Investigating the relation between persister formation and clinical outcome in Tuberculosis (TB) patients

By

Julian Luke Coetzee

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Molecular Biology at Stellenbosch University

Supervisor: Dr. Jomien Mouton

Co-supervisors: Prof. Samantha Sampson

Dr. Anzaan Dippenaar

March 2021

Declaration

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

Signature:

Date:

Abstract

Despite progressive research regarding *Mycobacterium tuberculosis*, Tuberculosis (TB) still remains the top cause of mortality worldwide, with South Africa being considered one of the top ten TB burdened countries. Once infected with *M. tuberculosis*, TB disease can progress to an active disease state, or in the majority of cases, to an asymptomatic infection state known as latency or Latent TB infection (LTBI). LTBI has been associated with recurrent TB infection after a cured TB treatment outcome was achieved as individuals with LTBI are considered reservoirs of active *M. tuberculosis*. A subpopulation of bacteria known as persisters is thought to contribute to the LTBI state. Persisters are viable but non-replicating (VBNR) bacteria, which are recalcitrant to antibiotic treatment. There are major knowledge gaps regarding VBNR bacteria and their role in TB treatment outcome. Previously it was observed that patients who underwent TB treatment had remaining lesion activity post-treatment and presence of *M. tuberculosis* mRNA suggested the presence of unculturable bacteria likely being persisters. Based on positron emission tomography – computed tomography (PET/CT) scans patients were characterized as cured, recurrent or failed.

In this study, we aimed to evaluate the correlation between persister formation and pulmonary TB (PTB) disease outcome. We exploited a dual fluorescence replication reporter plasmid, and assessed persister formation using a THP-1 infection model, which mimics the host environment pathogenic mycobacteria encounter upon infection. Whole genome sequencing (WGS) data of baseline and follow-up isolates was obtained to determine if isolates are genetically predisposed to persister formation. A total of eighteen baseline clinical *M. tuberculosis* isolates were selected for this study. Eight isolates represented bacteria from the cured patient group while ten isolates represented bacteria from the failed/recurrent patient group. Isolates were determined to be pure cultures and WGS data was obtained. In preparation for persister assay experiments, all eighteen isolates were transformed with the fluorescence dilution (FD) dual reporter plasmid pTiGc. Growth curves demonstrated that plasmid carriage had no impact on bacterial growth.

The infection model enriched for persister-like cells as reflected by a subpopulation of VBNR bacteria. We found that all bacterial isolates possessed a level of replication heterogeneity at baseline both *in vitro* and intracellularly. Furthermore, isolates from the cured patients showed a significantly lower frequency of persister cells compared to that of isolates from the

failed/recurrent patient group. This suggests that the inherent tendency to form persister-like cells may have an impact on PTB treatment outcome. Data suggests that persister-like cell formation may be strain dependent. However, WGS data analysis were inconclusive. Furthermore, we recognize that the sample size is a crucial limiting factor in this study and further investigation with a larger cohort would be essential.

This is the first study to use clinical strains of *M. tuberculosis*, obtained from failed/recurrent treatment outcome group, coupled with fluorescent reporters in combination with WGS data to investigate the relationship between persister formation and clinical outcome. Possible future work would be to to validate the phenotypic study findings in a murine model. Furthermore, future studies that determine the role of genetic variation in persister formation would greatly advance a patient-specific treatment regimen that could decrease the lengthy treatment duration.

