 2006). The East Asian lineage is widespread in the Indian subcontinent and includes strains of the *Beijing* family which is widely associated with drug resistance (Tsolaki, Gagneux et al. 2005, Ribeiro, Gomes et al. 2014). As expected, the East African lineage and two West African lineages are predominant in Africa whereas the Euro-American lineage predominates in different regions of America and Europe. Specific sub lineages of the Euro-American lineage are present in Africa and the Middle East as well. The East African/Indian lineage is predominant in East Africa and the Indian Ocean (Gagneux, DeRiemer et al. 2006). Recently, a new lineage was characterized which appears to be restricted to the Horn of Africa (Coll, McNerney et al. 2014, Firdessa, Berg et al. 2013). This lineage has been proposed to be an intermediate between the ancient and modern strains (Coll, McNerney et al. 2014)..

1. 2 History of tuberculosis

Tuberculosis is an illness that is probably as old as man himself and has been among the leading historical causes of mortality in humans. In Egypt, skeletons with apparent tubercular deformities were frequently unearthed by modern archaeologists suggesting the disease had been widespread in the area for a long time. Bones with deformations characteristic of TB were also uncovered in ancient sites in Italy, Denmark and part of the

Middle East indicating the disease was found throughout the world some 4000 years ago. Historians had written of patients coughing blood as early as the seventh century BC. Hippocrates also wrote of a disease associated with weight loss coughing and sometimes with blood in sputum (the Greek word is phthisis) which he called consumption (Smith 2003, Daniel, T. M. 2006). There is a common acceptance that Europeans of Indian origin brought the disease to Europe and Asia as they migrated. Europe, with population growth and urbanization became the epicenter of many TB epidemics starting in 16th and 17th centuries. At its peak, TB is believed to have killed more than 15% of the European population in the first half of the 19th century (Smith 2003). Recent evidence from archeological skeleton samples suggests that different *Mtb* lineages could have been present in Britain at the same time and even same location (Müller, Roberts et al. 2014). Most medical researchers in the 19th century associated TB with societal causes such as malnutrition, poor sanitation and overworking until Robert Koch conclusively showed that it was caused by a bacterium (Davies, P. D. O., Barnes et al. 2008, Smith 2003). Later, the tuberculin skin test (TST) was developed and is still used today as a test for latent TB infection (Daniel, T. M. 2006).

Schartz and Waksman discovered the antibiotic streptomycin in 1944 which was rapidly recognised as an effective treatment for TB (Woodruff 2014). Later isoniazid (INH), rifampicin (RIF) and pyrazinamide (PZA) were discovered and remain the key components of first-line treatment for TB (Ryan 1992). These drugs and the widespread use of the bacillus Calmette- Guerin (BCG) vaccine discovered in Paris in the 1920s have been unable to successfully eliminate the disease however (Andersen 2001) and new approaches are clearly needed in the fight against TB

1. 3 Tuberculosis as a global disease

TB remains a public health concern with 10.4 million new infections and 1.3 million fatalities each year. An additional 0.4 million deaths result from coinfection with human immunodeficiency virus (HIV) (WHO 2017b). Although the incidence of TB has been in decline since the early 2000s, the disease remains among the top ten leading causes of mortality globally (Figure 1.2). It is the 6th highest killer among low income countries and it is the biggest single killer of all infectious diseases (WHO 2017a). About 1.2 million (11%) of the new TB infections in 2015, were co-infected with HIV and 75% of these co-infections occurred in Africa (WHO 2016).

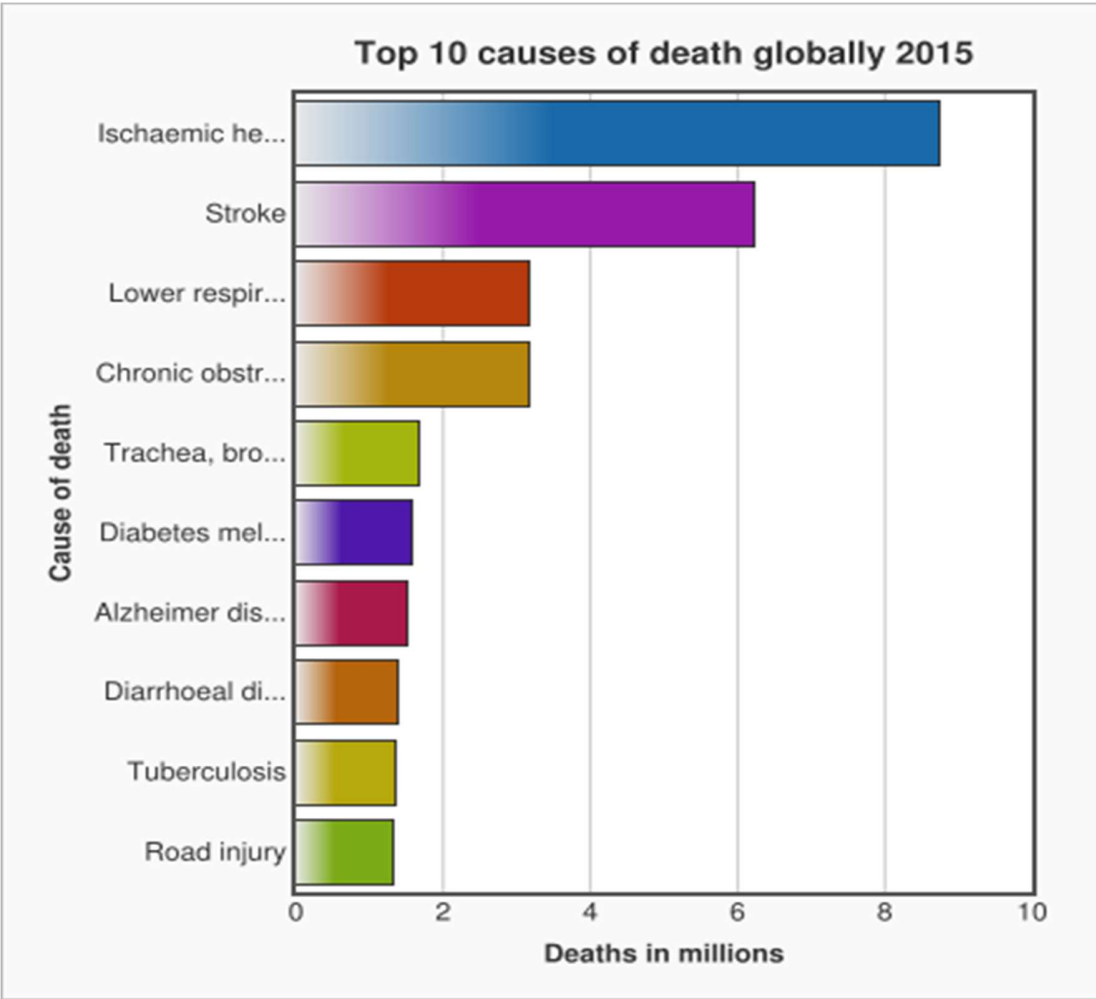


Figure 1. 2 Top 10 leading causes of death in globally

TB remains among the top ten leading causes of death globally (WHO 2017a)

With 5.9 million (56%) new cases among men compared to 3.5 million (34%) among women there appears to be a gender bias in new cases (WHO 2016). Although half of all TB infections are in the BRICS region (Brazil, Russia, India, China and South Africa), Africa remains the hardest hit in terms of rates of active TB per capita (Abdul-Aziz, Tsolaki 2013, Zumla, Raviglione et al. 2013, WHO 2017b). In addition the highest

proportion of TB cases co-infected with HIV were in Africa (31%) with southern Africa being the hardest hit (>50% of all cases) (WHO 2016)

1.4 Tuberculosis in sub-Saharan Africa

The burden of TB remains the highest in Africa in part due to high levels of poverty. According to the World Health Organization's (WHO) TB report for 2017, 30 high burden countries (HBCs) were low and middle-income countries and accounted for over 87% of active TB cases in the world (WHO 2017b). Fifty percent of these countries (South Africa, Zimbabwe, Mozambique, DR Congo, Kenya, Ethiopia, Nigeria, Zambia, Congo, Sierra Leone, Namibia, Liberia, Central African Republic and Tanzania) are in Africa. TB incidence rates in Lesotho, Mozambique, South Africa are above 500 per 100 000 populations compared to 150-300 cases per 100 000 populations for most HBCs (WHO 2017b). Seven of the top 10 countries with the highest TB incidence rates per 100 000 population in 2013 were in Africa (Figure 1.3) (Kaiser Family Foundation). Estimates show that in South Africa and Swaziland at least 1 in every 100 people (1000 per 100 000 population) become TB infected every year (Pham, Fattal et al. 2015). TB-HIV co-infection rates of more than 50% in all ages occurred in the Southern African countries of South Africa, Swaziland, Lesotho, Mozambique, Zimbabwe, Zambia and Malawi (where this study took place) although mortality rates among HIV positive people are on the decline (WHO 2016). It is estimated that in HBCs, 60% of all health expenditure is in the private sector and largely paid by patients out of their pockets (Lonnroth, Castro et al. 2010). The financial difficulties experienced by TB patients in these countries are huge. The costs due to TB have been described as catastrophic as they amount up to 10% of a patient's or household's income. TB also contributes to poverty by reducing the patient's

strength making them unable to work (Barter, Agboola et al. 2012). Care givers lose man hours looking after sick relatives and friends.

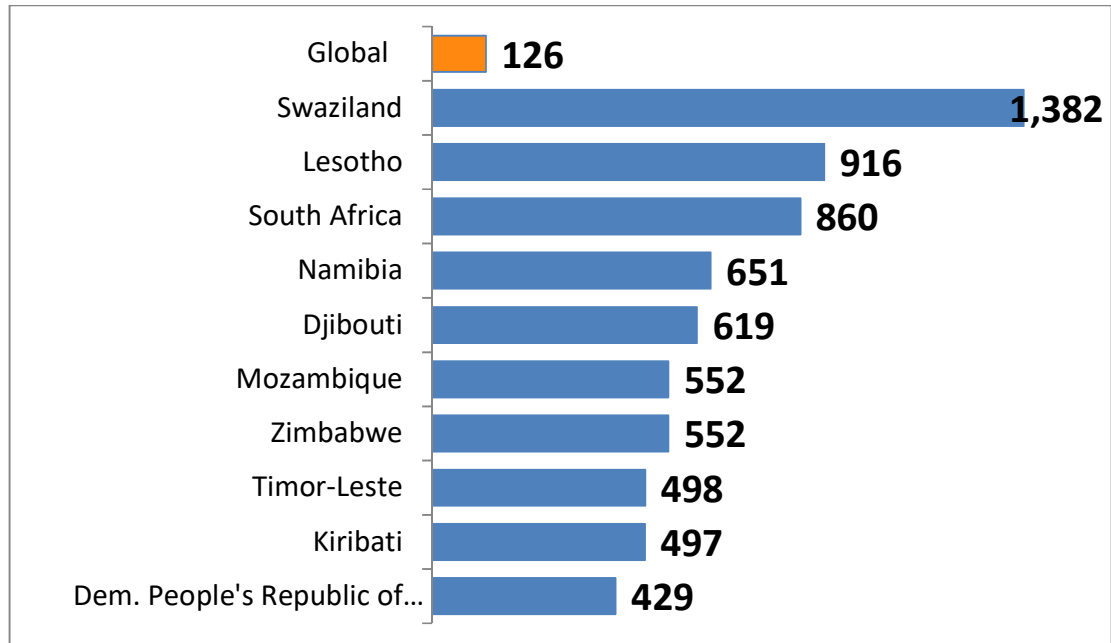


Figure 1. 3 Top 10 countries with new TB cases per 100,000 populations

Seven of these countries (Swaziland, South Africa, Namibia, Lesotho, Djibouti, Zimbabwe and Mozambique) are in Africa. Image adopted from (Kaiser Family Foundation).

1. 5 Tuberculosis pathogenesis

Mycobacteria are organisms that appear microscopically as straight or slightly curved rods, 1 to 4µm in length and 0.3 to 0.6 µm wide (Sakamoto 2012). They are non-motile, non-spourating, weakly Gram positive, acid fast bacilli. Mycobacteria express mycolic acids in their cell walls with unique characteristics for their physiological survival (Barry

III, Lee et al. 1998). *Mtb* are strictly parasitic and a member of the slow-growing pathogenic species of mycobacteria with a division rate of 12-24 hours and culture period on agar of up to 21 days (Sakamoto 2012).

Following entry into the lungs the bacilli proceed to the alveoli, and consequently come in contact with and are ingested by alveolar macrophages. As a consequence of this initial interaction, most are destroyed but some survive within the macrophage (Abdul-Aziz, Tsolaki 2013). Destruction of *Mtb* via phagocytosis involves a variety of receptors which can either bind non-opsonized *Mtb* or recognise opsonins on the mycobacterial surface. Several receptors have so far been implicated including C-type lectins eg. CD 207, complement receptors (CR), surfactant protein (SP) receptors, scavenger receptors, and glycosylphosphatidylinositol (GPI) - anchored receptors such as CD14 (Abdul-Aziz, Tsolaki 2013, Ehlers, Daffe 1998). *Mtb* entry into macrophages is facilitated by binding their mannose-capped lipoarabinomannan (ManLAM) to mannose receptors (Kang, Azad et al. 2005). Within resting macrophages, internalized *Mtb* reside in phagosomes where they block maturation, lysosomal fusion and acidification (Sakamoto 2012, Smith 2003). If normal phagosome maturation occurs, then phagosome-lysosome fusion occurs and bacteria encounter a hostile environment that includes acid pH, reactive oxygen intermediates (ROI), lysosomal enzymes and toxic peptides (Smith 2003). Additionally, this process produces a cytokine/chemokine response that initiates inflammation and attracts immune cell infiltration and tissue modification. First to be involved are neutrophils and monocytes which phagocytose the *Mtb* and secrete more cytokines and chemokines. Antigen presenting dendritic cells (DCs) also phagocytose bacilli and then migrate to regional lymph nodes to present the *Mtb* antigens to lymphocytes (Sakamoto

2012). Lymphocytes and infected macrophages become activated in response to inflammatory cytokines/ chemokines and migrate to sites of infection where they help to form granulomas (Kapoor, Pawar et al. 2013). Macrophages, T-cells, B-cells and DCs are all involved in this process (Figure 1. 4). Tumor Necrosis Factor- α (TNF- α) and Interferon- γ (INF- γ) are the chief signaling cytokines for tissue infiltration although they are not required to initiate granuloma formation (Abdul-Aziz, Tsolaki 2013). The role of TNF- α in granuloma formation is partly via recruitment of leukocytes but also TNF- α dependent chemokine production (Sakamoto 2012). A granuloma will therefore consist of central infected macrophages; foam cells, occasional multinucleated giant cells with recruited lymphocytes and a fibrous capsule (Figure 1.4). Studies in non-human primates have shown that there are a number of granulomas in the lungs during latent infection; either controlling or disseminating bacterial growth (Flynn, Capuano et al. 2003). The central core of the granuloma is believed to be hypoxic with little nutrient availability. Despite harsh conditions *Mtb* can persist within the granuloma for years. Reactivation of latent TB occurs when changes in the host immune system lead to rupture of the granuloma, disseminating the *Mtb* (Abdul-Aziz, Tsolaki 2013).

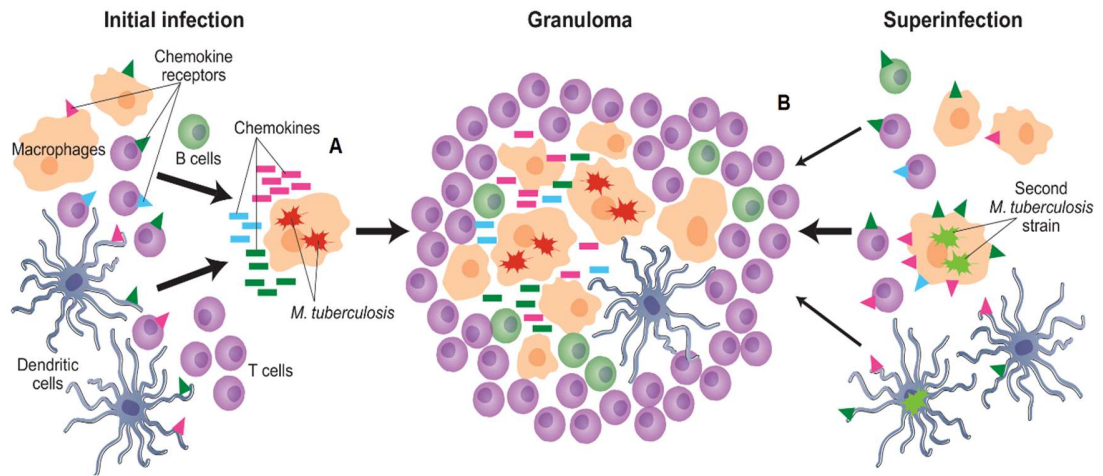


Figure 1. 4 Granuloma formation

Immune cells; macrophages, dendritic cells, T cells and B cells express a variety of receptors that can respond to the chemokines expressed by infected cells at the point of initial infection (A). This culminates in migration of the cells to the site of infection (B). After migration, the cells remain at the site of infection, probably in response to a variety of signals expressed in the granuloma (C). Adapted from (Flynn 2004)

1. 5. 1 The infectious process

The primary site of *Mtb* implantation forms a lesion called a *Ghon focus* generally referred to as the primary complex. The bacilli may be disseminated within the lung or to other organs during the early stages of the disease. Intriguingly, bacillary growth tends to favour the upper lobes of the human lung due to higher oxygen pressure and possibly delayed immune responses (Park, M. K., Myers et al. 1992). Despite replication and dissemination of the bacilli, it is estimated that only 5-10% of infected individuals develop clinically apparent and possibly life-threatening disease, while healthy individuals resolve the lesions without showing any symptoms (Dreher, Nicod 2002). Infected individuals may

remain without any symptoms for years with the bacilli in a “latent” state with only 5% ultimately developing active disease during their lifetime. The bacilli are presumably kept in check by the host immune system. Reactivation, or secondary disease, may occur as a consequence of any condition that may affect immune competence (Verver, Warren et al. 2005). The highest risk of reactivated disease occurs in HIV positive individuals with a rate of TB of approximately 10% per year (Corbett, Watt et al. 2003). Reactivated TB mainly affects the lungs forming lesions that are characterized by Bronchopneumonia and progress to caseous granulomas with necrosis and cavity formation following thereafter. Symptoms of primary TB are not easily noticeable and are often ignored. Secondary tuberculosis is usually associated with localized symptoms such as coughing, hemoptysis and pleuritic pain. Fever, anorexia, night sweats and cachexia may also be accompanying symptoms resulting in the patient’s health decline (Sakamoto 2012). This stage is crucial for disease transmission as pulmonary cavities are generated. Such cavities enable the bacilli to get access into sputum leading to dissemination of the pathogen through aerosols of respiratory secretions (Davies, P. D. O., Barnes et al. 2008) .

1. 6 The diagnosis of tuberculosis

One of the major barriers to elimination of TB is the absence of fast and effective diagnostic tools. In most parts of the world, diagnostic methods are the same as those developed over a century ago, such as the acid-fast smear of the expectorated sputum specimen. The sensitivity of the method in some cases is low especially in areas of high HIV prevalence. This results in as many as half of all the tuberculosis cases remaining undiagnosed (Perkins, Cunningham 2007, Davies, P. D. O., Barnes et al. 2008).

1. 6. 1 Smear microscopy

Microscopic sputum examination of smears for acid fast bacilli (AFB) remains the mainstay for TB diagnosis and is highly recommended by the WHO especially in resource poor settings. The technique uses Carbol-fuchsin Ziehl-Neelsen (ZN) or Kinyoun acid fast staining. The method however, suffers from poor sensitivity for detection of pulmonary TB (40-60%) (Perkins, Cunningham 2007, Verma, Dhole et al. 2013). Smear microscopy has the advantage of being simple, rapid, inexpensive and highly specific. Yield and identification of AFB with ZN sputum microscopy can be improved using specific methods of decontamination, liquefaction and concentration of the sputum sample (Saxena, Mathur et al. 2001, Verma, Dhole et al. 2013). One study reported of a sputum sample processing technique called ReaSLR that has shown to be more sensitive than the conventional method with sensitivity and specificity of 90.47% and 91.60% respectively. It employs ready-to-use tablets with a chaotrope and reducing agents that exert mucolytic activity resulting in effective release of trapped mycobacteria in sputum (Verma, Dhole et al. 2013). Contrastingly fluorescent microscopy (FM) uses a fluorescent stain (e.g., auramine O) that utilises a powerful source of light (Steingart, Henry et al. 2006) and has been shown to improve the sensitivity of sputum smears by up to 10% over conventional microscopy (Perkins, Cunningham 2007, Steingart, Henry et al. 2006, Wallis, Pai et al. 2010). However the cost of the fluorescent microscope alone (\$10,000-20,000) may be prohibitive for resource poor settings (Perkins, Cunningham 2007). In 2009 the WHO recommended the use of light emitting diode (LED) FM microscopy to replace conventional FM (Wallis, Pai et al. 2010). LED-FM microscopes have the advantage of being more sensitive, are inexpensive, more robust, use less electricity and require less

technician time than ZN microscopes. The performance of LED-FM is said to be equivalent to conventional FM, with added benefits of low cost, durability and the use without the need for a darkroom (Wallis, Pai et al. 2010). Overall microscopy as a diagnostic tool for TB has limited value in children and HIV positive individuals because by definition it does not identify smear negative tuberculosis (Steingart, Henry et al. 2006). Detecting HIV associated TB is extremely difficult and results in delays while diagnostic tests and treatment trials are being conducted. Studies show that in Malawi up to a third of patients with smear negative TB needed more than six visits to a health center before treatment was initiated (Harries, Nyirenda et al. 2003, Perkins, Cunningham 2007). It is therefore clear that many TB cases are missed when smear microscopy is used alone.

1. 6. 2 Culture methods

Mycobacterial culture on selective media remains the most sensitive diagnostic method and is the gold standard for *Mtb* detection in clinical samples and for drug susceptibility testing (Perkins, Cunningham 2007). A traditional egg based solid media culture known as Lowenstein-Jensen (LJ) takes 4-6 weeks (Cruciani, Scarparo et al. 2004, Nyendak, Lewinsohn et al. 2009). The long duration (4-6 weeks) of mycobacterial detection has remained an area of concern.

More sensitive and rapid alternatives to conventional solid culture are liquid culture systems (BACTEC and MGIT) which may detect growth in roughly half the time (1-3 weeks) compared to LJ (Cruciani, Scarparo et al. 2004, Nyendak, Lewinsohn et al. 2009). In these systems growth is not directly detectable but measured using an indirect proxy (Asmar, Drancourt 2015) via bacterial CO₂ production or O₂ consumption (Williams-Bouyer, Yorke et al. 2000). The Mycobacterial growth indicator tube (MGIT) uses a

fluorescent oxygen quenching sensor at the bottom of a tube embedded in silicon to detect bacterial growth directly. With MGIT it is possible to get results in less than eight days (Palomino, Juan Carlos, Martin et al. 2008). With the addition of critical concentrations of streptomycin, RIF, INH and ethambutol, an automated MGIT allows for drug susceptibility testing (Nyendak, Lewinsohn et al. 2009). Liquid culture systems use expensive culture vials and require large and costly incubators/readers. These plus the high risk of contamination when used by inexperienced laboratory staff has caused the limited use of liquid culture in TB endemic developing countries (Perkins, Cunningham 2007), although most laboratories in low income countries now make use of liquid culture systems.

With advancements in standard culture procedures including improved growth conditions, microaerophilic environment and use of auto-fluorescence detection, researchers have recently demonstrated that it is possible to dramatically shorten the culture period. In the best case, only 72 hours may be needed to achieve both primary culture and RIF susceptibility testing (Ghodbane, Raoult et al. 2014).

1. 6. 3 Nucleic acid amplification tests

Nucleic acid amplification tests (NAAT) are molecular based diagnostic tests that provide fast and accurate results (Nyendak, Lewinsohn et al. 2009). These tests use oligonucleotide primers and enzymes that catalyze reactions to amplify a target, a probe or a signal and yield results within minutes or hours. Some of the widely used assays include polymerase chain reaction (PCR; Roche Diagnostics), transcription mediated amplification (GenProbe) and strand displacement amplification (Becton Dickinson) (Piersimoni, Scarparo 2003, Huggett, McHugh et al. 2003). In the USA the Food and Drug

Administration (FDA) approved some NAAT tests for direct use on smear positive respiratory specimens; GenProbe Amplified *Mtb* Direct test (Nyendak, Lewinsohn et al. 2009), Roche Amplicor MTB test and BD-Probetec Direct (Becton Dickinson); second generation tests from GenProbe (E-AMTD) and Roche (AMTD2) on smear negative samples (Greco, Girardi et al. 2006). In the past few years, several of these NAATs have received WHO endorsement for widespread use (Wallis, Pai et al. 2010). In 2011 the WHO endorsed a novel, rapid automated NAAT called the Xpert MTB/RIF (Xpert) which can give results in under two hours (WHO 2011). This PCR- based test simultaneously detects TB and RIF resistance on the GeneXpert platform (Blakemore, Story et al. 2010, Helb, Jones et al. 2010). The assay PCR amplifies five overlapping probes complementary to the entire 81 bp RIF resistance-determining region of the *Mtb*, the *rpoB* gene and subsequently probes this region for mutations associated with RIF resistance (Blakemore, Story et al. 2010, Molicotti, Bua et al. 2014). Uniquely the Xpert, unlike conventional NAATs has sample processing and PCR amplification within a single self-enclosed test unit, the GeneXpert cartridge (Figure 1.5). Apart from sample loading, all steps are automated (Steingart, Schiller et al. 2014). As an initial test replacing smear microscopy, Xpert had shown a pooled sensitivity of 89% and pooled specificity of 99% whereas when used as an add-on test following a negative smear microscopy result, Xpert had a pooled sensitivity and specificity of 67% and 99% respectively. Further in comparison with smear microscopy, Xpert increased TB detection among culture-confirmed cases by 23%. For rifampicin resistance, Xpert had a pooled sensitivity of 95% and pooled specificity of 99% (Steingart, Schiller et al. 2014). Studies in India, South Africa and Uganda show that introduction of Xpert increased TB case finding in all the settings; from 72%-85% to 95%-

99% of the cohort of individuals with suspected TB compared to the base case of smear microscopy and clinical diagnosis of smear negative TB. These studies further suggest Xpert as a cost-effective method of TB diagnosis compared to the base case for low and middle income settings (Vassall, van Kampen et al. 2011). Despite the fact that the cost of the Xpert is considerably higher when compared with smear microscopy, roll-out commenced in more than 20 countries at a subsidized price of US \$9.98 (US) per cartridge and this price will be maintained until 2022 (Steingart, Schiller et al. 2014, Molicotti, Bua et al. 2014) . An improved version of the most widely used NAAT for TB diagnosis the Xpert MTB/RIF is now available as Xpert Ultra (Dorman, Schumacher et al. 2018)

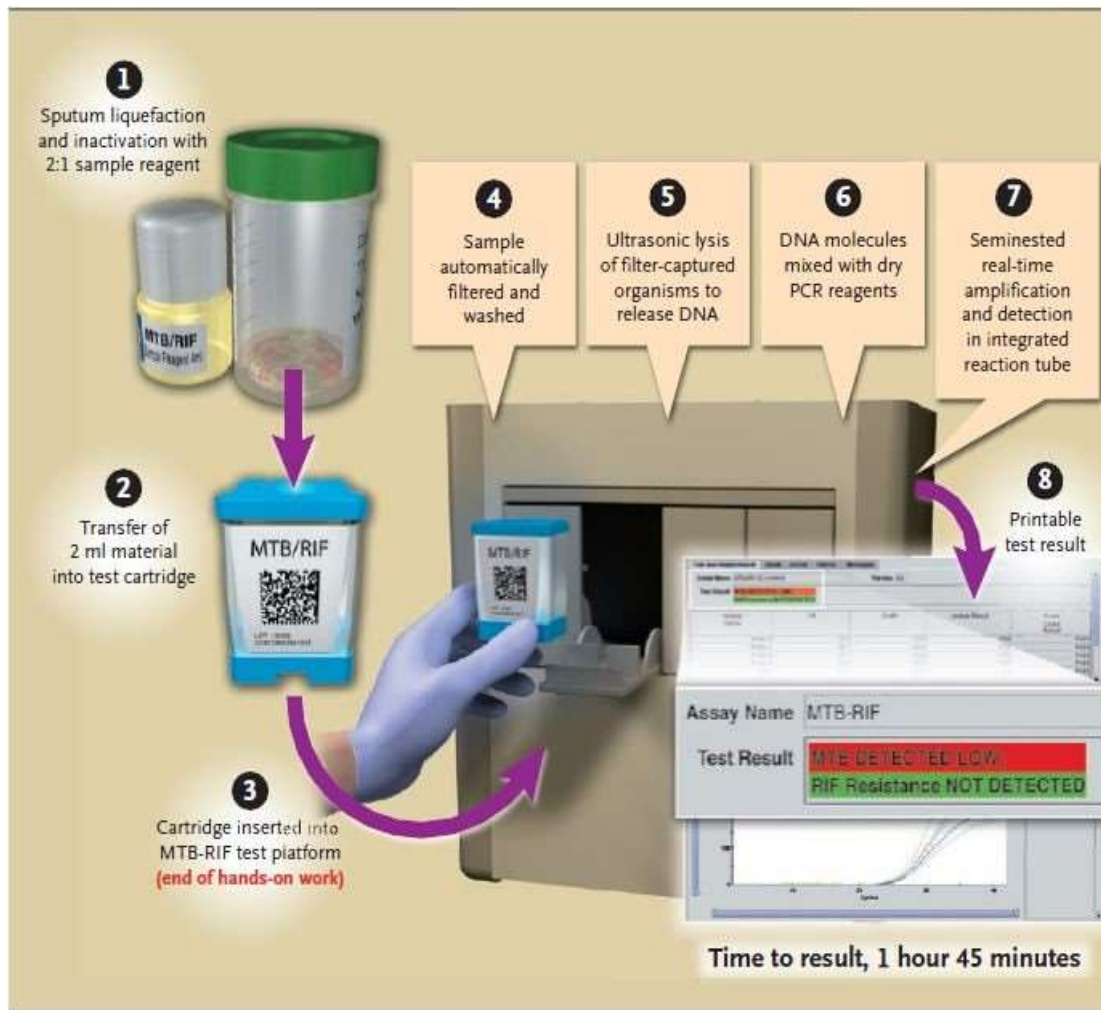


Figure 1. 5 The Xpert-MTB/RIF assay

Involves hands on (1-3) and automated steps (4-8); Sputum is liquefied and inactivated with sample reagent (1). Sample material is transferred into test cartridge (2). Cartridge is inserted into MTB-RIF test platform (3) Sample undergoes filtration and washing before organisms are lysed to release DNA. The DNA undergoes real-time amplification and detection within an integrated tube (4-8). Adapted (Boehme 2010).

1. 6. 4 Other diagnostic methods

There are several TB diagnostic tests at different stages of development or implementation.

In 2012 the WHO failed to proceed with policy recommendations for the replacement of the smear microscopy with the loop-mediated isothermal amplification (LAMP) platform due to insufficient evidence. Re-valuations of the TB LAMP in a multi-country setting have since been done (WHO 2014). In a recent evaluation in Malawi, a resource-poor setting where this study was conducted, TB LAMP showed sensitivity of 65% and did not vary by HIV status or smear status (Nliwasa, MacPherson et al. 2016). This sensitivity was significantly lower than that observed in Peru, Bangladesh and Tanzania (93%) (Boehme, Nabeta et al. 2007). Other studies of LAMP have reported higher sensitivities of 87%-97% (Adhikari, Pandey et al. 2009, Iwamoto, Sonobe et al. 2003). The study in Malawi also found the cost of LAMP to be high making it unsuitable in these settings (Nliwasa, MacPherson et al. 2016). However, in 2016 WHO gave a conditional recommendation for TB-LAMP as a replacement test for sputum smear microscopy for diagnosis of TB (WHO 2016).

New DNA probes that are capable of binding to rRNA, and enabling quick diagnosis of TB have been developed. Accuprobe (Gen-Probe, Bedford, USA) was the first and was able to identify the *Mtb* complex, *Mycobacterium avium* complex, *Mycobacterium kansasii* and *Mycobacterium gordonae* with a sensitivity and specificity of more than 90%. Line probe assays that detect different Mycobacterial species including *Mtb* on positive cultures were subsequently introduced. These include INNO-LiPA MYCOBACTERIA version 2 (Innogenetics NV, Ghent, Belgium) and the GenoType

MTBC and Genotype Mycobacterium (Hain Lifesciences, Nehren, Germany) (Molicotti, Bua et al. 2014). Interferon-gamma release assay (IGRA) is a new diagnostic method for latent TB infection based on detection of INF- γ response to TB antigens. These antigens include early secretory antigenic target-6 (ESAT-6), cultures filtrated protein-10 (CFP-10) and another antigen called TB7.7. Region of difference-1 (RD1) which is present in *Mtb* complex but absent in BCG and most non-tuberculous mycobacteria encodes these antigens (Molicotti, Bua et al. 2014, Taki-Eddin, Monem 2011). IGRAs are less prone to false positives caused by non-tuberculous mycobacteria or BCG but are unable to discriminate between latent infection and an active disease (Metcalf, Everett et al. 2011, Molicotti, Bua et al. 2014). Mycobacterial antigens are also detectable in urine of patients with pulmonary TB. One such antigen is the cell wall lipopolysaccharide lipoarabinomannan (LAM) and an enzyme linked immunosorbent assay (ELISA) that detects LAM has been developed. This assay could be crucial in detecting TB in HIV infected patients as it has shown moderate to high sensitivity in such individuals although its sensitivity in HIV negative individuals is poor (Molicotti, Bua et al. 2014). It was recently suggested that urine-LAM assays could be a key diagnostic tool for HIV positive patients with advanced immune-deficiency as LAM enters urine primarily via TB dissemination to the kidneys in blood (Lawn, Gupta-Wright 2016). Additionally, another systematic review demonstrated that LAM detection in urine could be positively correlated with increased risk of death during HIV associated TB treatment (Gupta-Wright, Peters et al. 2016).

The long culture period in *Mtb* remains an obstacle to quick diagnosis of TB. Researchers in the Gambia and UK have developed a TB diagnostic test that is able to detect and

characterize bacteria that cause TB from DNA obtained directly from sputum without the need for culture (Doughty, Sergeant et al. 2014). This approach has previously been applied enables useful clinical and genome-wide sequence data to be obtained (Kay, Sergeant et al. 2014, Wilson, Naccache et al. 2014). This new technique called metagenomics although not delivering sufficient depth of coverage for sequence based sensitivity testing, if improved has the potential to revolutionize TB diagnosis.

1. 7 Treatment of tuberculosis

The first line of treatment for TB is a multi-drug regimen consisting of rifampicin, isoniazid, pyrazinamide, and ethambutol (RHZE) (van den Boogaard, Kibiki et al. 2009) which if it is to be effective (over 85% cure rates for new cases) has to be taken for at least six months. The WHO recommendation is that a patient undergoes two months of RHZE therapy followed by RIF and INH for a further four months (Zumla, Raviglione et al. 2013, WHO 2014). The duration and the complexity of the treatment poses a major challenge for patients leading to non-adherence to treatment (van den Boogaard, Kibiki et al. 2009) . Non-adherence results in treatment failure and relapses and eventually development of drug resistance. In the 1990s, the WHO developed the directly observed therapy short course (DOTS) strategy to better improve response and adherence to TB treatment (van den Boogaard, Kibiki et al. 2009, WHO 1997). Despite being labour and resource intensive DOTS has been shown to produce the highest cure rates for drug sensitive TB (Balabanova, Drobniewski et al. 2006). As a result of this among other interventions, global TB mortality rates (deaths per 100 000 populations per year) and TB incidence rates are on the decline in most parts of the world. Through effective diagnosis and treatment, approximately 37 million lives were saved between 2000 and 2013. With

a goal of dramatically reducing the global burden of TB by 2015 and guided by *Millennium Development Goal 6*, the WHO developed the *Stop TB Strategy* for the period 2006-2015 (WHO 2014). Additionally, the World Health Assembly in 2014 approved the *WHO End TB* strategy with a call to cut TB deaths by 90% and TB incidence rates by 80% by the year 2030 (WHO 2016).

However, non-adherence to treatment (irregular and low dose) can also lead to the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) *Mtb* strains (Zhang, Ying, Yew et al. 2012). Drug resistant TB is usually managed using aminoglycosides and fluoroquinolones among other drugs (Flandrois, Lina et al. 2014). The WHO recommended regimens for MDR-TB may last up to 24 months with success rates of only 65% (WHO 2014). A short course of nine months (from 24 months) using up to seven drugs based on experience in Bangladesh (Aung, Van Deun et al. 2014) was recently approved for treatment of MDR-TB (WHO 2016) and up to 35 countries have since adopted the this for selected patients (WHO 2017b).

1. 7. 1 Drug resistant tuberculosis (DR-TB)

Mtb may undergo random spontaneous chromosomal mutations at a low frequency which renders them resistant to a specific drug or a drug class (Ando, H., Miyoshi-Akiyama et al. 2014, Flandrois, Lina et al. 2014, Zumla, Raviglione et al. 2013). Improper use of drugs such as low doses or poor-adherence to therapy enables selection pressures which may result in emergence of resistant mutants (acquired resistance). Infection with a drug resistant strain (primary resistance) may also result in infection and eventually disease (Andrews, Sarita Shah et al. 2007, Zumla, Raviglione et al. 2013). Chromosomal mutations in the *rpoB* gene confer resistance to RIF, while those in *katG* and the *mabA*

(*fabG1*)-*inhA* operon confer resistance to INH, in *pncA* to PZA, in *rrs* to amikacin (AMK) and in *gyrA* and *gyrB* to fluoroquinolones (FQs) (Flandrois, Lina et al. 2014, Musser, J. M. 1995) .

MDR-TB is defined as resistance to at least INH and RIF while XDR- TB (reported by up to 100 countries) is defined as, in addition to MDR-TB, resistance to any fluoroquinolone and at least one second line injectable drug-amikacin, kanamycin or capreomycin (Andrews, Sarita Shah et al. 2007, Dheda, Shean et al. 2010, Garzelli, Rindi 2012, Günther 2014, WHO 2014). These definitions of MDR-TB and XDR-TB are based on treatment outcomes (success rate compared to failure). In one study, Falzon et al. analyzed outcomes from 7000 patients and reported success rates of 64% for MDR-TB patients and 40% for XDR-TB patients (Falzon, Gandhi et al. 2013, Matteelli, Roggi et al. 2014). Available data from various studies in South Africa indicate that XDR-TB arises predominantly in HIV- infected individuals who have an increased susceptibility for TB (Dheda, Shean et al. 2010). In other countries, particularly the former Soviet states the epidemiological situation is different and the increased prevalence in XDR-TB relates to imprisonment and poor infections control and treatment practices (WHO 2017a).

1. 7. 2 Rifampicin resistance

As a semi-synthetic derivative of rifamycin with very high bactericidal activity, RIF is used as a first line drug for TB treatment (Ramaswamy, Musser 1998). Although the bactericidal activity of RIF is less than that of INH, RIF has longer-term sterilizing activity and is the key drug in preventing relapses during short-course therapy (Paramasivan, Sulochana et al. 2005). Extensive studies on RIF activity and resistance have been done on *Escherichia coli* (McClure, Cech 1978). RIF is a lipophilic ansamycin

that target the β -subunit of the RNA polymerase inhibiting transcription initiation. Resistance to RIF occurs as a result of mutations in the RNA polymerase beta subunit (*rpoB*) gene that encodes the β -subunit of the RNA polymerase. These mutations result in conformational changes that lead to low affinity for the drug (Shakya, Garg et al. 2012).. Data from various studies has shown that 96% of RIF resistant *Mtb* clinical isolates have mutations in the 81bp ‘hot spot’ region of the *rpoB* gene (Shakya, Garg et al. 2012). This region also called the rifampicin resistance determining region (RRDR) comprises *rpoB* codons 507-533 encoding 27 amino acids. (Almeida Da Silva, Palomino 2011, Engström, Morcillo et al. 2012, Ramaswamy, Musser 1998). The commonest changes however, are missense mutations occurring at codons 526 or 531 (Caws, Duy et al. 2006, Musser, J. M. 1995) and in one study 80% of resistant strains had mutations at these positions (Musser, J. M. 1995). Intriguingly some RIF resistant *Mtb* clinical isolates have resistance arising from mutations outside the hotspot region. A mutation, A191C in the gene, Rv2629 was initially associated with RIF resistance but has been found to be related to the *W-Beijing* genotype but not resistance to RIF. Furthermore, almost all RIF resistant clinical isolates have shown resistance to other drugs particularly INH suggesting RIF resistance could be used as a surrogate molecular marker for MDR (Almeida Da Silva, Palomino 2011, Caws, Duy et al. 2006).

1. 7. 3 Isoniazid resistance

INH (isonicotinic acid hydrazide) is a synthetic bactericidal antibiotic used worldwide as a first line drug to treat infections caused by *M. tuberculosis* complex members (Ramaswamy, Musser 1998). INH is highly effective against TB (minimum inhibitory concentrations (MIC) 0.02-0.2 μ g/mL) but is only active against metabolically active

replicating bacilli (Ando, H., Miyoshi-Akiyama et al. 2014, Zhang, Y., Yew 2009). INH is a prodrug that is activated by the catalase-peroxidase enzyme KatG, encoded by the *katG* gene (Ando, H., Miyoshi-Akiyama et al. 2014, Zhang, Y., Heym et al. 1992). In its active form INH inhibits production of essential mycolic acids by blocking the NAD-dependent enoyl-acyl carrier protein reductase InhA (Rawat, Whitty et al. 2003, Vilchèze, Av-Gay et al. 2008). Clinical isolates resistant to INH have often been shown to lose the catalase and peroxidase activities of KatG (Zhang, Y., Yew 2009). Some actively dividing *Mycobacterium smegmatis* cells express low levels of KatG, the enzyme that activates the pro-drug INH. Such cells are unable to activate INH and display classical resistance (Wakamoto, Dhar et al. 2013). They escape the killing effect of the drug. The S315T mutation in the *katG* is the most common mutation conferring INH resistance. In one study out of 85 INH resistant strains, 49 (85%) epidemiologically unassociated isolates had changes in amino acid 315 (Ser)(Musser, James M., Kapur et al. 1996). Furthermore this mutation has been consistently associated with high level resistance (MIC>1µg/mL) to INH (Andres, Hillemann et al. 2014, Palomino, J. C., Anandi 2014). Mutations in the *inhA* gene (lowering InhA affinity for the gene product InhA for activated INH) and mutations in the *mabA-inhA* promoter region (causing overexpression of InhA) also confer resistance to INH (Seddon, Furin et al. 2012). These are however, less frequent than *katG* mutations (Ando, H., Miyoshi-Akiyama et al. 2014, Hazbón, Brimacombe et al. 2006, Zhang, Y., Yew 2009). INH resistance is a rather complex process as several genes including *katG*, *ahpC*, *inhA*, *kasA* and *ndh* have all been associated with resistance (Zhang, Y., Heym et al. 1992). An analysis by Hazbón et al. of 240 alleles previously associated with INH resistance found that mutations in *katG*, *inhA* and *ahpC*, but not *kasA* were strongly

associated with resistance (Hazbón, Brimacombe et al. 2006). Such conflicting evidence suggests that INH resistance is an area requiring further investigation. By analyzing genes in the laboratory strain *H37Rv*, Ando et al. recently demonstrated that a silent mutation in the *mabA* also confers resistance to INH through upregulation of *inhA* (Ando, H., Miyoshi-Akiyama et al. 2014, Palomino, J. C., Anandi 2014). Ethionamide (ETH) is a structural analog of INH used as a second line drug against TB (Ramaswamy, Musser 1998, Vilchèze, Av-Gay et al. 2008). INH and ETH are said to share a molecular target as low level INH resistance has been correlated with co-acquisition of ETH resistance (Ramaswamy, Musser 1998, Winder, Collins 1970). Interestingly a recent study by Machado et al. found that mutations in the *inhA* regulatory region and in the *inhA* coding region confers high level INH resistance and cross-resistance to ETH in MDR-TB (Machado, Perdigão et al. 2013, Palomino, J. C., Anandi 2014).

1. 7. 4 Resistance to Fluoroquinolones

Fluoroquinolones (FQs) are an important class of broad spectrum antibacterial agents used as second line treatment against TB (Fàbrega, Madurga et al. 2009). Nalidixic acid, a compound discovered as a byproduct of the antimalarial drug chloroquine has been used to produce several generations of novel fluoroquinolones including ciprofloxacin, ofloxacin, moxifloxacin and levofloxacin (GOSS, DEITZ et al. 1965). Studies of quinolone activity in bacterial species have found that these antimicrobial agents inhibit topoisomerase II (DNA gyrase) and topoisomerase IV, two enzymes critical for bacterial viability (Fàbrega, Madurga et al. 2009, Palomino, J. C., Anandi 2014). In *Mtb* only topoisomerase II is present and is therefore the only target for fluoroquinolone activity (Aubry, Pan et al. 2004). DNA gyrase is composed of two A and two B subunits encoded

by *gyrA* and *gyrB* respectively (Musser, J. M. 1995) and its function is to catalyze DNA supercoiling (Takiff, Salazar et al. 1994). It uses energy from ATP to introduce negative supercoils into DNA. Introduction of negative supercoils in DNA is essential for RNA polymerase to introduce the transcription process (Fàbrega, Madurga et al. 2009). Resistance to fluoroquinolones in *Mtb* develops as a result of mutations in the Quinolone-resistance determining region (QRDR) of *gyrA* or *gyrB* (Palomino, J. C., Anandi 2014). Positions 90 and 94 of *gyrA* have the most frequent mutations but other studies have also reported mutations at positions 74, 88 and 91. Mutations in *gyrA* result in high level resistance against *Mtb* (Cheng, Yew et al. 2004). Contrastingly, mutations in *gyrB* are responsible for lower level resistance to *Mtb* (Takiff, Salazar et al. 1994, Maruri, Sterling et al. 2012). Although efflux determinants of FQ resistance in bacteria have previously been described (Pasca, Gugliera et al. 2004, Poole 2005), Lui et al. recently used expression profile analysis to demonstrate that high level FQ resistance may also be conferred by efflux pump mechanisms in *Mtb* (Lu, Liu et al. 2014).

1. 7. 5 Resistance to Kanamycin, Amikacin, Capreomycin and Viomycin

These four antibiotics are second-line drugs used in the management of MDR-TB. All four have the same mechanism of inhibiting protein synthesis by blocking normal function of ribosomes (Ramaswamy, Musser 1998). Kanamycin and amikacin bind to the 16S rRNA. Mutations at position 1400 and 1401 of the *rrs* gene are the most common in kanamycin resistant strains, conferring high level resistance to both kanamycin and amikacin (Palomino, J. C., Anandi 2014). Some studies have however reported mutations at position 1483 (Alangaden, Kreiswirth et al. 1998, Suzuki, Katsukawa et al. 1998). Up to 80% of clinical isolates in other studies have mutations showing low level resistance to

kanamycin but not amikacin. Such mutations occur at positions -10 and -35 of the *ies* promoter leading to overexpression of the protein (Campbell, Morlock et al. 2011, Palomino, J. C., Anandi 2014, Zaunbrecher, Sikes et al. 2009). The *ies* gene encodes an aminoglycoside acetyltransferase, *ies* (Zaunbrecher, Sikes et al. 2009). These data suggest that there is no complete cross-resistance between kanamycin and amikacin. Some resistant strains have not shown known mutations suggesting resistance to kanamycin and amikacin which could also be a result of as yet undetermined mechanisms (Sowajassatakul, Prammananan et al. 2014).

Capreomycin and viomycin share a similar structure and target the same site on the ribosome, causing interface of the small and large subunits (Palomino, J. C., Anandi 2014). They belong to the tuberactinomycin family of antibiotics, regarded among the most effective class of antibiotics against MDR-TB (Stanley, Blaha et al. 2010). Mutations in the *tylA* gene inhibit methylation of rRNA consequently conferring resistance to both capreomycin and viomycin. Full cross resistance between capreomycin and viomycin has also been previously reported (Palomino, J. C., Anandi 2014).

1. 8 The Mycobacterium tuberculosis genome

Sequencing of the entire *Mtb* genome was based on a reference strain, H37Rv which revealed that the genome contains 4 411 329 base pairs (Cole, S. T., Brosch et al. 1998) and around 4000 genes (Behr, Wilson et al. 1999, Davies, P. D. O., Barnes et al. 2008). In comparison to other bacteria, *Mtb* has a larger genome although it is smaller than that of *Escherichia coli* (*E. coli*). It has a uniformly high content of guanine + cytosine (65.6%) indicating there has been minimal incorporation of foreign DNA during its evolution. A large number of genes in *Mtb*, (250 compared to 50 in *E.coli*) code for lipid metabolism

enzymes. Studies have detected all known lipid biosynthetic pathways encountered elsewhere in nature within *Mtb*. The enzymes are largely involved in the synthesis of complex lipid-rich mycobacterial cell walls (Davies, P. D. O., Barnes et al. 2008). One unique feature of the *Mtb* genome is the large number of the genes it contains, up to 10% of the total coding potential containing polymorphic guanine-cytosine repetitive sequences (PGRS). Two unrelated families proline-glutamic acid (PE) and proline-proline glutamic acid (PPE) of the acidic glycine-rich proteins are coded for by these genes. Specific functions of the genes and their proteins remain unclear (Cole, S. T., Brosch et al. 1998, Cole, Stewart T. 2002) although they have recently been implicated in immune evasion and virulence (Fishbein, van Wyk et al. 2015). Consistent evidence has previously suggested that proteins located in the cell wall and cell membranes are responsible for diversity in antigenic structure and virulence. This greatly contributes to *Mtb* evolution and adaptation of its various members to different hosts (Brennan, Delogu 2002, Filliol, Motiwala et al. 2006).

1. 8. 1 Genotyping of Mycobacterium tuberculosis

In order to effectively control the transmission of *Mtb* there is a need to identify and treat individuals with active disease. Most transmission however may occur before such individuals are identified. It becomes crucial to track specific strains of *Mtb* as they are transmitted in a community and understand their transmission dynamics and how they may impact on TB control activities. Several methods are currently in use which can help to identify the source of infection including route of transmission and dissemination. Genotyping allows distinction between different strains of *Mtb*. Additionally, genomics has been critical in many areas including *Mtb* evolution (Gagneux 2012) drug discovery

(Lechartier, Rybniker et al. 2014) and more recently prediction of resistance (Helb, Jones et al. 2010). A project that seeks to use WGS to identify all drug resistance genes and correlate them with laboratory phenotype and clinical outcomes has been established at University of Oxford (University of Oxford 2016). Once completed the project will ensure not only quick TB diagnosis and treatment but also the right treatment for patients with drug resistant strains.

1. 8. 2 Genotyping methods

Some of the common methods that clinical researchers use to classify the *Mtb* complex into groups of related strains are IS6110 restriction fragment length polymorphism (IS6110 RFLP), mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) typing, spacer oligonucleotide typing (spoligotyping), targeting of large sequence polymorphisms (LSPs) and use of single nucleotide polymorphisms (SNPs) (Müller, Roberts et al. 2014).

1. 8. 3 IS6110-based restriction fragment length polymorphisms

The method uses restriction enzymes to cut mycobacterial DNA at sites of the common IS6110 insertion sequence once, resulting in DNA restriction fragments of varying lengths that can be separated by gel electrophoresis (van Soolingen, Hermans et al. 1991).

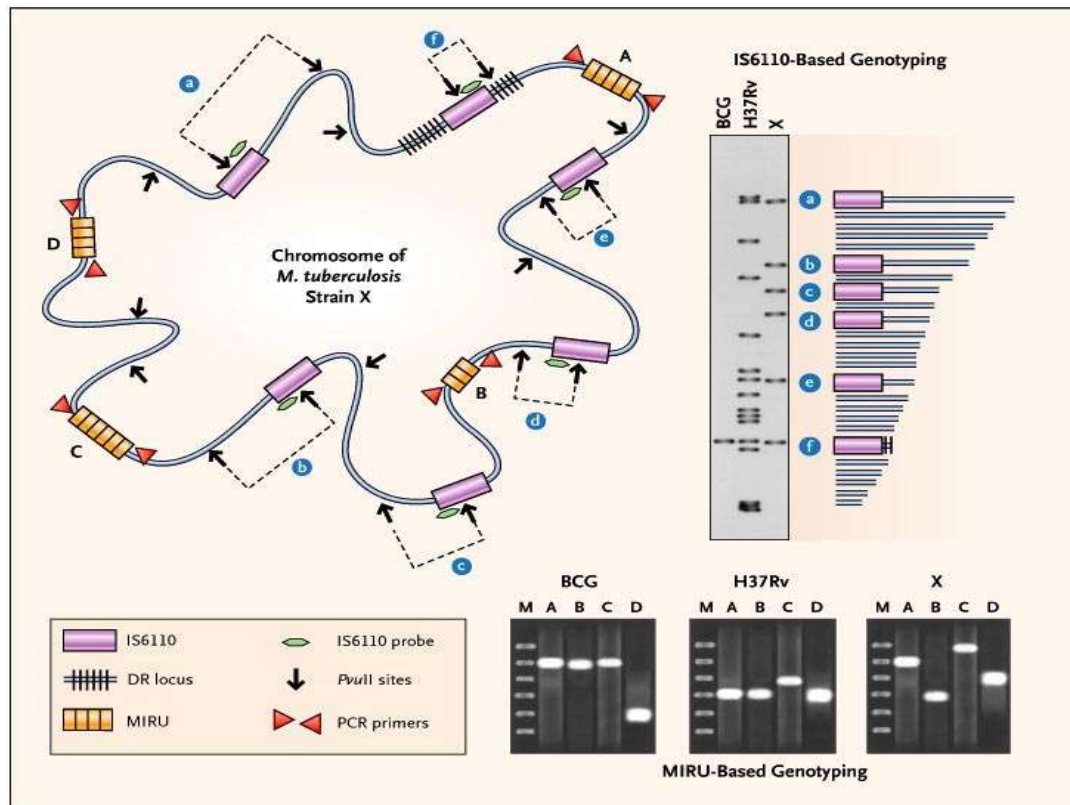


Figure 1. 6 Chromosome of *Mycobacterium tuberculosis* strain

IS6110 Insertion Sequences and Mycobacterial Interspersed Repetitive Units (MIRUs) typing of (BCG), the *M. tuberculosis* Reference organism H37Rv, and Strain X. Chromosome of Strain X (top left). Results of IS6110 genotyping (top right); Mycobacterial DNA is digested with the restriction enzyme *PvuII*. The IS6110 probe hybridizes to IS6110 DNA to the right of the *PvuII* site in IS6110. The size of each hybridizing fragment depends on the distance from this site to the next *PvuII* site in adjacent DNA (fragments a through f). Results are shown by gel electrophoresis of the DNA fragments of BCG, H37Rv, and X. Results of MIRU- based typing (bottom three panels). Sizes of molecular-weight markers (M) and PCR products for the loci A, B, C, and D in BCG, H37Rv, and X are shown. Adapted from (Barnes, Cave 2003)

IS6110 complementary probes hybridize to IS6110 DNA and these DNA complexes form a particular band pattern on the gel. The insertion sequence IS6110 is unique to *Mtb* and is dispersed randomly throughout the genome although areas of common insertion of IS6110 have been reported (Thierry, Cave et al. 1990). The method is highly discriminatory, reproducible and has historically been the gold standard for genotyping *Mtb* strains. One limitation of the method is that it requires large amounts of high quality DNA (2µg). Furthermore there is need for expensive computer software and experienced technical personnel. In addition the discriminatory power of this method is inadequate for strains with an IS6110 copy number of 6 or less (Jagielski, van Ingen et al. 2014).

1. 8. 4 Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU/VNTR)

MIRU/VNTR is a very accurate PCR-based method. It describes the number and size of the variable number tandem repeats (VNTR) in each of the 12 independent genetic loci termed mycobacterial independent repetitive units (MIRU) (Supply, P., Lesjean et al. 2001). Specific MIRU loci are PCR amplified and followed by determination of amplicon sizes by gel electrophoresis (Jagielski, van Ingen et al. 2014). The method can be used on strains regardless of their IS6110 RFLP copy number as MIRUs are more conserved. It is automation friendly and so can be used for genotyping large number of samples and is amenable to quick comparison of results between laboratories. MIRU-VNTR has lower discriminatory power than IS6110 based RFLP when using 12 loci but when analyzing strains with lower copy number of the IS6110 sequence the power is higher (Davies, P. D. O., Barnes et al. 2008).

1. 8. 5 Spacer oligonucleotide typing (Spoligotyping)

Spacer oligonucleotide (spoligotyping) is a PCR- based method based on a particular genomic region called the Direct Repeat (DR) locus which comprises a series of well conserved 36bp direct repeats. DRs are interspersed with unique non-repetitive spacer sequences of 34-41bp. Following PCR amplification, products of different sizes are hybridized to a membrane with 43 covalently bound synthetic oligonucleotides representing the polymorphic spacers identified in *Mtb* H37Rv (spacers 1-19,22-32,and 37-43) and in *M.bovis* BCG (spacer 20-22 and 33-36) (Kamerbeek, Schouls et al. 1997). Spoligotyping is simple, cheap, quick and easily reproducible. Although the method can identify *Mtb* complex isolates up to the (sub) species level (Jagielski, van Ingen et al. 2014), it is less discriminatory than IS6110 typing and is usually applied as a secondary or a primary method parallel with a more powerful genotyping method (Davies, P. D. O., Barnes et al. 2008). The reason for an inferior discriminatory power of spoligotyping is that it targets only a single genetic locus of less than 0.1% of the *Mtb* complex genome (Jagielski, van Ingen et al. 2014).

1. 8. 6 Other genotyping methods

Polymorphic GC-rich repetitive sequences (PGRS) occur abundantly and at multiple sites within the *Mtb* complex genome. Several repeats of a 9bp consensus sequence (5'-CGGCGGCAA-3') are arranged in segments of up to 1.5kb (Ross, Raios et al. 1992). The number and distribution of the PGRS is used to distinguish *Mtb* strains (Jagielski, van Ingen et al. 2014). Repetitive Sequence - Based-(rep-) PCR uses primers to generate amplified fragments that produce a fingerprint pattern upon gel electrophoresis (Sechi, Zanetti et al. 1998, Wojtasik, Majchrzak et al. 2011). Single Nucleotide Polymorphisms,

SNPs are a reliable marker for *Mtb* complex lineage classification. Non-synonymous SNPs (ns-SNPs) when present introduce amino acid changes to proteins and influence phenotype and selection pressures (drug resistance in *Mtb* is associated with ns-SNPs) while synonymous SNPs (s-SNPs) do not alter amino acid profiles (Ramaswamy, Musser 1998). The only major limitation of SNP typing is the large number of genes required to achieve satisfactory resolution (Homolka, Projahn et al. 2012). Genomic deletions within *Mtb* lead to large sequence polymorphisms (LSPs). Through sequencing, *Mtb* strains can be distinguished based on genomic deletions (Fleischmann, Alland et al. 2002). While LSP analysis targets only a small section of the genome usually <1%, sequencing covers the entire genome.

1.9 Bacterial persistence

The problem with antibiotic resistance in TB is that it has placed pressure on scientists in the pursuit of novel antimicrobials to tackle the issue. One major challenge however, is that there has not been attention paid to persister cells during TB therapy. Persisters have been described as cells that survive prolonged bacterial antibiotic exposure despite being genetically susceptible (Balaban, Gerdes et al. 2013). Persisters have largely been ignored in part due to the presumption that all pathogens that evade antibiotics eventually get eliminated by host immunity (Gefen, Balaban 2009). Furthermore, persisters have been difficult to study because of their low frequencies in a naïve bacterial population (10^{-4} to 10^{-6} in *Escherichia coli*) and the persister phenotype is not genetically inherited in passage (Dhar, McKinney 2007). As early as 1942 it was observed that *Staphylococcus aureus* infections could recur even following high doses of penicillin. Further penicillin could kill 99% of streptococcal culture leaving 1% intact (Hobby, Meyer et al. 1942). Joseph Bigger

later noted that within a clonal population of *Staphylococcus aureus*, there was phenotypic variation, with a subpopulation that remains tolerant to antibiotics (Bigger 1944). He termed this population of rare bacterial cells “persisters” and evaluated them at about one per million (Gefen, Balaban 2009). This phenotypic variation is not unique to bacteria as it has also been observed in fungi (Dhar, McKinney 2007). Unlike bacterial resistance which is attributable to heritable genetic mutations, bacterial persistence is a transient phenotype observed following exposure to antibiotics (Lewis 2010). When a bacterial population is exposed to a bactericidal antibiotic, quantification of viable bacteria at different time points can be achieved. The resulting killing curves are generally biphasic, with a majority of the cells rapidly killed before the killing rate decreases dramatically. The tail of the killing curve has a subpopulation of rare cells that survive antibiotic exposure (Lewis 2010). Persister cells do not proliferate in the presence of antibiotics but removal of the bactericidal agent leads to a random switch back to the growing state. Consequently removal of the antibiotic gives rise to a population that is as sensitive as the original and produces a small proportion of persisters (Maisonneuve, Etienne, Gerdes 2014). Persisters therefore possess non-heritable phenotypic antibiotic tolerance and produce progeny that are fully susceptible to antibiotics upon regrowth (Zhang, Ying, Yew et al. 2012). This suggests that persistence is an epigenetic rather than a genetic trait (Korch, S. B., Hill 2006, Lewis 2010). Evidence from other studies however suggests that persisters are present in a population prior to antibiotic exposure. In one study phenotypic switching occurred between normally growing *Escherichia coli* cells and persisters with reduced growth rates (Balaban, Merrin et al. 2004). It has now become clear that

persistence make treatment of chronic infections such as tuberculosis more difficult to manage and may be responsible for disease relapses (Gefen, Balaban 2009).

1. 9. 1 Mechanisms generating persister cells

Specific mechanisms responsible for persistence remain largely unknown. There have been questions as to whether persistence is indeed only as a result of antibiotic exposure or whether there may be other drug independent stochastic mechanisms. Balaban classified persisters into two groups: type I persisters (non-growing) that require an external trigger to differentiate into an antibiotic tolerant state and type II persisters (slowly growing) that are continuously generated in the absence of an external trigger (Balaban, Gerdes et al. 2013, Balaban, Merrin et al. 2004). Type II persisters are capable of regaining the non-persister phenotype and increase their population in the log phase (Zhang, Y. 2014). In type I persisters the suggestion is that unfavorable environmental conditions may be sensed by a small fraction of the cells which then shuts down into a dormant state. This is used as a bet-hedging strategy where one group of cells continues to proliferate and risk deaths while others shut down to avoid the unfavorable conditions (Balaban, Gerdes et al. 2013). Antibiotic exposure is just one of those stressful external triggers for type I persister formation. Zhang however suggests that type I and type II persisters are not easy to classify as both types contain heterogeneous persisters cells themselves. Furthermore both type I and type II persisters may interconvert (Zhang, Y. 2014)

1. 9. 2 Genetic basis of persistence

Moyed's pioneer study first suggested that persistence could have a genetic basis (Moyed, Bertrand 1983). Stable *hip* (high persister) mutants were generated by exposing a

population of chemically mutagenized bacteria to high doses of antibiotic (Moyed, Bertrand 1983, Wolfson, Hooper et al. 1990). A gain of function *hipA7* allele which enhanced persistence 1000 fold was identified. This demonstrated that a heritable mutation could increase the level of persistence (Maisonneuve, Etienne, Gerdes 2014). The *hipA7* allele carries two point mutations in *hipA*, a gene of 440 codons (Black, Kelly et al. 1991, Moyed, Broderick 1986). The *hipA* gene is upstream of *hipB* which is a transcription regulator of the *hipBA* operon (Black, Kelly et al. 1991). In *E.coli*, high production of HipA inhibited DNA replication, transcription and translation, consequently inhibiting cell growth and increasing tolerance to antibiotics (Korch, S. B., Hill 2006). Consistently HipB directly inhibited HipA activity suggesting *hipBA* constitutes a toxin-antitoxin (TA) locus (Korch, Shaleen B., Henderson et al. 2003). In addition to *hipA*, mutagenesis has been used to identify and isolate several other persistence related genes such as *relA*, *phoU*, *sucB* and *ubiF* (Zhang, Y. 2014). Mutations in persister related genes have caused either increased or decreased persistence. Mutations mapping to *hipA*, *metG*, *tktA*, and *glpD* cause increased persistence while those in *relA*, *phoU*, *sucB* and *ubiF* cause decreased persistence (Zhang, Y. 2014). A more recent study has demonstrated that tolerance mutations pave a way for rapid evolution of resistance (Levin-Reisman, Ronin et al. 2017). In the Levin-Reisman study, intermittent exposure of *E.coli* cultures to ampicillin led to development of genotypic resistance which was preceded by an increase in MIC, extended lag times and emergence of tolerance mutations.

1. 9. 3 Toxin and Antitoxin (TA) module

Several studies have demonstrated that persistence in bacteria may be based on the TA module (Dorr, Vulic et al. 2010, Maisonneuve, E., Shakespeare et al. 2011). TA modules

consist of gene pairs where one gene acts as a “toxin” by increasing the frequency of persisters while the other cancels out its effect (Rotem, Loinger et al. 2010). In a previous study, Keren et al. demonstrated that the frequency of persistence in *E.coli* was greatly increased due to expression of *relE* toxin in excess of antitoxin *relB*, suggesting that TA modules could be mediators of persistence (Keren, Shah et al. 2004). Additionally the number of TA systems in an organism directly determines the frequency of persisters in a population (Maisonneuve, E., Shakespeare et al. 2011). The HipA toxin and the HipB antitoxin form a tight complex and these two gene products are regulated by the HipBA module. Overproduction of the toxin inhibits cell growth and this is counteracted by activity of the antitoxin (Korch, S. B., Hill 2006, Keren, Shah et al. 2004, Gerdes, Maisonneuve 2012). Studies have shown that HipA is a serine/threonine kinase having the kinase activity necessary for enhancement of both cell growth arrest and persister cell formation (Correia, D'Onofrio et al. 2006). HipA phosphorylates elongation factor-thermal unstable (EF-TU) leading to growth arrest (Correia, D'Onofrio et al. 2006, Keren, Shah et al. 2004). It was previously discovered that HipA attenuated glutamyl tRNA synthetase (GltX) via phosphorylation (Germain, Castro-Roa et al. 2013). Inhibition of GltX triggers synthesis of guanosine tetra-phosphate (p)ppGpp, a signaling nucleotide. In turn overproduction of (p)ppGpp dramatically enhanced persistence (Germain, Castro-Roa et al. 2013). (p)ppGpp induces slow growth in rare cells by activation of TA loci through a regulatory cascade that depends on inorganic polyphosphate (PolyP) and Lon protease (Kåhrström 2013, Maisonneuve, E., Castro-Camargo et al. 2013). Consistent with this finding another study has demonstrated that the stringent response is activated due to attenuation of GltX by its phosphorylation. Inactivation of GltX leads to overexpression

of *RelA* and increases (p)ppGpp levels. Notably, expression of GltX reverses effects of HipA and prevents persister formation (Kaspy, Rotem et al. 2013). Others have suggested that (p)ppGpp synthesis may be triggered by a micro-starvation environment created in exponentially growing cells (Maisonneuve, Etienne, Gerdes 2014). TAs have been classified into types I to V based on nature of TA activity regulation and genetic architecture (Masuda, Tan et al. 2012). Of particular interest is the role of a type II TA locus which has been highly associated with persister cell formation. Type II loci encode an antitoxin that interacts with and attenuates the toxin (Masuda, Tan et al. 2012). Several studies have demonstrated that high production of type II toxins very efficiently lead to growth arrest in cells which could again be rapidly resuscitated by activation of related antitoxin genes (Christensen-Dalsgaard, Gerdes 2006, Pedersen, Christensen et al. 2002). Evidently the most up-regulated transcripts in the persister fraction obtained from the *hipA7* locus were TA mRNAs, further supporting the finding that TA genes regulate persistence (Keren, Shah et al. 2004). Interestingly TA modules have also been associated with tolerance by lag (*tbl*) genes. Notably *tbl* genes have been implicated in the extended lag time of bacteria before regrowth following exposure to antibiotics (Fridman, Goldberg et al. 2014).

1. 10 Mycobacterium tuberculosis persistence

Current TB therapy takes a minimum of six months. Some patients are unable or unwilling to complete treatment due to the prolonged duration (Sahbazian, Weis 2005). Poor adherence increases the risk of emergence of multi-drug resistant TB (MDR-TB) and extensively drug resistant TB (XDR-TB) (Zhang, Ying, Yew et al. 2012). In order to prevent emergence of resistance, this treatment must combine at least three drugs (Davies,

G. R. 2013). The prolonged treatment is however, necessitated in part because of Mycobacterial persistence (Zhang, Ying, Yew et al. 2012). Current TB drugs do not possess sufficient sterilizing ability against persister cells. Clearly new approaches are needed to develop ultra-short drug regimens to combat TB.

1. 10. 1 The special populations hypothesis and *Mtb* metabolic activity

Early work by Dennis Mitchison proposed that an *Mtb* lesion may constitute four bacterial sub-populations at varying metabolic states (Mitchison, D. A. 2004, Mitchison, D. A. 1979). These distinct populations are targeted individually by different drugs based on their physiological states resulting in a biphasic bacillary elimination. A metabolically active replicating population are rapidly killed by isoniazid during an ‘early bactericidal phase’. The surviving bacilli are in different states of non-replicating persistence (NRP) and isoniazid is ineffective against these (Mitchison, D. A. 1985, BLOCH, SEGAL 1956). During this phase, up to 2 months, NRPs with intermittent spurts of metabolism are targeted by pyrazinamide (Mitchison, D. A. 1985). Metabolically quiescent bacilli are targeted by rifampicin in a slow prolonged ‘sterilization phase’ suggesting a shorter regimen is achievable with an improved activity against NRP bacilli. Further studies revealed a synergy between pyrazinamide and rifampicin and that while pyrazinamide could be used only during the initial bacterial sterilizing phase, rifampicin was effective throughout the treatment period (Mitchison, D., Davies 2012). The modern short course chemotherapy is therefore credited to a series of British Medical Research Council (BMRC) clinical trials in East Africa, Hong Kong and Singapore (Fox, Ellard et al. 1999). A few more clinical trials have since been conducted in attempts to further shorten the treatment duration to four months.

Moxifloxacin and gatifloxacin are two antibiotics that have shown high sterilizing potential against mycobacteria *in vitro* (Johnson, Hadad et al. 2006). However, a recent clinical trial was abandoned prematurely as the two drugs failed to demonstrate non inferiority to the standard 6 month treatment (Jawahar, Banurekha et al. 2013). Similarly 4-month phase three clinical trials of two Moxifloxacin containing regimens failed to demonstrate non-inferiority with the standard 6-month regimen (Gillespie, Crook et al. 2014). Merle et al. compared the standard 6-month regimen to a four-month regimen in which gatifloxacin was substituted for ethambutol (Merle, Fielding et al. 2014). In this trial non-inferiority of the 4-month regimen to the standard regimen could not be demonstrated as it was characterized by higher recurrences as compared to the standard regimen. Development of shorter and more effective drugs against TB therefore, hinges on understanding mechanisms and pathways that underlie persister *Mtb* cells.

1. 10. 2 Laboratory investigations into *Mycobacterium tuberculosis* persistence

Well known animal or *in vitro* models to study *Mtb* persistence are not in abundance, making study of persistence more challenging. Models are however available that may simulate one or two aspects of persistence (Wang, Wang et al. 2011). McDermott demonstrated using the Cornell mouse model that persisters were responsible for the failure of drugs to completely kill supposedly *in vitro* drug susceptible *Mtb* in the body. This murine model has since been used in several other studies to demonstrate *Mtb* persistence (MCCUNE, R. M., Jr, TOMPSETT 1956, MCCUNE, R. M., Jr, MCDERMOTT et al. 1956, McCune, Robert M., Feldmann et al. 1966, McDermott 1958). Studies with guinea pigs demonstrate disease arrest during the initial acute phase of bacterial replication in *Mtb* infection. Notwithstanding the fact that guinea pigs display

pathological features of the disease seen in humans, they have extremely quick disease progression (McMurray 2001). Persistent infections have also been demonstrated in the cynomolgus monkey a non-human primate model (Flynn, Capuano et al. 2003, Walsh, Tan et al. 1996). Although infections closely mimic those of humans, use is limited by high costs (Monack, Mueller et al. 2004).

The hypoxic “Wayne model” is a well-recognised *in vitro* model that demonstrates entry into a non-replicating persistent state (NRP) under anaerobic conditions for an *Mtb* culture. The model is based on generating non-replicating bacilli by putting a low-inoculum culture in a sealed tube and stirred until all the oxygen is consumed (Wayne, Hayes 1996). Subsequently when such cultures are re-exposed to aerobic conditions, they resume synchronous replication, suggesting they could have been arrested at various stages of the cell cycle (Wayne, Sohaskey 2001). Several studies have attempted to demonstrate an *in vitro* model of the human granuloma, the hallmark of *Mtb* persistence (Marino, El-Kebir et al. 2011, Puissegur, Botanch et al. 2004, Segovia-Juarez, Ganguli et al. 2004). Others have used the Zebrafish-*Mycobacterium marinum* model to mimic the human granuloma (Clay, Davis et al. 2007). Infected zebrafish develop caseating granulomas with striking similarities in pathology to the human granuloma. These studies were however, unable to demonstrate resuscitation of *Mtb* (Kapoor, Pawar et al. 2013). Later an *in vitro* model in which *Mtb* goes into dormancy and subsequent resuscitation as observed in human TB was reported. Additionally the study demonstrated loss of acid fastness, lipid body accumulation, RIF resistance and gene expression changes for dormant bacilli (Kapoor, Pawar et al. 2013). A true *in vitro* model that fully demonstrates

the persistence phenotype requires multiple stress conditions similar to those encountered in the host (Deb, C., Lee et al. 2009).

Recent studies have however challenged the view that persisters are in a state of growth arrest. It has been suggested that some persisters are in an actively growing state but that efflux pumps detoxify the bacterial cells or lower antibiotic intake (Helaine, Kugelberg 2014). In another study low level production of *KatG* in replicating INH susceptible *Mtb* cells led to drug tolerance akin to persistence (Wakamoto, Dhar et al. 2013). This form of drug tolerance was however found to be drug specific and it is yet to be established if such cells could exhibit multi-drug tolerance.

1. 10. 3 Mechanisms of persister formation in *Mycobacterium tuberculosis*

Studies have identified several mechanisms and pathways believed to be responsible for persister formation and maintenance in *Mtb* (Zhang, Ying, Yew et al. 2012) .

It is now widely accepted that persistent *Mtb* use fatty acids as the major source of carbon and energy (McKinney, Honer zu Bentrup et al. 2000). During period of insufficient supply of pyruvate from glycolysis, β -oxidation of fatty acids becomes the major carbon source (Zahrt 2003). The glyoxylate shunt is necessary for carbon anaplerosis in the Krebs cycle, essential for survival of bacteria (Gomez, McKinney 2004, Gould, Van De Langemheen et al. 2006, Russell 2003). Isocitrate lyase (ICL) is an initial enzyme of the glyoxylate shunt essential for fatty acid metabolism (McKinney, Honer zu Bentrup et al. 2000). In *Mtb*, ICL is encoded by the *icl* gene and is responsible for catalytic conversion of isocitrate to glyoxylate and succinate (Zahrt 2003). There is up regulation of ICL during infection of macrophages by *Mtb* species, suggesting it could have a role in persistence (Parish, Smith et al. 2003, Zahrt, Deretic 2001). Furthermore, mouse studies demonstrated

a marked reduction in persistence and virulence when *icl* was disrupted while bacterial growth remained unaffected (Kumar, Bhakuni 2008).

However, other studies have disputed the fact that ICL plays a role in persistence and that fatty acids are not the sole carbon source (Wang, Wang et al. 2011). *In vitro* studies have shown that high expression of the *gcvB* gene was correlated with *Mtb* entry into persistence. Reductive amination of glyoxylate is catalyzed by glycine dehydrogenase, an enzyme encoded by the *gcvB* gene (Wayne, Lin 1982). It also proved not to be a drug target because there are two isoforms of the enzyme which cannot be completely simultaneously inhibited. A plethora of energy related pathways and genes associated with *Mtb* persistence have been identified. They include *cydC*, essential for energy production under anaerobic conditions and *in vivo* infection and encoding a transporter for cytochrome *bd* assembly (Dhar, McKinney 2010) and *tgs1* (triacylglycerol synthase 1) responsible for fatty acid storage in triglycerides (Deb, C., Lee et al. 2009) among others. *Tgs1* is the most active of enzymes encoded by a set of genes (*tgs1-tgs15*) that have been characterized in *Mtb*. Induction of *tgs1* resulted in both increased triacylglycerol and lipid body content (Garton, Waddell et al. 2008), phenotypes associated with persistence. The gene *tgs1* is a member of the *dosR* regulon which is discussed in more detail below.

In vitro studies using conditions of hypoxia have identified a set of at least 48 co-regulated genes, the *dosR* (dormancy survival regulator) regulon. The *dosR* regulon is activated by hypoxia and by pulmonary “poisons” including nitric oxide (Voskuil, Martin I., Schnappinger et al. 2003) and carbon monoxide (Shiloh, Manzanillo et al. 2008). *DosR* is a two-component transcription regulator encoded by Rv3331c and is necessary for induction of *dosR* regulon genes (Park, H. D., Guinn et al. 2003). Induction of *dosR* is via

phosphorylation by two sensor histidine kinases *dosS* and *dosT* (Roberts, D. M., Liao et al. 2004). Induced *dosR* genes are important for entry and maintenance of dormancy (Leistikow, Morton et al. 2010, Wang, Wang et al. 2011). One such gene which encodes a 16-kDa α -crystallin protein (also called HspX for heat shock protein X, Acr, and Rv2031c) is called the *hspX* gene (Yuan, Crane et al. 1996). *HspX* is the dominant antigen produced by stationary phase cultures *in vitro* (Yuan, Crane et al. 1996) and *in vivo* recognised by most TB patient sera (Park, H. D., Guinn et al. 2003). An *acr* mutant displayed increased growth in mice (*in vivo*) and both in activated and resting macrophages (*in vitro*). Consequently, when *acr* was constitutively expressed, *Mtb* growth slowed suggesting *hspX* plays an active role in slowing growth in *Mtb*. This evidence though remains rather controversial as no difference was observed in lung damage between a *hspX* deletion mutant and the wildtype. (Hu, Movahedzadeh et al. 2006). In other studies *dosR* has been associated with increased expression of *tgsI* and subsequent accumulation of intra cytoplasmic triacylglycerol (TAG) droplets (Daniel, Jaiyanth, Deb et al. 2004, Deb, C., Lee et al. 2009). Conversely deletion of *tgsI* led to a mutant's impaired ability to accumulate TAG (Kapoor, Pawar et al. 2013). Prolonged starvation induces expression of *lipY*, a gene encoding an enzyme responsible for hydrolysis of accumulated TAG stores back into Acetyl CoA for energy production (Deb, Chirajyoti, Daniel et al. 2006). Consistently, deletion mutants of *lipY* displayed compromised ability to mobilize stored TAG (Kapoor, Pawar et al. 2013). Interestingly, others have suggested that the *dosR* regulon is not involved in persister cell formation in *Mtb* as only eight of the ~48 genes were overexpressed in persisters (Keren, Minami et al. 2011).

Table 1. 1 Genes implicated in *Mtb* persistence.

Reference	Gene	Name	Function	Knockout/ expression studies
(McKinney, Honer zu Bentrup et al. 2000)	Rv0476	Icl	Isocitrate lyase	Reduced persistence in mice
(Wayne, Lin 1982)	Rv1832	gcvB	Glycine de carboxylase	Increased expression during hypoxia
(Fritz, Maass et al. 2002)	Rv1161-4	narGHJI	Nitrate reductase	Reduced persistence in mice
(Fritz, Maass et al. 2002)	Rv1736c	narX	Fused nitrate reductase	Increased expression during hypoxia and in human granulomas
(Hutter, Dick 2000)	Rv1737c	narK2	Nitrate/nitrite transport	Increased expression during hypoxia

(Glickman, Cox et al. 2000)	Rv0470c	pcaA	Cyclopropane synthase	Reduced persistence in mice
(Murphy, Stewart et al. 2005)	Rv0126	treS	Trehalose synthase	Prolonged time-to-death in mice
(Zahrt, Deretic 2001)	Rv0981	mprA	Two-component regulator	Reduced persistence in mice
(Parish, Smith et al. 2003)	Rv3132c	dosR	Two-component regulator	Increased lethality in mice
(Parish, Smith et al. 2003)	Rv3764/5c	trcY	Two-component regulator	Increased lethality in mice
(Parish, Smith et al. 2003)	Rv1032c	trcS	Two-component regulator	Increased lethality in mice
(Parish, Smith et al. 2003)	Rv1027/8c	kdpDE	Two-component regulator	Increased lethality in mice
(Chen, Ping, Ruiz et al. 2000)	Rv3286c	sigF	RNA polymerase σ factor	Prolonged time-to-death in mice
(Kaushal, Schroeder et al. 2002)	Rv3223c	sigH	RNA polymerase σ factor	Reduced pathology, prolonged time-to-death in mice

(Ando, M., Yoshimatsu et al. 2003)	Rv1221	sigE	RNA polymerase σ factor	Prolonged time-to-death in mice
(Steyn, Collins et al. 2002)	Rv3416	whiB3	Transcriptional regulator	Reduced pathology, prolonged time-to-death in mice
(Primm, Andersen et al. 2000)	Rv2583c	relA	(p)ppGpp synthase	Reduced persistence in mice
(Hu, Movahedzadeh et al. 2006)	Rv2031c	hspX	α -crystallin	Increased expression during hypoxia and in mice
(Stewart, Snewin et al. 2001)	Rv0353	hspR	HspR repressor	Reduced persistence in mice
(Stewart, Snewin et al. 2001)	Rv0350	hsp70	Heat shock protein 70	
(Li, Kelley et al. 1998)	Rv1908c	KatG	Catalase-peroxidase	Reduced persistence in mice

Adapted from (Wang, Wang et al. 2011)

The stringent response is a microorganism's overall alteration in metabolism when subjected to a nutrient-limited environment (Dahl, Kraus et al. 2003). The response is characterized by dramatic reduction in growth rate emanating from reduced levels of rRNA, tRNA and translation. This is often associated with reduction in metabolism of carbohydrates, amino acids and phospholipids. Dramatic increases in guanine nucleotides (p)ppGp, mediate the stringent response while low levels reverse it (Dahl, Kraus et al. 2003). (p)ppGpp production is regulated by RelA (ppGpp synthase) which has been implicated in *Mtb* persistence as *relA* mutant bacteria could not survive *in vitro* or in mice (Dahl, Kraus et al. 2003, Zhang, Ying, Yew et al. 2012).

TA modules are the most extensively studied persister related mechanism in *E.coli* (Zhang, Ying, Yew et al. 2012). Notably, over 30 TA modules have been identified in the *Mtb* genome (Pandey, Gerdes 2005). TA modules are highly redundant and others have suggested there could be over 80 in *Mtb* (Arcus, Rainey et al. 2005, Ramage, Connolly et al. 2009). Ten of these were overexpressed in persisters and *in vitro* one of the most overexpressed modules was Rv2866 a homologue of RelE (Keren, Minami et al. 2011). Toxin RelE is the most well studied in *Mtb* and its high expression was shown to cause persistence and drug tolerance (Singh, R., Barry et al. 2010). The role of other TA modules in *Mtb* remains poorly described (Zhang, Ying, Yew et al. 2012).

A recent study has reported novel genes unique to *Mtb* persistence notably lipid biosynthesis and feast or famine regulation (Torrey, Keren et al. 2016). In this study *Mtb* mutants were generated via mutagenesis and used to identify *Mtb* specific persister genes. Genes associated with synthesis of lipids, metabolism of carbon, toxin-antitoxin modules and regulation of transcription were identified.

1. 10. 4 Persistent *Mycobacterium tuberculosis* in sputum

The primary host cells for pulmonary *Mtb* are pulmonary macrophages (Russell 2001). The eventual formation of the granuloma is characterized by the presence of a specific population of macrophages enriched in lipid droplets (LD) known as foamy macrophages (Neyrolles 2014). Caire-Brändli and co-workers (Caire-Brändli, Papadopoulos et al. 2014) have described a model in which mycobacteria within foamy macrophages undergo reversible cell growth arrest. The mycobacterial phagosome fuses with the host cell LD and accumulates the TAG content of the LD to the phagosome. *In vitro*, *Mtb* in hypoxic macrophages were shown to accumulate lipid bodies and lose acid fastness (Daniel, J., Maamar et al. 2011). Loss of acid fastness is suggestive of organisms in dormant state. Furthermore non-replicating *Mtb in vitro* were previously shown to contain LD levels comparable to those found in sputum (Garton, Waddell et al. 2008), suggesting LD could be a putative marker of persistent *Mtb*. Interestingly LD accumulation in macrophages has also been associated with ESAT-6, the *Mtb* virulence factor (Singh, V., Jamwal et al. 2012).

Transcriptomic studies have shown that *Mtb* organisms are present in sputum which have gene expression signatures comparable to slow growing or non-replicating organisms. This sputum phenotype was also correlated with upregulation of the *dosR* regulon. Lipid body populations in sputum were negatively correlated with time to positivity (Garton, Waddell et al. 2008). Both lipid body formation and increased time to positivity are highly correlated with poor treatment outcomes in *Mtb* (Sloan, D. J., Mwandumba et al. 2015). Furthermore, transcription signatures of *Mtb* early in treatment can be used to predict efficacy of treatment later. This was demonstrated in a recent study where mRNA profiles

of bacilli at days 7 and 14 of treatment were consistent with pre-chemotherapy signatures (Honeyborne, McHugh et al. 2016). This suggests presence of a drug tolerant sub-population in sputum prior to treatment. Accumulating evidence suggests that drug tolerant *Mtb* cells while having low metabolic activity are still transcriptionally active (Walter, Dolganov et al. 2015). Taken together, these data suggest that drug tolerance or persistence plays a key role in *Mtb* treatment failure and disease relapse.

1. 11 Epigenetics in *Mycobacterium tuberculosis*

Conrad Waddington coined the term epigenetics in reference to factors above or in addition to genetics to explain the formation of tissues and organs during development (Waddington 1956). Currently epigenetic mechanisms are defined as stable and heritable changes in gene function with no associated change in genome sequence. These changes are characterized by modifications in the DNA and its associated histones. They include methylation, acetylation, phosphorylation, and ubiquitination (Yadav, Dwived et al. 2015). In humans epigenetic mechanisms have been shown to play a role in development of cancer via repression of tumor suppressor genes (Yadav, Dwived et al. 2015). In bacteria epigenetics works to aid the organism in adapting to varying environmental conditions (Chen, Poyin, Jeannotte et al. 2014). The most common and well characterized epigenetic mechanism in prokaryotes is DNA methylation (Clark, T. A., Murray et al. 2011, Sánchez-Romero, Cota et al. 2015). During DNA methylation a methyl group is transferred from S-adenosyl-methionine (SAM) to DNA by enzymes called methyltransferases (Jeltsch 2002, Clark, T. A., Murray et al. 2011). Methyltransferases (Mtases) recognise specific DNA sequences (motifs) methylating adenine and cytosine bases. The result is formation of N-6methyladenine (m6A), N-4methylcytosine (m4C) and N-5methylcytosine (m5C)

(Figure 1.7). Bacterial DNA may also undergo modification to the atoms between nucleotides, a phenomenon called phosphorothioation (Figure 1.7) (Chen, Poyin, Jeannotte et al. 2014).

New DNA methylation types such as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) have recently been discovered in eukaryotes although their biological functions remain unknown (Murray, Clark et al. 2012).

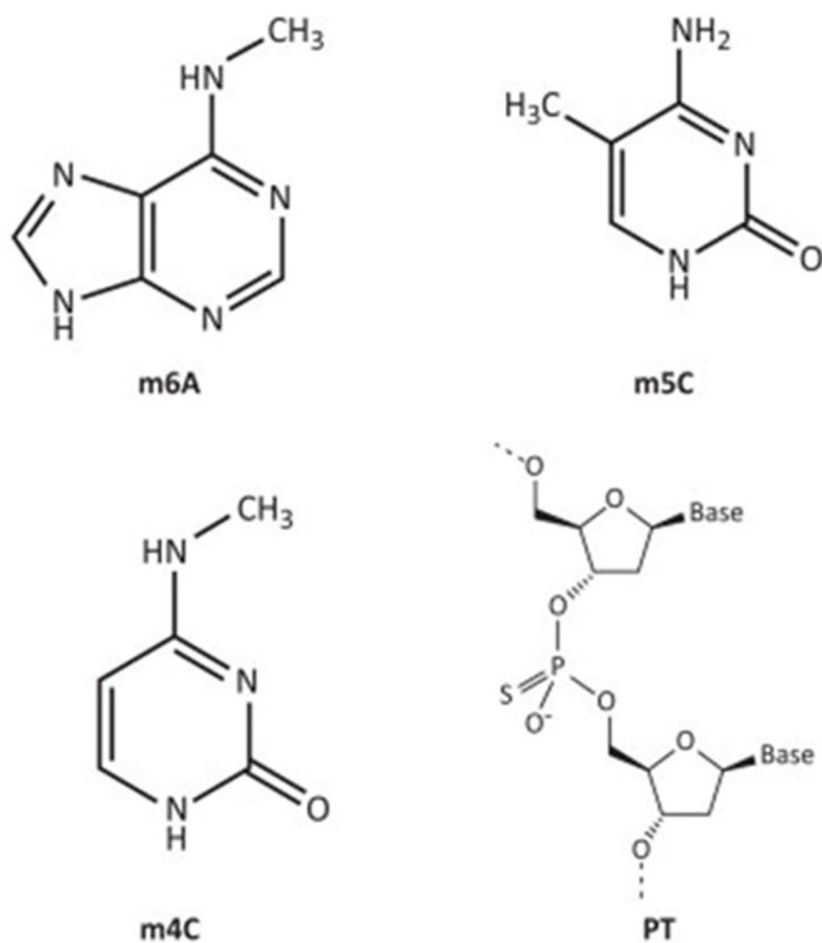


Figure 1. 7 Chemical structure of some methylated bases

Chemical structure of commonly methylated adenine and cytosine bases in *Mtb*.

6methyladenine(m6A), N-4methylcytosine(m4C), N-5methylcytosine(m5C) and phosphorothioation. Figure adapted from(Chen, Poyin, Jeannotte et al. 2014)

1. 11. 1 Role of DNA methylation

Bacterial *Mtases* form part of the restriction-modification (RM) system which also includes a restriction endonucleases (REs). Pairs of *Mtases* and REs target the same sequence motifs (Davis, Chao et al. 2013). Evolutionarily RM systems are believed to have offered protection to bacterial DNA against viruses. REs will cleave unmethylated viral DNA whereas methylation will protect host DNA from digestion (Wion, Casadesus

2006). This notwithstanding, ‘orphan’ *Mtases* with missing cognate REs have been described. These perform many other cellular functions. For example orphan *Mtases* may be involved in DNA mismatch repair and initiation of the cell cycle process (van der Woude, Hale et al. 1998). Interestingly REs that cleave methylated DNA have recently been described (Sánchez-Romero, Cota et al. 2015). In eukaryotes DNA methylation plays key roles in regulation of genes, chromatin remodeling, genomic integrity, parental imprinting and has been implicated in human disease (Wion, Casadesus 2006).

Three different types of RM systems have previously been described (types I, II and III). Types I and III have the R and M systems as part of a single polypeptide while in type II they are in separate polypeptides. In types II and III, the RE cleaves the DNA in proximity to the methylation target site whereas in type I it cleaves further from the site (Casadesus, Low 2006). A RM system was later discovered that recognises methylated DNA and phosphothioated DNA and was classed as a type IV system (Liu, Ou et al. 2010). The best evidence of the role of DNA methylation in gene expression comes from GATCs in *Ecoli* (Casadesus, Low 2006). Expression of pyelonephritis associated pilus (*Pap*) is under the control of a phase variation. Individual cells may be in the phase-on or phase-off state. When the GATC1028 is methylated the *pap* gene is in off state and the GATC1130 is unmethylated. Methylation of the GATC1130 means that the *pap* gene is on and the GATC1128 is unmethylated.

1. 11. 2 Detection of modified bases

Until recently, it has been challenging to detect modified DNA bases. It is possible to detect base modifications such as DNA methylation using Sanger sequencing although the method does not have enough throughput for this purpose (Clark, T. A., Murray et al.

2011). DNA amplification erases epigenetic modifications. To this end high throughput second generation sequencing (SGS) such as Illumina HiSeq that incorporate an amplification step cannot be used on native DNA for base modification detection (Davis, Chao et al. 2013). It has been possible to detect m5C methylation via bisulfite DNA sequencing. The method involves conversion of unmodified cytosine to uracil followed by sequencing using various platforms (Clark, S. J., Statham et al. 2006). Sample preparation is labour intensive and costly and bisulfite treatment causes degradation of DNA (Korlach, Turner 2012, Rhoads, Au 2015). Additionally bisulfite sequencing is unable to discriminate between C, mC, and hmC (Flusberg, Webster et al. 2010) although new methods can now circumvent this problem (Korlach, Turner 2012). It has also been possible to detect DNA methylation in *Mtb* using technologies such as chromatography and mass spectrometry (Zhu, Zhong et al. 2015). These methods are however unable to offer genome wide detection of epigenetic modifications (Korlach, Turner 2012).