Abstrak

Ten spyte van goeie vordering in *Mycobacterium tuberculosis* navorsing bly Tuberkulose (TB) een van die grootste oorsake van sterftes wêrelwyd met Suid-Afrika (SA) wat as een van die top tien mees geaffekteerde lande geag word. Sodra ‘n persoon geïnfekteer word met *M. tuberculosis* kan TB tot n aktiewe siekte toestand vorder of, soos in meeste gevalle, tot ‘n asimptomatiese infektiewe toestand ontwikkel, beter bekend as ‘n latente TB infeksie (LTBI). LTBI word geassosieer met ‘n herhalende TB infeksie nadat ‘n pasiënt genees is met behandeling omdat individue met LTBI as ‘n bron van aktiewe *M. tuberculosis* geag word. ‘n Subpopulasie van bakterieë bekend as persisters word as bydraende faktore van die LTBI toestand gesien. Persisters is lewendige maar nie-repliserende (VBNR) bakterieë wat anktibiotika behandeling kan weerstaan. Daar is groot gapings in ons kennis oor VBNR bakterieë, asook die rol van dié selle in die finale uitkoms van TB. Daar is voorheen waargeneem dat pasiënte wat TB behandeling ondergaan het steeds aktiewe letsels het na behandeling en die teenwoordigheid van *M. tuberculosis* mRNA gee aanduiding daartoe dat nie kultiveerbare bakterieë, moontlik persisters, steeds teenwoordig is. Gebaseer op resultate van positron emissie tomografie – berekende tomografie (PET/CT) skanderings is pasiënte in kategorieë genaamd genees, herhalend of misluk verdeel.

In die studie beoog ons om te evalueer wat die korrelasie is tussen persister vorming en pulmonêre TB (PTB). Ons maak gebruik van ‘n dubbele fluoreserende replikasie plasmied en asseseer persister vorming in ‘n THP-1 infeksie model wat die omstandighede naboots wat *M. tuberculosis* teëkom in die gasheer. Heel genoom volgorde (WGS) data was versamel van oorspronklike asook opvolg isolate om vas te stel of isolate geneties meer vatbaar is vir persister vorming. In totaal is agtien kliniese *M. tuberculosis* isolate gekies vir die studie. Agt isolate verteenwoordig die tenvolle herstelde patiënt groep, terwyl tien isolate pasiënte die herhalende/mislukte groep verteenwoordig. Isolate was geïdentifiseer as rein kulture en WGS data was verkry. Ter voorbereiding van persister vorming eksperimente was al agtien isolate getransformeer met ‘n dubbel fluoreserende replikasie (FD) plasmied, pTiGc. Groeikurwes het gedemonstreer dat die plasmied geen effek het op bakteriële groei nie.

Die infeksie model het verryk vir persister selle soos gereflekteer deur ‘n subpopulasie van VBNR bakterieë. Ons het gevind dat alle bakteriële isolate dui tot ‘n mate van heterogene replikasie, beide *in vitro* en intrasellulêr. Verder het die isolate van tne volle herstelde pasiënte

‘n aansienlike laer frekwensie persister selle gehad teenoor die isolate van die mislukte/herhalende groep. Dit dui daarop dat die natuurlike neiging van selle om persisters te vorm ‘n impak het op die uitkoms van TB behandeling. Data wys dat persister sel vorming spesifiek is tot kliniese isolate, alhoewel WGS data nie oortuigend was om hierdie observasie te ondersteun nie. Verder herken ons dat die klein aantal monsters ‘n belangrike beperkende faktor is in die studie en verdere ondersoek met ‘n groter monster poel noodsaaklik is.

Hierdie is die eerste studie van sy soort wat gebruik maak van kliniese *M. tuberculosis* selle verkry van mislukte/herhalende pasiënte groepe met behulp van fluoreserende plasmiede en WGS data om die verhouding tussen persister vorming en kliniese uitkomste te bepaal. Toekomstige werk moet daarop fokus om die fenotipiese uitkomste in ‘n muis model te bevestig. Verder sal studies wat fokus op die effek van genetiese variasie en persister vorming groot vordering maak in ‘n meer pasiënt gefokusde benadering tot behandeling, wat die verlengde behandelings tydperk wat tans nodig is moontlik kan verkort.

Acknowledgments

I would like to express my thanks to the following people for the constant support and guidance through my MSc and during the writing of this thesis:

- To my supervisor, Dr. Jomien Mouton (Infection queen). Words can't describe how grateful I am to be one of your students. Firstly, thank you for accepting me as a MSc candidate. You were an integral part of this project from the start of the interview process till the very end. Thank you for constantly being there, imparting your knowledge, assistance and guidance in all aspects of the study. Thank you for the caffeine you've provided through the years. Without you none of this would have been possible. Thank you for your constant availability especially considering my working times.
- To my co-supervisor and head of HPM, Prof. Samantha Sampson (HPM queen). Thank you for being you. Just for being an inspiration to all young scientists, for your patience, for your warmth, guidance and sometimes much needed sternness throughout the study duration. You are surely an incredible group leader and woman of science.
- To my co-supervisor, Dr. Anzaan Dippenaar (WGS queen). Thank you for always being there if I had any WGS questions. Thank you for imparting your knowledge, and guiding me through the rocky waters of the command line and anything sequence related!! You are without a doubt a dynamite in your field and I'm immensely grateful for all you've done for me!
- To Andrea Gutschmidt (Flow queen). It has definitely been amazing to get to know you, to learn from you. Thank you for your expert flow knowledge, thank you for constantly being available to run samples. I appreciate the time and energy you've put in. Thank you so much.
- To the HPM group members, thank you for welcoming me with all the positivity in the world and for the constant support.
- To my family and friends. Words can't describe how thankful I am for you. You have given me all the support I could've asked for. I am eternally grateful for all you've done for me. Thank you for believing in me completely. Thank you for tolerating me. Thank you for the jokes and needed distraction at times.
- I would like to say thank you to my funders, the National Research Foundation (NRF), Harry Crossley Foundation, and Validate Foundation for their contributions.

List of abbreviations

°C	Degrees Celsius
µg	Microgram
µl	Microliter
µm	Micrometer
ng	Nanogram
%	Percentage
7H9-OGT	7H9-OADC, Glycerol, and Tween 80
AA	Amino acid
AIDS	Acquired immune deficiency syndrome
Approx.	Approximately
bam	binary alignment map
bp	Base pair
BSL3	Biosafety level three
BWA	Burrows-Wheeler Aligner
CAF	Central Analytical Facility
CO_2	Carbon dioxide
D	Day
ddH_2O	Double distilled water
DNA	Deoxyribo-nucleic acid
DNase	Deoxyribonuclease
DR	Drug resistant, direct repeat
EDTA	Ethylenediaminetetraacetic acid
E. coli	*Escherichia coli*
et al.	*et al*ii (and others)
FACS	Fluorescence-Activated Cell Sorting
FCS	Forward side scatter
FD	Fluorescence dilution
GATK	Genome analysis toolkit
GFP	Green fluorescence protein
HGT	Horizontal gene transfer
HIV	Human immunodeficiency virus
i.e.	id est (that is)
In/del	Small insertions and deletions
INH	Isoniazid
kb	Kilobase
LAM	Latin-American Mediterranean
LTBI	Latent tuberculosis infection
MDR	Multi drug resistant
MDMs	Human monocyte-derived macrophages
MGIT	Mycobacterial growth indicator tube
MIC	Minimum inhibitory concentration
MIRU/VNTR	Mycobacterial interspersed repetitive units-variable tandem repeat
ml	Milliliter
mm	Milimeter
mM	Millimolar
mRNA	Messenger ribonucleic-acid
M. tuberculosis	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex

NHLS	National Health Laboratory Service
NO	nitric oxide
nt	Nucleotide
OADC	Oleic acid–albumin–dextrose– catalase
OD600nm	Optical density at 600 nanometers
ORF	Open reading frame
PAGE	Poly acrylamide gel electrophoresis
PET/CT	Positron Emission Tomography/Computed Tomography
pH	Potential of hydrogen
RD	Region of difference
RFLP	Restriction fragment length polymorphism
Rif	Rifampicin
RNA	Ribonucleic acid
rpm	Revolutions per minute
sam	sequence alignment map
SAMMtb	Severely Attenuated Mutant of *Mycobacterium tuberculosis*
SD	Standard deviation
SDS	Sodium dodecyl sulphate
sRNAs	Small ribonucleic acidsOADC
SMVs	Small-colony variants
SNP	Single nucleotide polymorphism
TAE	Tris Acetate EDTA (buffer)
TB	Tuberculosis
TCA	Tri-carboxylic acid
TDR	Totally drug resistant
TE	Tris EDTA
*T*m	Melting temperature
VBNR	Viable but not replicating
vcf	Variant call format
WGS	Whole genome sequencing
WHO	World Health Organization
www	Wold wide web
XDR	Extremely drug resistant
ZN	Ziehl-Neelsen