Single Molecule Real Time (SMRT) Sequencing developed by Pacific Biosciences is capable of determining the nucleic acid sequence together with the methylation status at single base resolution (Flusberg, Webster et al. 2010). Sequence information is captured in real time during incorporation of nucleotides which are conjugated with distinct fluorophores (Davis, Chao et al. 2013). When a sample of DNA template termed SMRTbell is loaded into a cell, it diffuses into a sequencing unit called the Zero Mode Waveguide (ZMW). The DNA polymerase attached to the bottom of ZMW initiates replication. A light pulse identifies the incorporated base and a 'movie' of light pulses records the replication process (Rhoads, Au 2015). Additionally, kinetic data regarding the time and width of nucleotide incorporation, inter pulse duration (IPD) and pulse width

(PW) respectively are obtained (Figure 1. 8). There are changes in kinetic signatures when a DNA polymerase encounters a modified base. This allows for identification of the type and position of the modified base (Clark, T. A., Murray et al. 2011, Flusberg, Webster et al. 2010). PacBio sequencing has the advantage of having longer read lengths than other second-generation sequencing (SGS) platforms. Typically the PacBio RS II machine with sixth generation of polymerase and fourth generation chemistry (P6/C4) offers average read lengths of over 10kb and is much faster (Rhoads, Au 2015). Contrastingly the Illumina Hiseq 2500 offers minimum read lengths of 250 bp (Illumina). The weaknesses of PacBio are less throughput and higher single pass error rate compared to other SGS techniques. Increasing the number of passes aids in increasing the accuracy. As an example sequencing coverage of 15 passes yields greater than 99% accuracy. Additionally, PacBio sequencing is costlier. Since strengths and weaknesses of PacBio and SGS are complementary, hybrid sequencing has been used as the best approach that integrates both methods (Rhoads, Au 2015).

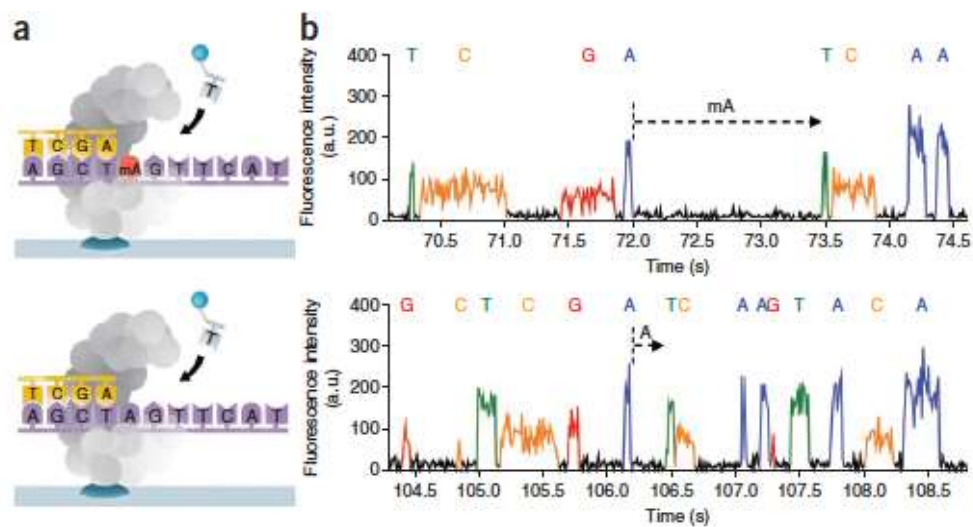


Figure 1.8 Detection of methylated DNA bases using Single Molecule Real Time Sequencing

(a) Addition of a nucleotide to a DNA strand containing a methylated (mA) and unmethylated (A) base (b) Kinetic variations during sequencing. Nucleotides being incorporated are identified using fluorescent pulses. When a nucleotide is incorporated opposite a modified base, the duration (IPD) taken is characteristically longer. In this case the IPD is $\sim 5x$ for mA than for A. Adapted from (Flusberg, Webster et al. 2010).

Circumstantial evidence points to the fact that epigenetic mechanisms could be responsible for the persistence phenotype in bacteria (Balaban, Gerdes et al. 2013, Korch, S. B., Hill 2006, Lewis 2010). DNA Methylation in host cells has been vigorously pursued. Methylation and gene expression profiles of dendritic cells (DCs) in individual patients were found to undergo differential changes following *Mtb* infection (Barreiro, Pacis et al. 2013). Additionally a recent study found striking differences in DNA methylation of granulocytes and monocytes between TB patients and health individuals (Esterhuysen, Weiner et al. 2015). These changes were all in host immune cells. However, within the pathogen DNA methylation has also been linked to survival of *Mtb* under

hypoxic conditions. This association was however found to be strain specific as methylation mediated by a mycobacterium adenine methyltransferase A (*MamA*) was shown to offer fitness to a Euro-American strain during hypoxia associated with infection. Such fitness could not be demonstrated in the *Beijing* lineage strain owing to point mutation within the *Mtase* (Shell, Prestwich et al. 2013). Since DNA methylation is the best characterized form of epigenetic gene regulation in prokaryotes (Shell, Prestwich et al. 2013) and that the persister phenotype is not genetically inherited, the hypothesis in this study is that methylation is responsible for the persistence phenotype in *Mtb*. Consistently, inactivation of SAM was associated with reduced DNA methylation in *Mtb* and rapid death of *Mtb* infected mice (Berney, Berney-Meyer et al. 2015) suggesting DNA methylation has a role in *Mtb* survival. To date, the role of DNA methylation in *Mtb* epigenetic inheritance and its association with the persister phenotype in *Mtb* remains to be evaluated.

1. 12 Aims of this thesis

A large body of evidence in part implicates persister *Mtb* in the need for prolonged chemotherapy for the effective treatment of TB. New ultra-short and effective drug regimens against TB that target persistent *Mtb* cells need to be developed. Such drug regimens should be able to eliminate *Mtb* in both replicating and non-replicating state. Additionally, they should be able to eliminate *Mtb* from tissues and granulomas. This will only be possible if scientists thoroughly understand mechanisms that underlie persistence *in vitro* and in clinically relevant *in vivo* conditions.

This thesis will report on a laboratory based investigation aimed at identifying genetic and epigenetic mechanisms responsible for *Mtb* persistence. Specific objectives include:

- i. To validate genomic DNA extraction methods compatible with DNA methylation analysis in *Mtb*
- ii. To analyze DNA methylation patterns in persistent *Mtb*
- iii. To assess stability of DNA methylation patterns in passage/strains
- iv. To evaluate occurrence of methylation within different *Mtb* strains and factors that lead to loss of methylation

1.13 Study hypothesis

The main hypothesis in this study was that persister *Mtb* cells would demonstrate a differential methylation profile compared to metabolically active mycobacteria and that specifically persister related genes would largely be unmethylated especially in their promoter regions.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Experimental procedures

All experiments involving *Mycobacterium tuberculosis* were performed in a Biosafety Level (BSL) 2 or 3 Laboratory as appropriate at University of Malawi-College of Medicine/ Malawi Liverpool Welcome Trust (CoM/MLW) TB Laboratory or at the Liverpool School of Tropical Medicine following local Standard Operating Procedures (SOPs). All reagents used were purchased from Sigma-Aldrich unless otherwise stated.

2.1.1 Mycobacterial strains and growth conditions

Mycobacterium tuberculosis strains used were clinical isolates from frozen stocks derived from a previously completed clinical study: Studying Persistence and Understanding Tuberculosis in Malawi (SPUTUM) or Monitoring and Evaluation (M & E) samples processed in the CoM/MLW TB laboratory. Samples were selected from a total of 133 stored strains derived from bacteriologically confirmed drug susceptible *Mtb* patients. Fifteen of these patients had poor treatment outcomes in the SPUTUM study (defined as confirmed culture positivity for *Mtb* positive at end of treatment, death attributable to TB or relapse within one year of finishing treatment).

Table 2. 1 Patient characteristics from whom bacterial isolates used in the study were obtained

Patient ID	Sex	HIV status	Age	Treatment	End of treatment outcome
S520R	F	Non-infected	27	2HRZE/4RH	Cured
S571T	M	Non-infected	43	2HRZE/4RH	Cured
S524R	M	Non-infected	61	2HRZE/4RH	Cured
S545R	M	Non-infected	28	2HRZE/4RH	Cured
S519T	F	Non-infected	36	2HRZE/4RH	Cured
S530Y	F	Non-infected	42	2HRZE/4RH	Cured
S528S	M	Infected	45	2HRZE/4RH	Cured
S555L	M	Infected	43	2HRZE/4RH	Cured
S621S	F	Non-infected	23	2HRZE/4RH	Cured
S610T	M	Non-infected	39	2HRZE/4RH	Cured

S616T	M	Infected	27	2HRZE/4RH	Cured
S625D	M	Non- infected	29	2HRZE/4RH	Cured
S657S	M	Non- infected	29	2HRZE/4RH	Cured
S714Y	F	Non- infected	37	2HRZE/4RH	Cured
S696E	M	Infected	38	2HRZE/4RH	Cured
S699W	M	Infected	45	2HRZE/4RH	Cured
S406T	F	Infected	28	2HRZE/4RH	Cured
S020Q	F	Infected	39	2HRZE/4RH	Cured
S07U	F	Non- infected	40	2HRZE/4RH	Cured
S533U	M	Non- infected	44	2HRZE/4RH	Cured
S616M	M	Non- infected	30	2HRZE/4RH	Cured
S677J	M	Non- infected	34	2HRZE/4RH	Cured
S566A	M	Non- infected	56	2HRZE/4RH	Cured
S634U	M	Non- infected	49	2HRZE/4RH	Cured

S532S	M	Non-infected	38	2HRZE/4RH	Cured
S010X	M	Non-infected	27	2HRZE/4RH	Cured
S334A	F	Non-infected	25	2HRZE/4RH	Cured
S060A	M	Infected	28	2HRZE/4RH	Cured
S137U	M	Infected	29	2HRZE/4RH	Cured
S181R	M	Infected	46	2HRZE/4RH	Cured
S005G	M	Infected	22	2HRZE/4RH	Not cured
S29R	M	Non-infected	28	2HRZE/4RH	Not cured
S180P	M	Infected	45	2HRZE/4RH	Cured
R730C	F	Infected	43	2HRZE/4RH	Not cured
S990F	F	Infected	52	2HRZE/4RH	Cured
S037Q	F	Non-infected	41	2HRZE/4RH	Cured
S004F	M	Non-infected	27	2HRZE/4RH	Cured
S147Q	F	Non-infected	29	2HRZE/4RH	Cured
S1025H	M	Non-infected	46	2HRZE/4RH	Cured

S994K	M	Non-infected	28	2HRZE/4RH	Not cured
S1027K	M	Non-infected	30	2HRZE/4RH	Cured
S006K	F	Non-infected	39	2HRZE/4RH	Cured
S081T	M	Infected	34	2HRZE/4RH	Cured
S003H	F	Non-infected	36	2HRZE/4RH	Cured
S249F	F	Non-infected	35	2HRZE/4RH	Cured
S300Q	F	Infected	47	2HRZE/4RH	Cured
S306H	M	Infected	45	2HRZE/4RH	Cured
S357K	F	Infected	41	2HRZE/4RH	Cured
S370M	M	Non-infected	43	2HRZE/4RH	Cured
S383Z	M	Non-Infected	29	2HRZE/4RH	Cured
S379J	M	Non-infected	45	2HRZE/4RH	Cured
S416M	F	Non-infected	22	2HRZE/4RH	Cured
S494T	F	Infected	28	2HRZE/4RH	Cured

S479A	M	Non-infected	35	2HRZE/4RH	Cured
S478B	M	Infected	37	2HRZE/4RH	Cured
S486T	M	Non-infected	53	2HRZE/4RH	Cured
S481L	M	Non-infected	36	2HRZE/4RH	Cured
S507T	F	Non-infected	27	2HRZE/4RH	Cured
S549Y	F	Non-infected	38	2HRZE/4RH	Cured
S510J	F	Infected	28	2HRZE/4RH	Cured
S539K	F	Infected	28	2HRZE/4RH	Cured
S511L	M	Non-infected	38	2HRZE/4RH	Cured
S589U	M	Infected	21	2HRZE/4RH	Cured
S662Y	M	Non-infected	51	2HRZE/4RH	Cured

Mtb strains were revived after storage in glycerol using liquid culture (MGIT) and later transferred to solid culture (LJ). Putative persister strains were selected based on their lipid body profile and treatment outcomes. Those with high lipid body droplet as assessed

during the SPUTUM study using LipTOX Red microscopy and poor treatment outcomes were classified as “persisters” while those with lower lipid body content and treatment success were classified as “non-persisters”. All samples with LipTOX Red microscopy counts of $\geq 50\%$ were classified as being high lipid body. Two putative “persister” and “non-persister” isolates respectively were cultured and sequenced. A further 14 isolates were cultured under varying conditions and sequenced.

2.1.2 Decontamination of sputum samples

It is obvious that in addition to *Mtb* cells sputum may contain non-tuberculous mycobacteria and fungi/bacterial contaminants. In view of this, it was imperative to decontaminate the sputum samples in order to rid it of contaminants. Frozen sputum samples were left to thaw before being processed. Samples were decontaminated using 3% NaOH following standard laboratory operating procedures.

2.1.3 The saline wash procedure

The major challenge of extracting genomic DNA directly from sputum is that of host DNA contamination and current methods do not effectively discriminate between host and bacterial DNA. A saline wash step was therefore employed to remove host cells before direct extraction from sputum as previously described (Votintseva, Pankhurst et al. 2015). A 1mL aliquot of sputum *Mtb* was heat-killed at 80 °C for 40 minutes and centrifuged at 3000xg for 15 minutes. Supernatant was carefully removed by pipetting before addition of 1mL of sterile saline. The pellet was re-suspended by pipetting up and down. This was followed by further centrifugation at 3000xg for 15 minutes. The final pellet was recovered by pipetting out the supernatant. A 700mL aliquot of distilled water was added

to the pellet followed by re-suspension by pipetting up and down. This was then used for extraction of total genomic DNA in the sample.

2. 1. 4 Solid plate preparation and determination of colony counts

Solid plates for determination of colony forming units (CFU) were prepared according to the manufacturer's instructions. Twenty-one grams of 7H11 medium powder (Becton Dickinson Difco, 212240) was suspended in 900mL of distilled water containing 5mL of glycerol. Following heating with frequent agitation, the mixture was boiled for a minute to dissolve the powder. A 2.5mL aliquot comprised of 2% suspension of carbendazim in 70% alcohol was added to the media before autoclaving at 121 °C for 15 minutes. After cooling to 45°C the media was supplemented with 100mL oleic acid-albumin, dextrose and catalase (OADC, Becton Dickinson). Antibiotics in the form of two selectatabs (MAST24) with a final concentration of polymyxin B, amphotericin B, nalidixic acid, trimetho prim and azlocillin (PANTA) were added. As the media was cooling 8mL was added to a tri-segment of each 100mm diameter petri dishes. Plates were dried overnight in an inverted position at 4°C until use. Sputum samples were homogenized by vortexing with glass beads to ensure even distribution of *Mtb* cells. The volume to be used for solid plates for CFU determination was incubated in an equal volume of dithiothreitol 1g/l (Oxoid) for one hour to liquefy the sample.

Phosphate buffered saline (PBS) was used to prepare three 10-fold dilutions of the concentrated homogenized sputum. A 50 µl aliquot of neat sputum and the three 10-fold dilutions from each sample were inoculated using fresh sterile loop onto tri-segmented plates. Culture plates were placed in Ziploc polythene bags and incubated at 37°C. Contamination was checked weekly and plate reading was done at 21 days. Any growth

in a segment preventing plate reading was regarded as contaminated and was excluded from the count. The optimum dilution for reading the count was selected by choosing the plate segment with between 10 and 100 colonies. Colony counts per mL were expressed as follows:

CFU/mL = number of colonies counted X 2 (dithiothreitol dilution) X 20 (50 µl was inoculated per segment) X dilution factor.

2. 1. 5 Liquid culture

Frozen clinical isolates were left to thaw and immediately after liquefaction of the sample a 0.5mL aliquot was added to a commercially available BBL MGIT™ 7ml tube. The tube contained Middlebrook 7H9 broth base supplemented with oleic acid-albumin, dextrose and catalase (OADC) and an antibiotic mixture of polymyxin B, amphotericin B, nalidixic acid, trimethoprim and azlocillin (PANTA). Tubes were incubated in a BACTEC MGIT 960 instrument at 37°C and automatically monitored with additional manual inspection for possible growth once a week for up to eight weeks.

2. 1. 6 Solid culture inoculation

After four weeks of growth 2mL of confirmed *Mtb* sample from each MGIT tube was inoculated into each of 3 Lowenstein-Jensen (LJ) slopes following laboratory SOP. LJ slopes were prepared by addition of each of the following: 1.2g Potassium dihydrogen phosphate anhydrous (KH₂PO₄), 0.12g Magnesium sulphate (MgSO₄ · 7H₂O), 0.3g Magnesium citrate, 1.8g Asparagine, 6mL Glycerol, 0.3g Malachite green to 300mL distilled water in a sterile reagent bottle. This was autoclaved at 121°C and 15lbs pressure for 15 minutes in order to make the mixture sterile and then left to cool to room temperature. Ten clean fresh eggs previously washed with a hand brush and plain alkaline

soap before being rinsed thoroughly in running water and soaked in ethanol for 15 minutes were used. The eggs were cracked with a sharp knife and beaten with a sterile glass rod for five minutes until watery. The salt solution was then poured into the egg solution and mixed to homogenize. While maintaining the pH at (6.8 – 7.2) 4mL of the solution was dispensed into 14mL McCartney bottles. Within 15 minutes, the medium was inspissated in a slanting position for six minutes at 85°C. The bottles were labeled and the media was stored at 2-8°C until use. Following inoculation of *Mtb* LJ slopes were incubated at 37°C and checked weekly for possible growth for up to eight weeks.

Mtb identification/confirmation was done using both the BD MGIT TBC ID test device (Becton Dickinson, Maryland U.S.A) following manufacturer's instructions and by ZN staining.

2. 1. 7 Ziehl Neelsen (ZN) staining

ZN staining was done according to laboratory SOP. Sputum smear were spread over the center of the slide using circular rotational movements. Slides were air dried for 30 minutes before being heat fixed by a flame. The smear was then covered with carbol fuchsin stain followed by further heating. The heated stain was allowed to stand on the slide for five minutes. The stain was then washed off with clean water and the smear covered with 3% v/v acid alcohol for three minutes or until the smear was sufficiently decolorized to a pale pink colour. The slide was washed with clean water followed by staining with methylene blue for up to 30 seconds and then further washed with clean water. The slide was air dried in a rack and then examined microscopically using the 100X oil immersion.

2. 1. 8 Nitric Oxide (NO) induction of H37Rv isolates

H37Rv isolates were grown onto solid plates of Middlebrook 7H11 (Diffico) agar supplemented with 10% v/v OADC. Following 28 days of growth cells were harvested and homogenized in pre-warmed Middlebrook 7H9 (Diffico) media supplemented with 10% v/v OADC and 0.05% v/v tween 80. The density of the homogenate was normalized to OD₅₈₀=0.25. The media temperature was maintained at 35-37°C to avoid cold shocking the cells. A 5ml aliquot of the suspension was exposed to 500 µM Spermine NONOate (Santa Cruz Biotechnology) and 500 µM Spermine-4HCL (XYZ) respectively for four hours at 37°C with shaking at 100rpm. Spermine NONOate is a rapid nitric oxide donor known to cause dose dependent bacteriastasis in *Mtb* with minimal apparent killing (Voskuil, M. I., Bartek et al. 2011). Spermine NONOate has a short half-life in aqueous solution and to avoid degradation weighing was done under nitrogen and the powder diluted in pre-warmed 7H9 immediately before use. After four hours, a 500 µl aliquot was drawn and used for genomic DNA extraction.

2. 1. 9 Drug susceptibility testing

Determination of MICs of the *Mtb* isolates was done using UKMYC3 sensititre custom made plates (Thermo Scientific) following manufacturer's instructions. Briefly, a loopful of colonies were scraped from LJ slants and emulsified in a McFarland tube containing saline, 0.2% tween and glass beads. The tube was vortexed for 30 seconds and turbidity adjusted to 0.5 McFarland standard. A 100µl aliquot was transferred into a tube of 10ml of 7H9 broth supplemented with 10% OADC to give an inoculum of $\sim 1 \times 10^5$ cfu/mL. Suspension was briefly vortexed and allowed to settle for 15 minutes. Another 100 µl aliquot was transferred into each well of the microtitre plate for inoculation. Wells were

covered with an adhesive seal and wiped with 5% Surfanios and 70% ethanol before being placed into a sealed plastic bag.

Plates were incubated at 37°C for 10 days after which visual inspection of growth was done. Plates with poor growth were re-checked using the positive control as a guide at days 14 and 21. Plate reading was done manually by visual inspection when growth was clearly visible in the positive control. If results of two plate readers differed a sample was re-tested using a fresh inoculum.

2. 1. 10 DNA extraction -from culture

To extract genomic DNA mycobacterial cultures, two in house methods both of which use a Cetyltrimethylammonium bromide (CTAB) step were employed; the Phenol/Chloroform/Isoamyl alcohol method and the Chloroform/Isoamyl alcohol method.

2. 1. 11 Phenol/Chloroform/Isoamyl alcohol

An adaptation of the previously described CTAB based method (Warren, de Kock et al. 2006) was used. In brief growth was scraped from LJ slants or pellet obtained from spun MGIT tubes and re-suspended in a sterile 50mL polypropylene Falcon tube containing 2mL TE buffer (0.01 M Tris-HCl, 0.001 M EDTA [pH 8.0]) with or without 20x4mm glass beads. The sample was heat killed at 80°C for 40 minutes. A 2mL aliquot of 2X extraction buffer was added followed by gentle inversion to mix the suspension. The sample was vigorously vortexed for two minutes and then a 500µl freshly prepared lysozyme (50mg/mL, Roche, USA) was added. Ten microliters of RNase, concentration at a concentration of 10mg/mL (Ambion, USA) was then added to the suspension followed by gentle inversion. A two-hour incubation in a 37°C preheated oven was followed by addition of 450 µl (1 x volume) of 10X Proteinase K buffer and 150 µl Proteinase K, at a

concentration of 10mg/mL (Qiagen, U.S.A). This was mixed by gentle inversion and then incubated at 45°C for 16 hours (overnight) in a preheated oven. An 820 µl aliquot of 5M NaCl and 710 µl of 10% CTAB (Amresco, USA) preheated to 65°C were added. The suspension was then incubated for 10 minutes at 65°C. After cooling to room temperature, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, Sigma-Aldrich, USA) was added (about 5 mL). This was followed by gentle mixing repeated every 30 minutes for two hours. The suspension was centrifuged at 3000xg for 20 minutes at room temperature. The top phase was carefully aspirated and transferred to a new labeled 15mL tube. This was followed by addition of an equal volume of chloroform/isoamyl alcohol solution (24:1, v/v, Sigma-Aldrich, USA). The suspension was gently mixed before being centrifuged at 3000xg for 20 minutes. The top phase was carefully aspirated to a new 15mL tube. A 0.1 volume of 3M sodium acetate, pH 5.2 final concentrations was then added followed by gentle inversion. To precipitate the DNA a 7mL aliquot of ice cold isopropanol was added followed by gentle inversion. DNA was pelleted by centrifugation at 3000xg for 20 minutes at 4°C. The supernatant was discarded and the DNA pellet washed in 1mL cold 70% ethanol for 10 minutes. The pellet was then air dried for 2-3 hours before being re-suspended in 50 µl x1 TE buffer. The DNA was re-dissolved by incubating at 4°C overnight. All DNA was stored at 4°C until use.

2. 1. 12 The Chloroform/ Isoamyl alcohol method

Genomic DNA was isolated using the traditional CTAB method as previously described (Somerville, Thibert et al. 2005). Growth was scraped from LJ slants and dispensed into 600µl of TE buffer. Growth from H37Rv was processed in a similar way as a positive control. The concentrated culture was heated to 80°C for 40 minutes to kill the

mycobacterial cells. Fifty microliters of 10mg/mL lysozyme (Roche, USA) was added to each tube which was vortexed and incubated at 37°C overnight. Ten microliters of 10mg/ml proteinase K (Qiagen, USA) and 35 µl of 20% (w/v) sodium dodecyl sulfate (SDS) were added and following vortexing incubated at 55°C for 30 minutes. One hundred microliters of pre-warmed (55°C) 5M NaCl and 100 µl pre-warmed (55°C) CTAB solution was added, vortexed and the tube incubated at 65°C for 15 minutes. After cooling, 700µl of chloroform: isoamyl alcohol (24:1, v/v) was added to the tube which vortexed for 10 seconds followed by centrifugation at 13000xg for 5 minutes. The top aqueous layer was pipetted out using a P200 pipette and a 0.6 volume of isopropanol added to precipitate the DNA. Tubes were kept at -20°C for a minimum of 30 minutes or overnight. After centrifugation at 13000xg for 15 minutes, the supernatant was discarded using one single motion. About 20 µl was left in the tube to which 1.0ml of 70% ethanol (stored at -20°C) was added. Following centrifugation at 13000xg for 7 minutes liquid was removed using a 1.0mL tip. After centrifugation at 13000xg for 2 minutes, the remaining liquid was removed using a 200 µl tip. The tube was left with the lid open at 37°C until the last traces of liquid had evaporated. A fifty microliter volume of TE buffer (0.01 M Tris-HCl, 0.001 M EDTA [pH 8.0]) was added and the tube left at 4°C - overnight for the pellet to dissolve. Extracted DNA was quantified using a Qubit 3.0 fluorometer (Life Technologies, USA) according to manufacturer's instructions. DNA purity was determined on a NanoDrop ND-1000 Spectrophotometer V3.7 (Thermo Scientific, Wilmington U.S.A) following manufacturer's instructions. DNA purity was checked at absorbance 260nm and 280nm by calculating a ratio of A260/A280. An A260/A280 ratio ranging 1.8 to 2.0 was considered a pure DNA preparation. For additional confirmation

of DNA quality, samples were also analyzed on 1.5% Agarose Gel electrophoresis following standard operating procedures. The gel was prepared by addition of 1.5g of agarose powder (Thermo Fisher Scientific, UK) to 200mL of TBE buffer. The mixture was heated in a microwave for not more than three minutes until the agarose had dissolved. After cooling to approximately 50°C, 10µl of ethidium bromide was added with continuous swirling. This was poured into casting dams and used for electrophoresis of the DNA. All extracted DNA was stored at 4°C until use.

2. 1. 13 DNA extraction -from sputum

Genomic DNA was extracted directly from sputum using the commercially available Primextract kit (Longhorn Vaccines and Diagnostics, USA) and an in house DNA extraction method as previously described (Amita, Vandana et al. 2002). All sputum samples used in this extraction were of bacteriologically confirmed 3+ smear status. For the Primextract kit 200 µL of the sputum sample was mixed with equal volumes of 100% ethanol and Primextract lysis buffer. The mixture was homogenized by a brief vortex followed by a quick spin. Following incubation at ambient temperatures for 5 minutes, entire contents of the sample were pipetted into an extraction column/collection tube. This was centrifuged at 7500xg for 60 seconds. The eluate was discarded and the collection tube reused. Following two buffer washes each of which preceded centrifugation at for 7500xg for 60 seconds a third centrifugation removed the remaining traces of wash buffer. The extraction column was placed into a clean 1.5mL micro centrifuge tube. Elution solution equal to 50 µl (preheated to 70°C) was pipetted directly onto the silica filter followed by incubation for 60 seconds. After a final centrifugation 7500xg for 60 seconds the column was discarded and the eluted DNA was stored at 4°C until use. Similarly, as a

positive control, H37Rv growth re-suspended in 200 µl of molecular biology grade water was processed. For the CTAB direct extraction of *Mtb* DNA from sputum sample was conducted as in section 2.1.12

2. 1. 14 Sample preparation

Sputum samples were prepared following laboratory standard operating procedures (SOPs) as follows; A 2mL aliquot of sputum sample was decontaminated by addition of 3% Sodium Hydroxide (NaOH) for 15 minutes. This was neutralized using an equal volume of phosphate buffer. The suspension was centrifuged at 16000xg for 20 minutes and the pellet re-suspended in 500 µL of phosphate buffered saline (PBS). The sample was incubated with an equal volume of dithiothreitol 1g/l (Oxoid) for one hour for the purpose of breaking sulfide bonds in mucoid sputum. Following vigorous vortexing for five minutes, 200 µL aliquots were drawn and used for DNA extraction using two different extraction methods. An aliquot of 50 µL and three tenfold dilutions were plated on 7H11 agar medium supplemented with oleic albumin dextrose complex (OADC, Becton and Dickinson, UK) with antibiotics and was used to determine cell counts.

2. 1. 15 Genotyping of *Mycobacterium tuberculosis* isolates using large sequence polymorphisms

Mtb strains used for genotyping were collected from patients presenting with sputum smear positive pulmonary TB at QECH in Blantyre during a period of 18 months (June 2010-December 2011). A singleplex PCR based method using specific primers targeting the occurrence of each of regions of difference RD239, RD105 and RD 750 was employed. These were selected based on *Mtb* molecular epidemiology studies in and around Malawi within the last 15 years. RD239 is a marker of Lineage 1(Indo-Oceanic),

RD105 is a marker of the Lineage 2 (East Asian/*W-Beijing*) and RD750 is a marker of Lineage 3 (East African Indian). All isolates not falling within these two lineages were considered Lineage 4 (Euro-American) based on the occurrence of these lineages within the region. PCR reactions were performed in a micro centrifuge tube in a final reaction volume of 15 μ L. For an RD239 reaction mixture, 15ng of DNA was added a PCR master-mix containing 1X buffer (Bioline, UK), 0.2mM dNTPs, 0.5 μ M of each primer (forward and reverse), 1.50mM MgCl₂, double distilled water and 0.5 μ M Taq polymerase (Bioline, UK). PCR reactions were performed using a denaturation step at 95°C /180 seconds, followed by five cycles of 95°C /30 seconds, 64°C/30 seconds, 72°C/30 seconds, and 72°C/360 seconds.

For the RD105 LSP the reaction mixture contained 15ng of DNA and a PCR master mix of 1X buffer without MgCl₂ (EHF), 0.2mM dNTPs, 1.5mM MgCl₂, 0.3 μ M each primer (forward and reverse) and 1 U EHF Taq polymerase. PCR reactions were performed using a denaturation step at 95°C/180 seconds, followed by 25 cycles at 95°C/15 seconds and 64°C/240 seconds and finally 72°C/360 seconds. The RD 750 PCR assay was run as previously described (Reed, Pichler et al. 2009) with modifications. In brief the reaction mixture composed of a 15 μ l volume containing 15ng of genomic DNA, 0.4 μ M of oligonucleotide, 200 μ M of dNTPs, 1.5mM of MgCl₂, 1X Eurofins Taq buffer (+KCl) and Taq Polymerase (Eurofins). A denaturation step at 95°C /120 seconds was followed by 25 cycles of 95 °C /15 seconds, 56°C /15 seconds, 72°C /240 seconds and a cycle of 72°C/360 seconds. Oligonucleotide primers used for RD239 were forward 5'-GGC CAA CAT CGA CCA CCT ACC C-3' and reverse 5'-ATC CTC GCT ACC GGC ACC TCA T-3'. For RD105 forward 5'-GGA GTC GTT GAG GGT GTT CAT CAG CTC AGT C-3' and

reverse 5'-CGC CAA GGC CGC ATA GTC ACG GTC G-3' were used. For RD750 forward 5'-GTC AAC TGC CGA TGG CTG AC-3' and reverse 5'-CGT CAG CGA TGA TCA CCT CG-3' were used.

2.1.16 Agarose gel electrophoresis

Five microliters of the extracted DNA solution were separated on 1.5% agarose gel at 120 volts for 40 minutes and visualized by staining with ethidium bromide. Gel DocTM EZ Gel Documentation System (Bio-Rad) was used to capture gel images. Image LabTM Software bundled in the Gel Doc System was used in image analysis.

2.2 DNA sequencing

DNA sequencing was performed by the staff at the Centre for Genomic Research (CGR), Institute of Integrative Biology, University of Liverpool, United Kingdom. Extracted DNA was purified with 1x cleaned AMPure beads (Agencourt) and the quantity and quality was assessed using the Qubit and NanoDrop assays respectively. In addition, the Fragment Analyzer (using a high sensitivity genomic kit) was used to determine the average size of the DNA and the extent of degradation. This procedure was also used at the steps indicated below to determine average fragment size of the DNA. DNA was used in the protocol without further shearing. DNA was treated with Exonuclease VII at 37 °C for 15 minutes. The ends of the DNA were repaired as described by the manufacturer (Pacific Biosciences, Menlo Park, CA, USA). Samples were incubated for 20 minutes at 37 °C with DNA damage repair mix supplied in the SMRTbell library kit (Pac Bio). This was followed by a five minute incubation at 25 °C with end repair mix. DNA was cleaned using 0.5x AMPure and 70% ethanol washes. DNA was ligated to adapter overnight at 25 °C. Ligation was terminated by incubation at 65°C for 10 minutes followed by

exonuclease treatment for one hour at 37°C. The SMRTbell library was purified with 0.5x AMPure beads. The library was size selected with 0.75% blue pippin cassettes in the range 7000-20000 bp. The recovered fragments were damage repaired again. The quantity of library and therefore the recovery was determined by a Qubit assay and the average fragment size determined by Fragment Analyzer. SMRTbell library was annealed to sequencing primer at values predetermined by the Binding Calculator (PacBio) and a complex made with the DNA polymerase (P6/C4 chemistry). The complex was bound to Magbeads and this was used to set up the required number of SMRT cells for the project (two for each sample). Sequencing was performed on Pacific Biosciences RSII sequencing system (Pacific Biosciences, Menlo Park, CA, USA) using 360-minute movie times per cell, yielding ~ 350x average genome coverage.

2. 2. 1 Bioinformatics analyses of SMRT sequence data

Analysis of Bioinformatic data was conducted using the RS_Modification_and_Motif_Analysis.1 protocol in SMRT Portal software (version 2.2.0). Sequence reads were assembled using Hierarchical Genome Assembly Process 2 (HGAP2) and mapped using the Basic Local Alignment with Successive Refinement (BLASR) (Chaisson, Tesler 2012) algorithm within the SMRT portal. The BLASR pipeline is specifically designed for long reads. Annotations were done by comparing with the well annotated reference genome H37Rv (GenBank accession NC_00962). Standard settings (QV of modified motifs >40) in SMRT Portal were used to detect base modifications in the genome. A modification QV is a score that gives the confidence to which a modified site is correctly predicted as being modified (Naidu 2014). A score of 20 represents a 99% chance that the site is indeed modified with a p value of 0.01.

Following identification of methylation sites downstream analysis was done to determine methylation patterns. The inter-pulse duration (IPD) ratio which is the difference between the observed and the expected IPD for an unmethylated position was measured. An IPD ratio of one represents no methylation while an IPD ratio of greater than one denotes methylation (Berney, Berney-Meyer et al. 2015). The methylated motif site ratio which is the proportion of modified motifs of all the motifs (modified + unmodified) was also determined. This was calculated automatically by SMRT portal (Zhu, Zhong et al. 2015)

2. 2. 2 Analysis of DNA methylation patterns using SMRT sequencing

To analyze methylation patterns in *Mtb* the reference file for H37Rv (GenBank accession NC_00962) containing all genomic annotations in general features format (gff) was downloaded. All rows in the genome containing m6A modifications were extracted using a custom script run in Python software.

The genomic ranges for the genes were obtained from ProteinTable166_159857.txt (http://www.ncbi.nlm.nih.gov/genome/proteins/166?genome_assembly_id=159857). If the gene was on the positive strand, the start site was reduced by 100 nucleotides while if the gene was on the negative strand, the start site was extended by 100 nucleotides. Any m6A modification found within the gene's genomic range was reported.

2. 2. 3 Prediction of promoter regions

Genes of prokaryotes occur in clusters called operons. Genes in an operon are co-transcribed as a single mRNA under the control of the promoter of the head gene (Lim, Lee et al. 2011). Operon gene organization is capable of altering gene expression in the presence of specific regulatory mechanisms (Lim, Lee et al. 2011). A promoter is a region between two genes and is usually found between 10 and 35 bases upstream of the

transcription start site. A promoter is the binding site of the RNA polymerase during transcription (Newton-Foot, Gey van Pittius 2013).

To predict the operons and promoter regions of *Mtb* a gff file of H37Rv (GenBank accession NC_00962) was uploaded and analyzed using a BPROM tool using soft berry software:

www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb

A tab separated values (tsv) file was generated with positions of start and end coordinates of operons. It was possible to identify the coordinates of promoter regions by subtracting 100 nucleotides upstream of the start codon of the first gene in the operon. Any methylation site occurring within greater than or equal to the promoter start site and less than or equal to the promoter end site was considered to be within the promoter region. The largest single operon consisted of eleven genes.

2. 2. 4 Comparative analysis of methylated sites within persisters and non-persisters

Methylation patterns including by strand were compared within persister and non-persister *Mtb* isolates. Methylated sites within genic, intergenic and promoter regions were determined. Having identified the methylated genes, a gene ontology (GO) enrichment analysis of these genes was conducted using a functional enrichment analysis tool (Panther 2018) to determine which gene categories were over-represented in persisters.

2. 2. 5 SNP calling and identification of mutations causing loss of methylation

MAFFT software (version 7) run in unix as previously described (Katoh, Standley 2013) was used in multiple sequence alignment of consensus fasta files against the H37Rv reference genome to identify SNPs and indels. MAFFT software was also used to identify large genomic deletions. The software was also used in multiple sequence alignment of

methyltransferase genes using specific reference sequences to identify possible mutations responsible for loss of methylation. A 7 bp deletion in the *pks15/1* gene characteristic of all lineage 4 *Mtb* isolates (Constant, Perez et al. 2002) was also identified in the same manner. A phylogenetic tree was constructed for the 18 sequences using RAxML (Randomized Accelerated Maximum Likelihood) a program for phylogenetic analysis in GTRCAT model (Stamatakis 2014) based on all polymorphic sites. *M. cannetti* was used as an outlier. The branch support values were determined using 1000 bootstrap replicates. To identify SNPs within genes implicated in persistence, gene sequences were extracted from the each of the four sequences (two failures and two successes). Gene ranges included were equal or less than the start site of the gene and equal to or greater than the gene end site. These sequences were aligned with the reference sequence to determine SNPs and indels. All data and analysis are freely available online at: <http://github.com/codemeleon/MtbMethylationProfile>.

2. 2. 6 Comparison of Malawian isolates against a global sample

In order to compare Malawian isolates against a global sample, 26 SMRT sequenced genomes were downloaded from <https://www.ebi.ac.uk/ena/data/view/PRJEB21888> and analyzed using SMRT portal version 5.1.0. After running SMRT analysis all genomes with analysis steps < 60 were considered to have low quality reads and were excluded from downstream analysis and therefore only 16 genomes were used. For the base modification analysis the reference file for H37Rv (GenBank accession NC_00962) containing all genomic annotations in general features format (gff) was downloaded and the genomes analyzed as detailed in sections 2.2.3 and 2.3.4. For the phylogenetic analysis, a fasta file

of the reference file for H37Rv (GenBank accession NC_00962) was used and analysis conducted as detailed in section

2. 2. 7 Ethical approval

The study from which these isolates were obtained had ethical approval previously granted by the College of medicine research ethics committee, University of Malawi (P.01/10/855) and LSTM ethics committee (09.67).

CHAPTER THREE:OPTIMIZATION OF A DNA EXTRACTION METHOD FOR SINGLE MOLECULE REAL TIME (SMRT) SEQUENCING

3. 1 Introduction

The arrival of high throughput sequencing platforms such as Pacific Biosciences Single Molecule Real Time (SMRT) Sequencing has led to unprecedented use of whole genome sequencing to better understand the genome and epigenome of *Mycobacterium tuberculosis* (Somerville, Thibert et al. 2005). Other whole genome sequencing platforms such as Illumina Hiseq are unable to completely sequence genomes with high GC content and repetitive sequences (Doughty, Sergeant et al. 2014). The *Mtb* genome is characterized by high GC content (~65%) and numerous repetitive PE/PPE regions (Cole, S. T., Brosch et al. 1998). While PCR based analyses are amenable to fragmented or heterogeneous DNA (Caws, Duy et al. 2006, Fresquet-Wolf, Haas et al. 1998, van Vollenhoven, Heyns et al. 1996), preparation of sequencing libraries requires good yields of high quality, pure DNA (Cole, S. T., Brosch et al. 1998). To obtain good yields of high molecular weight DNA from *Mtb*, a long culturing step is deemed necessary (Merker, Kohl et al. 2013, Witney, Gould et al. 2015). The detergent CTAB is useful in the chemical lysis of cells from organisms with large quantities of polysaccharides (Tan, Yiap 2009). The method that utilizes CTAB has been applied in several studies (Honore-Bouakline, Vincensini et al. 2003, van Soolingen, Hermans et al. 1991). This method has been shown to yield considerably higher amounts of pure DNA compared to other techniques (Ali, Al-Thwani et al. 2014). Mycobacterial DNA may be recovered from a number of sources including solid or liquid cultures and biological specimens such as sputum, urine, biopsies

and in vitro cultures (van Helden 2001). Recovery of pure *Mtb* DNA from liquid cultures such as MGIT is particularly problematic however due to human and bacterial DNA contamination (Votintseva, Pankhurst et al. 2015). Unlike other whole genome sequencing platforms SMRT sequencing does not incorporate an amplification step. This necessitates use of high molecular weight double stranded native DNA (Davis, Chao et al. 2013). Furthermore to prepare a 10kb SMRTbell library DNA should be a minimum of 5µg and free from RNA and insoluble materials (Pacific Biosciences 2015). Long read lengths and high sequencing coverage are critical due to the high GC content and the repetitive PE/PPE regions of the *Mtb* genome. Such stringent DNA requirements in SMRT sequencing mean that the extraction method should be harsh enough to lyse the tough mycobacteria cell wall and at the same time gentle enough so as not to shear the DNA. Sheared DNA will create problems during PacBio library preparation. Previous studies to sequence *Mtb* DNA using SMRT technology have largely used commercial kits (Naidu 2014, Zhu, Zhong et al. 2015). To date no published study has employed an in-house DNA extraction method for SMRT sequencing of *Mtb*.

3. 2 Aims

This chapter discusses optimization of a suitable DNA extraction method compatible with SMRT sequencing technology.

3. 3 Results

To optimize a method that would extract DNA compatible with SMRT sequencing, an evaluation of two DNA extraction techniques that incorporate a CTAB step; the phenol/chloroform/isoamyl alcohol and a chloroform/ isoamyl alcohol method was conducted.

3.3.1 The Chloroform/Phenol/Isoamyl alcohol method

Mycobacterium tuberculosis is made up of a tough cell wall consisting of various lipids and polysaccharides. To lyse this tough cell wall, enzymes and chemicals are usually used sometimes in combination with bead beating. In order to obtain sufficient DNA yield, an enzymatic chemical lysis method that utilizes chloroform: phenol/isoamyl alcohol in the purification step (Total time $11 \pm$ hours) was employed. In this study the DNA extraction method was evaluated with bead beating/ homogenization at cell optical density (OD) =1 at 540nm. Bead beating has been reported to be an efficient method of DNA isolation in *Mtb* (Amaro, Duarte et al. 2008). Therefore, the enzymatic/chemical lysis method with bead beating was evaluated at durations of one and five minutes (Materials and Methods 2.2.1.1). DNA was extracted using the same CTAB method at two different bead beating durations of one minute and five minutes. An equal loopful amount of sample was used for each extraction. Each extraction was done in duplicate. Extracted DNA was quantified using a qubit fluorometer and concentration expressed as $\mu\text{g/mL}$. The results were as follows: (Mean+SEM) 1 minute; 162.5 ± 114.6 , (n=3), five minutes; 181.1 ± 123.5 , (n=3), (P=0.92,95%CI -450.3 to 487.5 t-test with Welch's correction).The duration of bead beating did not significantly affect the DNA yield (Figure 3.1). It was therefore decided the shortest bead beating duration of one minute be adopted. The bead beating method was thereupon compared with the homogenization method in subsequent experiments. To ensure reproducibility each extraction of six samples was done in triplicate. Mean DNA concentrations ($\mu\text{g/mL}$) of 121.2 ± 55.81 (mean \pm SEM) and 81.32 ± 27.29 (mean \pm SEM) were obtained for bead beating and homogenization methods respectively (n=6) (Figure 3.1). The highest concentration obtained using the bead beating method was $389.2 \mu\text{g/mL}$

with the lowest being 24.7 $\mu\text{g/mL}$. Comparatively the homogenization method had 201.5 $\mu\text{g/mL}$ and 21.7 $\mu\text{g/mL}$ being the highest and lowest concentrations respectively. Although the mean concentration for the bead beating method was higher, no statistically significant difference in concentration between the two methods was observed ($P= 0.54$, 95% CI -185.7.3 to 105.9, t-test with Welch's correction). An assessment of the DNA purity using the A260/A280 ratio on a Spectrophotometer ND-1000 was done. Values of 1.518 ± 0.08616 (mean \pm SEM) and 1.532 ± 0.1258 (mean \pm SEM) for bead beating and homogenization respectively were obtained (Figure 3.3). Both these results were lower and outside the purity requirements for PacBio SMRT sequencing of 1.8-2.0. Analysis using gel electrophoresis revealed that most of the DNA obtained bead beating was degraded (Figure 3.2)

Comparison of bead beating duration on DNA concentration

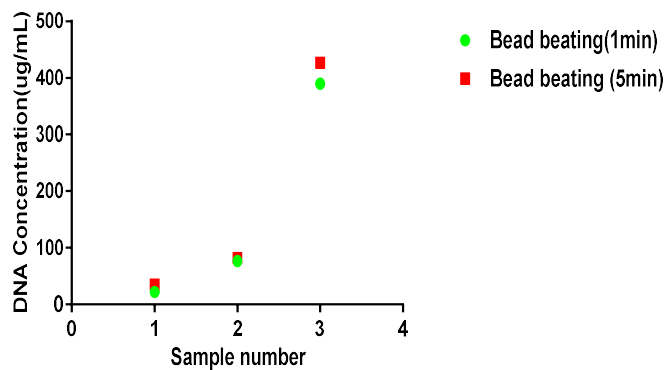


Figure 3. 1 Comparison of bead beating duration on DNA concentration

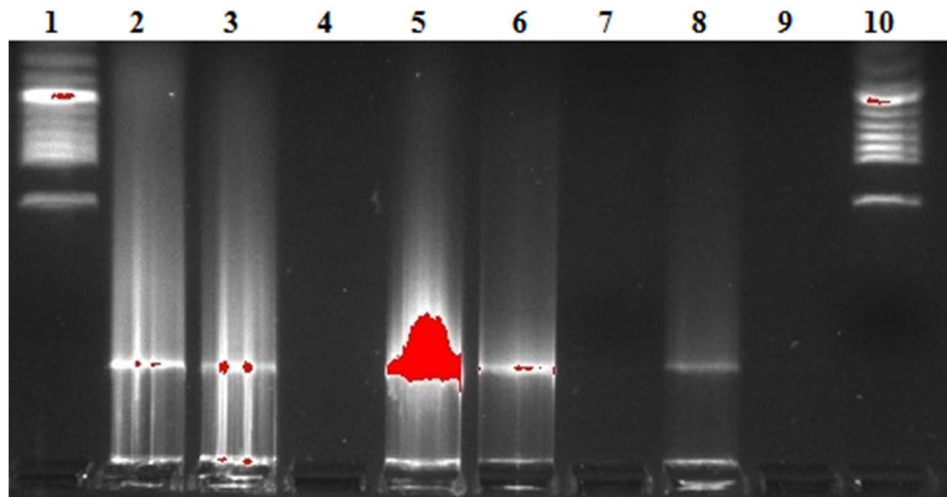


Figure 3.2 Gel electrophoresis image of genomic DNA obtained using the bead beating method

Lanes 1 and 10 Lambda HindIII DNA Ladder. Lanes 2-8 *Mtb* DNA from cultured clinical isolates showing mostly highly degraded DNA and lane 9 was blank (nuclease-free water) as a negative control. Lanes 4 and 7 were insufficiently loaded with sample.

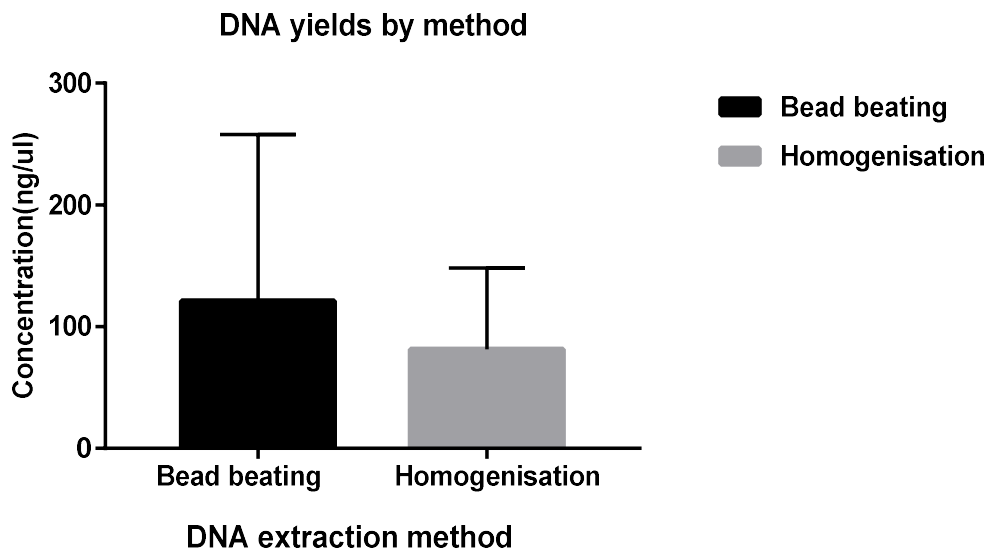


Figure 3. 3 Comparison of DNA concentration by extraction method.

Mtb culture was scrapped off from late exponential phase LJ slants and extracted using the chloroform; phenol; isoamyl alcohol method. Cells were heat killed at 80C for 40 minutes. Disruption of cells was done by chemical/enzymatic method with bead beating or homogenization. Error bars represent Mean+SEM. Results ($\mu\text{g}/\text{mL}$), bead beating: 121.2 ± 55.8 (n=6) and homogenization: 81.3 ± 27.3 (n=6) (P=0.5, CI, -185.7 to 105.9 t-test with Welch's correction).

3. 3. 2 The Chloroform/Isoamyl alcohol method

A decision was made to evaluate a faster and less complex method which would give fairly pure and sufficiently high yields of DNA. A chemolysis/enzymatic method without the use of phenol; (Materials and Methods 2.1.13) was evaluated. This method is less complicated and considerably shorter (Total time $7 \pm$ hours). The aim was to compare the yield and purity of DNA obtained from LJ slants with those from MGIT tubes. This comparison was necessary as a MGIT culture would offer a quicker alternative to conventional solid culture. Extraction of 11 samples in triplicate was conducted in order

to ensure reproducibility of the technique. Concentrations of ($\mu\text{g/ml}$) of 228.6 ± 54.1 , $n=11$ and 28.73 ± 10.83 , $n=11$ (Mean \pm SEM) for LJ and MGIT respectively were obtained (Figure 3.4). The results from LJ slants were significantly higher than those from MGIT ($P=0.004$, 95% CI -321.6 to -78.2). Purity values A260/A280 were 1.695 ± 0.1896 , $n=11$ (Mean \pm SEM) for LJ and 2.739 ± 0.4512 , $n=11$ (Mean \pm SEM) for MGIT. A 260/280 ratio of 1.8-2.0 was considered as a pure DNA preparation. These results were outside the purity range of 1.8-2.0 required for SMRT sequencing although LJ results were closer to the required purity value.

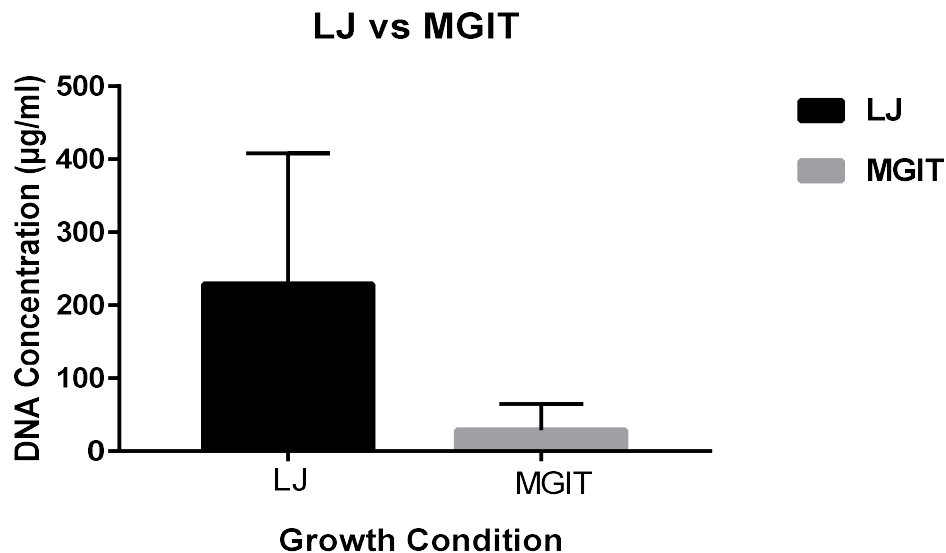


Figure 3. 4 Comparison of DNA yields from LJ slants and MGIT culture

Mtb cells were grown under two different growth conditions: LJ slants and MGIT tubes respectively. Cells were harvested at 6 and 8 weeks and DNA was extracted using the chloroform;isoamyl alcohol (CTAB) method. Extracted DNA was quantified using a qubit fluorometer and concentration expressed as $\mu\text{g} /\text{mL}$. Error bars represent Mean \pm SEM. Results, LJ: 228.6 ± 54.1 ($n=11$) and homogenization: 28.7 ± 10.8 ($n=11$) ($P=0.004$, CI, -321.6 to -78.2, t-test with Welch's correction).

Table 3. 1 General Summary of DNA extraction methods and yield/purity

Extraction Method	Procedure/Source	Concentration ($\mu\text{g/mL}$)	Purity 260/230
		Mean \pm SEM	Mean \pm SEM
CTAB (Phenol)	Bead beating	121.2 \pm 55.8	1.518 \pm 0.08616
	Homogenization	81.3 \pm 27.3	1.532 \pm 0.1258
	LJ	228.6 \pm 54.1	1.695 \pm 0.1896
CTAB (Without Phenol)	MGIT	28.73 \pm 10.83	2.739 \pm 0.4512

3. 4 Discussion

The success of downstream molecular applications is highly dependent on the quality and quantity of the DNA available. Choice of a DNA extraction method is therefore critical to success. One published study that has sequenced *Mtb* on the PacBio SMRT sequencing platform has used a TIANamp Bacteria Genomic DNA kit (Zhu, Zhong et al. 2015) while a more recent study used the CTAB method (Phelan, de Sessions et al. 2018). The amount of DNA obtained from a commercial kit is however limited by its design. This is due to the fact that most kits use 1.5-2.0 mL tubes or spin columns that take samples not exceeding 500 μl . The starting material is therefore usually low limiting the maximum amount of DNA that can be extracted. This study sought to find an extraction method that

would allow us to increase the input material while giving pure high molecular weight double stranded DNA. The use of CTAB in the extraction procedure helps to improve the purification process. CTAB has been recommended for extraction from solid cultures (Amaro, Duarte et al. 2008). Although the combined enzymatic/chemolysis bead beating method has been previously demonstrated to yield higher DNA quantities (Amaro, Duarte et al. 2008), in this study this was not the case. It was discovered that this was due to low starting cell density. The method was also found to be time and labour intensive (Total time $11 \pm$ hours). Mechanical disruption such as bead beating is one method used to lyse the tough cell wall of *Mtb*. Gel electrophoresis visibly showed highly sheared DNA produced with the bead beating method. This result was consistent with previous studies (Amaro, Duarte et al. 2008) where it was demonstrated that bead beating leads to massively degraded DNA. This notwithstanding, for downstream applications such as real-time PCR where DNA integrity and quality may be less critical, bead beating has been demonstrated to be the most efficient and reproducible method (Pan, Gu et al. 2013). However, sheared DNA is incompatible with SMRT sequencing and consequently this method was deemed unsuitable. A low purity value for this method of 1.518 ± 0.08616 (mean \pm SD) demonstrates RNA contamination. Although the method utilizes RNase to remove RNA, it is likely that it did not work efficiently. A longer incubation period than the two hours may have helped to improve the purity of the extracted yield.

The homogenization method has the advantage of not using mechanical disruption to lyse the cell thereby avoiding shearing of the DNA. Low DNA yield from the homogenization method could have resulted from failure to effectively weaken the hard cell wall of *Mtb* as there was no use of harsh physical treatment. Harsh pre-treatment such as freezing at -

20 °C for 20 minutes and boiling for 10 minutes help to weaken lipid linkages of the cell wall, freeing chromosomal DNA (Amita, Vandana et al. 2002).

Since both yield and purity of DNA obtained using the homogenization (CTAB-phenol) method did not significantly differ from that obtained using the bead beating method it was also deemed incompatible with SMRT sequencing. The two methods therefore, were not adopted for SMRT sequencing.

The enzymatic/chemolysis method without the use of phenol was adopted. The study compared yields and purity of DNA from LJs and MGIT tubes using enzymatic/chemolysis without phenol. Mean DNA yield for LJ was significantly higher than that from MGIT (Figure 3.2). This method was adopted for use as it was able to consistently recover >5µg of DNA from an LJ culture which is required to prepare a 10Kb insert SMRT library and up to 2 µg from a MGIT culture which is required to prepare a 5Kb insert library. LJ medium is selective to *Mtb* and efficiently inhibits growth of other bacteria and fungi. On the other hand, MGIT medium is more prone to contamination and may allow proliferation of various other bacterial and fungal organisms. It was therefore appropriate to infer that the DNA yield obtained from MGIT could also have contained fungal and other non-*Mtb* DNA. In addition, purification of DNA from a MGIT culture proved to be challenging as evidenced by both low yield and purity obtained by the method compared to LJ. An additional purification step was used to improve the purity of the DNA. Ultimately the chloroform:isoamylalcohol (CTAB) method was adopted and used for DNA extraction from solid cultures for SMRT sequencing.

It has been demonstrated in this chapter that the CTAB method remains the gold standard for the extraction of *Mtb* DNA from solid culture. Additionally, the CTAB

chloroform/isoamylalcohol method without phenol has been found to be faster and easier to use than the phenol method. The method is feasible for SMRT sequencing. Furthermore, the method can be applied to SMRT sequencing from a MGIT culture. The study has further demonstrated that the use of bead beating is insignificant to the final DNA yield.

CHAPTER FOUR: COMPARISON OF PRIMEXTRACT DNA EXTRACTION KIT AND AN IN- HOUSE CTAB METHOD

4. 1 Introduction

The purpose of this chapter was to evaluate two DNA extraction methods for direct extraction of DNA from *Mtb* sputum. This comparison would potentially aid in identifying an extraction method that is fast and efficient in isolating DNA directly from sputum and also compatible with SMRT sequencing.

Mtb is an agonizingly slow growing organism and conventional culture typically takes between two to eight weeks (Nolte, Metchock et al. 1993). However, WGS from a positive mycobacterial growth indicator tube (MGIT) culture is achievable in a mean time of two weeks (Fadzilah, Ng et al. 2009, Pfyffer, Welscher et al. 1997) although a later study was able to reduce it to three days (Koser, Bryant et al. 2013). As a consequence of the period required for culture, relevant genotyping results for molecular epidemiology can usually only be obtained in retrospect during outbreaks (Walker, Ip et al. 2013).

The long culture period has also hampered efforts to use WGS as a diagnostic tool for TB (Doughty, Sergeant et al. 2014).The ideal situation would be to efficiently sequence directly from clinical specimens such as sputum. However attempts to sequence directly from *Mtb* clinical samples have achieved very low coverage (less than 0.7X) due to high rates of human DNA contamination (Doughty, Sergeant et al. 2014).

Two published studies sequenced *Mtb* DNA extracted using commercial kits for SMRT sequencing but samples were from culture and not the primary sputum samples (Zhu, Zhong et al. 2015).To date, and to the best of our knowledge, no study has attempted

SMRT sequencing directly from clinical samples either through a commercial kit or an in-house DNA extraction method.

The CTAB method has been used extensively in the extraction of *Mtb* DNA from a variety of samples (Honore-Bouakline, Vincensini et al. 2003, van Soolingen, Hermans et al. 1991), one study demonstrated recovery of high yields of pure DNA directly from sputum (Amita, Vandana et al. 2002). In this study, the CTAB method was demonstrated to be superior to other in-house methods including some commercial kits. The Primextract DNA extraction kit has been claimed to recover high yields of *Mtb* DNA directly from sputum (Longhorn Vaccines 2017). No published study has evaluated the efficacy of this kit against the traditional CTAB method or any other DNA extraction method.

The purpose of this chapter is therefore to compare the efficiency of a commercially available DNA extraction kit (Primextract kit, Longhorn Vaccines and Diagnostics, USA) and an in-house CTAB based method for the extraction of DNA directly from confirmed *Mtb* positive sputum samples. The two methods were evaluated on the quantity and purity of the DNA and thus suitability of each method for SMRT sequencing. The aim of this chapter was to identify a DNA extraction method for *Mtb* compatible with SMRT sequencing.

4.2 Results

Forty sputum samples collected from the M&E study were used for the comparison. To evaluate the DNA extraction efficiency of the PrimeXtract kit against the CTAB method, an equal aliquot (200µl) of decontaminated sputum was simultaneously used for extraction. Cell density was determined by plating a sample on solid media to determine

CFU/mL. Since frozen sputum samples were used, those samples which were positive for DNA yield but yielded zero colony counts were recorded as <10CFU/mL

The two direct extraction methods were quantitatively and qualitatively evaluated. The efficiency of a method was evaluated based on the runtime, DNA yield, purity and labour intensity. Cell densities were obtained from colony counts using the formula: Number of colonies x 20 (50 µl was inoculated into each segment) x dilution factor. DNA concentrations were determined using a qubit fluorometer and the yield was calculated from the concentration in 50 µl of DNA sample with results presented as Mean ± SEM for each method. DNA purity was determined on a Nanodrop and results were also presented as Mean ± SEM. Significant differences were observed in the both the concentration and the yield of the different methods. The PrimeXtract kit produced both higher yield and concentration (Figure 4.1 and 4.2). Extraction with PrimeXtract resulted in a mean concentration (µg/mL) of 5.93 ± 0.94 (n=40) and DNA yield (µg) of 0.2975 ± 0.04723 , (n=40). By comparison the CTAB method produced a mean concentration (µg/mL) of 1.88 ± 0.38 (n=40) and DNA yield (µg) of 0.09 ± 0.02 (n=40). Both concentration (P=0.0002) and yield (P=0.0002) from PrimeXtract were significantly higher than those obtained using the CTAB method. The PrimeXtract kit had a DNA purity (260/280) ratio of 1.69 ± 0.09 compared to the CTAB's 1.73 ± 0.14 , (n=40). PrimeXtract's purity was lower than CTAB's, although this result was not statistically significant (P=0.76). Purity values for the CTAB method however, were closer to the target purity range of 1.8-2.0.

Comparison of two DNA extraction method

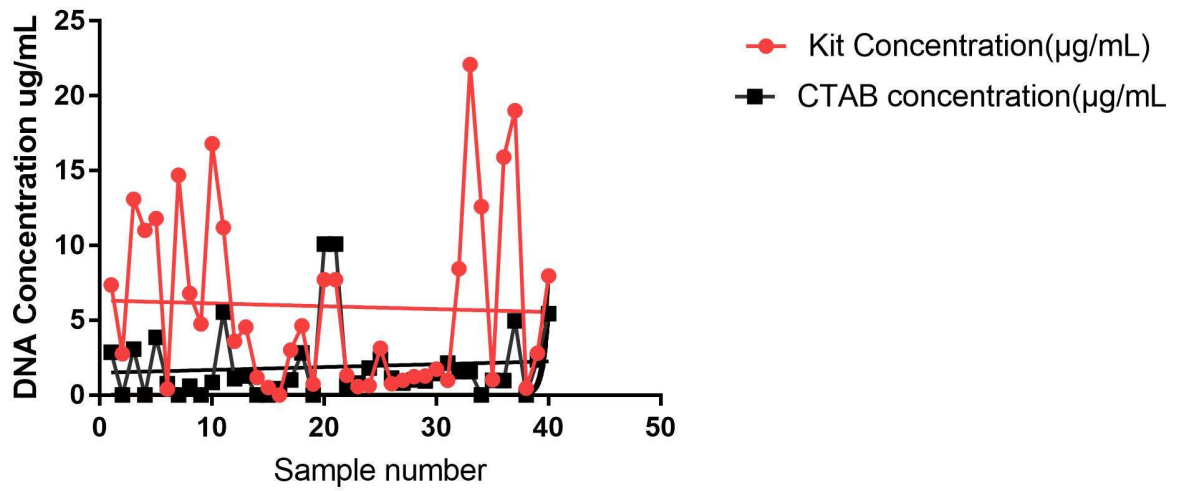


Figure 4. 1 Concentrations of DNA extracted using two different methods

DNA concentrations from two different extraction methods; the PrimeXtract kit (Longhorn Vaccines, USA) and an in- house CTAB method (n=40). Concentration was measured using a qubit fluorometer ($\mu\text{g/mL}$)

Comparison of DNA yield by method

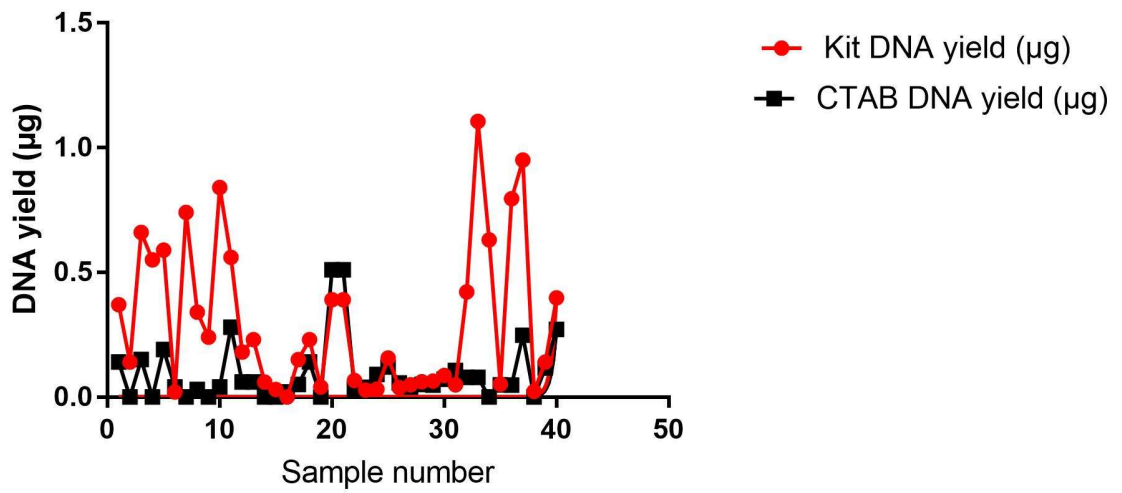


Figure 4. 2 Yields of DNA extracted using two different methods.

Yields of DNA obtained using two different extraction methods: The PrimeXtract kit (Longhorn vaccines, USA) and an in-house method CTAB (n=40). Yields were calculated from concentrations (ug)

Comparison of DNA purity by Extraction method

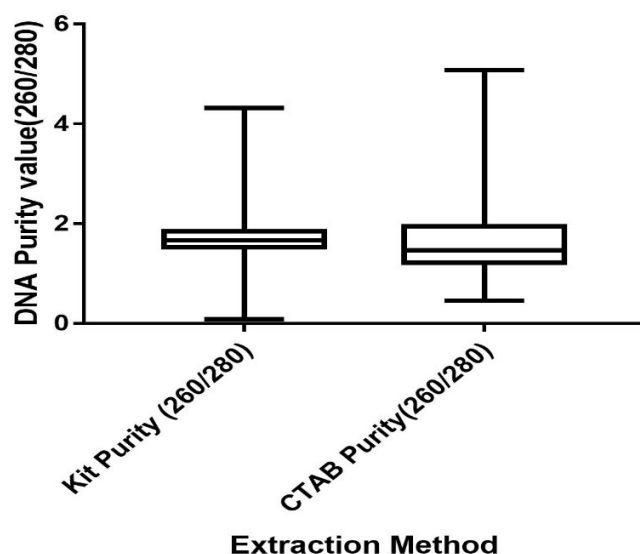


Figure 4. 3 Purity of DNA extracted using two different methods

Purity of DNA obtained using two different extraction method: The PrimeXtract kit (Longhorn Vaccines, USA) and an in-house CTAB method (n=40). Purity was obtained using an ND-1000 Spectrophotometer Nanodrop. A 260/280 value of 1.8-2.0 was considered a pure preparation. Error bars represent standard error of the mean (SEM)

Table 4. 1 Summary of results of DNA extraction using PrimeXtract and CTAB method

Method	Concentration($\mu\text{g/ml}$)	Yield (μg)	Purity	Cost	Total
			260/280	/Sample	Time
	Mean \pm SEM	Mean \pm SEM	Mean \pm SEM	\$	Hours
PrimeXtract	5.93 \pm 0.94	0.29 \pm 0.047	1.69 \pm 0.09	5.76	0.5
CTAB	1.88 \pm 0.38	0.09 \pm 0.02	1.73 \pm 0.14	0.39	7

n=40

One published study demonstrated that by employing a saline wash step in the extraction of *Mtb* DNA from sputum, it is possible to significantly remove host contamination from the bacterial DNA and improve the overall purity of the output (Votintseva, Pankhurst et al. 2015). Since it was found in this study that increasing the starting cell density in the CTAB method resulted in corresponding increase in DNA output, an attempt was made to increase the cell density but apply the saline wash step to improve the DNA purity. By extracting DNA from 500 μ l aliquots of 18 sputum samples it was revealed that the saline wash step significantly decreased the DNA output (Figure 4.4). Without the saline wash a mean concentration of 23.21 ± 6.154 , (n=18) was obtained whereas with the saline wash it was 2.002 ± 0.5754 , (n=18). There was a statistically significant difference (P=0.0016, unpaired t-test) between the concentrations of DNA obtained using the two methods. The mean DNA purity values (260/280) obtained by the two methods were 4.08 ± 0.985 (n=18) with the saline wash and 1.566 ± 0.1618 (n=18) with saline (Figure 4.5). Again there was a statistically significant difference between the two methods (P=0.02, unpaired t-test). The saline wash step was consequently abandoned.

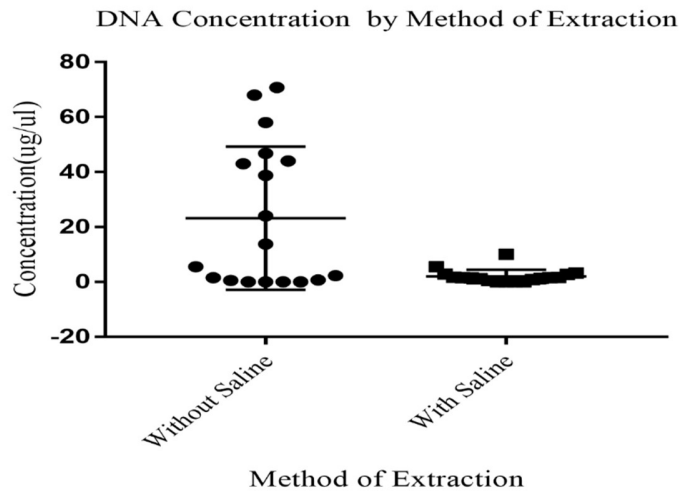


Figure 4. 4 Comparison of DNA concentrations obtained by CTAB method with and without a saline wash step

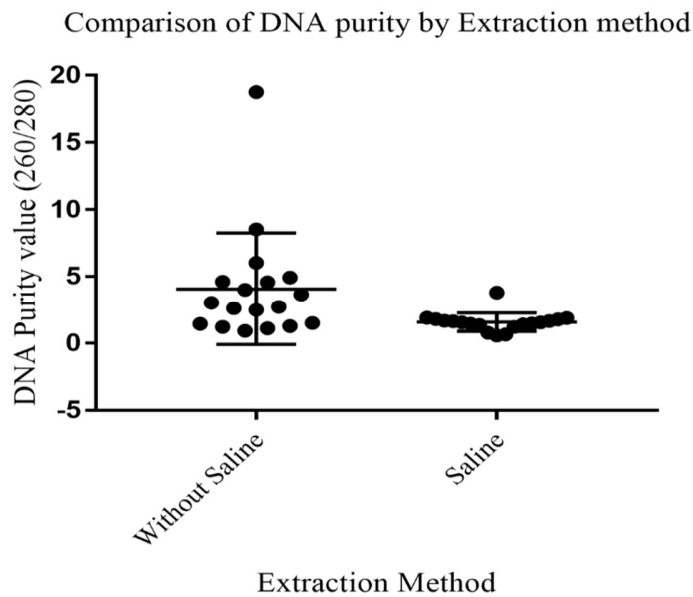


Figure 4. 5 Comparison of DNA purity obtained by CTAB method with and without a saline wash step

4.3 Discussion

In order to efficiently extract DNA for the purpose of sequencing choosing a suitable DNA extraction method is a critical step. With *Mtb* recovery of high molecular weight DNA is challenging even with relatively large volumes of cultured isolates.

Furthermore, unlike other WGS technologies, SMRT sequencing uses native DNA without an amplification step. In order to prepare a 10Kb insert library a minimum of 5µg of genomic DNA is typically required (Centre for Genomic Research, University of Liverpool). This amount cannot be easily recovered from *Mtb* without a culture step.

The PrimeXtract kit has been highly recommended for recovery of high molecular weight quality DNA directly from *Mycobacterium tuberculosis* clinical specimens (Longhorn Vaccines 2017). To date, no published study has evaluated the efficiency of the kit, either against other kits or an in-house extraction method.

In the present work, it has been demonstrated that the PrimeXtract DNA extraction kit is superior to the CTAB method for the isolation of *Mtb* genomic DNA directly from sputum. PrimeXtract was capable of recovering both higher quantities and better purity DNA than the comparator in the 200µl sample under consideration. The superiority of PrimeXtract is demonstrated in almost all the samples evaluated (Figures 4. 1-4. 3) where despite an equal amount of input material, it has consistently higher DNA yield and purity output. PrimeXtract uses a specialized lysis solution and wash buffers and aid to release the DNA and wash off any cell impurities. These are critical in release and clean-up of DNA. The sensitivity of the kit is enhanced by a silica-based spin column that traps the DNA during centrifugation though this format limits the amount of input material. The

other advantage of the kit is that it is easy to use with short centrifugation times (60 seconds each) and the total time to complete the extraction procedure is only \pm one hour . In the present study however, the kit was deemed unsuitable for SMRT sequencing. It could only recover less than the 5 μ g of extracted DNA which is the minimum the requirement for a 10Kb insert library preparation.

The CTAB method has been shown to recover higher yields of DNA with good purity from clinical samples compared to other in house methods (Amita, Vandana et al. 2002). In this study this was not the case. Some studies have suggested that the CTAB method is efficient for DNA extraction from solid *Mtb* cultures (Amaro, Duarte et al. 2008). Successful isolation of *Mtb* DNA, especially from clinical samples, requires harsh treatment to weaken the cell wall. In this study the use of heating and immediate freezing to disrupt connections holding lipid contents of the cell wall together was employed. It is obvious that this did not work very well or that the reagents used by the kit were just superior. Other studies propose heating at 100°C in an appropriate buffer to achieve better results (Cormican, Barry et al. 1992). It has further been suggested that phenol and chloroform aids in improving the yield (Sjobering, Mecklenburg et al. 1990). The study avoided using phenol as residual phenol is incompatible with PacBio library preparation. Chloroform aids in denaturation of proteins and formation of a separating layer between aqueous and organic phases (Amita, Vandana et al. 2002). From experience the protein removal and DNA precipitation steps were appropriate as these have previously been successfully applied. Isopropanol was used to selectively remove DNA leaving RNA and other impurities. The remaining impurities were washed off from the DNA then using 100% ethanol.

There were significant variations among the samples on the purity obtained using the two methods (Figure 4. 3). These variations are expected as these samples were likely to have had different levels of impurities/contaminations. The CTAB method however had better mean purity value of 1.73, which is closer to the required range of 1.8-2.0 compared to the kit's 1.69 which is lower. Since there was no use of phenol, the lower 260/280 ratio would most likely indicate protein contamination. This is suggestive of the fact that the chloroform step did not work properly.

The CTAB method however has the advantage of being flexible with regard to volume of material and capacity for further optimization. It is possible to increase the cell density and successfully recover higher amounts of DNA. To improve purity however a eukaryotic DNA, clean up step may be necessary. Another point to consider in the choice of an extraction method would be the cost. The PrimeXtract kit costs \$5.76 per extraction (Longhorn Vaccines 2017). Compared to other DNA extraction kits, \$2.30 for Qiagen and \$1.05 for Prepman (Aldous, Pounder et al. 2005) the PrimeXtract is much more expensive. Comparatively the CTAB method costs ~\$0.39 per extraction (Lickfeldt, Hofmann et al. 2002). Evidently the PrimeXtract despite being more efficient in terms of time, labour and quantity of DNA, is much more expensive than the CTAB method. It was observed that all samples had positive DNA concentrations despite some having zero CFUs. Though this discrepancy could simply be caused by release of mycobacterial DNA into the sample after cell death. It could also be as the result of a presence of persisters or resuscitation promotion factor (rpf) dependent cells which do not easily grow on solid media. Another challenge with the recovery of DNA using this approach is the use of a decontamination step during routine processing of the sputum samples. While decontamination greatly

reduces the risk of contamination from non-tuberculous bacterial and fungal organisms it is known to greatly compromise recovery of *Mtb*, reducing the load of organisms by 1-2 orders of magnitude (Sloan, D. 2013).

The results from this study suggest that the PrimeXtract kit is superior to the CTAB method for direct extraction of *Mtb* DNA from sputum samples. With proper optimization, the PrimeXtract kit holds the potential to be used in direct DNA extraction from sputum for the purpose of SMRT sequencing. However, the technique is proprietary and relatively expensive. This study has previously sequenced cultured isolates using the CTAB method on the SMRT platform. For resource-poor settings where kits are not readily available CTAB will likely remain the method of choice. Current data however suggests it is possible to recover adequate amounts of DNA from direct CTAB extraction by increasing the amount of input material but a clean-up step will necessarily be required in order to achieve the required DNA purity. The saline wash step that was tested in this study was found to significantly negatively affect the DNA yield. More studies will be required to develop a DNA clean-up method for *Mtb* which while improving the purity of the DNA also maintains the required yield.

The comparison of the two methods as presented here has some limitations. It is not possible to say with certainty that all the DNA recovered in this study was *Mtb* DNA as sputum will contain high levels of human and non-tuberculous DNA. The optimal clean-up method for SMRT sequencing is an area requiring further research.

CHAPTER FIVE: GENETIC DIVERSITY OF MYCOBACTERIUM TUBERCULOSIS ISOLATES IN BLANTYRE, MALAWI

5.1 Introduction

Despite the high burden of TB worldwide, specific factors responsible for disease transmission remain elusive. Long term epidemiological studies provide evidence of relative fitness of *Mtb* strains but few such studies are available (Cowley, Govender et al. 2008). Several methods are currently in use for molecular typing and lineage assignment of *Mtb* strains. Some studies have used insertion sequence 6110 restriction fragment length polymorphism (IS6110 RFLP) typing in *Mtb* epidemiology and the method was once regarded as the gold standard (Stavrum, Mphahlele et al. 2009). The validity of data from IS6110 RFLP typing has however been hampered by the fast molecular clock with a short half-life of this method ~3.5 years (Glynn, JR, Alghamdi et al. 2010). Additionally, the method is labour intensive and suffers from poor inter-laboratory reproducibility. Mycobacterial interspersed repetitive units- variable number tandem repeats (MIRU-VNTR) is a PCR based method which is faster than IS6110 RFLP and offers comparatively higher discriminatory power (Supply, Philip, Mazars et al. 2000). For long term molecular trends, MIRU-VNTR in combination with spoligotyping has been efficiently used in the study of transmission and phylogenetic changes in *Mtb*. Spoligotyping is a hybridization assay that relies on absence or presence of unique spacer sequences separating copies of conserved 36 bp sequences (Gagneux, Small 2007). However with spoligotyping, homoplasmy (convergent evolution of phylogenetically unrelated strains) has been reported (Flores, Van et al. 2007, Rindi, Medici et al. 2014). . Additionally, it has been possible to describe *Mtb* lineages based on deletions and single

nucleotide polymorphisms (SNPs). Lineage classification from such studies has been closely consistent with families described from use of spoligotyping (Jagielski, van Ingen et al. 2014)

Genomic deletions and rearrangements result in large sequence polymorphisms (LSPs) and in *Mtb* these are irreversible owing to non-existence of horizontal gene transfer between strains (Mathema, Kurepina et al. 2006, Coscolla, Gagneux 2014). LSPs are a robust molecular marker and are feasible as an epidemiological investigative tool. IS transposition has been associated with some deletions and such deletions are termed regions of difference (RD). A genomic deletion analysis of 875 global samples from 80 countries revealed six *Mtb* phylogenetic lineages and 15 sub lineages. There was an association between each of the six lineages and a particular geographic region (Gagneux, DeRiemer et al. 2006, Reed, Pichler et al. 2009). A novel lineage restricted to the horn of Africa has been characterized in Ethiopia and is classified as lineage 7 (Coll, McNerney et al. 2014). These lineages were comprehensively discussed in section 1.1. Important RD loci described in *Mtb* include RD9, RD105, RD207, RD239 and RD750 among others (Flores, Van et al. 2007). These deleted fragments can be used to discriminate between *Mtb* strains by use of a simple PCR method (Desikan, Narayanan 2015). The advantage of LSP or RD-PCR is that it can be used to screen large numbers of strains in a high-throughput manner (Gagneux, DeRiemer et al. 2006). Although all the above genotyping methods are useful, whole genome sequencing (WGS) has proved to be superior for many purposes (Brown, Bryant et al. 2015). WGS is the only tool capable of definitively assigning lineages to all *Mtb* strains as well as analyzing genomic diversity at all levels (RW.ERROR - Unable to find reference:644). Additionally, it can be used to track the

source of disease outbreaks and monitor transmission patterns (Brown, Bryant et al. 2015, Roetzler, Diel et al. 2013).

Genotyping *Mtb* strains is also important in discriminating between disease relapses and new episodes, particularly in clinical trials (Stucki, Malla et al. 2012). In Malawi a few *Mtb* molecular epidemiological studies have been reported but these have mostly been from the Northern Region of the country under the Karonga Prevention Study (KPS) (Glynn, JR, Alghamdi et al. 2010, Glynn, J. R., Crampin et al. 2005). Lack of laboratories with molecular tools has hampered efforts to conduct *Mtb* molecular epidemiology studies in Malawi and in most other low resource settings. *Mtb* molecular epidemiological data for the rest of the country is therefore currently lacking. The currently prevailing *Mtb* lineages for Malawi as a whole remain to be established and the impact of transmission patterns over time elucidated.

This chapter of the thesis aimed at performing a preliminary characterization of the genetic diversity of *Mtb* clinical isolates amongst patients presenting at Queen Elizabeth Central Hospital, Blantyre, Malawi the largest referral hospital in the country. Additionally, the study sought to correlate the identified lineages to clinical phenotypes observed in *Mtb* patients including phenotypic drug tolerance.

5. 2 Results

5. 2. 1 Large sequence polymorphism (LSP)-PCR

Following a literature search of *Mtb* molecular epidemiological studies conducted in Malawi and the surrounding countries of Tanzania, Mozambique and South Africa over the last 15 years, primers were designed targeting the RDs of circulating lineages within the region (Geurra-Assuncao, Crampin et al. 2015, Cowley, Govender et al. 2008,

Mulenga, Shamputa et al. 2010, Viegas, Machado et al. 2013). These were RD239 for the Indo-Oceanic lineage, RD105 for the *W-Beijing* lineage and RD 750 for the East African/Indian lineage. Sixty-four bacteriologically culture confirmed *Mtb* positive samples were selected from a total of 133 isolates based on the ones to be successfully resuscitated from frozen state and subjected to genomic deletion analysis, as outlined in Chapter 2. The analysis revealed that six strains (9%) were of the *Beijing* lineage (RD 105 deleted) generating a deleted product ~785bp on gel electrophoresis (Figure 5. 1), ten strains (16%) were of the Indo-Oceanic lineage (RD 239 deleted) with a deleted fragment of ~888bp (Figure 5. 2), two strains (3%) were of the East African/Indian lineage (RD750 deleted) producing a deleted product of ~743bp and forty six strains (72%) were of the Euro-American lineage (Figure 5. 3).

It was not possible to design lineage specific primers for the Euro-American lineage, a common previously reported lineage in Malawi. Twelve out of the 46 isolates were therefore confirmed as Euro-American using whole genome sequencing. All other strains lacking deletions stated above were hence assumed to belong to the Euro-American lineage owing to the fact that this lineage was reported to be the most predominant in the region and has previously been reported at a high prevalence in Karonga (Geurra-Assuncao, Crampin et al. 2015). Drug sensitivity testing of all the 64 isolates revealed that one sample had INH resistance while one another lineage was multi-drug resistant.

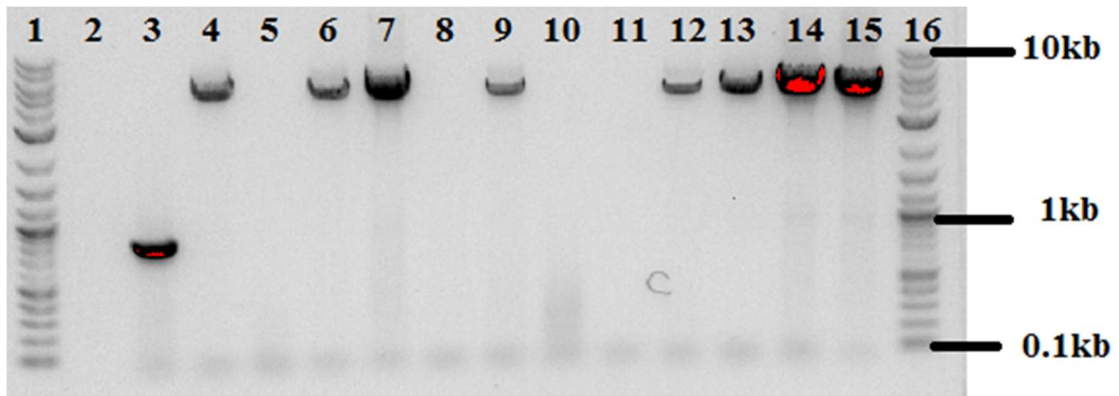


Figure 5. 1 Gel electrophoresis of RD-105 deletion PCR

Lanes 1 and 16, 2-Log DNA ladder (New England Biolabs). Lane 2 Blank (nuclease free water) as a negative control. Lanes 3-13, Clinical Mycobacterium tuberculosis DNA samples. Lanes 14-15 H37Rv DNA as a positive control. A deletion ~785bp observed in sample in lane 3 with a strong band. Samples in lane 5,8,10 and 11 were re-run to confirm that there was no deletion.

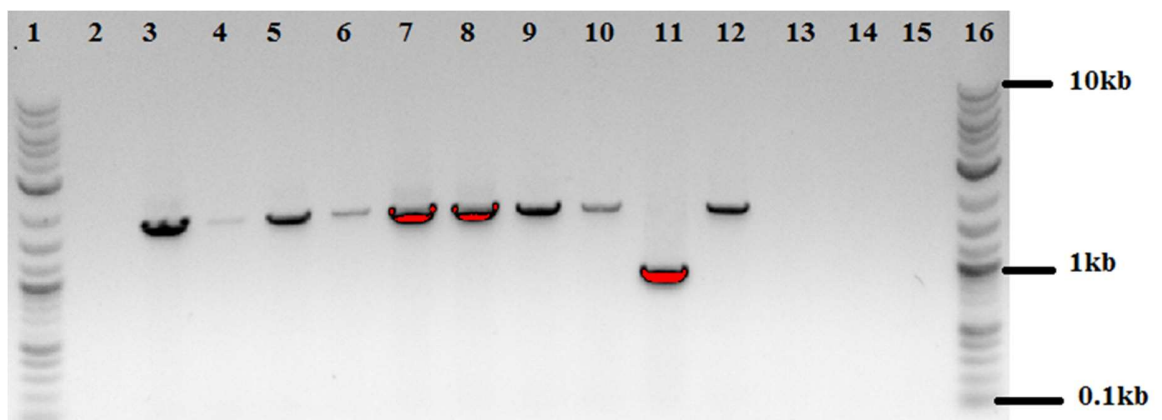


Figure 5. 2 Gel electrophoresis of RD-239 deletion PCR

Lanes 1 and 16, 2-Log DNA ladder (New England Biolabs). Lanes 2 and 3 H37Rv DNA as a positive control. Lanes 4-13 clinical Mycobacterium tuberculosis DNA samples. Lanes 14 and 15 blank (nuclease free water). A deletion ~888bp was observed in lane 11 with a strong band. Lanes 2,4, 10 and 13 were re-run to confirm there was no deletion as the bands were either faint or missing.

Diversity of *Mycobacterium tuberculosis* lineages in Blantyre, Malawi

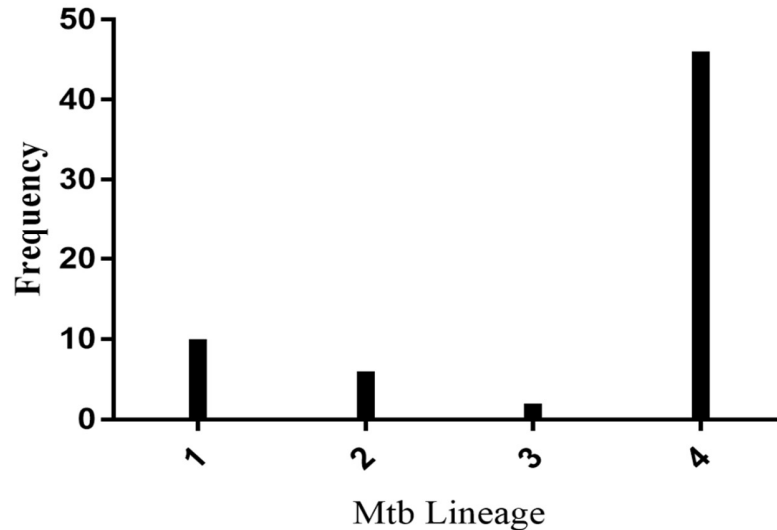


Figure 5. 3 Diversity of *Mycobacterium tuberculosis* in Blantyre, Malawi

Proportions of *Mtb* lineages occurring in Blantyre. 10 (16%) lineage 1 (Indo-Oceanic), 6 (9%) lineage 2 (*Beijing*), 2 (3%) lineage 3 (East African/Indian) and 46 (72%) lineage 4 (Euro-American).