List of figures

Figure 1.1. The TB disease paradigm……..2
Figure 1.2. The yin-yang model…….....3
Figure 1.3. *M. tuberculosis* complexes……...4
Figure 2.1. Bi-phasic killing curve and the difference between antibiotic resistance and antibiotic persistence……..9
Figure 2.2. The overlapping characteristics of persistence, resistance and dormancy……...11
Figure 2.3. Fundamental tools utilized in understanding bacterial persistence…….13
Figure 3.1. The principle of FD……28
Figure 3.2.1. Theophylline induction during macrophage infection…….34
Figure 3.2.2. Gating strategy for flow cytometry……..35
Figure 3.2.3. Scematic representation of the gating strategy for accurate bead population identification…….36
Figure 3.3.1. Ziehl-Neelsen(ZN) staining of H37Rv representing pure *M. tuberculosis* bacteria…….38
Figure 3.3.2. OD_{600} -based growth curve assessing fitness of bacterial isolates from the cured group carrying the pTiGc plasmid. All time points represent four technical replicates…….41
Figure 3.3.3. OD_{600} -based growth curve assessing fitness of bacterial isolates from the failed/recurrent group carrying the pTiGc plasmid. All time points represent four technical replicates……...42
Figure 3.3.4. Generation time of all strains from cured and failed/recurrent treatment groups[wild-type(wt) vs pTiGc transformed isolates (pTiGc)**]** The asterisk depicts 120h time interval while the rest are at 96h time interval……...43
Figure 3.3.5. Uptake percentage of cured versus failed/recurrent patient groups 0h post infection……44
Figure 3.3.6. Percentage intracellular survival from cured and failed/recurrent patient groups 120h post infection……...45
Figure 3.3.7. Population-wide replication dynamics from representatives of both cured and failed/recurrent patient groups upon macrophage infection……....46
Figure 3.3.8. Population wide replication dynamics of S126dx upon macrophage infection.47
Figure 3.3.9. Persister analysis gating strategy……48

Figure 3.3.10. Persister frequency in isolates obtained from the cured and failed/recurrent patient groups following macrophage infections……..49
Figure 4.2.1. A typical read from a FASTQ file generated by the Illumina sequencing platform version 1.5……...59
Figure 4.2.2. Workfow of the computational analysis of WGS data……61
Figure 4.3.1 The per base quality of the sequencing reads of a representative strain S105dx from the Illumina NextSeq platform……71
Figure 4.3.2. The per base quality of the sequencing reads of a representative strain S5 from Miseq platform……72
Figure 4.3.3. Molecular phylogenetic analysis by maximum likelihood method with 1000 bootstrap replicates showing the relationship of the included *M. tuberculosis* isolates to other members of the *Mycobacterium tuberculosis* complex……...77
Figure S3.1. Population wide dynamics of baseline isolates obtained from cured treatment group……99
Figure S3.2. Population wide dynamics of baseline isolates obtained from failed/recurrent treatment group……..100

List of tables

Table 2.1. Diseases known to be influenced by bacterial persisters ……16

Table 2.2. High throughput methods for identifying genes related to persister formation and their advantages and disadvantages……..25

Table 3.2.1. Plasmids and strains……..30

Table 3.2.2. Mycobacterial clinical isolates selected for this study ……32

Table 3.31. Mycobacterium catalysis clinical patient information……38

Table 3.3.2. Transformation information of clinical strains……39

Table 4.2.1. Examples of phred scores and the correlating ASCII characters…….....66

Table 4.3.1. Information of *M. tuberculosis* clinical strains……..69

Table 4.3.2. Average percentage of mapped reads and depth of coverage calculated based on Qualimap results from alignments produced by BWA, Novoalign and SMALT……73-75