5. 2. 2 Whole genome sequence analysis

In this study, 18/64 *Mtb* clinical isolate genomes were completed using SMRT sequencing and it was therefore possible to compare and confirm the results of genomic deletion analysis in this subset of the strains. The average sequencing coverage was \sim x174 (range x35-460) with a mean read length of \sim 8kb (range 3.5-12kb) and a N50 read length of \sim 12kb (range 5-24kb). In comparison with the *Mtb* H37Rv reference genome (NC_000962) (Cole, S. T., Brosch et al. 1998) the 18 genomes showed >99.8% identity at the nucleotide level. In the phylogenetic analysis a total of 14353 SNPs were found in

the non-repetitive regions with one sample having 6909 SNPs not present in any other sample. These SNPs were used to construct a genome-wide maximum likelihood tree (Figure 5.3). The tree comprised of three of the four *Mtb* lineages known to be present in Malawi (L1, L2 and L4). Three of the 18 strains belonged to lineage 1 (Indo-oceanic), 3/18 to lineage 2 (East-Asian-*Beijing*) and 12/18 belonged to lineage 4 (Euro-American). No lineage 3 strain was found among the 18 sequenced strains. All lineage 4 *Mtb* strains have been reported to possess a 7 bp deletion in the *pks15/1* gene (Gagneux, DeRiemer et al. 2006). Screening for this deletion in the 18 strains confirmed that indeed 12 of the 18 strains belonged to lineage 4 (Figure 5. 4). Additionally, membership of lineages 1 and 2 was confirmed by the presence of lineage specific mutations causing loss of methyltransferase activity (C758T and A809C respectively) as previously reported (Phelan, de Sessions et al. 2018). This analysis has been comprehensively described in section 6. 2. 8

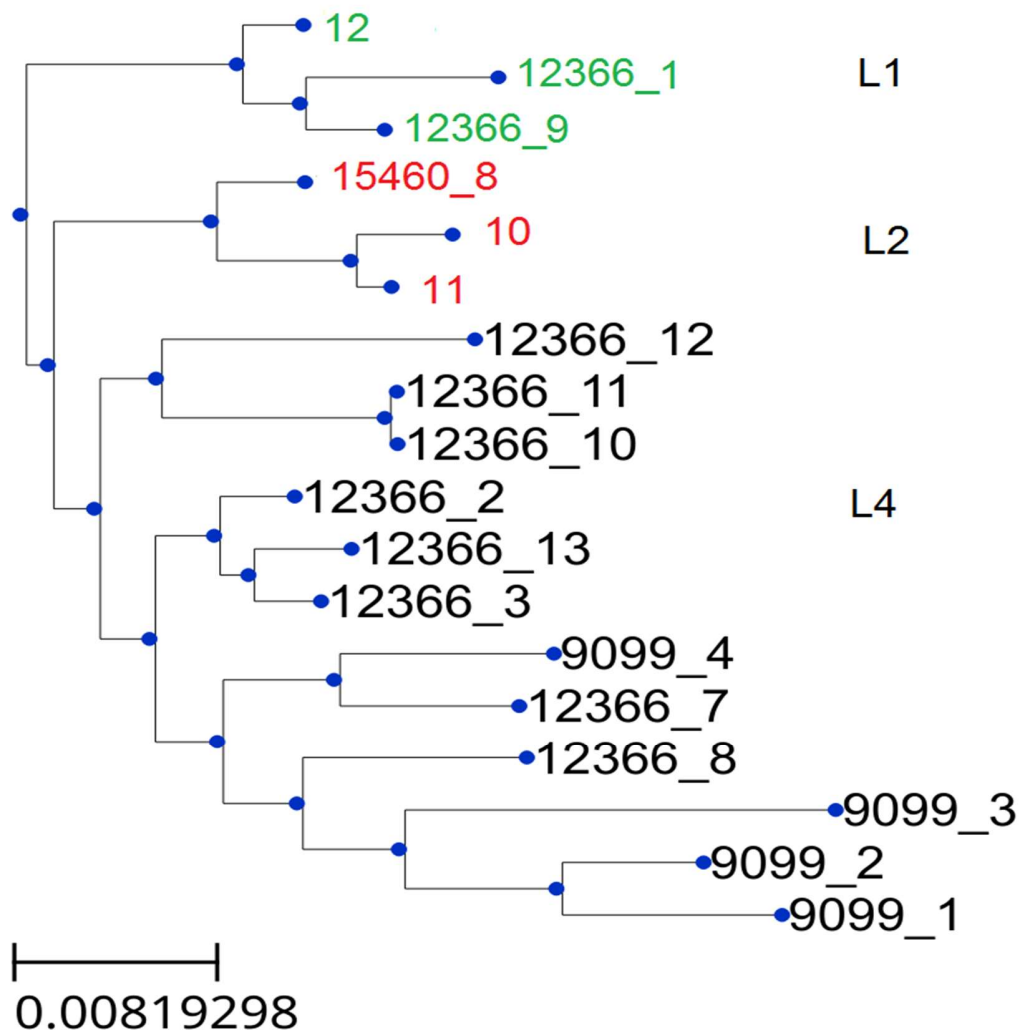


Figure 5. 4 Phylogenetic analysis of the 18 sequenced *Mycobacterium tuberculosis* strains from Blantyre Malawi. Strains clustered into 3 lineages (L1,2 and 4). Green indicates that strains belong to lineage 1 (Indo-Oceanic). Red indicates that strains belong to lineage 2 (East Asian-*Beijing*). Black indicates strains belong to lineage 4 (Euro-American). *M.canettii* was used as an outlier and the branch support values determined by 1000 bootstrap replicates. There are large deletions in all samples except L1 samples 12366_1, 12366_9 and 12 (18501-13222 and 18605-18877)

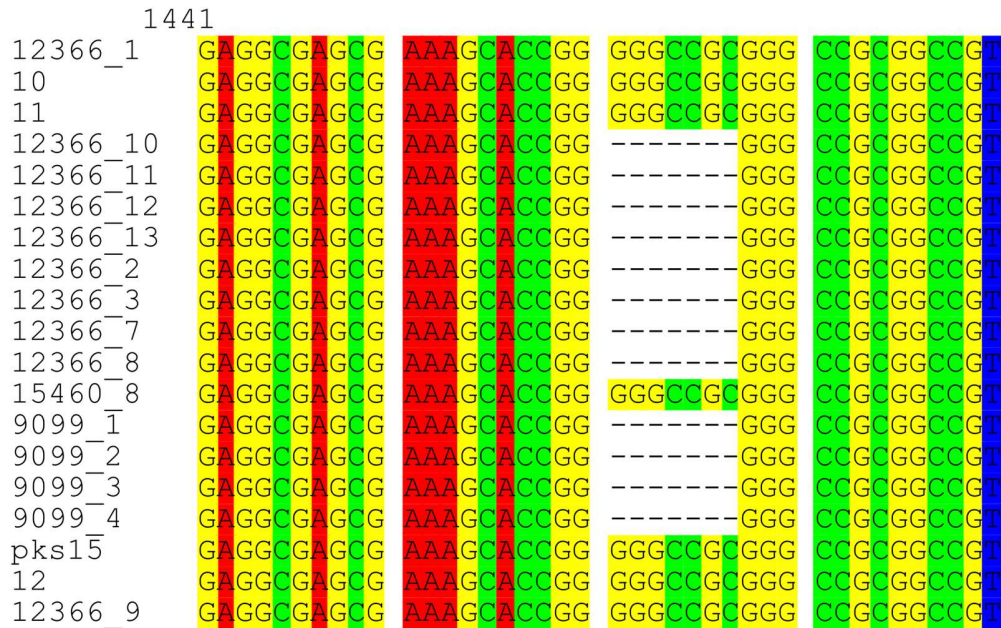


Figure 5. 5 Presence of a 7 bp deletion in the polyketide synthase (*pks15/1*) gene
Six out of 18 samples 12366_1, 10,11,15460_8,12 and 12366_9 have an intact *pks15/1* gene responsible for production of phenol glycolipids. Twelve out of 18 samples have a 7 base pair deletion at position 1461-1467 in the *pks15/1* gene characteristic of lineage 4 (Euro-American lineage). Sample that has been identified as pks15 in the diagram was the reference obtained from an Indo-Oceanic lineage sample known to have an intact *pks15/1* gene.

5. 3 Discussion

In this chapter 64/133 clinical isolates from patients presenting at a referral facility in Blantyre, Malawi were genotyped using genomic deletion analysis and 18 of these compared with classification according to whole genome sequencing. Four of the seven

known *Mtb* lineages were detected as circulating in Blantyre. The Euro-American lineage (lineage 4) is the most predominant of the *Mtb* lineages in Europe and America but has specific sub-lineages dominating in Africa (Rindi, Medici et al. 2014, Gagneux, DeRiemer et al. 2006). Results in this study and those done previously in Malawi (Glynn, JR, Alghamdi et al. 2010, Geurra-Assuncao, Crampin et al. 2015) demonstrate complete concordance with those done in this region of Africa (Stucki, Malla et al. 2012). In the present study the lineage was identified as a default based on the absence of deletions RD 105, 239 and 750 representing the East-Asian, Indo-Oceanic and East African/Indian lineages respectively which were previously identified in Malawi. The fact that no deletion analysis was done to confirm the number of isolates belonging to lineage 4 does not exclude the possibility that a novel lineage not previously reported in Malawi could be present. The probability of such a lineage being detected however appears low as Malawi is not presently a high burden TB country. In addition membership of lineage 4 can be confirmed by the presence of a short 7bp deletion (GCCGCGG) in the polyketide synthase (*pks 15/1*) gene and a CTG to CGG substitution in *katG* gene at position/codon 463 (Gagneux, DeRiemer et al. 2006, Rindi, Medici et al. 2014, Marmiesse, Brodin et al. 2004, Constant, Perez et al. 2002). The deletion in the *pks15/1* gene results in a frameshift mutation in the gene inactivating production of phenol glycolipids in a large group of *Mycobacterium tuberculosis* strains (Constant, Perez et al. 2002). Bioinformatic analysis of the 18 sequenced isolates in this study revealed the presence of this deletion in only 12 of the 18 sequences further confirming the findings from the LSP assay. However, results from the study being reported here are largely in agreement with previous studies done in Karonga (Glynn, JR, Alghamdi et al. 2010, Glynn, J. R., Crampin et al. 2005). In one

particular study from this region, out of 781 patients with a first episode of TB, 76% were Euro-American lineage (Glynn, JR, Alghamdi et al. 2010) while in another the lineage was found to occur at 68% (Geurra-Assuncao, Crampin et al. 2015). A proportion of lineage 4 isolates could not be definitively confirmed using *pks15/1* deletion as these were not sequenced but they are very unlikely to be any other lineage. The Euro-American lineage appears to be the predominant lineage in Malawi based on data from these few studies although more studies are necessary to increase the power of these data.

Lineage 1 (Indo-Oceanic) was found to be present in 16% of patients in this study, consistent with reports from Karonga (Geurra-Assuncao, Crampin et al. 2015). Similarly, phylogenetic analysis (SNPs and indels) confirmed that three of the isolates that were identified as belonging to lineage 1 (Indo-Oceanic) using LSP-PCR were indeed of this lineage. These isolates also possessed less deleted sequences characteristic of ancient *Mtb* strains compared to the other sequences as previously documented (Brosch, Gordon et al. 2002). Again, membership of this lineage could further be confirmed by loss of a methyltransferase *mamB* attributed to a missense mutation in the Rv2024c gene (Zhu, Zhong et al. 2015) as discussed in section 6.2.8. Although lineage 1 has global prevalence, it is more common in East Africa and the Indian sub-continent (Duarte, Nery et al. 2017). Karonga is easily accessible from the East African country of Tanzania which has a higher prevalence of this lineage and so the expectation might have been that a higher prevalence of this lineage would have been observed than in Blantyre. However, the relatively small number of isolates available means that significant uncertainty about the prevalence and transmission of lineage 1 within Malawi remains.

The *Beijing* strain (East Asian lineage) has been reported to be widespread globally and may be associated with increasing drug resistance in some regions (Stavrum, Mphahlele et al. 2009). The RD 105 LSP is the universally accepted marker of the *Beijing* lineage although this deletion has also been reported in some isolates with non-*Beijing* spoligoprofiles (Flores, Van et al. 2007). This lineage was previously specifically associated with an intact *pks15/1* sequence and production of phenolic glycolipid, which has been proposed as an important virulence factor, until an intact sequence was reported in non-*Beijing Mtb* clinical isolates in Thailand (Alonso, Borrell et al. 2008). A later study suggested that the intact *pks15/1* sequence could be a marker of isolates originating from Asia rather than a general marker of the *Beijing* lineage (RW.ERROR - Unable to find reference:639). The study being reported here did not use the *pks15/1* marker to specifically discriminate isolates belonging to the *Beijing* lineage. Rather, RD105 which is believed to be a more robust marker of the *Beijing* lineage was used. In WGS analysis, the phylogenetic tree constructed from SNPs and indels confirmed that 3/18 isolates belonged to lineage 2. Additionally, methylation analysis revealed that these three samples lacked the *mamA* methyltransferase owing to a point mutation in the Rv3263c gene (see section 6.2.8). This mutation is known to be present only in lineage 2 clinical isolates (Shell, Prestwich et al. 2013).

In Malawi, the *Beijing* genotype has previously been reported with proportions of patients at ~2% and no association with drug resistance has been reported (Glynn, JR, Alghamdi et al. 2010). All the *Beijing* strains from the current study were drug sensitive which is consistent with the study from Karonga (Geurra-Assuncao, Crampin et al. 2015). While the *Beijing* strain has been associated with increasing drug resistance in South Africa

(Cowley, Govender et al. 2008), such an association is yet to be established in Malawi, possibly because the overall prevalence of drug resistance is low by regional standards and lineage 2 remains relatively rare to date. The *Beijing* lineage was also found to be more prevalent among HIV positive individuals in South Africa and Mozambique (Middelkoop, Bekker et al. 2009, Viegas, Machado et al. 2013). The results of the published data in Malawi have differed in that no association with HIV could be established (Geurra-Assuncao, Crampin et al. 2015). The results presented in the current study may appear to suggest a higher prevalence of the *Beijing* genotype at 9% in Blantyre compared to 4% in Karonga (Geurra-Assuncao, Crampin et al. 2015) however it should be recognised that the sample size in this study was quite small. A larger sample size would be necessary to definitively confirm these findings. However, the apparently higher prevalence of the *Beijing* lineage in Blantyre compared to Karonga could be attributed to the fact that Blantyre is an urban setting where transmission rates are likely to be higher. Furthermore Blantyre has a higher geographical accessibility to South Africa and Mozambique where the *Beijing* lineage has been reported at a higher prevalence (Middelkoop, Bekker et al. 2009, Viegas, Machado et al. 2013).

A study of strains from histological samples from a 76-year period in South Africa demonstrated that the *Beijing* strain was a recent arrival in the country. The *Beijing* strain was found to be absent in samples from 1930-1965 and more common in samples from 1996-2005 (Cowley, Govender et al. 2008). From these data, it would be reasonable to infer that a similar trend could be occurring in Malawi as there is significant regional migration between Malawi and South Africa. A study from Karonga suggests that the *Beijing* lineage may have only recently arrived in Malawi compared to other

Mycobacterium tuberculosis lineages (Glynn, JR, Alghamdi et al. 2010). The spread of *Beijing* in Malawi could also be attributed to recent influx of people of Eastern-Asian ethnic background involved in road construction and trade. Additionally, there has been an increase of Malawian business people travelling to China over the past few years. A country-wide study on the diversity of *Beijing* strain in Malawi will be needed to fully confirm the current prevalence of this lineage and its drug sensitivity profile and to monitor trends in its representation among cases of tuberculosis in Malawi in the future. (Zhu, Zhong et al. 2015)(Zhu, Zhong et al. 2015)(Zhu, Zhong et al. 2015)(Zhu, Zhong et al. 2015)

Overall the data presented in this chapter reveal that results from lineage assignment using WGS of the 18 sequences were in complete concordance with those obtained from LSP-PCR. It could therefore be reasonably inferred that the rest of the lineage assignment from LSP-PCR was reliable.

Although *Mtb* lineages are associated with a specific geographical location, Africa is the only region where all the lineages are represented consistent with the evolutionary theory that the pathogen *Mtb* originated from Africa (Gagneux, DeRiemer et al. 2006).

Although LSP-PCR analysis appears to be consistent with WGS analysis, the method has a number of limitations and newer methods are needed. The power of WGS in TB transmission surveillance and control has previously been documented (Roetzer, Diel et al. 2013). With increasingly widespread use of WGS sequencing, it is only proper that definitive molecular characterization of *Mtb* make use of WGS in all areas where such tools exist.

CHAPTER SIX: CHARACTERIZATION OF DNA METHYLATION IN CULTURED *MTB* ISOLATES USING SMRT SEQUENCING

6.1 Introduction

The aim of this chapter of the thesis was to investigate whether epigenetic mechanisms specifically, DNA methylation could play a role in *Mtb* persistence. Although *Mtb* strains have been shown to exhibit nucleotide level similarity of >99% (Hershberg, Lipatov et al. 2008) such similarity is rarely replicated in the phenotype. This phenotypic heterogeneity has been seen in the virulence of the *Beijing* strain which has been associated with increasing MDR-TB (van der Spuy, Kremer et al. 2009) whereas the East African Indian (EAI) lineage has been associated with lower rates of transmission compared to other lineages (Albanna, Reed et al. 2011). Similarly, the Euro-American lineage is the most geographically successful strain (Coscolla, Gagneux 2014) but specific reasons for this phenotype remain unknown. Phenotypic heterogeneity in *Mtb* has been associated with epigenetic inheritance (Balaban, Merrin et al. 2004) and the most common epigenetic mechanism in *Mtb* is DNA methylation (Shell, Prestwich et al. 2013). This chapter therefore sought to characterize DNA methylation patterns within the *Mtb* genome. This characterization is important as some studies have demonstrated that DNA methylation plays a critical role in gene expression and long term survival of *Mtb* under hypoxic conditions (Shell, Prestwich et al. 2013). Furthermore, the link between DNA methylation and *Mtb* persistence has not yet been explored (Shell, Prestwich et al. 2013). The persister phenotype is not genetically inherited (Balaban, Gerdes et al. 2013) suggesting that there could be some other as yet unknown mechanisms responsible for this phenotype. The characterization of the *Mtb* methylome in the current study was specifically done on

clinical isolates, this enabled selection of strains and samples most closely related to a clinical definition of persistence and also because if DNA methylation indeed has a role in *Mtb* persistence, this could have important implications for treatment outcomes. Understanding the biology of persisters remains among the most important barriers to shortening of *Mtb* treatment regimen.

DNA methylation in prokaryotes is widespread and is identified by diverse modification types such as 6methyladenine(m6A), N-4methylcytosine(m4C), N-5methylcytosine(m5C) among others. DNA methylation is regulated by methyltransferases which target specific sequence motifs (Sater, Lamelas et al. 2015).

DNA methylation has been associated with pathogenicity and virulence in *Salmonella enterica* (Wion, Casadesus 2006). In *Mtb* it was previously proposed that DNA methylation had a role in virulence as cytosine methylation was found to be present in the virulent strain H37Rv but absent in the avirulent H37Ra (Srivastava, Gopinathan et al. 1981). However later reports seem to be inconsistent with this finding (Naidu 2014) as no compelling evidence was found to associate methylation to virulence in *Mtb*. A recent study has established a possible role of DNA methylation in the long term survival of *Mtb* under anaerobic conditions (Shell, Prestwich et al. 2013) and this link requires further investigation. The advent of SMRT sequencing has made it possible to determine novel DNA modifications in prokaryotes and more importantly to study it globally across the genome and not just in specific genes. In resequencing of six bacteria using SMRT sequencing, new DNA methylation sites were discovered and their cognate methyltransferases assigned (Murray, Clark et al. 2012). Although the study of DNA methylation in *Mtb* using SMRT sequencing is becoming an area of active research (Zhu, Zhong et al.

2015), to date no published study has so far investigated its association with persistence in *Mtb*.

In this chapter, the aim was to characterize the complete DNA methylome of *Mtb* using Pacific Biosciences SMRT sequencing with the specific objectives:

- To identify and characterize the different types of DNA base modification (DNA methylation) in *Mtb*
- To identify the DNA sequence motifs in *Mtb* and their cognate methyltransferases
- To compare DNA methylation patterns in genic and intergenic regions between treatment failures and successes (“persisters” and “non-persisters”).
- To compare DNA methylation patterns between promoter regions of “persisters” and “non-persisters” and stability of methylation under different growth conditions including within different lineages.

6.2 Results

In order to perform a preliminary characterization of *Mtb* methylation patterns four confirmed drug susceptible *Mtb* clinical isolates from a previous study (Sloan, Mwandumba et al. 2015) were selected. Two of the isolates had an outcome of treatment failure and were rich in lipid body content (putative “persisters”), while the other two had treatment success with lipid poor in body content (non-persisters). Phenotypic drug tolerance and accumulation of lipid body content are putative markers of *Mtb* persistence. The aim was to conduct a global comparative analysis of the methylation patterns between treatment failures (“persisters”) and successes (“non-persisters”) and try to establish a possible role of DNA methylation in the persistence phenotype. It was not possible to

resuscitate the targeted number of failures (“persisters”) from frozen samples and therefore comparison of “persisters” and “non-persisters” was only limited to four clinical isolates. Strains were resuscitated from a frozen state in liquid culture media before transferring to solid culture media. All the data and the analysis from this study are available online at <https://github.com/codemeleon/MtbMethylationProfile>.

6. 2. 1 Comparative bioinformatics analysis of “persister” and “non-persister” *Mtb* isolates

Four clinical isolates were sequenced using SMRT sequencing technology at the Centre for Genomic Research (CGR), University of Liverpool. The average sequencing genome coverage for the four isolates was ~ 352X. All genomes were analyzed using the SMRT portal software provided by the manufacturer (version 2.2.0). Through sequence alignment with the reference genome H37Rv and kinetic data analysis, it was possible to accurately detect m6A methylated DNA sequence motifs in the genomes of the four isolates (Table 6. 1). Some weak m4C signals could be detected but with very low IPD ratio and no specific pattern. No m5C could be detected. Some bases could only be reported as “modified” without a specific methylation pattern. The bases reported as only “modified” could not be experimentally verified, but previous studies have confirmed using WGS based on multiple displacement amplification that such modifications are in actual fact false positives (Zhu, Zhong et al. 2015). Additionally, such modified bases had low IPD ratios (average <5). This might indicate that they could have been as a result of sequencing errors (Naidu 2014). All modifications with bases reported only as “modified” or with low IPD ratios (<5) were therefore excluded from further downstream analysis. Two m6A motifs (CTCCAG and CACGCAG) 5’ to 3’ direction; (a base with methylation

was indicated by A) were detected using SMRT portal analysis in all the four isolates (Table 6. 1). The motif CTGGAG is a partner motif for CTCCAG. The modifications in these motifs had apparently higher IPD ratios (average >5) compared to those reported only as “modified” at <5. Additionally, they had high motif site ratios of ~0.80-0.98 (Table 6. 1). There were some exceptions to this and have are discussed below (Section 6.2.2).

Table 6. 1 Motifs and Modifications within “persisters” and “non-persisters”

Motif	Modification	Motif site ratio*		Mean IPD ratio
		“persisters”	“non-persisters”	
CTCC <u>A</u> G	6mA	0.97	0.98	6.5
C <u>A</u> CGC <u>A</u> G	6mA	0.97	0.98	6.2
CTCC <u>A</u> GB	6mA	0	0.69	2.9
CTCC <u>A</u> GA	6mA	0	0.31	2.1
AGGCMGYA	Modified	0.43	0.38	1.9
GARVDTKGV	Modified	0.19	0.9	1.9
TNNNNNNH	Modified	0.4	0.5	1.7
TVNVNNG	Modified	0.2	0.2	1.7

*Motif site ratios were computed as the average of two isolates

6.2.2 Occurrence of methylated motifs within persister and non-persister isolates

The occurrence of methylated motifs within “persisters” and “non-persisters” was first investigated. The two “persister” isolates (9909_1 and 9909_2) were sequenced at a mean motif coverage of ~145X and 225X respectively. Two predicted m6A motifs were detected at a similar proportion of ~97% and an IPD ratio of ~6-7 (CTCCAG, and CACGCAG) 5’ to 3’ direction. Motif CTGGAG (5’ to 3’ direction) a palindrome of CTCCAG, was detected at an average proportion of 95% and a lower IPD ratio of ~5 only (Table 6.1 and 6.2). This suggests that some sites with this motif remained unmethylated. A search for genes homologous to known methyltransferases in the database of prokaryotic enzymes REBASE (Roberts, Vincze et al. 2015) was conducted. This search led to identification of the gene Rv3263 which encodes the enzyme *Mycobacterium Adenine Methyltransferase A (MamA)* termed M. MtuHIII (according to standard nomenclature) which target the motif CTCCAG. This sequence motif was also detected by Shell in the *Mtb* reference strain H37Rv of the Euro-American lineage (Shell, Prestwich et al. 2013) and *MamA* was implicated. An asymmetric m6A motif CACGCAG was also detected in all of the four isolates and predicted to occur 820 times in the genome. It was detected as being methylated >97% across the four isolates (Table 6.2). CACGCAG is a non-palindromic methylated Type II RM system motif (Naidu 2014). A search for this motif in the REBASE database for H37Rv yielded nothing. However, the motif was found in the REBASE database for *Beijing* strain with its target methyltransferase termed MtuBEORF5175P according to standard methyltransferase nomenclature (Roberts, Vincze et al. 2015). The two “non-persister” (9909_3 and 9909_4) isolates were sequenced at mean motif coverages of ~179X and 145X respectively. While one of them

(9909_4) had similar proportion of methylated motifs and IPD ratios to the “persisters” the other (9909_3) was different in that it had an additional unique motif. Notwithstanding the fact that the average motif sequencing coverage was high (~ 179X), a rare motif CTCCAGB was detected, predicted to occur 1738 times in the genome and detected to be methylated only 1194 (69%) times in 9909_3

. This motif could not be detected in the other three isolates. Furthermore, the IPD ratio for this motif was a low (2.9) signifying a weak signal. This motif is similar to a type II H37Rv Restriction Modification motif CTCCAG but with just an extra seventh base at the end. Another motif similar to CTCCAGB but having the seventh base B replaced by an A (CTCCAGA) was detected in this isolate. It was predicted to occur in 1947 sites within the genome and was detected as methylated only 209 (11%) times in the genome of the isolate. This isolate was shown to be methylated at 64 sites (30%) compared to ~98% methylation in the other three isolates for motif CTCCAG. Additionally, this motif (CTCCAGA) had a low IPD ratio of ~2, compared to >6 in the other three isolates.

Table 6. 2 General genome methylation for 4 *Mtb* clinical isolates

Sample ID	Sequence Motif			
	CTGGAG		CACGCAG	
	No. of motifs ^a	% modified motif	No. of motifs	% modified motif
9909_1	1947	95	820	97
9909_2	1947	97	820	98
9909_3	1947	58	820	97
9909_4	1947	98	820	98

^a The number of motifs only represents a single strand

Across all the four isolates, an unknown motif (AGGCMGY/AKGCMSY) predicted to occur a variable number of times (average 0.38-0.42) was detected (Table 6.1). No further analysis for this motif was conducted due to a very low IPD ratio of ~2 and unspecified methylation pattern. Another unknown motif ADDTRGC was detected in one “persister” (9909_2) and one “non-persister” (9909_3). Again due to a low IPD ratio and non-specific methylation pattern, no further analysis was done.

Since the aim was to conduct a comparative analysis of the methylation patterns between “persisters” and “non-persisters”, the first to be investigated was methylation occurrence by gene strand. Some motifs occurred as methylated palindromes (CTGGAG) on both strands while others were found only on a single strand (CACGCAG). Occurrence of methylation across these strands was variable. The most highly methylated gene was found to be Rv2524c with 21 sites methylated with a large number of genes having zero

methylation. According to Table 6.3, no significant differences could be detected between “persisters” and “non-persisters” in total adenine methylation within the genic (coding) regions by strand (χ^2 test with Yates correction, $p = 0.20$). Strand specific adenine methylation ranged between 49%-51% of the genic methylation in both “persisters” and “non-persisters”. Genic methylation comprised an average ~80% of the total genome methylation in both “persisters” and “non-persisters”. Evidently there is sharing of methylation sites between “persisters” and “non-persisters” and only a few sites are uniquely methylated.

Table 6. 3 Methylation within genic regions in “persisters” and “non persisters” by strand

Sample ID.	Positive strand No. of sites (%)	Negative strand No. of sites (%)	%Total
9909_1 (“Persister”)	1533 (51)	1487 (49)	81%
9909_2 (“Persister”)	1711 (50)	1692 (50)	80%
9909_3(“Non-”)	1262 (49)	1315 (50)	79%
9909_4(“Non-”)	1566 (50)	1572 (50)	81%

Next was the analysis of methylation patterns within the intergenic regions of the four isolates. Intergenic regions were defined as any non-coding region between two gene regions. An analysis of the intergenic methylation revealed that 19%-21% of the whole genome methylation occurred within the intergenic regions across the four isolates (Table 6.4). Strand specific methylation showed that the positive strand was slightly more methylated at 51% compared to the negative strand at 49% across all the four isolates although this difference was not statistically significant across the isolates (Fisher’s exact

p=0.90). This difference could be as a result of some sites of the strand remaining hemimethylated following cell division.

Table 6. 4 Methylation within intergenic regions in persisters and non persisters by strand

(Fisher's exact p = 0.90)

Sample ID.	Positive strand	Negative strand	Total Intergenic Methylation	%Intergenic Methylation
	No.of sites, (%)	No. of sites, (%)		
9909_1("Persister")	350 (51)	347 (49)	697	19%
9909_2("Persister")	418 (51)	416 (49)	834	20%
9909_3 ("Non-Persister")	356 (51)	346 (49)	702	21%
9909_4 ("Non-Persister")	372 (51)	371 (49)	743	19%

Promoter regions are normally regarded as regions in the genome 10bp to 35bp upstream of the transcription start site (TSS)(Newton-Foot, Gey van Pittius 2013). It is in the promoter sequence where the RNA polymerase enzyme interacts during transcription thereby regulating the transcription process. Promoters are variable in length and in this

study promoter regions were defined as the region 100bp upstream of the TSS. *Mtb* genes occur in clusters called operons and each operon is under the control of promoter of the head gene (Lim, Lee et al. 2011). In this study promoters were considered as part of the intergenic sites. Methylation within promoter regions affects gene expression and eventually phenotype (Wion, Casadesus 2006). Adenine methylation sites have been shown to overlap with promoters of some genes thereby regulating expression of such genes (Shell, Prestwich et al. 2013). Shell et al. found that *mamA* sites in promoters overlapped with the sigma factor -10 binding sites regulating the expression of the involved genes. Additionally, *mamA* deletion mutants did not display transcription changes suggesting a significant role of DNA methylation.

The effect of gene promoter methylation specifically on *Mtb* persistence has not been previously characterized using PacBio SMRT sequencing technology or any other technology. The main interest was to investigate if methylation between “persisters” and “non-persisters” occurred differentially within promoter regions and if there was any differential occurrence by strand. Methylation of promoter regions was analyzed in proportion to overall methylation within intergenic regions (Table 6.5). The results reveal that promoter methylation in all the four isolates ranged between 39% to 41% of the total intergenic methylation and there was no differential methylation by persistence (Fishers exact, $p=0.8$) (Table 6.5). Strand specific promoter methylation ranged between 46% and 54% within the four isolates. This methylation again was highly variable between the strands and there was no overrepresentation by strand (Fishers exact, $p=0.5$) for “persisters” and (Fishers exact, $p=0.9$) for “non-persisters” respectively.

Table 6. 5 Methylation within promoter regions by strand for persisters and non-persisters compared to the rest of intergenic region.

(Fishers exact, p= 0.8)

Sample ID.	Positive	Negative	Total	% Promoter
	strand	strand		
	No. of sites (%)	No.of sites (%)	Promoter	Methylation
9909_1				
("Persister")	144/276 (52)	132/276 (48)	276/690	40%
9909_2				
("Persister")	158/323 (49)	165/323 (51)	323/828	39%
9909_3				
("Non-Persister")	159/273 (51)	134/273 (49)	273/700	39%
9909_4				
("Non-Persister")	164/306 (54)	142/306 (46)	306/746	41%

The assumption is that methylation occurring in intergenic region has restriction modification function rather than a regulatory function. Although there was no significant difference in overall promoter methylation between “persisters” and “non-persisters”, it was observed that the exact location of this methylation varied between the two classes of isolates. An average 67 (21%) out of 323 m6A sites within promoter regions were uniquely methylated to “persisters” while an average of 50 (16%) out of 306 m6A sites were uniquely methylated to “non-persisters”. Genes whose operon sites were uniquely

methyated in both “persisters” and “non-persisters” were investigated further. The study focuses on the persister induced genes.

6. 2. 3 “Persister” gene and promoter methylation

An analysis of SNPs and indels within persister and non persister isolates revealed 6700 variants between the genomes of the two groups. A total of 697 only were SNPS suggesting a large number of variants were deletions. Multiple sequence analysis of the selected 15 genes previously associated with persistence (Keren, Minami et al. 2011, Sala, Bordes et al. 2014) against the reference genome H37Rv revealed perfect match. No SNPs and indels were identified in either persisters or non-persisters.

Several genes have been implicated in the entry and maintenance of persister phenotype within the *Mtb* cells (Georgiades, Raoult 2011, Sala, Bordes et al. 2014). These include toxin-antitoxin modules whose transcriptome was shown to be induced under persistence (Keren, Minami et al. 2011) and members of the *dosR* regulon which have shown upregulation during the dormant infection model (Voskuil, Schnappinger et al. 2003). It has been previously reported that methylation of some motifs within the -10 Sigma factor binding site promoter region led to differential expression of the genes concerned in response to hypoxia (Shell, Prestwich et al. 2013). In this study, methylation was associated with long term survival of *Mtb* under hypoxia. Interestingly hypoxia has also been associated with activation of persistence genes including the *dosR* regulon (Voskuil, Schnappinger et al. 2003). Genes and pathways that have previously been implicated in persistence were of interest to this study. The aim was to investigate if promoter regions of such genes could be differentially methylated in the “persister” isolates in comparison to “non-persisters”. The analysis revealed that in general, promoter regions of “persister”

genes had low methylation (39-41%) in comparison to the rest of intergenic region (Table 6.5). Could persistence be due to methylation within promoters of key genes in “persister” cells? To answer this question, it was necessary to pinpoint the specific genes whose operons had the promoter region methylated.

6. 2. 4 Specific location of differentially methylated gene promoters in persisters

Although it was found that the four isolates were proportionally equally methylated, further analysis revealed that this methylation was differentially distributed along the genomes of “persisters” and “non-persisters”. Sixty seven out of 323 promoter sites were found to be differentially methylated in “persisters” as opposed to “non-persisters” and it was important to further investigate the specific genes under the regulation of these promoters. The analysis revealed that all *Mtb* genes whose promoters were differentially methylated in “persisters” could be clustered into: cell wall and cell wall processes (33%), virulence detoxification and adaptation (22%), intermediary metabolism and respiration (25%), transcription and regulatory (1%), information pathways (9%), and conserved hypotheticals and unknown (4%) (Figure 6. 1). A gene ontology (GO) pathway enrichment analysis of 67 differentially methylated genes in persisters revealed that none of the categories were over-represented (p-value <0.05 Fishers exact with FDR multiple test correction). Previously an analysis of the *Mtb* persister transcriptome had shown a downshift in metabolic and biosynthetic pathway, characteristic of dormancy (Keren, Minami et al. 2011). In that study, *Mtb* cells in log phase were exposed to D-cycloserine and a transcriptome of the surviving cells obtained. There was a marked downshift in metabolic and biosynthetic pathways accompanied by induction of a characteristic set of genes. Could this massive downshift be as a result of promoter methylation in genes

regulating these pathways? It was important to investigate if methylation in these genes could be correlated to the characteristic metabolic downshift observed in the *Mtb* transcriptome under conditions of persistence. Analysis of promoters revealed that both the *dosR* genes and the TA modules were largely non-methylated. Out of all the genes induced in persistence (Keren, Minami et al. 2011, Sala, Bordes et al. 2014), *hspX* (the heat shock protein) was the only gene found to be methylated at a single site in each of all the four isolates while its homologue *acr2* was methylated in only one non-persister isolate. Other genes with a single methylation were *gntR* (Rv1152) and Rv3290c possessing a site each in 9909_2 and 9909_3 respectively (Table 6. 6). Low promoter methylation could be indicative of unperturbed gene expression. In absence of corresponding transcription data, it would be impossible to determine if any of the methylation patterns in the promoters may have affected gene expression in either persisters or non-persisters. Naidu found the promoter for operon of *mymA* (Rv3083-Rv3089) to be differentially methylated in the hyper-virulent strain of *Beijing* strain and suggested it could play a key role in adaptation to acidic stress and persistence within macrophages (Naidu 2014). This operon was found to be unmethylated in all the four strains in this study.

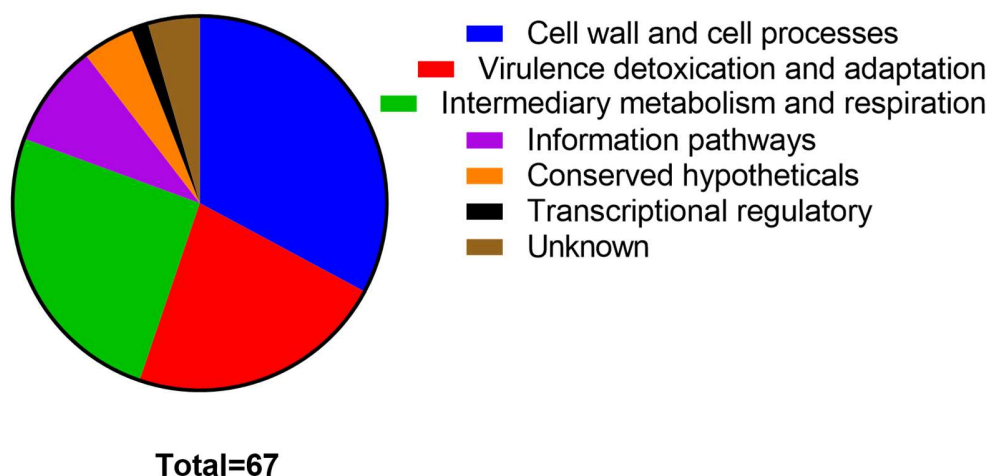


Figure 6. 1 Pathways differentially methylated in persisters

Cell wall and cell processes 22(33%),virulence detoxication and adaptation 15(22%),intermediary metabolism 17(25%),information pathways 6(9%) conserved hypotheticals 3(4%), transcriptional regulatory 1(1%) and unknown 3(4%)

Table 6. 6 Methylation within promoter regions of selected persister related genes

Genes	Persisters		Non-Persisters	
	9909_1	9909_2	9909_3	9909_4
RelBE1	0	0	0	0
RelBE2	0	0	0	0
MazEF1	0	0	0	0
MazEF5	0	0	0	0
MazEF6	0	0	0	0
hspX	1	1	1	1
dosR	0	0	0	0
dosS	0	0	0	0
acr	0	0	0	0
gntr	0	0	1	0
Rv3290c	0	0	0	0
pkdA	0	0	0	0
Rv2517	0	0	0	0
Rv3290c	0	1	0	0
RelEB3	0	0	0	0

6. 2. 5 Differential methylation within persister related genes

Analysis of sites in gene coding regions that have previously shown to be induced during *Mtb* persistence (Sala, Bordes et al. 2014, Keren, Minami et al. 2011) revealed that a few genes were differentially methylated in persisters as opposed to non-persisters (Table 6.7). A total of 180 different sites were found to be methylated within the 15 genes under analysis. One hundred and twelve sites were shared between persisters and non-persisters. *RelBE1* (*Rv1246c-1247c*), a TA module had a single site differentially methylated in non-persisters while two other different sites were common between persisters and non-persisters. In *RelBE2* (*Rv2865-2866*) another TA module, no single site was found to be differentially methylated in both persisters and non-persisters. *Rv2517*, whose product is a hypothetical protein with unknown function (DeJesus, Gerrick et al. 2017) had only one methylated site shared between persisters and non-persisters. *Rv3134c*, which encodes a universal stress protein, *acr2* which encodes a heat shock protein and *dosR* which encodes a two component regulatory protein (DeJesus, Gerrick et al. 2017) were each found to have a single site differentially methylated in persisters. *Rv3290c* which encodes a L-lysine-epsilon aminotransferase possessed two sites which were methylated in persisters. Other persister induced genes including TA modules *RelBE3*, *MazF1*, *MazF5*, *MazF6* and *gntR* were non-methylated in both persisters and non-persisters.

Table 6. 7 Sites differentially methylated within selected persister related genes

Gene	Sites methylated in persisters only	Sites methylated in non-persisters only	Sites shared
RelBE1	0	1	2
RelBE2	0	0	0
MazF1	0	0	1
MazF5	0	0	4
MazF6	0	0	0
hspX	0	0	0
dosR	1	1	0
dosS	0	1	2
Rv3134c	1	0	0
Acr2	1	0	1
gntR	0	0	0
bkdA	1	0	2
Rv2517	0	0	1
Rv3290c	2	0	0
RelBE3	0	0	0
Total	38	30	112

6. 2. 6 Stability of methylation patterns in *Mycobacterium tuberculosis* genome under different growth conditions

Since the gene expression profiles of *in vitro* *Mtb* was shown to be different from that found in sputum (Garcia, Loxton et al. 2016), it was hypothesized that the methylome of the cultured *Mtb* clinical isolates could as well be different to that of similar isolates during infection. It has been suggested that culturing *Mtb* cells may introduce mutations in the genome (Bjørn-Mortensen, Zallet et al. 2015) consequently impacting on the methylation patterns. In order to know for sure, an attempt was made to sequence directly from sputum and determine if the methylome profile of sputum *Mtb* would be different to that in culture. A PrimeXtract DNA extraction kit was used for this purpose as it has been previously claimed to achieve sufficient amounts of DNA for WGS (Longhorn Vaccines 2017). Sufficient DNA for SMRT sequencing was obtained but expert advice compelled the study not to proceed with sequencing due to possibility of high host DNA contamination (data not shown). As an alternative, a decision was made to interrogate if the methylation patterns observed within *Mtb* would remain stable under different growth conditions. To do this, in addition to isolates 9909_1- 9909_4, a further 14 cultured clinical isolates were SMRT sequenced (Table 6.9). These were again selected based on the one that grew first. The average genome sequencing coverage for the 18 isolates being ~173X. Two of these (12366_10,12366_11) were cultured in a MGIT tube for up to six weeks while the rest were cultured on LJ slopes and harvested after eight weeks. Lineage assignment revealed that 12 samples belonged to lineage 4 while 3 each belonged to lineage 1 and 2 respectively. The aim was to compare DNA methylation patterns between MGIT cultured isolates and those cultured on LJ slopes. It was not possible to efficiently sequence from

the rest of the MGIT cultured isolates as the DNA failed to pass the sequencing quality control stage and therefore only two were selected for analysis. Since it had been established in section 6.2.2 that methylation patterns within *Mtb* strains were largely similar, four isolates were selected for comparison, two MGIT cultured (12366_10,12366_11) and two LJ cultured (12366_1,12366_13). The selection was based on which one grew first and had sufficient DNA yield for SMRT sequencing.

6.2.7 Comparison of MGIT and LJ cultured isolates

Since the persister phenotype has been reported to be in part a result of the bacterial stress response (Balaban, Gerdes et al. 2013) and if methylation has a role in persistence then the hypothesis would be that methylation itself would be an environment dependent phenomenon. *Mtb* strains grown under different growth conditions would then have variable methylation patterns. This part of the study was aimed at establishing whether growth conditions would affect the *Mtb* genome methylation profile. It was important to compare methylation patterns of liquid culture isolates and those from solid culture as these remain the commonly used culture methods in *Mtb*. Since only two MGIT grown isolates could be sequenced, it was decided that these be compared with two selected LJ samples (12366_10, 12366_11 against 12366_1, 12366_13). Selection was based on those that would grow first and had sufficient yield for SMRT sequencing. Three of the four isolates were lineage 4 (12366_10,12366_11 and 12366_13) while one (12366_1) belonged to lineage 1 having been confirmed using LSP PCR (Chapter 5).

The two methylated motifs CTGGAG, and CACGCAG were detected in all the four isolates under comparison. Although GATN₄RTAC was detected in only 12366_1, it was excluded from the comparison as it was absent in some other LJ cultured isolates

signifying the difference was not growth condition associated. Furthermore this motif was also found to be present in some other non-LJ cultured isolates. Motif CTGGAG was methylated in an average of 95% in solid cultured isolates compared to an average 94% in liquid cultured isolates. No significant difference could be established between liquid and solid culture isolates for methylated motif CTGGAG (Fishers exact test $p=0.76$). As for motif CACGCAG solid cultured isolates were methylated at an average 76% while liquid cultures were methylated an average 97%. It was found that liquid cultured isolates were significantly more methylated than solid cultures (Fisher's exact $p=0.02$) for motif CACGCAG. It was observed that this difference was as a result of sample 12366_1 being lowly methylated at 56% compared to the rest at >95% (Table 6.9). This has been comprehensively discussed in section 6.2.6.

Methylation within the gene regions and promoter regions of genes was next investigated. Methylation within gene regions ranged between 49% to 51% in each strand and there was no over representation of methylation by strand (Chi squared test with Yates correction $P=0.44$). On the other hand, methylation within promoter region of genes ranged from 37% to 62% by strand of the promoter methylation. Again there were no significant statistical differences observed by strand (Fishers exact test $P=0.19$) (Table 6.8)

Table 6. 8 Comparison of methylation patterns between MGIT and LJ cultured isolates

Sample	Sites with genic methylation (%)		Sites with promoter methylation	
	Positive Strand	Negative strand	Positive Strand	Negative strand
12366_1 (LJ)	1632/3210(51)	1578/3210(49)	188/302(62)	114/302(37)
12366_13 (LJ)	1223/2438(50)	1215/2438(50)	120/221(54)	101/221(46)
12366_10	1218/2454(50)	1236/2454(50)	105/200(53)	95/200(47)
(MGIT) 12366_11	1358/2721(50)	1363/2721(50)	138/245(56)	107/245(44)

6. 2. 8 Lineage specific methylation patterns across 18 *Mtb* strains

It has been previously reported that DNA methylation in *Mtb* occurs in a strain specific manner (Phelan, de Sessions et al. 2018, Zhu, Zhong et al. 2015, Shell, Prestwich et al. 2013) although the power of such studies remains suboptimal. In the current study, it was decided that this finding be investigated further. A sum total of 18 *Mtb* clinical isolates were therefore sequenced for this purpose including the four from section 6.2.1.

In addition to the two motifs detected in the first four isolates sequenced (CTGGAG, and CACGCAG) 5' to 3' direction, detection was made of a characteristic type I palindromic motif GATN₄RTAC (Table 6.9). Similar to the first four isolates sequenced, a few other motifs could be detected in these isolates. Again these either had a non-specific methylation pattern or very low IPD ratios <5 signifying they could have been as a result of sequencing errors. They were excluded from further downstream analysis.

Motif CTGGAG was detected as methylated in the 15/18 isolates in various proportions ranging from 34-98% although methylation was mostly close to 100%. The target *Mtase* for this motif, *MamA* has been predicted to be active in all *Mtb* complex strains (Zhu, Zhong et al. 2015). Motif CACGCAG was detected as methylated in 16/18 isolates although in one isolate it was methylated at 56% with most of them being close to 100%. Methylation for this motif ranged from 56-98% across all the genomes. Motif GATN4RTAC was methylated in only 6/18 isolates with a range of 49-94% with methylation mostly below 90%. There was a strong association between presence or absence of methylation and strain lineage (Table 6.9). Strain lineage of each of the isolates had previously been assigned using RD-PCR and confirmed using phylogenetic analysis of SMRT sequenced genomes of the 18 isolates.

Table 6. 9 Genome-wide methylation patterns among the 18 sequenced *Mtb* isolates

Sequence motifs

Sample ID(*)	CTGGAG ^a		CACGCAG		GATN4RTAC/GTAYN4ATC ^a	
	No.of motifs	%modified motif	No of motifs	%modified motif	No.of motifs	%modified motif
9909_1 (4)	1947	95	820	97	363	0
9909_2 (4)	1947	97	820	98	363	0
9909_3 (4)	1947	58	820	97	363	0
9909_4 (4)	1947	98	820	98	363	0
12366_1 (1)	1947	97	820	56	363	94
12366_2 (4)	1947	94	820	97	363	0
12366_3 (4)	1947	69	820	74	363	0
12366_7 (4)	1947	95	820	97	363	0
12366_8 (4)	1947	95	820	97	363	0
12366_9 (1)	1947	43	820	0	282	51
12366_10 (4)	1947	92	820	96	363	0
12366_11(4)	1947	96	820	97	363	0
12366_12(4)	1947	95	820	97	363	90
12366_13(4)	1947	93	820	95	363	0
10 (2)	1947	0	820	96	363	87
11 (2)	1947	0	820	96	363	88
12 (1)	1947	34	820	0	363	0
15460_8 (2)	1947	0	820	65	363	49

^a Represents methylation in only a single DNA strand

(*) Represents strain lineage

Loss of at least a single methyltransferase was found to be common across all the isolates.

Out of the 18 isolates 16 had lost at least a single methylation type with only two isolates

retaining a complete set of methylation. There was great variability in methylation across

the isolates with some being either close to 100% or zero with a few in between. It was

important to investigate this variability in methylation including the cause of loss of methylation. Multiple sequence alignment revealed that absence of *mamA* methylation (motif CTGGAG) in three isolates (10,11 and 15460_8) could be associated with a point mutation A809C (E270A) in the *mamA* gene (Rv3263) as previously documented (Shell, Prestwich et al. 2013). This mutation was found in all these three isolates belonging to lineage 2 (Figure 6. 2). Additionally one lineage 4 isolate (9909_3) possessed a previously uncharacterized missense mutation G454A (G152S). This was again revealed via multiple sequence alignment with the *mamA* gene using the laboratory strain H37Rv as a reference.

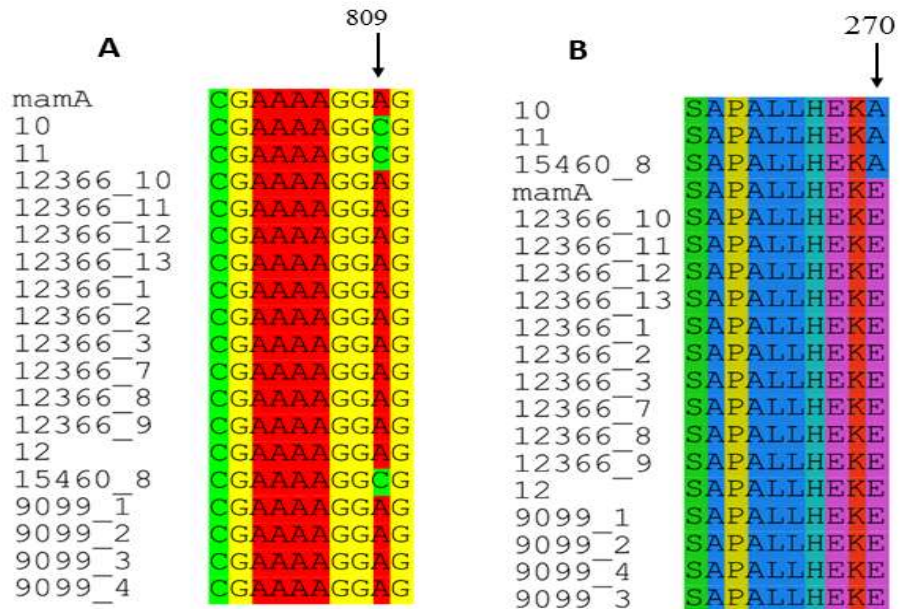


Figure 6. 2 Occurrence of mutations within the gene Rv3263 encoding the methyltransferase *MamA* and targeting the motif CTGGAG leading to partial and complete loss of methylation in three isolates

(A) Nucleotide sequence: Missense mutations A809C in samples 10, 11 and 15460_8

(B) Amino acid sequence: Amino acid changes resulting from mutations (A) in samples 10, 11 and 15460_8 respectively, leading to complete loss of methyltransferase in those samples. *mamA*_Rv3263 (H37Rv) was used as the reference sequence.

Similarly, partial or total loss of *mamB* methyltransferase activity (motif CACGAG) in 12366_1,12366_9 and 12, could be ascribed to a novel missense mutation C758T

resulting in amino acid change S253L in the *mamB* gene (Rv2420c) (Figure 6. 3A and B). This mutation has recently been characterized as being present only in lineage 1 *Mtb* isolates (Phelan, de Sessions et al. 2018). Surprisingly, this missense mutation was observed in all three isolates sample 12366_1 had 56% methylation with the other two (12366_9 and 12) being methylated at 0%. This is the first time this mutation has been reported as leading to complete loss of methylation in lineage 1. A previously characterized (Zhu, Zhong et al. 2015) synonymous mutation C696T was also detected in all three lineage 2 isolates in the *mamB* gene. Additionally, sample 12366_1 possessed a 5bp frameshift deletion at position 1512, sample 12366_12 had a 3bp frameshift deletion at position 1509 and there were two 2bp frameshift deletions at position 1509 in isolates 9909_1 and 9909_3

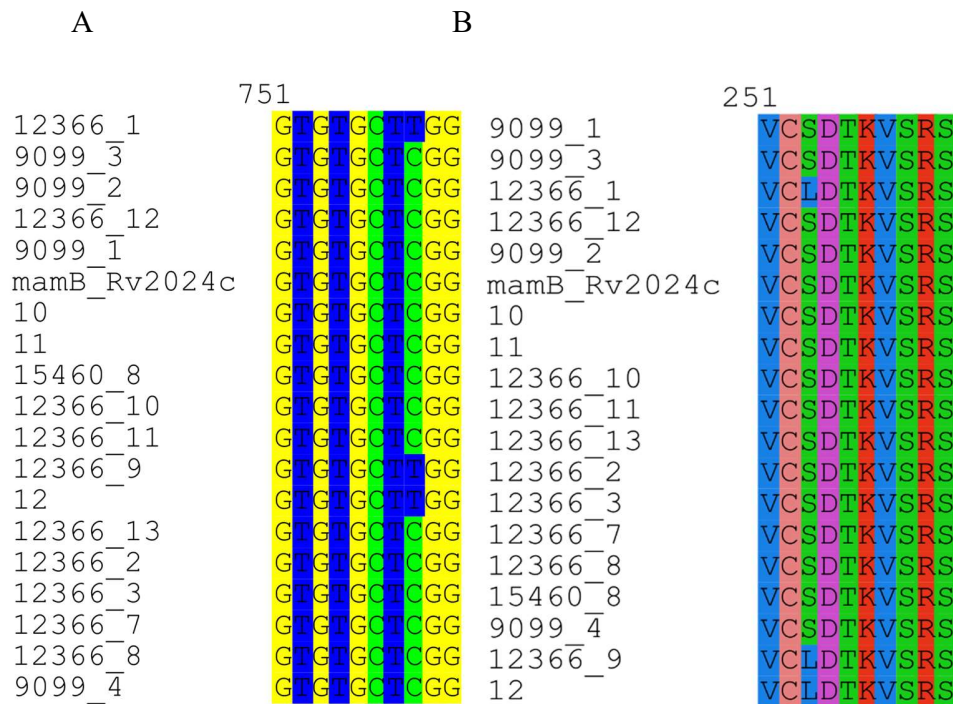


Figure 6. 3 Occurrence of mutations within the gene Rv2024c encoding the methyltransferase *MamB* and targeting the motif CACGCAG leading to partial and complete loss of methylation.

(A) Nucleotide sequence: Missense mutations C758T in samples 12366_1 and 12366_9 and 12 (B) Amino acid sequence: Amino acid changes resulting from mutations and deletions Ser253Lue in samples 12366_1 and 12366_9, led to partial loss of methyltransferase in 12366_1 and complete loss in 12366_9 and 12. mamB_Rv2024c was the reference sequence.

Further analysis revealed that 12 isolates had completely lost *HsdM* methylation (motif GATN₄RTAC) and possessed the missense mutation C917T leading to amino acid change

Pro306Leu (Figure 6.4A and B). Eleven out of the 12 isolates belonged to lineage 4. (Table 6.8). Interestingly, one lineage 1 isolate (12) that had lost *HsdM* methylation did not have the Pro306Leu mutation and also belonged to lineage 1 (Figure 6.4). Further analysis revealed that the *hsdM* gene in this isolate did not possess any other mutation. The reference sequence H37Rv used in the analysis belongs to the Euro-American (L4) and most Euro-American strains possess this mutation. Conversely all isolates without this mutation possessed a methylated GATN₄RTAC motif in various proportions (Table 6. 8).

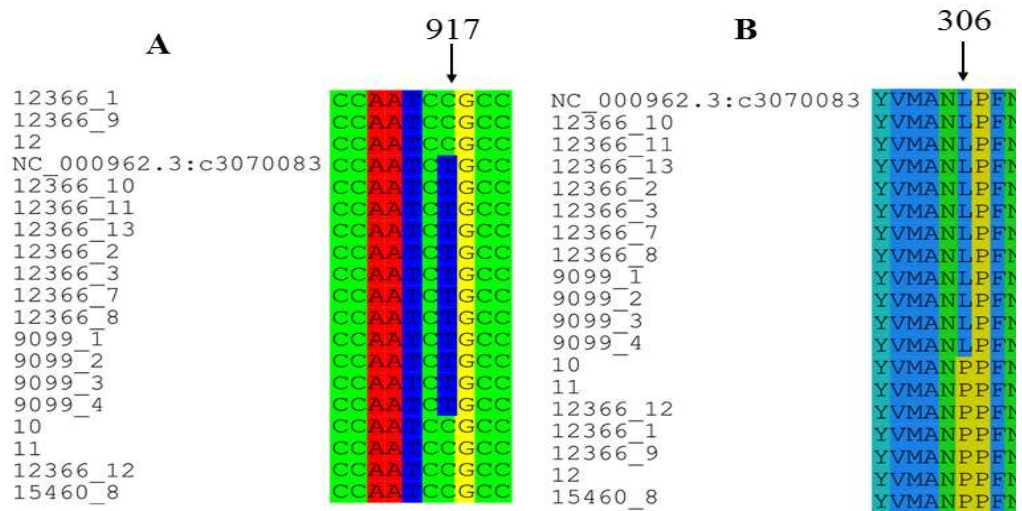


Figure 6. 4 Occurrence of mutations within the gene Rv2756c encoding the methyltransferase *HsdM* which is known to methylate the GATN4RTAC motif leading to loss of methyltransferase activity.

Occurrence of mutations within the gene Rv2756c encoding the methyltransferase *HsdM* which is known to methylate the GATN4RTAC motif. (A) Nucleotide sequence: Missense mutation C917T in samples 12366_2 and 12366_3, 12366_7, 12366_8, 12366_10, 12366_11, and 12366_13, 9099_1, 9099_2, 9099_3, 9099_4 that resulted in amino acid change. (B) Amino acid sequence: Amino acid change at Pro306Leu in the same samples resulting from the mutation C917T leading to loss methyltransferase activity for *HsdM*. The laboratory strain H37Rv (NC_000962.3:c3070083-3068461) was used as a reference sequence.

6.2.9 Comparative analysis of Malawian isolates against a global sample

A comparison was conducted of the 18 sequences from the current study against a global sample of 16 SMRT sequences downloaded from <https://www.ebi.ac.uk/ena/data/view/PRJEB21888> of which a subset was used in a published study (Phelan, de Sessions et al. 2018). These samples originated from the Europe, South Africa, west Africa and Asia. After aligning all the 34 genomes to the reference H37Rv, a total of 60,477 variant sites were identified with one sample having 51746 variants not found in any other sample. A maximum likelihood tree (Figure 6.5) constructed using the variants revealed that the samples clustered by lineage as expected. The three lineage 1 isolates from the present study (green in figure 6.5) clustered separately from the others. Two samples from the comparative study (yellow in figure 6.5) clustered separately as lineage 6. These were additionally identified by a lineage specific synonymous SNP C/G at position 1816587 (Rv1617) as previously described (Coll, McNerney et al. 2014). Eight clustered as lineage 5 (red) and were further identified using a lineage specific synonymous SNP C/A at position 1799921 in Rv1599 (Coll, McNerney et al. 2014). All the lineage 5 isolates also possessed a synonymous mutation G1374T in the *mamA* gene. Three lineage 2 isolates (blue) from the current study clustered with two isolates from the comparative study whereas 12 lineage 4 isolates clustered correctly with 4 samples from the comparative study.

The base modifications and motif analysis of the global sample of 16 genomes revealed three confidently identified motifs CTCCAG, and CACGCAG, GATN₄RTAC in complete concordance with the isolates from Malawi. All these had an IPD ratio of >5. Three samples (A03, A20 and A25) lacked the *MamA* methyltransferase. Multiple

sequence alignment against the reference H37Rv revealed that consistent with the three Malawian lineage samples, these possessed the E270A mutation in the *mamA* gene. Two of these (A03 and A20) clustered correctly as lineage 2 together with Malawian lineage 2 isolates while one (A25) clustered as lineage 6. Sample A25 did not possess the E270A mutation but possessed a mutation G1378A (A460T) only in the *mamA* gene. The rest of the samples (13) were methylated in the CTCCAG motif at proportions of the range 96-98%. All the 16 samples possessed the *MamB* Mtase the motif in proportions ranging 63%-99% with most of them close to 100%. Four samples (A05,A06,A07 and A17) lacked the *hsdM* Mtase whereas the rest were methylated in the range 84%-98%. While three (A05,A06 and A07) possessed the C917T (P306L) mutation A17 did not. Intriguingly A17 did not possess any other mutation in the *hsdM* gene.

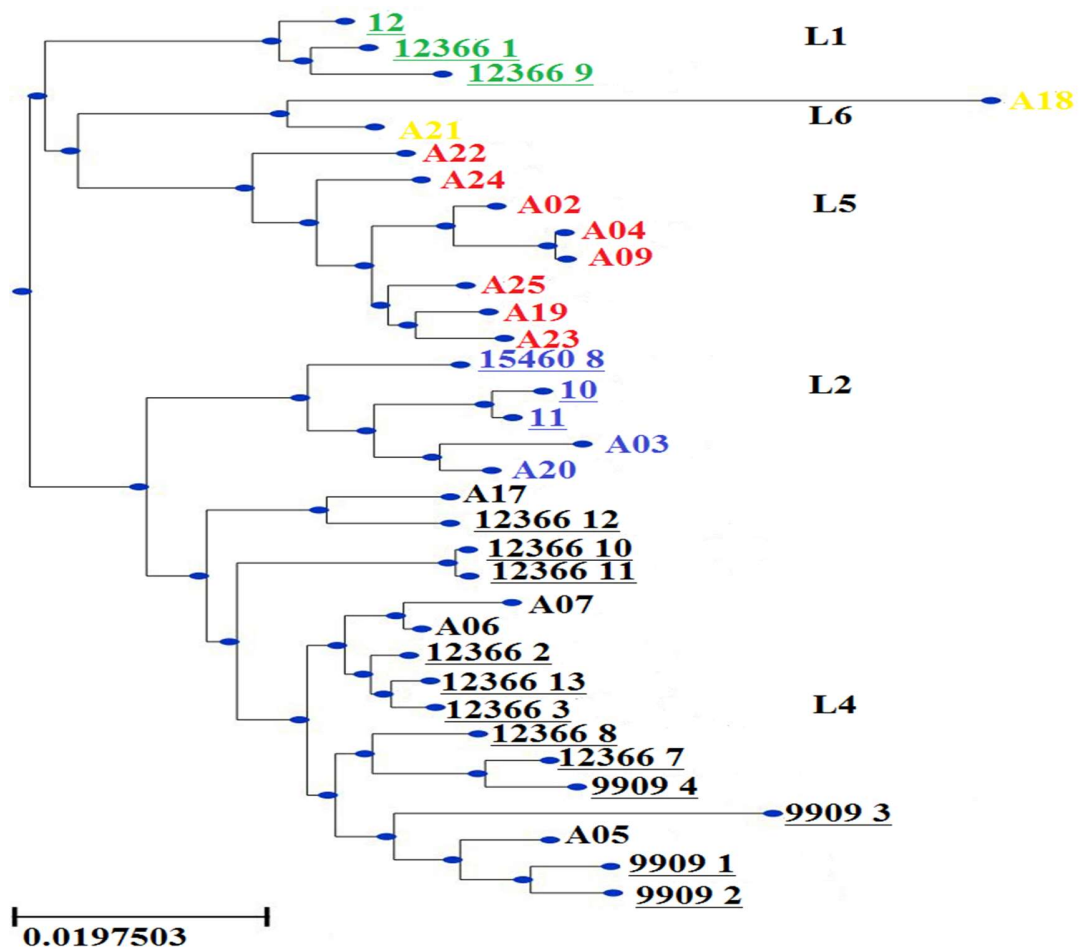


Figure 6.5 Comparative phylogenetic analysis of Malawian isolates against a global sample
 Phylogenetic analysis of 34 *Mtb* genomes run in MAFFT software. A phylogenetic tree was constructed for the sequences using RAxML (Randomized Axelerated Maximum Likelihood) a program for phylogenetic analysis in GTRCAT model. Three lineage one isolates (green) from the present study clustered separately from the others. Two samples from the comparative study (yellow) clustered separately as lineage 6 whereas eight clustered as lineage 5 (red). Three lineage 2 isolates (blue) from the current study clustered with two from the previous study whereas 12 lineage 4 isolates (black)

clustered correctly with 4 samples from the reference study. All samples from the present study are underlined.

6.3 Discussion

6.3.1 Detection of methylated motifs in *Mycobacterium tuberculosis* strains

PacBio's SMRT sequencing offers accurate consensus long reads and is especially capable of resolving repetitive regions like the *PE/PPE* genes. In this study 18 new *Mtb* genomes were completed, the largest number to be completed using PacBio sequencing within a single study and had their methylome profiles analyzed. Although all these strains belonged to lineages 1-4 and originated from Malawi, this is the first time that strains from this region have undergone methylome profile analysis. It is of interest to compare these new data to *Mtb* genomes and methylomes from other regions.

Three confidently identified motifs (CTGGAG, CACGCAG and GATN₄RTAC) were detected across the 18 genomes and were detected as methylated (6mA) to varying degrees. Evidently this is not the first study to demonstrate this pattern as others have described these same motifs although with variability in occurrence within similar strains (Zhu, Zhong et al. 2015). A high IPD ratio of >5 in all cases is evidence of the strength of a 6mA signal within the genome. Additionally an average genome sequencing coverage of ~173X is evidence of the confidence in the accuracy of the sequencing process. The presence of m4C in *Mtb* is an area requiring further investigation as in this study weak m4C signals were detected. So far these results seem to suggest presence of this methylation pattern, but with a relatively low IPD ratio and non-specificity of methylation type. This is not the first study to report presence of m4C in *Mtb* as others (Naidu 2014)

have previously reported it suggesting that it could have a role to play in gene regulation. The methylation type m5C has not been detected in *Mtb* using SMRT sequencing technology. It has been suggested that m5C has a weak signal owing to the fact that it is located on the major groove of the DNA strand and is therefore less sensitive to polymerase kinetics (Flusberg, Webster et al. 2010). The methylation m5C was previously detected however in the virulent strain of H37Rv using methyl labelling and thin layer chromatography (Srivastava, Gopinathan et al. 1981). In this chapter no overall compelling correlation between methylation pattern and persistence phenotype has been established. To a certain extent this may reflect the fact that isolation of true persister *Mtb* cells remains a challenge even in vitro. The phenotype of the isolates used in this study was defined by clinical treatment failure (despite being drug susceptible) and high lipid body content, putative biomarkers of persistence. However, it is unknown as to whether at the time of DNA isolation, these isolates were indeed still in persistent state as persistence has been reported to be reversible. Improvements in single cell technologies should make it possible to isolate true persister cells in future studies. Although it has been demonstrated in this study that MGIT culturing had some effect on methylation, this has been attributed to the lineage of the concerned isolate (see section 6.3.3). Under the current experimental conditions, it was not practically possible to confirm if the isolates tested in this study were in a true persister state. Several attempts were made to obtain pure *Mtb* DNA directly from sputum for SMRT sequencing. This was based on the assumption that persister-like *Mtb* organisms have previously been detected in sputum (Garton, Waddell et al. 2008, Honeyborne, McHugh et al. 2016). An attempt to sequence directly from sputum was largely unsuccessful due to low DNA yields and host DNA contamination.

A few methods were attempted to extract pure high molecular weight DNA. One method employed was the CTAB method with which one study claimed to achieve high yields of DNA for WGS (Amita, Vandana et al. 2002). In a separate study, researchers at the University Warwick were able to sequence directly from *Mtb* sputum on an illumina Miseq platform (Doughty, Sergeant et al. 2014) albeit at a low coverage. In that study a method known as metagenomics was applied. Unfortunately this method remains unsuitable for PacBio's SMRT sequencing as it has limitations of sequencing all organisms within a sample and will include organisms other than *Mtb*. Sequencing coverage was overall rather low as a result of probable host DNA contamination. Researchers at the University of Oxford recently devised a DNA clean-up method that greatly reduces host DNA contamination in a MGIT culture for WGS (Votintseva, Pankhurst et al. 2015). In the study a saline wash step of DNA clean-up achieved 83% successful extractions resulting in >90% sequencing coverage. A few attempts were made to apply this method to *Mtb* sputum in the present study. While the clean-up step was efficient in removing host DNA and other non-tuberculous organisms, it greatly affected the *Mtb* DNA yield. It is evident that this method is not currently feasible for DNA extraction from sputum especially for SMRT sequencing. Additionally an attempt was made to generate *in-vitro* persister *Mtb* isolates via NO exposures. In vitro, NO has been shown to inhibit *Mtb* growth in a dose-dependent manner and induce the 48-gene *dosR* regulon (Voskuil, Schnappinger et al. 2003). Additionally, higher levels of NO led to dose dependent bacteriostasis with little apparent killing (Voskuil, Bartek et al. 2011). DNA was isolated from the resulting cells and submitted for sequencing. Unfortunately, samples were unable to pass the PacBio quality control stage due to high degradation of

DNA. This experiment was not repeated owing to time and resource constraints. With advances in host DNA clean-up methods future studies should focus on sequencing directly from sputum and possibly from *in vitro* generated persisters.

6.3.2 Methylation within promoters and genes of persister *Mtb* strains

It has been long proposed that persistence might be as a result of toxin-antitoxin modules that may contribute to *Mtb* pathogenesis (Georgiades, Raoult 2011). It has further been revealed that pathway metabolic shutdown during *Mtb* persistence is accompanied by upregulation of at least 10 TA systems (Sala, Bordes et al. 2014). In *E. coli* it was shown that demobilization of the *RelB* antitoxin led to overexpression of the toxin *RelE* leading to a delayed relaxed stringent response in a starvation model (Christensen, Gerdes 2004). Two *RelBE* systems have been characterized in *Mtb*, namely *RelBE1* (*Rv1947c-Rv1246c*) and *RelBE2* (*Rv2865-Rv2866*) and *RelBE2* was one of the 10 most upregulated TA modules in the *Mtb* persister transcriptome (Keren, Minami et al. 2011). These two modules were found to impact growth, protein function and morphology in *Mtb* cells (Korch, Malhotra et al. 2015). These TA systems were of interest to this study. However, this module was not found to be differentially methylated in the “persister” isolates suggesting methylation did not have a role in the expression of this modules (*RelBE1* and *RelBE2*) and hence in persistence.

Similarly, overexpression of *MazF1* (*Rv2801c*), *MazF5* (*Rv1942c*) or *MazF6* (*Rv1102c*) has been reported to increase “persister” formation in *Mtb*, while deletion of either of them decreased “persister” formation (Sala, Bordes et al. 2014). *MazF1*, *MazF5* and *MazF6* are members of the *MazEF* TA system. Results from the current study indicate that none of these three *MazEF* TA system members were uniquely methylated in “persisters”. Since

no other TA modules were methylated in “persisters”, it could be concluded that methylation was in no way related to TA mediated “persister” formation.

The *dosR* regulon has consistently been associated with dormant infection where bacilli are in a non-replicating persistence state (Voskuil, Schlesinger 2015, Voskuil, Schnappinger et al. 2003, Parish, Smith et al. 2003). A previous analysis of the “persister” *Mtb* transcriptome revealed that only eight of *dosR* genes were upregulated suggesting that the *dosR* regulon was not involved in persistence (Keren, Minami et al. 2011). However, *hspX* (*acr1*), a member of the *dosR* regulon, is one of the most upregulated genes in the hypoxic dormancy model. It was of interest to interrogate the methylation status of *hspX* as this might have gene expression implications. Results revealed that the promoter of *hspX* was methylated at only a single site across all the four isolates. It was found that *Rv0251c* (*acr2*), a homologue of *hspX*, was unmethylated in the four isolates. *Rv0251c* encodes a heat shock protein and was found to be overexpressed in all “persister” models (Keren, Minami et al. 2011, Torrey, Keren et al. 2016). Interestingly, *acr2* was the only one of the five genes overexpressed in all models (persister, starvation, enduring hypoxic response (EHR) and non-replicating persistence) whose promoter was found to be differentially methylated in persisters. In addition to *acr2*, the other four genes upregulated in all models were *Rv1152* (*gntR*), *Rv2497* (*bkdA*), *Rv2517c* and *Rv3290c* (Keren, Minami et al. 2011). Consistently apart from *gntR* and *Rv3290c* possessing a single methylation in the promoter of a non-persister and a persister respectively, no differential methylation within “persisters” for promoters of all these genes was observed. If dormancy is consistent with downshift of some metabolic pathways (Keren, Minami et al. 2011), it was hypothesized that methylation sites may overlap with TSS of the

concerned genes thereby inhibiting transcription and regulating expression of such genes. The results revealed that while other members of the *dosR* regulon remained unmethylated in all the four isolates. The researcher felt that the methylation in *dosR* regulon did not warrant further investigation as any methylation observed was insignificant and was most likely a random event. It is not possible to speculate with certainty in the absence of transcriptomic data whether the of presence or absence of methylation impacted on gene expression. Future studies integrating transcriptome and methylome data may need to probe this finding further. The gene regions of Rv2517, which codes for a conserved hypothetical protein were the most highly methylated although specific reasons for this are unclear. Furthermore, although methylation within gene regions (Furuta, Namba-Fukuyo et al. 2014) has been reported to affect gene expression specific mechanisms remain to be elucidated.

6.3.3 Lineage specific methylation in *Mtb* strains from Malawi

Several lines of evidence have suggested that methylation in *Mtb* occurs in a lineage specific manner (Shell, Prestwich et al. 2013, Phelan, de Sessions et al. 2018). In one study *mamA* was found to influence gene expression in *Mtb* and to play a critical role in the strain specific survival of the pathogen during hypoxia. In another study, the presence or absence of methylation in the genomes of 16 *Mtb* isolates were correlated to mutations that occurred within specific lineages. Studies of *Mtb* DNA methylation using SMRT sequencing have focused on strains originating from the United States of America (Shell, Prestwich et al. 2013), Asia (Zhu, Zhong et al. 2015) and more recently a global sample that included Europe, West Africa and South Africa and Asia (Phelan, de Sessions et al. 2018). To date, no *Mtb* samples from Malawi and the surrounding region have been

subjected to either PacBio SMRT sequencing technology or DNA methylation analysis. To this end, a decision was made to investigate methylation patterns within the commonly occurring lineages in Malawi and determine whether methylation could have a role in lineage propensity to cause disease within this geographical region. In this study, SMRT sequencing of 18 *Mtb* clinical isolates revealed three methyltransferases (*MamA*, *MamB* and *Hsdm*) across the three lineages under study. The activity of these methyltransferases was inactivated by three different mutations in a lineage specific manner. *MamA* was found to be active in all isolates except three lineage 2 (*Beijing*) isolates putatively due to a point mutation A809C (E270A). This point mutation has been previously characterized (Shell, Prestwich et al. 2013). This mutation was also implicated in the failure of *Beijing* strains to survive during hypoxia, a condition in which lineage 4 (H37Rv) isolates were able to survive. Interestingly, lineage 2 (*Beijing*) strains are the most virulent and have been associated with increasing drug resistance in some areas (Cowley, Govender et al. 2008, Tsolaki, Gagneux et al. 2005, Ribeiro, Gomes et al. 2014). A recent study however failed to establish a possible role of methylation in virulence of *Beijing* strains (Naidu 2014). *MamA* was shown to play a key role in the long term survival of H37Rv *Mtb* strain during hypoxia (Shell, Prestwich et al. 2013), a condition under which the *dosR* regulon is also known to be induced (Leistikow, Morton et al. 2010, Voskuil, Schnappinger et al. 2003). The *dosR* regulon is also induced in *Beijing* strains and whether these two conditions are related is an area requiring further investigation (Phelan, de Sessions et al. 2018). However, the novel mutation G454A (G152S) in *mamA* was found to be dispensable in impacting methylation. Similarly, the *mamB* methyltransferase (motif CACGCAG) was absent in two lineage 1 isolates. This was attributed to a novel missense

mutation C758T (S253L), recently characterized elsewhere (Phelan, de Sessions et al. 2018) and confirmed in this study. Additionally, while this mutation appeared to lead to partial methylation (50-60%) this study found that it could also lead to complete loss of methyltransferase activity as two of the lineage 2 isolates had 0% methylation. This is the first time the mutation C758T (S253L) has been reported as resulting in complete loss of *MamB* methylation. This mutation is present only in the EAI6 family of lineage 1 which have been shown to be responsible for recent TB outbreaks globally (Duarte, Nery et al. 2017). The occurrence of the EAI6 family of lineage 1 in Malawi is not surprising as it has previously been reported in Malawi, South Africa, Germany, the Republic of Congo, Zambia, Zimbabwe and French Guiana (Duarte, Nery et al. 2017). One study from Karonga reported occurrence of variety in the lineage 1 (EAI) spoligotype families (Glynn, Alghamdi et al. 2010). It is still unknown whether this global prevalence could be correlated with fitness advantages associated with the S253L mutation. Although other frameshift deletions were detected in the *mamB* gene in some samples (5bp deletion at 1512 in 12366_1), these were found not to lead to functional inactivation of the gene. Other studies have reported a truncated of the *mamB* gene (positions 1520-48210) in some Euro-American lineage strains (H37Rv and H37Ra) which has been suspected to lead to loss of methyltransferase activity (Zhu, Zhong et al. 2015, Phelan, de Sessions et al. 2018). No reference strains (H37Rv and H37Ra) were sequenced in the current study and no such deletion in the *mamB* gene was identified in any of the 18 clinical isolates sequenced. This suggests that a truncated *mamB* gene is dispensable in terms of loss of methyltransferase activity at least in these clinical isolates (L1,2 and 4). This is consistent with previous studies (Zhu, Zhong et al. 2015, Phelan, de Sessions et al. 2018) as this deletion was not

reported in any of the clinical isolates. As for *HsdM*, 11/12 lineage 4 clinical isolates lacked this methyltransferase which could be attributed to the C917T (P306L) mutation present in all the 11 isolates. Similarly, these results are consistent with previous studies which seem to suggest that the P306L mutation is very common in lineage 4 strains (Phelan, de Sessions et al. 2018, Zhu, Zhong et al. 2015). In one study, the mutation was found to be present in 35 out of 37 lineage 4 clinical isolates (Zhu, Zhong et al. 2015). No cognate restriction enzyme for *HsdM* has been identified suggesting it could be an orphan methyltransferase (Zhu, Zhong et al. 2015). Its main function could therefore be more related to gene regulation as opposed to restriction modification. Lineage 4 isolates have a higher global prevalence than any other lineage and more studies will be required to establish whether loss of *HsdM* methylation could be associated with this global success. One lineage 1 isolate that was found to lack the *HsdM* methylation did not have the Pro306Leu mutation. Lack of methylation therefore could not be attributed to the mutation in all lineage 4 isolates. However, it is suspected that failure to detect *HsdM* methylation in this isolate could be attributed to the low average genome sequencing coverage of ~27X compared to the overall average genome coverage of ~173X as no methylation disrupting mutation was identified elsewhere in the genome of the strain. Elsewhere, lineage 3 (East African/Indian) clinical isolates were found to be associated with reduced transmissibility compared to other lineages (Albanna, Reed et al. 2011, Reed, Pichler et al. 2009) although evidence supporting this phenotype is very limited. In animal studies Indian lineage 1 isolates were associated with reduced virulence (Narayanan, Gagneux et al. 2008). The conviction in this study is that DNA methylation could help provide clues to some of these phenotypes which have been associated with lineage variation. The fact

that mutations leading to loss of methylation were so common (16/18) in these isolates may seem to suggest that *Mtb* uses loss of methylation as a form of achieving fitness advantage such as immune evasion or persistence in the hostile host environment.

In addition, this study has revealed the importance of using lineage specific references to identify these mutations. For example, in the identification of the mutation C917T which is common in the Euro-American strain, a reference H37Rv belonging to Euro-American lineage was used. This particular strain is a mutant at position 917 in this gene. This created problems in identifying the mutation as the reference strain was supposed to be the wild type for this locus. Similarly, all Euro-American strains possess a 7bp deletion in the *pks15/1* gene. Consequently during identification of all lineage 4 strains a reference belonging to lineage 1 (Indo-Oceanic) was used as lineage 1 strains have an intact *pks15/1* gene as opposed to H37Rv.

Comparison of isolates from this study against a global sample of isolates doubles the number of isolates available for analysis and expands the representation of lineages prevalent globally. While access to the twenty six genomes from <https://www.ebi.ac.uk/ena/data/view/PRJEB21888> was granted, the identities of the samples that had been used in the (Phelan, de Sessions et al. 2018) study could not be provided. Consequently, the 16 samples that were used for comparison in this study were selected based on read quality of the genome and might not have been congruent with those in the study. This analysis revealed that *Mtases* are highly conserved in all *Mtb* complex isolates and that loss of *Mtase* may be used as a means to increase fitness advantage. While methylation disrupting mutations have been identified and appear to be highly lineage specific, some unresolved issues remain regarding correlation of genotype

with loss of methylation. For instance, it is not currently possible to explain why some lineage 1 isolates lose activity of *HsdM Mtase* in the absence of the C917T mutation or any other mutation in the *hsdM* gene. This has been demonstrated in both Malawi isolates and the global sample. This may suggest that there could be some other currently unknown additional mechanism that is responsible for disrupting *Mtase* activity. While the comparative study had three EAI6 lineage 1 isolates, these were not found during the comparison due to differences in the choice of the 16 samples used. It was not possible to definitively identify the samples in absence of sample identities. It is likely that the three lineage one isolates used in the Phelan et al. study were not included in the comparison analysis.

Overall, the data presented in this chapter shows the potential of SMRT sequencing to help in better understanding the complete biology of *Mycobacterium tuberculosis* and its relationship to epidemiology and clinical phenotype.

CHAPTER SEVEN: FINAL DISCUSSIONS AND CONCLUSIONS

7.1 Introduction

The thesis started by evaluating different DNA extraction methods that can be used for *Mtb* DNA extraction compatible with SMRT sequencing. The advent of WGS has offered an opportunity to characterize the entire genome of *Mtb*. Pacific Biosciences SMRT sequencing offers additional benefits of being able to completely characterize the entire methylome of *Mtb* in addition to the genome. Although a few extraction methods have been applied for other WGS platforms, SMRT sequencing poses additional challenges. SMRT sequencing has more stringent DNA requirements than those required in other WGS platforms. One of the barriers to extracting high molecular weight DNA from *Mtb* is the tough cell wall made of numerous lipids and polysaccharides. It was important to find a DNA extraction method that would effectively lyse the *Mtb* cell wall and also efficiently separate the DNA from the rest of the impurities.

This DNA would be subjected to SMRT sequencing to determine base sequences and compare methylation patterns in various *Mtb* clinical isolates. Additionally the clinical isolates were subjected to molecular characterization and their specific lineages assigned using an alternative approach.

The main aims of the study have been restated hereunder in advance of the final discussion:

Primary aim: To identify genetic and epigenetic factors that could be responsible for *Mtb* persistence.

Secondary aims:

- i. To validate genomic DNA extraction methods compatible with DNA methylation analysis in *Mtb*
- ii. To analyze DNA methylation patterns in persistent *Mtb*
- iii. To assess stability of DNA methylation patterns in passage/strains
- iv. To evaluate occurrence of methylation within different *Mtb* strains and factors that lead to loss of methylation

7. 2 Isolation of persisters and DNA extraction

The definition of a true persister *Mtb* cell has not been completely elucidated. In this study persister cell selection was based on putative markers of persistence. These are lipid body formation, increased time to positivity and poor treatment outcomes (Sloan, Mwandumba et al. 2015). The cells used in this study were cultured in LJ and MGIT. It has been suggested that persistence is a transient phenotype and that persisters may switch back into a non-persistent state if a stressful environmental condition is removed (Balaban, Gerdes et al. 2013). One of the limitations in this study may have been to confirm whether the cells were indeed in true a persister state when they were isolated. Culturing may have exposed them to a different environmental condition that may affect this phenotype. The study attempted to generate *in vitro* persister *Mtb* cells via NO exposure of the laboratory strain H37Rv. DNA obtained could not be successfully sequenced as it failed to pass the quality control stage. This assay could not be repeated owing to time and resource constraints. Future studies should explore interrogating persisters generated via this method as it would guarantee a characteristic *in vitro* persister phenotype. The alternative was to use cells directly from sputum as persisters have been demonstrated in respiratory samples (Garton, Waddell et al. 2008, Honeyborne, McHugh et al. 2016).The main

challenge with sputum was to find a method of separating non-tuberculous and host cells from *Mtb*. A few methods were attempted including one that employed use of a saline wash step as previously documented (Votintseva, Pankhurst et al. 2015). Although this method has successfully been demonstrated to work with other WGS platforms, it was found to be incompatible with SMRT sequencing. The saline wash step, while improving the purity, was found to significantly affect the DNA yield. Extraction of *Mtb* DNA has always been problematic. Successful extraction of *Mtb* DNA was dependent on finding the optimum extraction method that would be compatible with SMRT sequencing. This study has explored a number of DNA extraction methods both in culture and sputum in Chapters 3 and 4. The first to be explored in culture was the CTAB based chloroform-phenol method (Warren, de Kock et al. 2006). This method may or may not employ a bead beating step to lyse the tough mycobacterial cell wall. In this study, bead beating was found not to significantly affect DNA yield. Bead beating was therefore abandoned as it only poses an additional safety risk in the laboratory due to aerosols from the beating process. Overall the phenol/chloroform CTAB method has a long turn-around time and was therefore dismissed.

The alternative CTAB method without phenol (Somerville, Thibert et al. 2005) was preferred and was successfully used for SMRT sequencing. While it has been demonstrated in this study that the CTAB method of DNA extraction from culture is amenable to SMRT sequencing future studies should focus on developing DNA a clean-up method for extraction of *Mtb* DNA from sputum. The DNA clean-up method attempted in this study was found to be incompatible with SMRT sequencing as it could not generate significantly sufficient quantities of DNA. One method that offers hope has used

biotinylated RNA baits to capture *Mtb* genomes from sputum (Brown, Bryant et al. 2015). In this study, 20 smear positive sputa achieved > 20X sequencing depth and >90% genome coverage on an Illumina sequencing following targeted oligonucleotide enrichment. Whether this method can be successfully applied to SMRT sequencing should be a question answered in future studies.

7.3 Lineage assignment and SMRT sequencing

Limited data on *Mtb* molecular epidemiology in Malawi compelled this study to investigate this area. Currently the only available data on *Mtb* strains circulating in Malawi comes from Karonga in the Northern Region (Glynn, Crampin et al. 2005, Glynn, Alghamdi et al. 2010). Although Malawi is a relatively small country the two areas (north Malawi and Blantyre) are geographically distinct. This may suggest that distinct *Mtb* strains may be circulating in these areas. Chapter 4 covers the molecular characterization of 64 *Mtb* clinical isolates obtained from patients presenting with bacteriologically confirmed pulmonary TB at the main referral hospital in Blantyre, Malawi. Results from this study show that the diversity of *Mtb* lineages in Malawi may follow a similar pattern. The sample size in this study was relatively small, more larger studies are required to confirm results obtained in this study. This study has for the first time compared phylogenies and base modifications of SMRT sequences between Malawian isolates and a global sample. This is the first time that such an analysis has included SMRT sequences of lineages 5 and 6 which are restricted to west Africa. This will aid in increasing understanding the methylome diversity of *Mtb* complex.

The principal objective of this study was to establish a correlation between DNA methylation and *Mtb* persistence using SMRT sequencing. SMRT sequencing is capable

of providing the link between a particular change at single base pair resolution to genes and ultimately to biological function (Chen, Jeannotte et al. 2014). Since the *Mtases* in *Mtb* have so far not been shown to have cognate restriction enzymes, the restriction function is therefore dispensable. It is logical to suggest that DNA methylation indeed has gene regulation function in *Mtb* (Shell, Prestwich et al. 2013). While the role of *MamA* during *Mtb* hypoxia has been established (Shell, Prestwich et al. 2013), studies should now focus on finding the role of other *Mtases* such as *MamB* and *HsdM* in gene expression. A comparison of two persisters and non-persisters in the current study failed to establish a compelling correlation. While a few sites were found to be differentially methylated in persisters as opposed to non-persisters, these were considered not to be significant. A further attempt to interrogate differences in methylation within persister specific genes was not informative. The *dosR* regulon among many other genes has been widely associated with the persister phenotype (Voskuil, Schnappinger et al. 2003) and was specifically targeted in this study. In the absence of any gene expression data, it is not possible to determine whether any of these small differences in methylation may have had an impact. Future studies should combine methylation data with transcriptomic data to put this question to rest. Additionally, an enrichment analysis of genes and pathways differentially methylated found none of the categories were over-represented in persisters.

7. 4 Key findings of the study

The study has established that four principal *Mtb* strains (L1-4) circulate in Blantyre and that lineage 4 is predominant. These results are consistent with previous studies from Karonga (Glynn, Crampin et al. 2005, Glynn, Alghamdi et al. 2010). Studies with larger sample sizes will be required in Blantyre to confirm the proportion of these strains.

Lineage assignments obtained by LSP-PCR were consistent with those from SMRT sequencing for the 18 *Mtb* isolates. SMRT sequencing has also confirmed three confidently detected motifs (CTGGAG, CACGCAG and GATN₄RTAC) in *Mtb* as previously documented (Zhu, Zhong et al. 2015, Phelan, de Sessions et al. 2018). Three methylation disrupting mutations (A809C, C758T and C917T) were detected in which a novel mutation (C758T) recently characterized (Phelan, de Sessions et al. 2018) was confirmed. Additionally, for the first time, mutation C758T was found to cause complete loss of methylation in lineage 1 strains and not only partial loss as previously believed. These mutations occur in a lineage specific manner. While methylation plays a role in gene regulation, complete loss of one or more *Mtases* has been shown to be a common phenomenon of *Mtb*. This may suggest that loss of methylation could play a critical role in the fitness advantage of different strains of the pathogen, perhaps in key characteristics such as transmissibility, immune evasion and persistence *in vivo*. It is possible that different strains adapt to different environments through loss of specific *Mtases*. While *Mama* was shown to play a key role in survival of H37Rv (L4) during hypoxia, it is possible that it could have a completely different role in a different strain in the presence or absence of other *Mtases*. Future studies should explore the link between loss of *Mtase* and pathogen survival *in vivo* in different lineages. This study has completed the largest number of *Mtb* clinical isolates (18) from a single study in a single location using SMRT sequencing. This study also completed the largest number of Euro-American strains (12) and has confirmed the methylation profile of this lineage with much greater confidence. All lineage 2 clinical isolates to be completed using SMRT sequencing (Zhu, Zhong et al. 2015, Phelan, de Sessions et al. 2018) in the present study have shown the presence of

the A809C mutation in the *mamA* gene and the mutation could be used as an alternative means of lineage assignment. As for lineage 1 although all clinical isolates sequenced using SMRT sequencing (Phelan, de Sessions et al. 2018) and the present study possess the C758T mutation in the *mamB* gene, others have suggested that this mutation is only present in the EAI6 family of lineage 1 globally (Duarte, Nery et al. 2017). Whether this methylation disrupting mutation can be used for lineage assignment of all lineage 1 isolates will require more studies. At present however, it can be used to classify EAI6 family. While the C917T mutation is the commonest occurring in the *hsdM* gene within lineage 4 isolates, and has been associated with loss of *HsdM* *Mtase* activity, some lineage 1 isolates have been shown to lose the *Mtase* activity in absence of this *Mtase*. Future studies will need to investigate other possible factors responsible for such loss of *Mtase* activity such as gene silencing via small nuclear RNAs. At present this mutation may only be used for confirmatory purposes of lineage 4 isolates.

Overall this work has helped to confirm the results of two key prior studies to have SMRT-sequenced *Mtb* isolates and will help in increasing the power and representativeness of such studies. Furthermore, it has demonstrated the power of SMRT sequencing to completely elucidate the entire methylome of *Mtb* in an efficient way, enabling global analyses of its impact on important biological functions in representative sets of strains. This tool will aid in understanding the complete biology of *Mtb* and could assist in epidemiological studies and development of more efficient and ultra-short drug regimens.

7. 5 Limitations of the study

The study analyzed 64 *Mtb* clinical isolates and assigned them into 4 different lineages. This study was unable to design specific primers for lineage 4 and as a result all isolates

not belonging to three lineages (L1-3) were assumed to belong to lineage 4. This does not discount the fact that a novel strain could exist. Although the study has managed to sequence the largest number (18) of *Mtb* isolates to date using SMRT sequencing, only three lineages have been covered. A more diverse coverage of the seven *Mtb* lineages would have given the study more power. This however was due to the fact that these are the commonly circulating lineages in Malawi as the study targeted strains from this country alone. The study has however managed to conduct a comparative analysis of a global sample of isolates to compare the isolates from Malawi to those from other geographical areas. The study failed to successfully sequence directly from sputum samples in line with the objectives. There are a few methods of cleaning up DNA for sequencing that have come up. One such method makes use of biotinylated RNA baits to trap *Mtb* genomes (Brown, Bryant et al. 2015). Whether such a method would be compatible with SMRT sequencing is a question that needs to be answered in future studies. Recently SMRT sequencing has become an area of active research and similar studies will add and increase the power of these studies.

7. 6 Future direction

Future work should explore the possibility of SMRT sequencing directly from sputum without culture by using more efficient extraction methods such as capturing *Mtb* genomes using biotinylated RNA baits (Brown, Bryant et al. 2015). Further work comparing methylation patterns under stress conditions in vitro and in clinical isolates may be helpful in understanding the role of methylation profiles in determining the persistence phenotype. Finally, methods that can directly isolate persister cells from clinical samples for

characterization of their methylome would help to overcome many of the limitations encountered in this study.

8. REFERENCES

- ABDUL-AZIZ, M. and TSOLAKI, A.G., 2013. Mycobacterium Tuberculosis: Innate Immune Recognition and Persistence. In: U. KISHORE and A. NAYAK, eds, *Microbial Pathogenesis: Infection and Immunity*. London: Landes Bioscience, pp. 63-82.
- ADHIKARI, B.R., PANDEY, B.D., GHIMIRE, P., SHRESTHA, B., KHADKA, M., YODA, T. and SUZUKI, Y., 2009. Loop-mediated isothermal amplification (LAMP) for the direct detection of human pulmonary infections with environmental (nontuberculosis) mycobacteria. *Japanese journal of infectious diseases*, **62**(3), pp. 212-214.
- ALANGADEN, G.J., KREISWIRTH, B.N., AOUAD, A., KHETARPAL, M., IGNO, F.R., MOGHAZEH, S.L., MANAVATHU, E.K. and LERNER, S.A., 1998. Mechanism of Resistance to Amikacin and Kanamycin in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*, **42**(5), pp. 1295-1297.
- ALBANNA, A.S., REED, M.B., KOTAR, K.V., FALLOW, A., MCINTOSH, F.A., BEHR, M.A. and MENZIES, D., 2011. Reduced transmissibility of East African Indian strains of Mycobacterium tuberculosis. *PloS one*, **6**(9), pp. e25075.
- ALDOUS, W.K., POUNDER, J.I., CLOUD, J.L. and WOODS, G.L., 2005. Comparison of six methods of extracting Mycobacterium tuberculosis DNA from processed sputum for testing by quantitative real-time PCR. *Journal of clinical microbiology*, **43**(5), pp. 2471-2473.
- ALI, M.R., AL-THWANI, A.N. and DHAER, F.H., 2014. Comparison between different methods for DNA Extraction for detection of Mycobacterium tuberculosis from sputum using Polymerase chain reaction. *Diyala Journal for Pure Sciences*, **10**(1), pp. 51-58.
- ALMEIDA DA SILVA, P.E. and PALOMINO, J.C., 2011. Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs. *The Journal of antimicrobial chemotherapy*, **66**(7), pp. 1417-1430.
- ALONSO, M., BORRELL, S., LIROLA, M.M., BOUZA, E. and GARCÍA DE VIEDMA, D., 2008. A proposal for applying molecular markers as an aid to identifying potential cases of imported tuberculosis in immigrants. *Tuberculosis*, **88**(6), pp. 641-647.
- AMARO, A., DUARTE, E., AMADO, A., FERRONHA, H. and BOTELHO, A., 2008. Comparison of three DNA extraction methods for Mycobacterium bovis, Mycobacterium tuberculosis and Mycobacterium avium subsp. avium. *Letters in applied microbiology*, **47**(1), pp. 8-11.
- AMITA, J., VANDANA, T., GULERIA, R.S. and VERMA, R.K., 2002. Qualitative Evaluation of Mycobacterial DNA Extraction Protocols for Polymerase Chain Reaction. *Molecular Biology Today*, **3**, pp. 43-50.