Table of Contents

Declaration i

Acknowledgments vi

List of abbreviations vii

List of figures ix

List of tables xi

Chapter 1 1

Introduction 1

1.1. Background 1

1.1.1. The prevailing tuberculosis pandemic 1

1.1.2. Understanding the disease and bacterium 2

1.1.3. Persister formation and the mechanisms of antibiotic tolerance. 4

1.2. Problem Statement 6

1.3. Hypothesis 6

1.4. Aim and Objectives 6

1.5. Thesis Overview 7

Chapter 2 8

Literature review: Persisters and genetic contributors to their formation 8

2.1. What is persistence? 8

2.1.1. Persistent infections vs antibiotic persistence 8

2.1.2. Antibiotic resistance vs antibiotic persistence 8

2.1.3. Antibiotic persistence vs tolerance and dormancy 10

2.2. Tools used to study persistence 11

2.3. Relevance of persisters in pathogenic diseases 14

2.4. Triggers of persisters 16

2.4.1. Nutritional stresses 17

2.4.2. Oxidative, acidic and antibiotics 17

2.4.3. Host 18

2.5. Pathways and genes involved in persister formation in *M. tuberculosis* 18

2.5.1. The Stringent Response 18

2.5.2. SOS response 19

2.5.3. Metabolic slowdown/shifting 19

2.5.4. Transcriptional and post-transcriptional gene regulation 20

2.5.4.1. WhiB-like family genes 20

2.5.4.2. Sigma (σ) factors 21

2.5.4.3. Toxin-antitoxin (TA) modules 22
2.5.4.4. Small RNAs (sRNAs) 22
2.5.4.5. Protein post-translation modifications 22
2.5.5. Genetic adaptations 23
2.6. Identifying genes associated with persister formation 24
2.7. Conclusion 25
Chapter 3 27
Assessment of persister proportions in baseline clinical *M. tuberculosis* isolates from cured and recurrent/failed patient groups. 27
3.1. Introduction 27
3.2. Methods and materials 29
3.2.1. Plasmid constructs 29
3.2.2. Bacterial strains and culturing 30
3.2.3. Growth curve analysis of transformed clinical isolates 32
3.2.4. Infection of transformed isolates into THP-1 33
3.2.5. Flow cytometry sample preparation, acquisition and analysis. 34
3.2.6. Determination of bacterial uptake and survival within macrophages utilizing counting beads. 35
3.2.7. Statistical analysis 36
3.3.1. Rationale 37
3.3.2. Patient and isolate information 37
3.3.3. Confirmation of clinical isolates, transformation with replication reporter plasmid and growth of transformed strains. 38
3.3.4. Intracellular mycobacterial uptake and survival following macrophage infection. 43
3.3.5. Fluorescence dilution (FD) analysis assessing persister formation within patient groups at baseline. 45
3.4. Discussion 50
3.4.1. Patient information and PTB outcome 50
3.4.2. *In vitro* mycobacterial growth 50
3.4.3. Uptake percentage and survival of mycobacterial strains within THP-1 macrophages 52
3.4.4. Replication dynamics of intracellular *M. tuberculosis* clinical isolates reveals population heterogeneity on multiple levels 53
3.4.5. Persister-like cell formation between cured and failed/recurrent patient groups 53
3.4.6. Adaptation to host environment 54
3.5. Limitations 56