ANDERSEN, P., 2001. TB vaccines: progress and problems. *Trends in immunology*, **22**(3), pp. 160-168.

ANDO, H., MIYOSHI-AKIYAMA, T., WATANABE, S. and KIRIKAE, T., 2014. A silent mutation in *mabA* confers isoniazid resistance on *Mycobacterium tuberculosis*. *Molecular microbiology*, **91**(3), pp. 538-547.

ANDO, M., YOSHIMATSU, T., KO, C., CONVERSE, P.J. and BISHAI, W.R., 2003. Deletion of *Mycobacterium tuberculosis* Sigma Factor E Results in Delayed Time to Death with Bacterial Persistence in the Lungs of Aerosol-Infected Mice. *Infection and immunity*, **71**(12), pp. 7170-7172.

ANDRES, S., HILLEMANN, D., RÜSCH-GERDES, S. and RICHTER, E., 2014. Occurrence of *rpoB* Mutations in Isoniazid-Resistant but Rifampin-Susceptible *Mycobacterium tuberculosis* Isolates from Germany. *Antimicrobial Agents and Chemotherapy*, **58**(1), pp. 590-592.

ANDREWS, J.R., SARITA SHAH, N., GANDHI, N., MOLL, T. and FRIEDLAND, G., 2007. Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis: Implications for the HIV Epidemic and Antiretroviral Therapy Rollout in South Africa. *Journal of Infectious Diseases*, **196**(Supplement 3), pp. S482-S490.

ARCUS, V.L., RAINEY, P.B. and TURNER, S.J., 2005. The PIN-domain toxin-antitoxin array in mycobacteria. *Trends in microbiology*, **13**(8), pp. 360-365.

ASMAR, S. and DRANCOURT, M., 2015. Rapid culture-based diagnosis of pulmonary tuberculosis in developed and developing countries. *Frontiers in microbiology*, **6**, pp. 1184.

AUBRY, A., PAN, X.S., FISHER, L.M., JARLIER, V. and CAMBAU, E., 2004. *Mycobacterium tuberculosis* DNA gyrase: interaction with quinolones and correlation with antimycobacterial drug activity. *Antimicrobial Agents and Chemotherapy*, **48**(4), pp. 1281-1288.

AUNG, K.J., VAN DEUN, A., DECLERCQ, E., SARKER, M.R., DAS, P.K., HOSSAIN, M.A. and RIEDER, H.L., 2014. Successful '9-month Bangladesh regimen' for multidrug-resistant tuberculosis among over 500 consecutive patients. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, **18**(10), pp. 1180-1187.

BALABAN, N.Q., GERDES, K., LEWIS, K. and MCKINNEY, J.D., 2013. A problem of persistence: still more questions than answers? *Nature reviews. Microbiology*, **11**, pp. 587-591.

BALABAN, N.Q., MERRIN, J., CHAIT, R., KOWALIK, L. and LEIBLER, S., 2004. Bacterial Persistence as a Phenotypic Switch. *Science*, **305**(5690), pp. 1622-1625.

BALABANOVA, Y., DROBNIEWSKI, F., FEDORIN, I., ZAKHAROVA, S., NIKOLAYEVSKYY, V., ATUN, R. and COKER, R., 2006. The Directly Observed Therapy Short-Course (DOTS) strategy in Samara Oblast, Russian Federation. *Respiratory Research*, **7**(1), pp. 44.

BARNES, P.F. and CAVE, M.D., 2003. Molecular Epidemiology of Tuberculosis. *N Engl J Med*, **349**(12), pp. 1149-1156.

BARREIRO, L.B., PACIS, A., TAILLEAUX, L., YOTOVA, V., GRENIER, J.C., PIQUE-REGI, R., HANSEN, K.D. and GILAD, Y., 2013. Widespread Changes at CpG island shores and distal regulatory regions in response to a bacterial infection, *American Society of Human Genetics*, October 22-26 2013.

BARRY III, C.E., LEE, R.E., MDLULI, K., SAMPSON, A.E., SCHROEDER, B.G., SLAYDEN, R.A. and YUAN, Y., 1998. Mycolic acids: structure, biosynthesis and physiological functions. *Progress in lipid research*, **37**(2-3), pp. 143-179.

BARTER, D., AGBOOLA, S., MURRAY, M. and BARNIGHAUSEN, T., 2012. Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa - a systematic review. *BMC Public Health*, **12**(1), pp. 980.

BEHR, M.A., WILSON, M.A., GILL, W.P., SALAMON, H., SCHOOLNIK, G.K., RANE, S. and SMALL, P.M., 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science (New York, N.Y.)*, **284**(5419), pp. 1520-1523.

BERNEY, M., BERNEY-MEYER, L., WONG, K.W., CHEN, B., CHEN, M., KIM, J., WANG, J., HARRIS, D., PARKHILL, J., CHAN, J., WANG, F. and JACOBS, W.R., Jr, 2015. Essential roles of methionine and S-adenosylmethionine in the autarkic lifestyle of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(32), pp. 10008-10013.

BIGGER, J., 1944. TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. *The Lancet*, **244**(6320), pp. 497-500.

BJØRN-MORTENSEN, K., ZALLET, J., LILLEBAEK, T., ANDERSEN, A.B., NIEMANN, S., RASMUSSEN, E.M. and KOHL, T.A., 2015. Direct DNA extraction from *Mycobacterium tuberculosis* frozen stocks as a re-culture independent approach to whole genome sequencing. *Journal of clinical microbiology*, .

BLACK, D.S., KELLY, A.J., MARDIS, M.J. and MOYED, H.S., 1991. Structure and organization of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. *Journal of Bacteriology*, **173**(18), pp. 5732-5739.

BLAKEMORE, R., STORY, E., HELB, D., KOP, J., BANADA, P., OWENS, M.R., CHAKRAVORTY, S., JONES, M. and ALLAND, D., 2010. Evaluation of the analytical

performance of the Xpert MTB/RIF assay. *Journal of clinical microbiology*, **48**(7), pp. 2495-2501.

BLOCH, H. and SEGAL, W., 1956. Biochemical differentiation of *Mycobacterium tuberculosis* grown in vivo and in vitro. *Journal of Bacteriology*, **72**(2), pp. 132-141.

BOEHME, C.C., NABETA, P., HENOSTROZA, G., RAQIB, R., RAHIM, Z., GERHARDT, M., SANGA, E., HOELSCHER, M., NOTOMI, T., HASE, T. and PERKINS, M.D., 2007. Operational feasibility of using loop-mediated isothermal amplification for diagnosis of pulmonary tuberculosis in microscopy centers of developing countries. *Journal of clinical microbiology*, **45**(6), pp. 1936-1940.

BRENNAN, M.J. and DELOGU, G., 2002. The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends in microbiology*, **10**(5), pp. 246-249.

BROSCH, R., GORDON, S.V., MARMIESSE, M., BRODIN, P., BUCHRIESER, C., EIGLMEIER, K., GARNIER, T., GUTIERREZ, C., HEWINSON, G., KREMER, K., PARSONS, L.M., PYM, A.S., SAMPER, S., VAN SOOLINGEN, D. and COLE, S.T., 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(6), pp. 3684-3689.

BROWN, A.C., BRYANT, J.M., EINER-JENSEN, K., HOLDSTOCK, J., HOUNIET, D.T., CHAN, J.Z.M., DEPLEDGE, D.P., NIKOLAYEVSKYY, V., BRODA, A., STONE, M.J., CHRISTIANSEN, M.T., WILLIAMS, R., MCANDREW, M.B., TUTILL, H., BROWN, J., MELZER, M., ROSMARIN, C., MCHUGH, T.D., SHORTEN, R.J., DROBNIIEWSKI, F., SPEIGHT, G. and BREUER, J., 2015. Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples. *Journal of clinical microbiology*, **53**(7), pp. 2230-2237.

CAIRE-BRÄNDLI, I., PAPADOPOULOS, A., MALAGA, W., MARAIS, D., CANAAN, S., THILO, L. and DE CHASTELLIER, C., 2014. Reversible Lipid Accumulation and Associated Division Arrest of *Mycobacterium avium* in Lipoprotein-Induced Foamy Macrophages May Resemble Key Events during Latency and Reactivation of Tuberculosis. *Infection and immunity*, **82**(2), pp. 476-490.

CAMPBELL, P.J., MORLOCK, G.P., SIKES, R.D., DALTON, T.L., METCHOCK, B., STARKS, A.M., HOOKS, D.P., COWAN, L.S., PLIKAYTIS, B.B. and POSEY, J.E., 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), pp. 2032-2041.

CASADESUS, J. and LOW, D., 2006. Epigenetic gene regulation in the bacterial world. *Microbiology and molecular biology reviews : MMBR*, **70**(3), pp. 830-856.

CAWS, M., DUY, P.M., THO, D.Q., NGOC LAN, N.T., HOA, D.V. and FARRAR, J., 2006. Mutations Prevalent among Rifampin- and Isoniazid-Resistant Mycobacterium tuberculosis Isolates from a Hospital in Vietnam. *Journal of Clinical Microbiology*, **44**(7), pp. 2333-2337.

CENTRE FOR GENOMIC RESEARCH, UNIVERSITY OF LIVERPOOL, , Preparation of PacBio libraries. Available: <http://cgrlms.liv.ac.uk/gslc/mainPage> [January 25, 2017].

CHAISSON, M.J. and TESLER, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC bioinformatics*, **13**, pp. 238-2105-13-238.

CHEN, P., RUIZ, R.E., LI, Q., SILVER, R.F. and BISHAI, W.R., 2000. Construction and Characterization of a Mycobacterium tuberculosis Mutant Lacking the Alternate Sigma Factor Gene, sigF. *Infection and immunity*, **68**(10), pp. 5575-5580.

CHEN, P., JEANNOTTE, R. and WEIMER, B.C., 2014. Exploring bacterial epigenomics in the next-generation sequencing era: a new approach for an emerging frontier. *Trends in microbiology*, **22**(5), pp. 292-300.

CHENG, A.F.B., YEW, W.W., CHAN, E.W.C., CHIN, M.L., HUI, M.M.M. and CHAN, R.C.Y., 2004. Multiplex PCR Amplimer Conformation Analysis for Rapid Detection of gyrA Mutations in Fluoroquinolone-Resistant Mycobacterium tuberculosis Clinical Isolates. *Antimicrobial Agents and Chemotherapy*, **48**(2), pp. 596-601.

CHRISTENSEN, S.K. and GERDES, K., 2004. Delayed-relaxed response explained by hyperactivation of RelE. *Molecular microbiology*, **53**(2), pp. 587-597.

CHRISTENSEN-DALSGAARD, M. and GERDES, K., 2006. Two higBA loci in the Vibrio cholerae superintegron encode mRNA cleaving enzymes and can stabilize plasmids. *Molecular microbiology*, **62**(2), pp. 397-411.

CLARK, S.J., STATHAM, A., STIRZAKER, C., MOLLOY, P.L. and FROMMER, M., 2006. DNA methylation: bisulphite modification and analysis. *Nature protocols*, **1**(5), pp. 2353-2364.

CLARK, T.A., MURRAY, I.A., MORGAN, R.D., KISLYUK, A.O., SPITTLE, K.E., BOITANO, M., FOMENKOV, A., ROBERTS, R.J. and KORLACH, J., 2011. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic acids research*, .

CLAY, H., DAVIS, J.M., BEERY, D., HUTTENLOCHER, A., LYONS, S.E. and RAMAKRISHNAN, L., 2007. *Dichotomous Role of the Macrophage in Early Mycobacterium marinum Infection of the Zebrafish*.

COLE, S.T., BROSCHE, R., PARKHILL, J., GARNIER, T., CHURCHER, C., HARRIS, D., GORDON, S.V., EIGLMEIER, K., GAS, S., BARRY, C.E., 3rd, TEKAIA, F., BADCOCK, K., BASHAM, D., BROWN, D., CHILLINGWORTH, T., CONNOR, R., DAVIES, R., DEVLIN, K., FELTWELL, T., GENTLES, S., HAMLIN, N., HOLROYD, S., HORNSBY, T., JAGELS, K., KROGH, A., MCLEAN, J., MOULE, S., MURPHY, L., OLIVER, K., OSBORNE, J., QUAIL, M.A., RAJANDREAM, M.A., ROGERS, J., RUTTER, S., SEEGER, K., SKELTON, J., SQUARES, R., SQUARES, S., SULSTON, J.E., TAYLOR, K., WHITEHEAD, S. and BARRELL, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393**(6685), pp. 537-544.

COLE, S.T., 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology*, **148**(10), pp. 2919-2928.

COLL, F., MCNERNEY, R., GUERRA-ASSUNCAO, J.A., GLYNN, J.R., PERDIGAO, J., VIVEIROS, M., PORTUGAL, I., PAIN, A., MARTIN, N. and CLARK, T.G., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nature communications*, **5**, pp. 4812.

CONSTANT, P., PEREZ, E., MALAGA, W., LANEELLE, M., SAUREL, O., DAFTE, M. and DAFTE, M., 2002. Role of *pks15/1* gene in the Biosynthesis of Phenoglycolipids in the *Mycobacterium tuberculosis* complex. EVIDENCE THAT ALL STRAINS SYNTHESIZE GLYCOSYLATED p-HYDROXYBENZOIC METHYL ESTERS AND THAT STRAINS DEVOID OF PHENOLGLYCOLIPIDS HABOUR A FRAMESHIFT MUTATION IN THE *pks15/1* GENE. *Journal of Biological Chemistry*, **277**, pp. 38148-38158.

CORBETT, E.L., WATT, C.J., WALKER, N., MAHER, D., WILLIAMS, B.G., RAVIGLIONE, M.C. and DYE, C., 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Archives of Internal Medicine*, **163**(9), pp. 1009-1021.

CORMICAN, M.G., BARRY, T., GANNON, F. and FLYNN, J., 1992. Use of polymerase chain reaction for early identification of *Mycobacterium tuberculosis* in positive cultures. *Journal of clinical pathology*, **45**(7), pp. 601-604.

CORREIA, F.F., D'ONOFRIO, A., REJTAR, T., LI, L., KARGER, B.L., MAKAROVA, K., KOONIN, E.V. and LEWIS, K., 2006. Kinase Activity of Overexpressed *HipA* Is Required for Growth Arrest and Multidrug Tolerance in *Escherichia coli*. *Journal of Bacteriology*, **188**(24), pp. 8360-8367.

COSCOLLA, M. and GAGNEUX, S., 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Seminars in Immunology*, **26**(6), pp. 431-444.

COWLEY, D., GOVENDER, D., FEBRUARY, B., WOLFE, M., STEYN, L., EVANS, J., WILKINSON, R.J. and NICOL, M.P., 2008. Recent and rapid emergence of W-Beijing

strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, **47**(10), pp. 1252-1259.

CRUCIANI, M., SCARPARO, C., MALENA, M., BOSCO, O., SERPELLONI, G. and MENGOLI, C., 2004. Meta-Analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without Solid Media, for Detection of Mycobacteria. *Journal of clinical microbiology*, **42**(5), pp. 2321-2325.

DAHL, J.L., KRAUS, C.N., BOSHOFF, H.I., DOAN, B., FOLEY, K., AVARBOCK, D., KAPLAN, G., MIZRAHI, V., RUBIN, H. and BARRY, C.E., 3rd, 2003. The role of RelMtb-mediated adaptation to stationary phase in long-term persistence of *Mycobacterium tuberculosis* in mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(17), pp. 10026-10031.

DANIEL, J., MAAMAR, H., DEB, C., SIRAKOVA, T.D. and KOLATTUKUDY, P.E., 2011. *Mycobacterium tuberculosis* uses host triacylglycerol to accumulate lipid droplets and acquires a dormancy-like phenotype in lipid-loaded macrophages. *PLoS pathogens*, **7**(6), pp. e1002093.

DANIEL, J., DEB, C., DUBEY, V.S., SIRAKOVA, T.D., ABOMOELAK, B., MORBIDONI, H.R. and KOLATTUKUDY, P.E., 2004. Induction of a Novel Class of Diacylglycerol Acyltransferases and Triacylglycerol Accumulation in *Mycobacterium tuberculosis* as It Goes into a Dormancy-Like State in Culture. *Journal of Bacteriology*, **186**(15), pp. 5017-5030.

DANIEL, T.M., 2006. The History of Tuberculosis. *Respiratory Medicine*, **100**, pp. 1862-1870.

DAVIES, P.D.O., BARNES, P.F. and GORDON, S.B., eds, 2008. *Clinical Tuberculosis*. 4th edn. London: Hodder Arnold.

DAVIES, G.R., 2013. Bridging the gap in the fight against tuberculosis. *Drug Discovery Today: Technologies*, **10**(3), pp. e359-e364.

DAVIS, B.M., CHAO, M.C. and WALDOR, M.K., 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Current opinion in microbiology*, **16**(2), pp. 192-198.

DEB, C., LEE, C.M., DUBEY, V.S., DANIEL, J., ABOMOELAK, B., SIRAKOVA, T.D., PAWAR, S., ROGERS, L. and KOLATTUKUDY, P.E., 2009. A novel in vitro multiple-stress dormancy model for *Mycobacterium tuberculosis* generates a lipid-loaded, drug-tolerant, dormant pathogen. *PloS one*, **4**(6), pp. e6077.

DEB, C., DANIEL, J., SIRAKOVA, T.D., ABOMOELAK, B., DUBEY, V.S. and KOLATTUKUDY, P.E., 2006. A Novel Lipase Belonging to the Hormone-sensitive

Lipase Family Induced under Starvation to Utilize Stored Triacylglycerol in *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, **281**(7), pp. 3866-3875.

DEJESUS, M.A., GERRICK, E.R., XU, W., PARK, S.W., LONG, J.E., BOUTTE, C.C., RUBIN, E.J., SCHNAPPINGER, D., EHRT, S., FORTUNE, S.M., SASSETTI, C.M. and IOERGER, T.R., 2017. Comprehensive Essentiality Analysis of the *Mycobacterium tuberculosis* Genome via Saturating Transposon Mutagenesis. *mBio*, **8**(1), pp. 10.1128/mBio.02133-16.

DESIKAN, S. and NARAYANAN, S., 2015. Genetic markers, genotyping methods & next generation sequencing in *Mycobacterium tuberculosis*. *The Indian journal of medical research*, **141**(6), pp. 761-774.

DHAR, N. and MCKINNEY, J.D., 2010. *Mycobacterium tuberculosis* persistence mutants identified by screening in isoniazid treated-mice. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(27), pp. 12275-12280.

DHAR, N. and MCKINNEY, J.D., 2007. Microbial phenotypic heterogeneity and antibiotic tolerance. *Current opinion in microbiology*, **10**(1), pp. 30-38.

DHEDA, K., SHEAN, K., ZUMLA, A., BADRI, M., STREICHER, E.M., PAGE-SHIPP, L., WILLCOX, P., JOHN, M.A., REUBENSON, G., GOVINDASAMY, D., WONG, M., PADANILAM, X., DZIWIECKI, A., VAN HELDEN, P.D., SIWENDU, S., JARAND, J., MENEZES, C.N., BURNS, A., VICTOR, T., WARREN, R., GROBUSCH, M.P., VAN DER WALT, M. and KVASNOVSKY, C., 2010. Early treatment outcomes and HIV status of patients with extensively drug-resistant tuberculosis in South Africa: a retrospective cohort study. *Lancet*, **375**(9728), pp. 1798-1807.

DORMAN, S.E., SCHUMACHER, S.G., ALLAND, D., NABETA, P., ARMSTRONG, D.T., KING, B., HALL, S.L., CHAKRAVORTY, S., CIRILLO, D.M., TUKVADZE, N., BABLISHVILI, N., STEVENS, W., SCOTT, L., RODRIGUES, C., KAZI, M.I., JOLOBA, M., NAKIYINGI, L., NICOL, M.P., GHEBREKRISTOS, Y., ANYANGO, I., MURITHI, W., DIETZE, R., LYRIO PERES, R., SKRAHINA, A., AUCHYNKA, V., CHOPRA, K.K., HANIF, M., LIU, X., YUAN, X., BOEHME, C.C., ELLNER, J.J. and DENKINGER, C.M., 2018. Xpert MTB/RIF Ultra for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre diagnostic accuracy study. *The Lancet Infectious Diseases*, **18**(1), pp. 76-84.

DORR, T., VULIC, M. and LEWIS, K., 2010. Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. *PLoS biology*, **8**(2), pp. e1000317.

DOUGHTY, E.L., SERGEANT, M.J., ADETIFA, I., ANTONIO, M. and PALLAN, M.J., 2014. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. *PeerJ*, **2**, pp. 585.

DREHER, D. and NICOD, L.P., 2002. Dendritic cells in the mycobacterial granuloma are involved in acquired immunity. *American journal of Respiratory and Critical Care Medicine*, **165**(12), pp. 1577-1578.

DUARTE, T.A., NERY, J.S., BOECHAT, N., PEREIRA, S.M., SIMONSEN, V., OLIVEIRA, M., GOMES, M.G.M., PENHA-GONÇALVES, C., BARRETO, M.L. and BARBOSA, T., 2017. *A systematic review of East African-Indian family of Mycobacterium tuberculosis in Brazil*.

EHLERS, M.R. and DAFPE, M., 1998. Interactions between Mycobacterium tuberculosis and host cells: are mycobacterial sugars the key? *Trends in microbiology*, **6**(8), pp. 328-335.

ENGSTRÖM, A., MORCILLO, N., IMPERIALE, B., HOFFNER, S.E. and JURÉEN, P., 2012. Detection of First- and Second-Line Drug Resistance in Mycobacterium tuberculosis Clinical Isolates by Pyrosequencing. *Journal of clinical microbiology*, **50**(6), pp. 2026-2033.

ESTERHUYSE, M.M., WEINER, J., CARON, E., LOXTON, A.G., IANNACCONE, M., WAGMAN, C., SAIKALI, P., STANLEY, K., WOLSKI, W.E., MOLLENKOPF, H., SCHICK, M., AEBERSOLD, R., LINHART, H., WALZL, G. and KAUFMANN, S.H.E., 2015. Epigenetics and Proteomics Join Transcriptomics in the Quest for Tuberculosis Biomarkers. *mBio*, **6**(5),.

FÀBREGA, A., MADURGA, S., GIRALT, E. and VILA, J., 2009. Mechanism of action of and resistance to quinolones. *Microbial Biotechnology*, **2**(1), pp. 40-61.

FADZILAH, M.N., NG, K.P. and NGEOW, Y.F., 2009. The manual MGIT system for the detection of M. tuberculosis in respiratory specimens: an experience in the University Malaya Medical Centre. *The Malaysian journal of pathology*, **31**(2), pp. 93-97.

FALZON, D., GANDHI, N., MIGLIORI, G.B., SOTGIU, G., COX, H.S., HOLTZ, T.H., HOLLM-DELGADO, M.G., KESHAVJEE, S., DERIEMER, K., CENTIS, R., D'AMBROSIO, L., LANGE, C.G., BAUER, M., MENZIES, D. and COLLABORATIVE GROUP FOR META-ANALYSIS OF INDIVIDUAL PATIENT DATA IN MDR-TB, 2013. Resistance to fluoroquinolones and second-line injectable drugs: impact on multidrug-resistant TB outcomes. *The European respiratory journal*, **42**(1), pp. 156-168.

FILLIOL, I., MOTIWALA, A.S., CAVATORE, M., QI, W., HAZBÓN, M.H., BOBADILLA DEL VALLE, M., FYFE, J., GARCÍA-GARCÍA, L., RASTOGI, N., SOLA, C., ZOZIO, T., GUERRERO, M.I., LEÓN, C.I., CRABTREE, J., ANGIUOLI, S., EISENACH, K.D., DURMAZ, R., JOLOBA, M.L., RENDÓN, A., SIFUENTES-OSORNIO, J., PONCE DE LEÓN, A., CAVE, M.D., FLEISCHMANN, R., WHITTAM, T.S. and ALLAND, D., 2006. Global Phylogeny of Mycobacterium tuberculosis Based on Single Nucleotide Polymorphism (SNP) Analysis: Insights into Tuberculosis Evolution, Phylogenetic Accuracy of Other DNA Fingerprinting Systems, and

Recommendations for a Minimal Standard SNP Set. *Journal of Bacteriology*, **188**(2), pp. 759-772.

FIRDESSA, R., BERG, S., HAILU, E., SCHELLING, E., GUMI, B., ERENDO, G., GADISA, E., KIROS, T., HABTAMU, M., HUSSEIN, J., ZINSSTAG, J., ROBERTSON, B.D., AMENI, G., LOHAN, A.J., LOFTUS, B., COMAS, I., GAGNEUX, S., TSCHOPP, R., YAMUAH, L., HEWINSON, G., GORDON, S.V., YOUNG, D.B. and ASEFFA, A., 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerging infectious diseases*, **19**(3), pp. 460-463.

FISHBEIN, S., VAN WYK, N., WARREN, R.M. and SAMPSON, S.L., 2015. Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. *Molecular microbiology*, **96**(5), pp. 901-916.

FLANDROIS, J., LINA, G. and DUMITRESCU, O., 2014. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. *BMC Bioinformatics*, **15**(1), pp. 107.

FLEISCHMANN, R.D., ALLAND, D., EISEN, J.A., CARPENTER, L., WHITE, O., PETERSON, J., DEBOY, R., DODSON, R., GWINN, M., HAFT, D., HICKEY, E., KOLONAY, J.F., NELSON, W.C., UMayAM, L.A., ERMOLAEVA, M., SALZBERG, S.L., DELCHER, A., UTTERBACK, T., WEIDMAN, J., KHOURI, H., GILL, J., MIKULA, A., BISHAI, W., JACOBS, J., W.R., VENTER, J.C. and FRASER, C.M., 2002. Whole-Genome Comparison of Mycobacterium tuberculosis Clinical and Laboratory Strains. *Journal of Bacteriology*, **184**(19), pp. 5479-5490.

FLORES, L., VAN, T., NARAYANAN, S., DERIEMER, K., KATO-MAEDA, M. and GAGNEUX, S., 2007. Large sequence polymorphisms classify Mycobacterium tuberculosis strains with ancestral spoligotyping patterns. *Journal of clinical microbiology*, **45**(10), pp. 3393-3395.

FLUSBERG, B.A., WEBSTER, D.R., LEE, J.H., TRAVERS, K.J., OLIVARES, E.C., CLARK, T.A., KORLACH, J. and TURNER, S.W., 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature methods*, **7**(6), pp. 461-465.

FLYNN, J.L., 2004. Mutual attraction: does it benefit the host or the bug? *Nature Immunology*, **5**(8), pp. 778-779.

FLYNN, J.L., CAPUANO, S.V., CROIX, D., PAWAR, S., MYERS, A., ZINOVIK, A. and KLEIN, E., 2003. Non-human primates: a model for tuberculosis research. *Tuberculosis*, **83**(1-3), pp. 116-118.

FOX, W., ELLARD, G.A. and MITCHISON, D.A., 1999. Studies on the treatment of tuberculosis undertaken by the British Medical Research Council tuberculosis units, 1946-1986, with relevant subsequent publications. *The international journal of tuberculosis and*

lung disease : the official journal of the International Union against Tuberculosis and Lung Disease, **3**(10 Suppl 2), pp. S231-79.

FRESQUET-WOLF, C., HAAS, J., WILDEMANN, B. and STORCH-HAGENLOCHER, B., 1998. Value of polymerase chain reaction (PCR) for diagnosis of tuberculoid meningitis. *Der Nervenarzt*, **69**(6), pp. 502-506.

FRIDMAN, O., GOLDBERG, A., RONIN, I., SHORESH, N. and BALABAN, N.Q., 2014. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature*, **513**(7518), pp. 418-421.

FRITZ, C., MAASS, S., KREFT, A. and BANGE, F.C., 2002. Dependence of *Mycobacterium bovis* BCG on anaerobic nitrate reductase for persistence is tissue specific. *Infection and immunity*, **70**(1), pp. 286-291.

FURUTA, Y., NAMBA-FUKUYO, H., SHIBATA, T.F., NISHIYAMA, T., SHIGENOBU, S., SUZUKI, Y., SUGANO, S., HASEBE, M. and KOBAYASHI, I., 2014. Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS genetics*, **10**(4), pp. e1004272.

GAGNEUX, S., 2012. Host-pathogen coevolution in human tuberculosis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**(1590), pp. 850-859.

GAGNEUX, S., DERIEMER, K., VAN, T., KATO-MAEDA, M., DE JONG, B.C., NARAYANAN, S., NICOL, M., NIEMANN, S., KREMER, K., GUTIERREZ, M.C., HILTY, M., HOPEWELL, P.C. and SMALL, P.M., 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(8), pp. 2869-2873.

GAGNEUX, S. and SMALL, P.M., 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *The Lancet. Infectious diseases*, **7**(5), pp. 328-337.

GARCIA, B.J., LOXTON, A.G., DOLGANOV, G.M., VAN, T.T., DAVIS, J.L., DE JONG, B.C., VOSKUIL, M.I., LEACH, S.M., SCHOOLNIK, G.K., WALZL, G., STRONG, M. and WALTER, N.D., 2016. *Sputum is a surrogate for bronchoalveolar lavage for monitoring Mycobacterium tuberculosis transcriptional profiles in TB patients.*

GARTON, N.J., WADDELL, S.J., SHERRATT, A.L., LEE, S.M., SMITH, R.J., SENNER, C., HINDS, J., RAJAKUMAR, K., ADEGBOLA, R.A., BESRA, G.S., BUTCHER, P.D. and BARER, M.R., 2008. Cytological and transcript analyses reveal fat and lazy persister-like bacilli in tuberculous sputum. *PLoS medicine*, **5**(4), pp. e75.

GARZELLI, C. and RINDI, L., 2012. Molecular epidemiological approaches to study the epidemiology of tuberculosis in low-incidence settings receiving immigrants. *Infection*,

genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, **12**(4), pp. 610-618.

GEFEN, O. and BALABAN, N.Q., 2009. The importance of being persistent: heterogeneity of bacterial populations under antibiotic stress. *FEMS microbiology reviews*, **33**(4), pp. 704-717.

GEORGIADES, K. and RAOULT, D., 2011. Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PloS one*, **6**(3), pp. e17962.

GERDES, K. and MAISONNEUVE, E., 2012. Bacterial persistence and toxin-antitoxin loci. *Annual Review of Microbiology*, **66**, pp. 103-123.

GERMAIN, E., CASTRO-ROA, D., ZENKIN, N. and GERDES, K., 2013. Molecular mechanism of bacterial persistence by HipA. *Molecular cell*, **52**(2), pp. 248-254.

GEURRA-ASSUNCAO, J., CRAMPIN, A., HOUBEN, R., MZEMBE, T., MALLAD, K., COLL, F., KHAN, P., BANDA, L., CHIWAYA, A., PERREIRA, R., MCNERNEY, R., FINE, P.E., PARKHILL, J., CLARK, T. and GLYNN, J., 2015. Large-scale whole genome sequencing of *M.tuberculosis* provides insights into transmission in a high prevalence area. *Elife*, **3**(4),.

GHODBANE, R., RAOULT, D. and DRANCOURT, M., 2014. Dramatic reduction of culture time of *Mycobacterium tuberculosis*. *Scientific reports*, **4**, pp. 4236.

GILLESPIE, S.H., CROOK, A.M., MCHUGH, T.D., MENDEL, C.M., MEREDITH, S.K., MURRAY, S.R., PAPPAS, F., PHILLIPS, P.P.J. and NUNN, A.J., 2014. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. *N Engl J Med*, **371**(17), pp. 1577-1587.

GLICKMAN, M.S., COX, J.S. and JACOBS, W.R., Jr, 2000. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Molecular cell*, **5**(4), pp. 717-727.

GLYNN, J.R., CRAMPIN, A.C., TRAORE, H., YATES, M.D., MWAUNGULU, F.D., NGWIRA, B.M., CHAGULUKA, S.D., MWAFULIRWA, D.T., FLOYD, S., MURPHY, C., DROBNIEWSKI, F.A. and FINE, P.E., 2005. *Mycobacterium tuberculosis* Beijing genotype, northern Malawi. *Emerging infectious diseases*, **11**(1), pp. 150-153.

GLYNN, J., ALGHAMDI, S., MCNERNEY, R., NDHLOVU, R., MUNTHALI, L., HOUBEN, R., FINE, P., FRENCH, N. and CRAMPIN, A., 2010. Changes in *Mycobacterium tuberculosis* genotype families over 20 years in a population-based study in Northern Malawi. *PloS one*, **5**(8),.

GOMEZ, J.E. and MCKINNEY, J.D., 2004. M. tuberculosis persistence, latency, and drug tolerance. *Tuberculosis (Edinburgh, Scotland)*, **84**(1-2), pp. 29-44.

GOSS, W.A., DEITZ, W.H. and COOK, T.M., 1965. Mechanism of Action of Nalidixic Acid on Escherichia Coli.Ii. Inhibition of Deoxyribonucleic Acid Synthesis. *Journal of Bacteriology*, **89**, pp. 1068-1074.

GOULD, T.A., VAN DE LANGEMHEEN, H., MUÑOZ-ELÍAS, E.J., MCKINNEY, J.D. and SACCHETTINI, J.C., 2006. Dual role of isocitrate lyase 1 in the glyoxylate and methylcitrate cycles in Mycobacterium tuberculosis. *Molecular microbiology*, **61**(4), pp. 940-947.

GRECO, S., GIRARDI, E., NAVARRA, A. and SALTINI, C., 2006. Current evidence on diagnostic accuracy of commercially based nucleic acid amplification tests for the diagnosis of pulmonary tuberculosis. *Thorax*, **61**(9), pp. 783-790.

GÜNTHER, G., 2014. Multidrug-resistant and extensively drug-resistant tuberculosis: a review of current concepts and future challenges. *Clinical Medicine*, **14**(3), pp. 279-285.

GUPTA-WRIGHT, A., PETERS, A., FLACH, C. and LAWN, S., 2016. Detection of lipoarabinomannan(LAM) in urine as an independent predictor of mortality risk in patients receiving treatment for HIV associated tuberculosis in sub-Saharan Africa: a systematic review and meta-analysis. *BMC medicine*, **14**(53),.

GUTIERREZ, M.C., BRISSE, S., BROSCHE, R., FABRE, M., OMAIS, B., MARMIESSE, M., SUPPLY, P. and VINCENT, V., 2005. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS pathogens*, **1**(1), pp. e5.

HARRIES, A.D., NYIRENDA, T.E., GODFREY-FAUSSETT, P. and SALANIPONI, F.M., 2003. Defining and assessing the maximum number of visits patients should make to a health facility to obtain a diagnosis of pulmonary tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, **7**(10), pp. 953-958.

HAYMAN, J., 1984. Mycobacterium ulcerans: an infection from Jurassic time? *Lancet*, **2**, pp. 1015-1016.

HAZBÓN, M.H., BRIMACOMBE, M., BOBADILLA DEL VALLE, M., CAVATORE, M., GUERRERO, M.I., VARMA-BASIL, M., BILLMAN-JACOB, H., LAVENDER, C., FYFE, J., GARCÍA-GARCÍA, L., LEÓN, C.I., BOSE, M., CHAVES, F., MURRAY, M., EISENACH, K.D., SIFUENTES-OSORNIO, J., CAVE, M.D., PONCE DE LEÓN, A. and ALLAND, D., 2006. Population Genetics Study of Isoniazid Resistance Mutations and Evolution of Multidrug-Resistant Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*, **50**(8), pp. 2640-2649.

HELAINÉ, S. and KUGELBERG, E., 2014. Bacterial persisters: formation, eradication, and experimental systems. *Trends in microbiology*, **22**(7), pp. 417-424.

HELB, D., JONES, M., STORY, E., BOEHME, C., WALLACE, E., HO, K., KOP, J., OWENS, M.R., RODGERS, R., BANADA, P., SAFI, H., BLAKEMORE, R., LAN, N.T.N., JONES-LÓPEZ, E.C., LEVI, M., BURDAY, M., AYAKAKA, I., MUGERWA, R.D., MCMILLAN, B., WINN-DEEN, E., CHRISTEL, L., DAILEY, P., PERKINS, M.D., PERSING, D.H. and ALLAND, D., 2010. Rapid Detection of Mycobacterium tuberculosis and Rifampin Resistance by Use of On-Demand, Near-Patient Technology. *Journal of clinical microbiology*, **48**(1), pp. 229-237.

HERSHBERG, R., LIPATOV, M., SMALL, P.M., SHEFFER, H., NIEMANN, S., HOMOLKA, S., ROACH, J.C., KREMER, K., PETROV, D.A., FELDMAN, M.W. and GAGNEUX, S., 2008. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS biology*, **6**(12), pp. e311.

HOBBY, G.L., MEYER, K. and CHAFFEE, E., 1942. Observations on the Mechanism of Action of Penicillin. *Experimental biology and medicine*, **50**(2), pp. 281-285.

HOMOLKA, S., PROJAHN, M., FEUERRIEGEL, S., UBBEN, T., DIEL, R., NUBEL, U. and NIEMANN, S., 2012. High resolution discrimination of clinical Mycobacterium tuberculosis complex strains based on single nucleotide polymorphisms. *PloS one*, **7**(7), pp. e39855.

HONEYBORNE, I., MCHUGH, T.D., KUITTINEN, I., CICHONSKA, A., EVANGELOPOULOS, D., RONACHER, K., VAN HELDEN, P.D., GILLESPIE, S.H., FERNANDEZ-REYES, D., WALZL, G., ROUSU, J., BUTCHER, P.D. and WADDELL, S.J., 2016. Profiling persistent tubercule bacilli from patient sputa during therapy predicts early drug efficacy. *BMC medicine*, **14**(1), pp. 68-016-0609-3.

HONORE-BOUAKLINE, S., VINCENSINI, J.P., GIACUZZO, V., LAGRANGE, P.H. and HERRMANN, J.L., 2003. Rapid diagnosis of extrapulmonary tuberculosis by PCR: impact of sample preparation and DNA extraction. *Journal of clinical microbiology*, **41**(6), pp. 2323-2329.

HU, Y., MOVAHEDZADEH, F., STOKER, N.G. and COATES, A.R., 2006. Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo. *Infection and immunity*, **74**(2), pp. 861-868.

HUGGETT, J.F., MCHUGH, T.D. and ZUMLA, A., 2003. Tuberculosis: amplification-based clinical diagnostic techniques. *The international journal of biochemistry & cell biology*, **35**(10), pp. 1407-1412.

HUTTER, B. and DICK, T., 2000. Analysis of the dormancy-inducible narK2 promoter in Mycobacterium bovis BCG. *FEMS microbiology letters*, **188**(2), pp. 141-146.

ILLUMINA, . HiSeq 2500 specifications. Available: http://www.illumina.com/systems/hiseq_2500_1500/performance_specifications.html [April 18, 2016].

IWAMOTO, T., SONOBE, T. and HAYASHI, K., 2003. Loop-Mediated Isothermal Amplification for Direct Detection of Mycobacterium tuberculosis Complex, M. avium, and M. intracellulare in Sputum Samples. *Journal of clinical microbiology*, **41**(6), pp. 2616-2622.

JAGIELSKI, T., VAN INGEN, J., RASTOGI, N., DZIADEK, J., MAZUR, P.K. and BIELECKI, J., 2014. Current methods in the molecular typing of Mycobacterium tuberculosis and other mycobacteria. *BioMed research international*, **2014**, pp. 645802.

JAWAHAR, M.S., BANUREKHA, V.V., PARAMASIVAN, C.N., RAHMAN, F., RAMACHANDRAN, R., VENKATESAN, P., BALASUBRAMANIAN, R., SELVAKUMAR, N., PONNURAJA, C., ILIAYAS, A.S., GANGADEVI, N.P., RAMAN, B., BASKARAN, D., KUMAR, S.R., KUMAR, M.M., MOHAN, V., GANAPATHY, S., KUMAR, V., SHANMUGAM, G., CHARLES, N., SAKTHIVEL, M.R., JAGANNATH, K., CHANDRASEKAR, C., PARTHASARATHY, R.T. and NARAYANAN, P.R., 2013. Randomized clinical trial of thrice-weekly 4-month moxifloxacin or gatifloxacin containing regimens in the treatment of new sputum positive pulmonary tuberculosis patients. *PloS one*, **8**(7), pp. e67030.

JELTSCH, A., 2002. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem : a European journal of chemical biology*, **3**(4), pp. 274-293.

JOHNSON, J.L., HADAD, D.J., BOOM, W.H., DALEY, C.L., PELOQUIN, C.A., EISENACH, K.D., JANKUS, D.D., DEBANNE, S.M., CHARLEBOIS, E.D., MACIEL, E., PALACI, M. and DIETZE, R., 2006. Early and extended early bactericidal activity of levofloxacin, gatifloxacin and moxifloxacin in pulmonary tuberculosis. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, **10**(6), pp. 605-612.

KÅHRSTRÖM, C.T., 2013. Bacterial Physiology A Persistent Magic Spot. *Nat.Rev.Microbiol.*, **11**, pp. 739.

KAISER FAMILY FOUNDATION, . Available: <http://kff.org/global-health-policy/slide/top-10-countries-new-tuberculosis-tb-cases-per-10000-population/> [February 19, 2017].

KAMERBEEK, J., SCHOULS, L., KOLK, A., VAN AGTERVELD, M., VAN SOOLINGEN, D., KUIJPER, S., BUNSCHOTEN, A., MOLHUIZEN, H., SHAW, R., GOYAL, M. and VAN EMBDEN, J., 1997. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *Journal of clinical microbiology*, **35**(4), pp. 907-914.

KANG, P.B., AZAD, A.K., TORRELLES, J.B., KAUFMAN, T.M., BEHARKA, A., TIBESAR, E., DESJARDIN, L.E. and SCHLESINGER, L.S., 2005. *The human macrophage mannose receptor directs Mycobacterium tuberculosis lipoarabinomannan-mediated phagosome biogenesis.*

KAPOOR, N., PAWAR, S., SIRAKOVA, T.D., DEB, C., WARREN, W.L. and KOLATTUKUDY, P.E., 2013. Human granuloma in vitro model, for TB dormancy and resuscitation. *PloS one*, **8**(1), pp. e53657.

KAPUR, V., WHITTAN, T.S. and MUSSER, J.S., 1994. Is Mycobacterium Tuberculosis 15000 years old. *Journal of Infectious Disease*, **170**, pp. 1348-1349.

KASPY, I., ROTEM, E., WEISS, N., RONIN, I., BALABAN, N.Q. and GLASER, G., 2013. HipA-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase. *Nature communications*, **4**, pp. 3001.

KATOH, K. and STANDLEY, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**(4), pp. 772-780.

KAUSHAL, D., SCHROEDER, B.G., TYAGI, S., YOSHIMATSU, T., SCOTT, C., KO, C., CARPENTER, L., MEHROTRA, J., MANABE, Y., FLEISCHMANN, R.D. and BISHAI, W.R., 2002. Reduced immunopathology and mortality despite tissue persistence in a Mycobacterium tuberculosis mutant lacking alternative sigma factor, SigH. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, pp. 8330-8335.

KAY, G.L., SERGEANT, M.J., GIUFFRA, V., BANDIERA, P., MILANESE, M., BRAMANTI, B., BIANUCCI, R. and PALLEEN, M.J., 2014. Recovery of a Medieval Brucella melitensis Genome Using Shotgun Metagenomics. *mBio*, **5**(4),.

KEREN, I., MINAMI, S., RUBIN, E. and LEWIS, K., 2011. Characterization and Transcriptome Analysis of Mycobacterium tuberculosis Persisters. *mBio*, **2**(3),.

KEREN, I., SHAH, D., SPOERING, A., KALDALU, N. and LEWIS, K., 2004. Specialized Persister Cells and the Mechanism of Multidrug Tolerance in Escherichia coli. *Journal of Bacteriology*, **186**(24), pp. 8172-8180.

KORCH, S.B. and HILL, T.M., 2006. Ectopic overexpression of wild-type and mutant hipA genes in Escherichia coli: effects on macromolecular synthesis and persister formation. *Journal of Bacteriology*, **188**(11), pp. 3826-3836.

KORCH, S.B., MALHOTRA, V., CONTRERAS, H. and CLARK-CURTISS, J.E., 2015. The Mycobacterium tuberculosis relBE toxin:antitoxin genes are stress-responsive modules that regulate growth through translation inhibition. *Journal of microbiology (Seoul, Korea)*, **53**(11), pp. 783-795.

- KORCH, S.B., HENDERSON, T.A. and HILL, T.M., 2003. Characterization of the hipA7 allele of Escherichia coli and evidence that high persistence is governed by (p)ppGpp synthesis. *Molecular microbiology*, **50**(4), pp. 1199-1213.
- KORLACH, J. and TURNER, S.W., 2012. Going beyond five bases in DNA sequencing. *Current opinion in structural biology*, **22**(3), pp. 251-261.
- KOSER, C.U., BRYANT, J.M., BECQ, J., TOROK, M.E., ELLINGTON, M.J., MARTIRENOM, M.A., CARMICHAEL, A.J., PARKHILL, J., SMITH, G.P. and PEACOCK, S.J., 2013. Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis. *The New England journal of medicine*, **369**(3), pp. 290-292.
- KUMAR, R. and BHAKUNI, V., 2008. Mycobacterium tuberculosis isocitrate lyase (MtbIcl): role of divalent cations in modulation of functional and structural properties. *Proteins*, **72**(3), pp. 892-900.
- LAWN, S.D. and GUPTA-WRIGHT, A., 2016. Detection of lipoarabinomannan (LAM) in urine is indicative of disseminated TB with renal involvement in patients living with HIV and advanced immunodeficiency: evidence and implications. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **110**(3), pp. 180-185.
- LECHARTIER, B., RYBNIKER, J., ZUMLA, A. and COLE, S.T., 2014. Tuberculosis drug discovery in the post-post-genomic era. *EMBO molecular medicine*, **6**(2), pp. 158-168.
- LEISTIKOW, R.L., MORTON, R.A., BARTEK, I.L., FRIMPONG, I., WAGNER, K. and VOSKUIL, M.I., 2010. The Mycobacterium tuberculosis DosR regulon assists in metabolic homeostasis and enables rapid recovery from nonrespiring dormancy. *Journal of Bacteriology*, **192**(6), pp. 1662-1670.
- LEVIN-REISMAN, I., RONIN, I., GEFEN, O., BRANISS, I., SHORESH, N. and BALABAN, N.Q., 2017. Antibiotic tolerance facilitates the evolution of resistance. *Science (New York, N.Y.)*, **355**(6327), pp. 826-830.
- LEWIS, K., 2010. Persister cells. *Annual Review of Microbiology*, **64**, pp. 357-372.
- LI, Z., KELLEY, C., COLLINS, F., ROUSE, D. and MORRIS, S., 1998. Expression of katG in Mycobacterium tuberculosis Is Associated with Its Growth and Persistence in Mice and Guinea Pigs. *Journal of Infectious Diseases*, **177**(4), pp. 1030-1035.
- LICKFELDT, D.W., HOFMANN, N.E., JONES, J.D., HAMBLIN, A.M. and VOIGT, T.B., 2002. Comparing Three DNA Extraction Procedures for Cost, Efficiency and DNA Yield. *HortScience*, **37**(5), pp. 822-825.

LIM, H.N., LEE, Y. and HUSSEIN, R., 2011. Fundamental relationship between operon organization and gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(26), pp. 10626-10631.

LIU, G., OU, H.Y., WANG, T., LI, L., TAN, H., ZHOU, X., RAJAKUMAR, K., DENG, Z. and HE, X., 2010. Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS genetics*, **6**(12), pp. e1001253.

LONGHORN VACCINES, 2017-last update, PrimeXtract. Available: <http://www.lhnvd.com/page/primextract> [01/20, .

LONNROTH, K., CASTRO, K.G., CHAKAYA, J.M., CHAUHAN, L.S., FLOYD, K., GLAZIOU, P. and RAVIGLIONE, M.C., 2010. Tuberculosis control and elimination 2010-50: cure, care, and social development. *Lancet*, **375**(9728), pp. 1814-1829.

LU, J., LIU, M., WANG, Y., PANG, Y. and ZHAO, Z., 2014. Mechanisms of fluoroquinolone monoresistance in *Mycobacterium tuberculosis*. *FEMS microbiology letters*, **353**(1), pp. 40-48.

MACHADO, D., PERDIGÃO, J., RAMOS, J., COUTO, I., PORTUGAL, I., RITTER, C., BOETTGER, E.C. and VIVEIROS, M., 2013. High-level resistance to isoniazid and ethionamide in multidrug-resistant *Mycobacterium tuberculosis* of the Lisboa family is associated with *inhA* double mutations. *Journal of Antimicrobial Chemotherapy*, **68**(8), pp. 1728-1732.

MAISONNEUVE, E., CASTRO-CAMARGO, M. and GERDES, K., 2013. (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell*, **154**(5), pp. 1140-1150.

MAISONNEUVE, E., SHAKESPEARE, L.J., JORGENSEN, M.G. and GERDES, K., 2011. Bacterial persistence by RNA endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(32), pp. 13206-13211.

MAISONNEUVE, E. and GERDES, K., 2014. Molecular Mechanisms Underlying Bacterial Persisters. *Cell*, **157**(3), pp. 539-548.

MARINO, S., EL-KEBIR, M. and KIRSCHNER, D., 2011. A hybrid multi-compartment model of granuloma formation and T cell priming in Tuberculosis. *Journal of theoretical biology*, **280**(1), pp. 50-62.

MARMIESSE, M., BRODIN, P., BUCHRIESER, C., GUTIERREZ, C., SIMOES, N., VINCENT, V., GLASER, P., COLE, S.T. and BROSCHE, R., 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology (Reading, England)*, **150**(Pt 2), pp. 483-496.

MARURI, F., STERLING, T.R., KAIGA, A.W., BLACKMAN, A., VAN DER HEIJDEN, Y.F., MAYER, C., CAMBAU, E. and AUBRY, A., 2012. A systematic review of gyrase mutations associated with fluoroquinolone-resistant *Mycobacterium tuberculosis* and a proposed gyrase numbering system. *Journal of Antimicrobial Chemotherapy*, **67**(4), pp. 819-831.

MASUDA, H., TAN, Q., AWANO, N., WU, K.P. and INOUE, M., 2012. YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Molecular microbiology*, **84**(5), pp. 979-989.

MATHEMA, B., KUREPINA, N.E., BIFANI, P.J. and KREISWIRTH, B.N., 2006. Molecular epidemiology of tuberculosis: current insights. *Clinical microbiology reviews*, **19**(4), pp. 658-685.

MATTEELLI, A., ROGGI, A. and CARVALHO, A.C., 2014. Extensively drug-resistant tuberculosis: epidemiology and management. *Clinical epidemiology*, **6**, pp. 111-118.

MCCLURE, W.R. and CECH, C.L., 1978. On the mechanism of rifampicin inhibition of RNA synthesis. *Journal of Biological Chemistry*, **253**(24), pp. 8949-8956.

MCCUNE, R.M., Jr, MCDERMOTT, W. and TOMPSETT, R., 1956. The fate of *Mycobacterium tuberculosis* in mouse tissues as determined by the microbial enumeration technique. II. The conversion of tuberculous infection to the latent state by the administration of pyrazinamide and a companion drug. *The Journal of experimental medicine*, **104**(5), pp. 763-802.

MCCUNE, R.M., Jr and TOMPSETT, R., 1956. Fate of *Mycobacterium tuberculosis* in mouse tissues as determined by the microbial enumeration technique. I. The persistence of drug-susceptible tubercle bacilli in the tissues despite prolonged antimicrobial therapy. *The Journal of experimental medicine*, **104**(5), pp. 737-762.

MCCUNE, R.M., FELDMANN, F.M. and MCDERMOTT, W., 1966. MICROBIAL PERSISTENCE: II. CHARACTERISTICS OF THE STERILE STATE OF TUBERCLE BACILLI. *The Journal of experimental medicine*, **123**(3), pp. 469-486.

MCDERMOTT, W., 1958. Microbial Persistence. *Yale Journal of Biology and Medicine*, **30**(4), pp. 257-291.

MCKINNEY, J.D., HONER ZU BENTRUP, K., MUNOZ-ELIAS, E.J., MICZAK, A., CHEN, B., CHAN, W.T., SWENSON, D., SACCHETTINI, J.C., JACOBS, W.R., Jr and RUSSELL, D.G., 2000. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature*, **406**(6797), pp. 735-738.

MCMURRAY, D.N., 2001. Disease model: pulmonary tuberculosis. *Trends in molecular medicine*, **7**(3), pp. 135-137.

MERKER, M., KOHL, T.A., ROETZER, A., TRUEBE, L., RICHTER, E., RUSCHGERDES, S., FATTORINI, L., OGGIONI, M.R., COX, H., VARAINE, F. and NIEMANN, S., 2013. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients. *PloS one*, **8**(12), pp. e82551.

MERLE, C.S., FIELDING, K., SOW, O.B., GNINAFON, M., LO, M.B., MTHIYANE, T., ODHIAMBO, J., AMUKOYE, E., BAH, B., KASSA, F., N'DIAYE, A., RUSTOMJEE, R., DE JONG, B.C., HORTON, J., PERRONNE, C., SISMANIDIS, C., LAPUJADE, O., OLLIARO, P.L. and LIENHARDT, C., 2014. A Four-Month Gatifloxacin-Containing Regimen for Treating Tuberculosis. *N Engl J Med*, **371**(17), pp. 1588-1598.

METCALFE, J.Z., EVERETT, C.K., STEINGART, K.R., CATTAMANCHI, A., HUANG, L., HOPEWELL, P.C. and PAI, M., 2011. Interferon-gamma release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *The Journal of infectious diseases*, **204 Suppl 4**, pp. S1120-9.

MIDDELKOOP, K., BEKKER, L., MATHEMA, B., SHASHKINA, E., KUREPINA, N., WHITELAW, A., FALLOWS, D., MORROW, C., KREISWIRTH, B., KAPLAN, G. and ROBIN WOOD, G., 2009. Molecular Epidemiology of Mycobacterium tuberculosis in a South African Community with High HIV Prevalence. *Journal of Infectious Diseases*, **200**(8), pp. 1207-1211.

MITCHISON, D. and DAVIES, G., 2012. The chemotherapy of tuberculosis: past, present and future. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, **16**(6), pp. 724-732.

MITCHISON, D.A., 2004. The search for new sterilizing anti-tuberculosis drugs. *Frontiers in bioscience : a journal and virtual library*, **9**, pp. 1059-1072.

MITCHISON, D.A., 1985. The action of antituberculosis drugs in short-course chemotherapy. *Tubercle*, **66**(3), pp. 219-225.

MITCHISON, D.A., 1979. Basic mechanisms of chemotherapy. *Chest*, **76**(6), pp. 771-781.

MOLICOTTI, P., BUA, A. and ZANETTI, S., 2014. Cost-effectiveness in the diagnosis of tuberculosis: choices in developing countries. *Journal of infection in developing countries*, **8**(1), pp. 24-38.

MONACK, D.M., MUELLER, A. and FALKOW, S., 2004. Persistent bacterial infections: the interface of the pathogen and the host immune system. *Nature reviews.Microbiology*, **2**(9), pp. 747-765.

- MOYED, H.S. and BERTRAND, K.P., 1983. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of Bacteriology*, **155**(2), pp. 768-775.
- MOYED, H.S. and BRODERICK, S.H., 1986. Molecular cloning and expression of hipA, a gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of Bacteriology*, **166**(2), pp. 399-403.
- MULENGA, C., SHAMPUTA, I.C., MWAKAZANGA, D., KAPATA, N., PORTAELS, F. and RIGOUTS, L., 2010. Diversity of Mycobacterium tuberculosis genotypes circulating in Ndola, Zambia. *BMC infectious diseases*, **10**, pp. 177-2334-10-177.
- MÜLLER, R., ROBERTS, C.A. and BROWN, T.A., 2014. Genotyping of ancient Mycobacterium tuberculosis strains reveals historic genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1781),.
- MURPHY, H.N., STEWART, G.R., MISCHENKO, V.V., APT, A.S., HARRIS, R., MCALISTER, M.S.B., DRISCOLL, P.C., YOUNG, D.B. and ROBERTSON, B.D., 2005. The OtsAB Pathway Is Essential for Trehalose Biosynthesis in Mycobacterium tuberculosis. *Journal of Biological Chemistry*, **280**(15), pp. 14524-14529.
- MURRAY, I.A., CLARK, T.A., MORGAN, R.D., BOITANO, M., ANTON, B.P., LUONG, K., FOMENKOV, A., TURNER, S.W., KORLACH, J. and ROBERTS, R.J., 2012. The methylomes of six bacteria. *Nucleic acids research*, **40**(22), pp. 11450-11462.
- MUSSER, J.M., 1995. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clinical microbiology reviews*, **8**(4), pp. 496-514.
- MUSSER, J.M., KAPUR, V., WILLIAMS, D.L., KREISWIRTH, B.N., VAN SOOLINGEN, D. and VAN EMBDEN, J.D.A., 1996. Characterization of the Catalase-Peroxidase Gene (katG) and inhA Locus in Isoniazid-Resistant and -Susceptible Strains of Mycobacterium tuberculosis by Automated DNA Sequencing: Restricted Array of Mutations Associated with Drug Resistance. *Journal of Infectious Diseases*, **173**(1), pp. 196-202.
- NAIDU, A.G., 2014. *Computational Characterization of DNA Methylomes in Mycobacterium tuberculosis Beijing hyper- and hypo-virulent strains*. PhD edn. Capetown: University of Western Cape.
- NARAYANAN, S., GAGNEUX, S., HARI, L., TSOLAKI, A.G., RAJASEKHAR, S., NARAYANAN, P.R., SMALL, P.M., HOLMES, S. and DERIEMER, K., 2008. Genomic interrogation of ancestral Mycobacterium tuberculosis from south India. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, **8**(4), pp. 474-483.

- NEWTON-FOOT, M. and GEY VAN PITTIUS, N.C., 2013. The complex architecture of mycobacterial promoters. *Tuberculosis (Edinburgh, Scotland)*, **93**(1), pp. 60-74.
- NEYROLLES, O., 2014. Mycobacteria and the greasy macrophage: getting fat and frustrated. *Infection and immunity*, **82**(2), pp. 472-475.
- NLIWASA, M., MACPHERSON, P., CHISALA, P., KAMDOLOZI, M., KHUNDI, M., KASWASWA, K., MWAPASA, M., MSEFULA, C., SOHN, H., FLACH, C. and CORBETT, E.L., 2016. The Sensitivity and Specificity of Loop-Mediated Isothermal Amplification (LAMP) Assay for Tuberculosis Diagnosis in Adults with Chronic Cough in Malawi. *PloS one*, **11**(5), pp. e0155101.
- NOLTE, F.S., METCHOCK, B., MCGOWAN, J.E., EDWARDS, A., OKWUMABUA, O., THURMOND, C., MITCHELL, P.S., PLIKAYTIS, B. and SHINNICK, T., 1993. Direct detection of Mycobacterium tuberculosis in sputum by polymerase chain reaction and DNA hybridization. *Journal of clinical microbiology*, **31**(7), pp. 1777-1782.
- NYENDAK, M.R., LEWINSOHN, D.M. and LEWINSOHN, D.A., 2009. New Diagnostic Methods for Tuberculosis. *Current Opinions in Infectious Diseases*, **22**(2), pp. 174-182.
- PACIFIC BIOSCIENCES, December 18, 2015-last update, Procedures and Checklist-10kb to 20kb Template Preparation and Sequencing with Low (100ng) input DNA. Available: <http://www.pacb.com/documentation/procedure-checklist-10-kb-20-kb-template-preparation-sequencing-low-100-ng-input-dna/> [June 17, 2016].
- PALOMINO, J.C. and ANANDI, M., 2014. Drug Resistance Mechanisms in Mycobacterium tuberculosis. *Antibiotics*, **3**, pp. 317-340.
- PALOMINO, J.C., MARTIN, A., VON GROLL, A. and PORTAELS, F., 2008. Rapid culture-based methods for drug-resistance detection in Mycobacterium tuberculosis. *Journal of microbiological methods*, **75**(2), pp. 161-166.
- PAN, S., GU, B., WANG, H., YAN, Z., WANG, P., PEI, H., XIE, W., CHEN, D. and LIU, G., 2013. Comparison of four DNA extraction methods for detecting *Mycobacterium tuberculosis* by real-time PCR and its clinical application in pulmonary tuberculosis. *Journal of Thoracic Disease*, **5**(3),.
- PANDEY, D.P. and GERDES, K., 2005. Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic acids research*, **33**(3), pp. 966-976.
- PANTHER, 2018-last update, GO Enrichment Analysis. Available: <http://geneontology.org/page/go-enrichment-analysis> [June 20, 2018].

- PARAMASIVAN, C.N., SULOCHANA, S., KUBENDIRAN, G., VENKATESAN, P. and MITCHISON, D.A., 2005. Bactericidal Action of Gatifloxacin, Rifampin, and Isoniazid on Logarithmic- and Stationary-Phase Cultures of *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, **49**(2), pp. 627-631.
- PARISH, T., SMITH, D.A., KENDALL, S., CASALI, N., BANCROFT, G.J. and STOKER, N.G., 2003. Deletion of two-component regulatory systems increases the virulence of *Mycobacterium tuberculosis*. *Infection and immunity*, **71**(3), pp. 1134-1140.
- PARK, H.D., GUINN, K.M., HARRELL, M.I., LIAO, R., VOSKUIL, M.I., TOMPA, M., SCHOOLNIK, G.K. and SHERMAN, D.R., 2003. Rv3133c/dosR is a transcription factor that mediates the hypoxic response of *Mycobacterium tuberculosis*. *Molecular microbiology*, **48**(3), pp. 833-843.
- PARK, M.K., MYERS, R.A. and MARZELLA, L., 1992. Oxygen tensions and infections: modulation of microbial growth, activity of antimicrobial agents, and immunologic responses. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, **14**(3), pp. 720-740.
- PASCA, M.R., GUGLIERAME, P., ARCESI, F., BELLINZONI, M., DE ROSSI, E. and RICCARDI, G., 2004. Rv2686c-Rv2687c-Rv2688c, an ABC Fluoroquinolone Efflux Pump in *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, **48**(8), pp. 3175-3178.
- PEDERSEN, K., CHRISTENSEN, S.K. and GERDES, K., 2002. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Molecular microbiology*, **45**(2), pp. 501-510.
- PERKINS, M.D. and CUNNINGHAM, J., 2007. Facing the Crisis: Improving the Diagnosis of Tuberculosis in the HIV Era. *Journal of Infectious Diseases*, **196**(Supplement 1), pp. S15-S27.
- PFYFFER, G.E., WELSCHER, H.M., KISSLING, P., CIESLAK, C., CASAL, M.J., GUTIERREZ, J. and RUSCH-GERDES, S., 1997. Comparison of the *Mycobacteria* Growth Indicator Tube (MGIT) with radiometric and solid culture for recovery of acid-fast bacilli. *Journal of clinical microbiology*, **35**(2), pp. 364-368.
- PHAM, D., FATTAL, E. and TSAPIS, N., 2015. Pulmonary drug delivery systems for tuberculosis treatment. *International journal of pharmaceutics*, **478**(2), pp. 517-529.
- PHELAN, J., DE SESSIONS, P.F., TIENCHEU, L., PERDIGAO, J., MACHADO, D., HASAN, R., HASAN, Z., BERGVAL, I.L., ANTHONY, R., MCNERNEY, R., ANTONIO, M., PORTUGAL, I., VIVEIROS, M., CAMPINO, S., HIBBERD, M.L. and CLARK, T.G., 2018. Methylation in *Mycobacterium tuberculosis* is lineage specific with associated mutations present globally. *Scientific reports*, **8**(1), pp. 160-017-18188-y.

- PIERSIMONI, C. and SCARPARO, C., 2003. Relevance of commercial amplification methods for direct detection of *Mycobacterium tuberculosis* complex in clinical samples. *Journal of clinical microbiology*, **41**(12), pp. 5355-5365.
- POOLE, K., 2005. Efflux-mediated antimicrobial resistance. *Journal of Antimicrobial Chemotherapy*, **56**(1), pp. 20-51.
- PRIMM, T.P., ANDERSEN, S.J., MIZRAHI, V., AVARBOCK, D., RUBIN, H. and BARRY, C.E., 2000. The Stringent Response of *Mycobacterium tuberculosis* Is Required for Long-Term Survival. *Journal of Bacteriology*, **182**(17), pp. 4889-4898.
- PUISSEGUR, M., BOTANCH, C., DUTEYRAT, J., DELSOL, G., CARATERO, C. and ALTARE, F., 2004. An in vitro dual model of mycobacterial granulomas to investigate the molecular interactions between mycobacteria and human host cells. *Cellular microbiology*, **6**(5), pp. 423-433.
- RAMAGE, H.R., CONNOLLY, L.E. and COX, J.S., 2009. Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. *PLoS genetics*, **5**(12), pp. e1000767.
- RAMASWAMY, S. and MUSSER, J.M., 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tubercle and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, **79**(1), pp. 3-29.
- RAWAT, R., WHITTY, A. and TONGE, P.J., 2003. The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: Adduct affinity and drug resistance. *Proceedings of the National Academy of Sciences*, **100**(24), pp. 13881-13886.
- REED, M.B., PICHLER, V.K., MCINTOSH, F., MATTIA, A., FALLOW, A., MASALA, S., DOMENECH, P., ZWERLING, A., THIBERT, L., MENZIES, D., SCHWARTZMAN, K. and BEHR, M.A., 2009. Major *Mycobacterium tuberculosis* Lineages Associate with Patient Country of Origin. *Journal of clinical microbiology*, **47**(4), pp. 1119-1128.
- RHOADS, A. and AU, K.F., 2015. PacBio Sequencing and Its Applications. *Genomics, proteomics & bioinformatics*, **13**(5), pp. 278-289.
- RIBEIRO, S.C., GOMES, L.L., AMARAL, E.P., ANDRADE, M.R., ALMEIDA, F.M., REZENDE, A.L., LANES, V.R., CARVALHO, E.C., SUFFYS, P.N., MOKROUSOV, I. and LASUNSKAIA, E.B., 2014. *Mycobacterium tuberculosis* strains of the modern sublineage of the Beijing family are more likely to display increased virulence than strains of the ancient sublineage. *Journal of clinical microbiology*, **52**(7), pp. 2615-2624.

- RINDI, L., MEDICI, C., BIMBI, N., BUZZIGOLI, A., LARI, N. and GARZELLI, C., 2014. Genomic variability of Mycobacterium tuberculosis strains of the Euro-American lineage based on large sequence deletions and 15-locus MIRU-VNTR polymorphism. *PloS one*, **9**(9), pp. e107150.
- ROBERTS, D.M., LIAO, R.P., WISEDCHAISRI, G., HOL, W.G.J. and SHERMAN, D.R., 2004. Two Sensor Kinases Contribute to the Hypoxic Response of Mycobacterium tuberculosis. *Journal of Biological Chemistry*, **279**(22), pp. 23082-23087.
- ROBERTS, R.J., VINCZE, T., POSFAI, J. and MACELIS, D., 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic acids research*, **43**(D1), pp. D298-D299.
- ROETZER, A., DIEL, R., KOHL, T.A., RUCKERT, C., NUBEL, U., BLOM, J., WIRTH, T., JAENICKE, S., SCHUBACK, S., RUSCH-GERDES, S., SUPPLY, P., KALINOWSKI, J. and NIEMANN, S., 2013. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS medicine*, **10**(2), pp. e1001387.
- ROSS, B.C., RAIOS, K., JACKSON, K. and DWYER, B., 1992. Molecular cloning of a highly repeated DNA element from Mycobacterium tuberculosis and its use as an epidemiological tool. *Journal of clinical microbiology*, **30**(4), pp. 942-946.
- ROTEM, E., LOINGER, A., RONIN, I., LEVIN-REISMAN, I., GABAY, C., SHORESH, N., BIHAM, O. and BALABAN, N.Q., 2010. Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proceedings of the National Academy of Sciences*, **107**(28), pp. 12541-12546.
- RUSSELL, D.G., 2003. Phagosomes, fatty acids and tuberculosis. *Nature cell biology*, **5**(9), pp. 776-778.
- RUSSELL, D.G., 2001. Mycobacterium tuberculosis: here today, and here tomorrow. *Nature reviews.Molecular cell biology*, **2**(8), pp. 569-577.
- RYAN, F., 1992. *The Forgotten Plague: How the Battle Against Tuberculosis was Won-And Lost*. Boston: Liitle,Brown and Company.
- SAHBAZIAN, B. and WEIS, S.E., 2005. Treatment of active tuberculosis: challenges and prospects. *Clinics in chest medicine*, **26**(2), pp. 273-82, vi.
- SAKAMOTO, K., 2012. The Pathology of Mycobacterium tuberculosis Infection. *Veterinary Pathology Online*, **49**(3), pp. 423-439.
- SALA, A., BORDES, P. and GENEVAUX, P., 2014. Multiple toxin-antitoxin systems in Mycobacterium tuberculosis. *Toxins*, **6**(3), pp. 1002-1020.

SÁNCHEZ-ROMERO, M.A., COTA, I. and CASADESÚS, J., 2015. DNA methylation in bacteria: from the methyl group to the methylome. *Current opinion in microbiology*, **25**, pp. 9-16.

SATER, M.R., LAMELAS, A., WANG, G., CLARK, T.A., ROLTGEN, K., MANE, S., KORLACH, J., PLUSCHKE, G. and SCHMID, C.D., 2015. DNA Methylation Assessed by SMRT Sequencing Is Linked to Mutations in *Neisseria meningitidis* Isolates. *PloS one*, **10**(12), pp. e0144612.

SAXENA, S., MATHUR, M. and TALWAR, V.K., 2001. Detection of tubercle bacilli in sputum: application of sodium hypochlorite concentration method. *The Journal of communicable diseases*, **33**(4), pp. 241-244.

SECHI, L.A., ZANETTI, S., DUPRÉ, I., DELOGU, G. and FADDA, G., 1998. Enterobacterial Repetitive Intergenic Consensus Sequences as Molecular Targets for Typing of *Mycobacterium tuberculosis* Strains. *Journal of clinical microbiology*, **36**(1), pp. 128-132.

SEDDON, J.A., FURIN, J.J., GALE, M., DEL CASTILLO BARRIENTOS, H., HURTADO, R.M., AMANULLAH, F., FORD, N., STARKE, J.R., SCHAAF, H.S. and SENTINEL PROJECT ON PEDIATRIC DRUG-RESISTANT TUBERCULOSIS, 2012. Caring for children with drug-resistant tuberculosis: practice-based recommendations. *American journal of respiratory and critical care medicine*, **186**(10), pp. 953-964.

SEGOVIA-JUAREZ, J.L., GANGULI, S. and KIRSCHNER, D., 2004. Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *Journal of theoretical biology*, **231**(3), pp. 357-376.

SHAKYA, N., GARG, G., AGRAWAL, B. and KUMAR, R., 2012. Chemotherapeutic interventions against tuberculosis. *Pharmaceuticals (Basel, Switzerland)*, **5**(7), pp. 690-718.

SHELL, S.S., PRESTWICH, E.G., BAEK, S.H., SHAH, R.R., SASSETTI, C.M., DEDON, P.C. and FORTUNE, S.M., 2013. DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. *PLoS pathogens*, **9**(7), pp. e1003419.

SHILOH, M.U., MANZANILLO, P. and COX, J.S., 2008. *Mycobacterium tuberculosis* Senses Host-Derived Carbon Monoxide during Macrophage Infection. *Cell Host & Microbe*, **3**(5), pp. 323-330.

SINGH, R., BARRY, C.E. and BOSHOFF, H.I.M., 2010. The Three RelE Homologs of *Mycobacterium tuberculosis* Have Individual, Drug-Specific Effects on Bacterial Antibiotic Tolerance. *Journal of Bacteriology*, **192**(5), pp. 1279-1291.

- SINGH, V., JAMWAL, S., JAIN, R., VERMA, P., GOKHALE, R. and RAO, K.V., 2012. Mycobacterium tuberculosis-driven targeted recalibration of macrophage lipid homeostasis promotes the foamy phenotype. *Cell host & microbe*, **12**(5), pp. 669-681.
- SJOBRING, U., MECKLENBURG, M., ANDERSEN, A.B. and MIORNER, H., 1990. Polymerase chain reaction for detection of Mycobacterium tuberculosis. *Journal of clinical microbiology*, **28**(10), pp. 2200-2204.
- SLOAN, D.J., 2013. *Investigation into mycobacterial persistence and early markers of outcome in the treatment of pulmonary tuberculosis*. PhD edn. Liverpool: Ethos.Uk.
- SLOAN, D.J., MWANDUMBA, H.C., GARTON, N.J., KHOO, S.H., BUTTERWORTH, A.E., ALLAIN, T.J., HEYDERMAN, R.S., CORBETT, E.L., BARER, M.R. and DAVIES, G.R., 2015. Pharmacodynamic Modeling of Bacillary Elimination Rates and Detection of Bacterial Lipid Bodies in Sputum to Predict and Understand Outcomes in Treatment of Pulmonary Tuberculosis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, **61**(1), pp. 1-8.
- SMITH, I., 2003. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews*, **16**(3), pp. 463-496.
- SOMERVILLE, W., THIBERT, L., SCHWARTZMAN, K. and BEHR, M.A., 2005. Extraction of Mycobacterium tuberculosis DNA: a Question of Containment. *Journal of clinical microbiology*, **43**(6), pp. 2996-2997.
- SOWAJASSATAKUL, A., PRAMMANANAN, T., CHAIPRASERT, A. and PHUNPRUCH, S., 2014. Molecular characterization of amikacin, kanamycin and capreomycin resistance in M/XDR-TB strains isolated in Thailand. *BMC Microbiology*, **14**(1), pp. 165.
- SRIVASTAVA, R., GOPINATHAN, K.P. and RAMAKRISHNAN, T., 1981. Deoxyribonucleic acid methylation in mycobacteria. *Journal of Bacteriology*, **148**(2), pp. 716-719.
- STAMATAKIS, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, **30**(9), pp. 1312-1313.
- STANLEY, R.E., BLAHA, G., GRODZICKI, R.L., STRICKLER, M.D. and STEITZ, T.A., 2010. The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nature structural & molecular biology*, **17**(3), pp. 289-293.
- STAVRUM, R., MPHAHLELE, M., OVREAS, K., MUTHIVHI, T., FOURIE, P.B., WEYER, K. and GREWAL, H.M., 2009. High diversity of Mycobacterium tuberculosis genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *Journal of clinical microbiology*, **47**(6), pp. 1848-1856.

STEAD, W.W., EISENACH, K.D., CAVE, M.D., BEGGS, M.L., TEMPLETON, G.L., THOEN, C.O. and BATES, J.H., 1995. When did Mycobacterium tuberculosis infection first occur in the New World? An important question with public health implications. *American journal of respiratory and critical care medicine*, **151**(4), pp. 1267-1268.

STEINGART, K.R., HENRY, M., NG, V., HOPEWELL, P.C., RAMSAY, A., CUNNINGHAM, J., URBANCZIK, R., PERKINS, M., AZIZ, M.A. and PAI, M., 2006. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *The Lancet. Infectious diseases*, **6**(9), pp. 570-581.

STEINGART, K.R., SCHILLER, I., HORNE, D.J., PAI, M., BOEHME, C.C. and DENDUKURI, N., 2014. Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *The Cochrane database of systematic reviews*, **1**, pp. CD009593.

STEWART, G.R., SNEWIN, V.A., WALZL, G., HUSSELL, T., TORMAY, P., O'GAORA, P., GOYAL, M., BETTS, J., BROWN, I.N. and YOUNG, D.B., 2001. Overexpression of heat-shock proteins reduces survival of Mycobacterium tuberculosis in the chronic phase of infection. *Nature medicine*, **7**(6), pp. 732-737.

STEYN, A.J., COLLINS, D.M., HONDALUS, M.K., JACOBS, W.R., Jr, KAWAKAMI, R.P. and BLOOM, B.R., 2002. Mycobacterium tuberculosis WhiB3 interacts with RpoV to affect host survival but is dispensable for in vivo growth. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(5), pp. 3147-3152.

STUCKI, D., MALLA, B., HOSTETTLER, S., HUNA, T., FELDMANN, J., YEBOAH-MANU, D., BORRELL, S., FENNER, L., COMAS, I., COSCOLLA, M. and GAGNEUX, S., 2012. Two new rapid SNP-typing methods for classifying Mycobacterium tuberculosis complex into the main phylogenetic lineages. *PloS one*, **7**(7), pp. e41253.

SUPPLY, P., LESJEAN, S., SAVINE, E., KREMER, K., VAN SOOLINGEN, D. and LOCHT, C., 2001. Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. *Journal of clinical microbiology*, **39**(10), pp. 3563-3571.

SUPPLY, P., MAZARS, E., LESJEAN, S., VINCENT, V., GICQUEL, B. and LOCHT, C., 2000. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Molecular microbiology*, **36**(3), pp. 762-771.

SUZUKI, Y., KATSUKAWA, C., TAMARU, A., ABE, C., MAKINO, M., MIZUGUCHI, Y. and TANIGUCHI, H., 1998. Detection of Kanamycin-Resistant Mycobacterium tuberculosis by Identifying Mutations in the 16S rRNA Gene. *Journal of Clinical Microbiology*, **36**(5), pp. 1220-1225.

TAKI-EDDIN, L. and MONEM, F., 2011. Utility of an interferon-gamma release assay as a potential diagnostic aid for active pulmonary tuberculosis. *The Journal of Infection in Developing Countries*, **6**(01),.

TAKIFF, H.E., SALAZAR, L., GUERRERO, C., PHILIPP, W., HUANG, W.M., KREISWIRTH, B., COLE, S.T., JACOBS, W.R., Jr and TELENTI, A., 1994. Cloning and nucleotide sequence of Mycobacterium tuberculosis gyrA and gyrB genes and detection of quinolone resistance mutations. *Antimicrobial Agents and Chemotherapy*, **38**(4), pp. 773-780.

TAN, S.C. and YIAP, B.C., 2009. DNA, RNA, and protein extraction: the past and the present. *Journal of biomedicine & biotechnology*, **2009**, pp. 574398.

THIERRY, D., CAVE, M.D., EISENACH, K.D., CRAWFORD, J.T., BATES, J.H., GICQUEL, B. and GUESDON, J.L., 1990. IS6110, an IS-like element of Mycobacterium tuberculosis complex. *Nucleic acids research*, **18**(1), pp. 188-188.

TORREY, H.L., KEREN, I., VIA, L.E., LEE, J.S. and LEWIS, K., 2016. High Persister Mutants in Mycobacterium tuberculosis. *PloS one*, **11**(5), pp. e0155127.

TSOLAKI, A.G., GAGNEUX, S., PYM, A.S., GOGUET DE LA SALMONIERE, YVES-OLIVIER L., KREISWIRTH, B.N., VAN SOOLINGEN, D. and SMALL, P.M., 2005. Genomic Deletions Classify the Beijing/W Strains as a Distinct Genetic Lineage of Mycobacterium tuberculosis. *Journal of clinical microbiology*, **43**(7), pp. 3185-3191.

UNIVERSITY OF OXFORD, 24 March, 2016-last update. Available: <http://www.ox.ac.uk/news/2016-03-24-global-team-aim-faster-more-effective-tb-diagnosi> [3 August, 2017].

VAN DEN BOOGAARD, J., KIBIKI, G.S., KISANGA, E.R., BOEREE, M.J. and AARNOUTSE, R.E., 2009. New drugs against tuberculosis: problems, progress, and evaluation of agents in clinical development. *Antimicrobial Agents and Chemotherapy*, **53**(3), pp. 849-862.

VAN DER SPUY, G.D., KREMER, K., NDABAMBI, S.L., BEYERS, N., DUNBAR, R., MARAIS, B.J., VAN HELDEN, P.D. and WARREN, R.M., 2009. Changing Mycobacterium tuberculosis population highlights clade-specific pathogenic characteristics. *Tuberculosis (Edinburgh, Scotland)*, **89**(2), pp. 120-125.

VAN DER WOUDE, M., HALE, W.B. and LOW, D.A., 1998. Formation of DNA Methylation Patterns: Nonmethylated GATC Sequences in gut and papOperons. *Journal of Bacteriology*, **180**(22), pp. 5913-5920.

VAN HELDEN, P.D., 2001. Mycobacterium Tuberculosis Protocols. In: T. PARISH and N.G. STOKER, eds, *Methods in Molecular Medicine*. Totowa NJ: Humana Press Inc., pp. 19-30.

VAN SOOLINGEN, D., HERMANS, P.W., DE HAAS, P.E., SOLL, D.R. and VAN EMBDEN, J.D., 1991. Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *Journal of clinical microbiology*, **29**(11), pp. 2578-2586.

VAN VOLLENHOVEN, P., HEYNS, C.F., DE BEER, P.M., WHITAKER, P., VAN HELDEN, P.D. and VICTOR, T., 1996. Polymerase chain reaction in the diagnosis of urinary tract tuberculosis. *Urological research*, **24**(2), pp. 107-111.

VASSALL, A., VAN KAMPEN, S., SOHN, H., MICHAEL, J.S., JOHN, K.R., DEN BOON, S., DAVIS, J.L., WHITELAW, A., NICOL, M.P., GLER, M.T., KHALIQOV, A., ZAMUDIO, C., PERKINS, M.D., BOEHME, C.C. and COBELENS, F., 2011. Rapid diagnosis of tuberculosis with the Xpert MTB/RIF assay in high burden countries: a cost-effectiveness analysis. *PLoS medicine*, **8**(11), pp. e1001120.

VERMA, S., DHOLE, T.N., KUMAR, M. and KASHYAP, S., 2013. Novel approach for improving sensitivity of microscopic detection of acid-fast bacilli (AFB) by use of the ReaSLR method. *Journal of clinical microbiology*, **51**(11), pp. 3597-3601.

VERVER, S., WARREN, R.M., BEYERS, N., RICHARDSON, M., VAN DER SPUY, G.D., BORGDORFF, M.W., ENARSON, D.A., BEHR, M.A. and VAN HELDEN, P.D., 2005. Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *American journal of respiratory and critical care medicine*, **171**(12), pp. 1430-1435.

VIEGAS, S.O., MACHADO, A., GROENHEIT, R., GHEBREMICHAEL, S., PENNHAG, A., GUDO, P.S., CUNA, Z., LANGA, E., MIOTTO, P., CIRILLO, D.M., RASTOGI, N., WARREN, R.M., VAN HELDEN, P.D., KOIVULA, T. and KALLENIOUS, G., 2013. Mycobacterium tuberculosis Beijing genotype is associated with HIV infection in Mozambique. *PloS one*, **8**(8), pp. e71999.

VILCHÈZE, C., AV-GAY, Y., ATTARIAN, R., LIU, Z., HAZBÓN, M.H., COLANGELI, R., CHEN, B., LIU, W., ALLAND, D., SACCHETTINI, J.C. and JACOBS JR, W.R., 2008. Mycothiol biosynthesis is essential for ethionamide susceptibility in Mycobacterium tuberculosis. *Molecular microbiology*, **69**(5), pp. 1316-1329.

VOSKUIL, M.I., BARTEK, I.L., VISCONTI, K. and SCHOOLNIK, G.K., 2011. The response of mycobacterium tuberculosis to reactive oxygen and nitrogen species. *Frontiers in microbiology*, **2**, pp. 105.

VOSKUIL, M.I. and SCHLESINGER, L.S., 2015. Toward Resolving the Paradox of the Critical Role of the DosR Regulon in Mycobacterium tuberculosis Persistence and Active Disease. *Am J Respir Crit Care Med*, **191**(10), pp. 1103-1105.

VOSKUIL, M.I., SCHNAPPINGER, D., VISCONTI, K.C., HARRELL, M.I., DOLGANOV, G.M., SHERMAN, D.R. and SCHOOLNIK, G.K., 2003. Inhibition of Respiration by Nitric Oxide Induces a Mycobacterium tuberculosis Dormancy Program. *The Journal of experimental medicine*, **198**(5), pp. 705-713.

VOTINTSEVA, A.A., PANKHURST, L.J., ANSON, L.W., MORGAN, M.R., GASCOYNE-BINZI, D., WALKER, T.M., QUAN, T.P., WYLLIE, D.H., DEL OJO ELIAS, C., WILCOX, M., WALKER, A.S., PETO, T.E. and CROOK, D.W., 2015. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. *Journal of clinical microbiology*, **53**(4), pp. 1137-1143.

WADDINGTON, C.H., 1956. Genetic Assimilation of the Bithorax Phenotype. *Evolution*, **10**(1), pp. 1-13.

WAKAMOTO, Y., DHAR, N., CHAIT, R., SCHNEIDER, K., SIGNORINO-GELO, F., LEIBLER, S. and MCKINNEY, J.D., 2013. Dynamic persistence of antibiotic-stressed mycobacteria. *Science (New York, N.Y.)*, **339**(6115), pp. 91-95.

WALKER, T.M., IP, C.L., HARRELL, R.H., EVANS, J.T., KAPATAI, G., DEDICOAT, M.J., EYRE, D.W., WILSON, D.J., HAWKEY, P.M., CROOK, D.W., PARKHILL, J., HARRIS, D., WALKER, A.S., BOWDEN, R., MONK, P., SMITH, E.G. and PETO, T.E., 2013. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *The Lancet. Infectious diseases*, **13**(2), pp. 137-146.

WALLIS, R.S., PAI, M., MENZIES, D., DOHERTY, T.M., WALZL, G., PERKINS, M.D. and ZUMLA, A., 2010. Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice. *Lancet*, **375**(9729), pp. 1920-1937.

WALSH, G.P., TAN, E.V., DELA CRUZ, E.C., ABALOS, R.M., VILLAHERMOSA, L.G., YOUNG, L.J., CELLONA, R.V., NAZARENO, J.B. and HORWITZ, M.A., 1996. The Philippine cynomolgus monkey (*Macaca fascicularis*) provides a new nonhuman primate model of tuberculosis that resembles human disease. *Nature medicine*, **2**(4), pp. 430-436.

WALTER, N.D., DOLGANOV, G.M., GARCIA, B.J., WORODRIA, W., ANDAMA, A., MUSISI, E., AYAKAKA, I., VAN, T.T., VOSKUIL, M.I., DE JONG, B.C., DAVIDSON, R.M., FINGERLIN, T.E., KECHRIS, K., PALMER, C., NAHID, P., DALEY, C.L., GERACI, M., HUANG, L., CATTAMANCHI, A., STRONG, M., SCHOOLNIK, G.K. and DAVIS, J.L., 2015. Transcriptional adaptation of drug-tolerant Mycobacterium tuberculosis during treatment of human tuberculosis. *Journal of Infectious Diseases*, **212**(6), pp. 990-998.

WANG, X., WANG, H. and XIE, J., 2011. Genes and regulatory networks involved in persistence of Mycobacterium tuberculosis. *Science China. Life sciences*, **54**(4), pp. 300-310.

WARREN, R., DE KOCK, M., ENGELKE, E., MYBURGH, R., GEY VAN PITTIUS, N., VICTOR, T. and VAN HELDEN, P., 2006. Safe Mycobacterium tuberculosis DNA Extraction Method That Does Not Compromise Integrity. *Journal of clinical microbiology*, **44**(1), pp. 254-256.

WAYNE, L.G. and LIN, K.Y., 1982. Glyoxylate metabolism and adaptation of Mycobacterium tuberculosis to survival under anaerobic conditions. *Infection and immunity*, **37**(3), pp. 1042-1049.

WAYNE, L.G. and HAYES, L.G., 1996. An in vitro model for sequential study of shift down of Mycobacteria tuberculosis through two stages of nonreplicating persistence. *Infection and immunity*, **64**(6), pp. 2062-2069.

WAYNE, L.G. and SOHASKEY, C.D., 2001. Nonreplicating persistence of mycobacterium tuberculosis. *Annual Review of Microbiology*, **55**, pp. 139-163.

WHO, 17 February, 2017a-last update. Available: <http://www.who.int/en/> [February/18, 2017].

WHO, 2017b. *Tuberculosis Global Report*. Geneva: World Health Organization.

WHO, 2016. *Tuberculosis Global Report*. Geneva: .

WHO, 2014. *Tuberculosis Global Report*. Geneva: .

WHO, 2011. *Policy Xpert MTB/RIF*. Geneva: World Health Organization.

WHO, 1997. *Treatment of tuberculosis: guidelines for national programmes*. Geneva: World Health Organization.

WILLIAMS-BOUYER, N., YORKE, R., LEE, H.I. and WOODS, G.L., 2000. Comparison of the BACTEC MGIT 960 and ESP culture system II for growth and detection of mycobacteria. *Journal of clinical microbiology*, **38**(11), pp. 4167-4170.

WILSON, M.R., NACCACHE, S.N., SAMAYOA, E., BIAGTAN, M., BASHIR, H., YU, G., SALAMAT, S.M., SOMASEKAR, S., FEDERMAN, S., MILLER, S., SOKOLIC, R., GARABEDIAN, E., CANDOTTI, F., BUCKLEY, R.H., REED, K.D., MEYER, T.L., SEROOGY, C.M., GALLOWAY, R., HENDERSON, S.L., GERN, J.E., DERISI, J.L. and CHIU, C.Y., 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *The New England journal of medicine*, **370**(25), pp. 2408-2417.

WINDER, F.G. and COLLINS, P.B., 1970. Inhibition by Isoniazid of Synthesis of Mycolic Acids in Mycobacterium tuberculosis. *Journal of general microbiology*, **63**(1), pp. 41-48.

- WION, D. and CASADESUS, J., 2006. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature reviews.Microbiology*, **4**(3), pp. 183-192.
- WITNEY, A.A., GOULD, K.A., ARNOLD, A., COLEMAN, D., DELGADO, R., DHILLON, J., POND, M.J., POPE, C.F., PLANCHE, T.D., STOKER, N.G., COSGROVE, C.A., BUTCHER, P.D., HARRISON, T.S. and HINDS, J., 2015. Clinical Application of Whole-Genome Sequencing To Inform Treatment for Multidrug-Resistant Tuberculosis Cases. *Journal of clinical microbiology*, **53**(5), pp. 1473-1483.
- WOJTASIK, A., MAJCHRZAK, M., ADAMUS-BIALEK, W., AUGUSTYNOWICZ-KOPEC, E., ZWOLSKA, Z., DZIADEK, J. and PARNIEWSKI, P., 2011. Trinucleotide repeat sequence-based PCR as a potential approach for genotyping *Mycobacterium gordonae* strains. *Journal of microbiological methods*, **85**(1), pp. 28-32.
- WOLFSON, J.S., HOOPER, D.C., MCHUGH, G.L., BOZZA, M.A. and SWARTZ, M.N., 1990. Mutants of *Escherichia coli* K-12 exhibiting reduced killing by both quinolone and beta-lactam antimicrobial agents. *Antimicrobial Agents and Chemotherapy*, **34**(10), pp. 1938-1943.
- WOODRUFF, H.B., 2014. Selman A. Waksman, Winner of the 1952 Nobel Prize for Physiology or Medicine. *Applied and Environmental Microbiology*, **80**(1), pp. 2-8.
- YADAV, V., DWIVED, V.P., BHATTACHARYA, D., MITTAL, A., MOODLEY, P. and DAS, G., 2015. Understanding the Host Epigenetics in *Mycobacterium tuberculosis* Infection. *Journal of Genetics and Genome Research*, **2**(1),.
- YUAN, Y., CRANE, D.D. and BARRY, C.E., 1996. Stationary phase-associated protein expression in *Mycobacterium tuberculosis*: function of the mycobacterial alpha-crystallin homolog. *Journal of Bacteriology*, **178**(15), pp. 4484-4492.
- ZAHRT, T.C., 2003. Molecular mechanisms regulating persistent *Mycobacterium tuberculosis* infection. *Microbes and infection / Institut Pasteur*, **5**(2), pp. 159-167.
- ZAHRT, T.C. and DERETIC, V., 2001. *Mycobacterium tuberculosis* signal transduction system required for persistent infections. *Proceedings of the National Academy of Sciences of the United States of America*, **98**(22), pp. 12706-12711.
- ZAUNBRECHER, M.A., SIKES, R.D., METCHOCK, B., SHINNICK, T.M. and POSEY, J.E., 2009. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase *eis* confers kanamycin resistance in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, **106**(47), pp. 20004-20009.
- ZHANG, Y., 2014. Persisters, persistent infections and the Yin-Yang model. *Emerging microbes & infections*, **3**(1), pp. e3.

ZHANG, Y., HEYM, B., ALLEN, B., YOUNG, D. and COLE, S., 1992. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature*, **358**, pp. 391-393.

ZHANG, Y. and YEW, W.W., 2009. Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *International Journal of Tuberculosis and Lung Disease*, **13**(11), pp. 1320-1330.

ZHANG, Y., YEW, W.W. and BARER, M.R., 2012. Targeting Persisters for Tuberculosis Control. *Antimicrobial Agents and Chemotherapy*, **56**(5), pp. 2223-2230.

ZHU, L., ZHONG, J., JIA, X., LIU, G., KANG, Y., DONG, M., ZHANG, X., LI, Q., YUE, L., LI, C., FU, J., XIAO, J., YAN, J., ZHANG, B., LEI, M., CHEN, S., LV, L., ZHU, B., HUANG, H. and CHEN, F., 2015. Precision methylome characterization of *Mycobacterium tuberculosis* complex (MTBC) using PacBio single-molecule real-time (SMRT) technology. *Nucleic acids research*, .

ZUMLA, A., RAVIGLIONE, M., HAFNER, R. and FORDHAM, V.R., 2013. Tuberculosis. *N Engl J Med*, **368**(8), pp. 745-755.