3.6. Future work 56
3.7. Conclusion 56
Chapter 4 58
Whole genome sequencing analyses of clinical isolates. 58
4.1. Introduction 58
4.2. Materials and Methods 59
4.2.1 Genomic DNA extraction 59
4.2.2. Next-generation sequence analysis 59
4.2.3. FASTQ file format 60
4.2.4. Phred-scaled quality values 60
4.2.5. Automated WGS Data Analysis Pipeline (USAP) 61
4.2.6. Quality control 62
4.2.7. Trimming of sequences 63
4.2.8. Alignment and mapping 63
4.2.8.1. Novoalign 64
4.2.8.2. Burrows-Wheeler Aligner (BWA) 64
4.2.8.3. SMALT 64
4.2.9. SAM File Validation 65
4.2.10. Converting the Sequencing Alignment Map (SAM) File Format to Binary Alignment (BAM) File Format 65
4.2.11. Alignment Statistics 65
4.2.12. Post Alignment Processing of BAM Files 65
4.2.12.1. Coordinate sorting and indexing of BAM files 65
4.2.12.2. Realignment focused on in/dels (insertions and deletions) 66
4.2.12.3. Coordinate sorting and indexing of realigned BAM files 66
4.2.12.4. Removal of PCR duplicates 66
4.2.13. Variant calling 66
4.2.13.1. GATK 66
4.2.13.2. SAMTools 67
4.2.14. Annotation of variants obtained from the different aligners 67
4.2.15. Comparison of annotations obtained from aligners 67
4.2.16. Filtering of unique variants after pairwise comparison 68
4.2.17. Drug susceptibility and lineage prediction of isolates obtained from cured and failed/recurrent patient group 68
4.2.18. Phylogenetic tree construction 69
4.3. Results 70

4.3.1 Introduction 70
4.3.2. Read Assessment and Trimming 71
4.3.3. Read Alignment and Mapping Statistics 73
4.3.4. High Confidence Variants 76
4.3.5. Phylogenetic Tree Construction 77
4.4.1. Data clean-up and quality control 79
4.4.2. Antibiotic susceptibility and lineage specificity 80
4.4.3. Unique variant recognition 80
4.5. Limitations 82
4.6. Future work: 82
4.7. Conclusion 83
Chapter 5 84
General conclusion 84
Chapter 6 88
References 88
Supplementary material 103
Appendices 105
Appendix A: Recipes and Protocols 105
Appendix B: Commands 108
Appendix C: Scripts 111

Chapter 1

Introduction

1.1. Background

1.1.1. The prevailing tuberculosis pandemic

Mycobacterium tuberculosis (*M. tuberculosis*) the etiologic agent of tuberculosis (TB) – which is a contagious infection that has tormented mankind for millennia. *M. tuberculosis* was identified in the 19th century (Barberis *et al.*, 2017). However, this pathogen continues to cause a worldwide epidemic, where new TB cases were estimated to reach 10 million in 2019 (WHO, 2020). Factors that advance the high TB burden include co-morbidities like diabetes and human immunodeficiency virus (HIV) co-infection (Singh *et al.*, 2020). However, factors such as the prevalence of totally drug-resistant (TDR), extremely drug-resistant (XDR)-, multidrug-resistant (MDR)-, drug-resistant (DR)- and persistent *M. tuberculosis* decrease the efficacy of TB treatment resulting in longer treatment regimens (Millet *et al.*, 2013; Seung, Keshavjee and Rich, 2015; Singh *et al.*, 2020; Yam *et al.*, 2020). DR is defined as an organism that is resistant to one drug e.g. isoniazid, MDR refers to *M. tuberculosis* which is resistant to two first-line TB drugs namely, isoniazid and rifampicin (Rif), XDR-TB a kind of MDR-TB which is resistant to all fluoroquinolones including at least one Group A drug. Group A drugs are are the foremost potent group within the second-line drug class for treatment against drug-resistant *M. tuberculosis* consisting of bedaquiline, levofloxacin, linezolid and moxifloxacin (WHO, 2021). Mycobacterial persisters are a subpopulation of bacteria that survives environmental stressors and antibiotic concentrations which are lethal to phenotypically non-persister mycobacteria (Goossens, Sampson and Rie, 2021).

The TB burden in Africa, which encompasses South Africa, is high compared to the rest of the world (WHO, 2020). In 2019 South Africa was considered one of the top ten high TB burdened countries world-wide with 360 000 incidences. Mortality due to TB in South Africa (excluding patients with HIV co-infection) was 22 000 during 2019 (WHO., 2020). It is estimated that ~80% of the South African population has TB, of which ~ 24% to 88% have latent TB infection (LTBI) (Mahomed *et al.*, 2011; Ncayiyana *et al.*, 2016; Drain *et al.*, 2018).

1.1.2. Understanding the disease and bacterium

The TB disease paradigm is a dynamic continuum defined by various states between active and latent infection, including incipient and subclinical states (Barry *et al.*, 2009). An incipient disease state occurs when viable *M. tuberculosis* infects a host and there is a likelihood of progression to active disease without intervention but is asymptomatic, has not induced radiographic irregularities, and is culture-negative. A subclinical TB state is caused by viable *M. tuberculosis* which results in abnormalities besides clinical TB symptoms and can be detected with the use of existing radiological and microbiological tests. LTBI is defined as causing an immune response following *M. tuberculosis* antigen stimulation by immunological tests [TST (Tuberculin Skin Test) or IFN-γ (Interferon-gamma) release assay (IGRA), QuantiFERON-TB Plus, and Statens Serum Institut, Copenhagen, Denmark C-Tb] without clinical symptoms of the disease and a normal chest radiograph e.g. positron emission tomography-computed tomography (PET/CT) scan. Following the establishment of LTBI, there are multiple pathways through which the disease can progress; (i) latency (consisting of a persistent disease burden) (ii) eliminated infection, (iii) fast or (iv) slow reactivation through the subclinical and incipient disease to active disease, or (v) cycling between incipient and subclinical states which may lead to symptomatic disease or disease resolutions (Fig 1.1). Individuals with LTBI are considered reservoirs for active TB cases as reactivation occurs in approximately 5%-15% of these individuals (Kiazyk and Ball, 2017; Jeon, 2020).

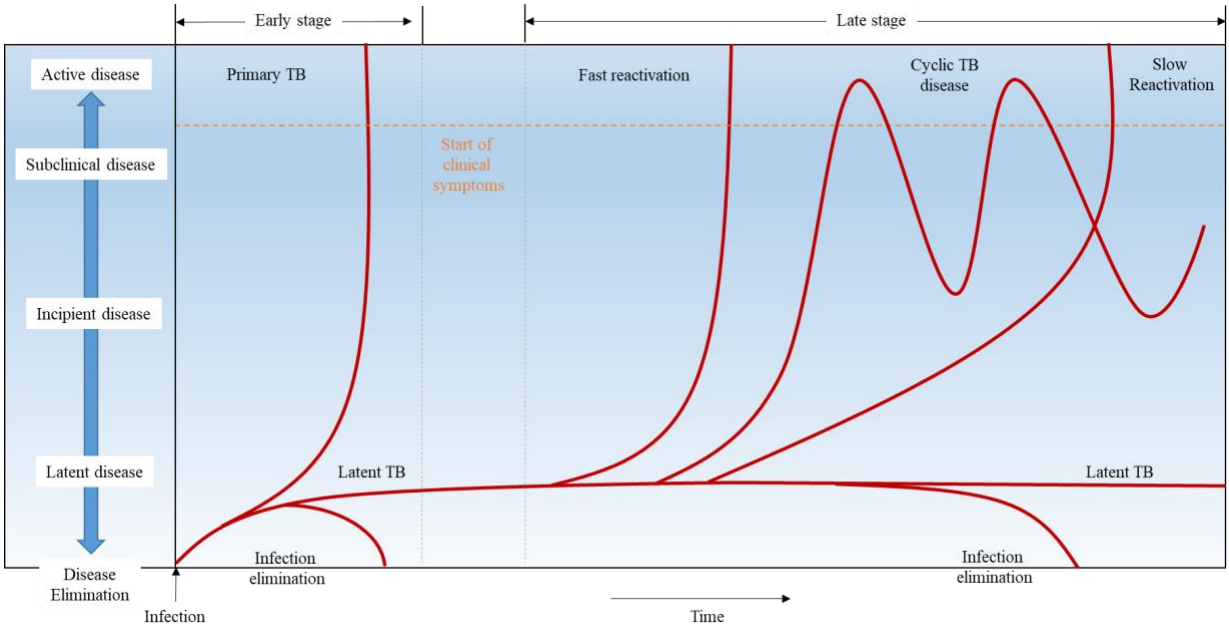


Figure 1.1. The TB disease paradigm adapted from Drain *et al.*, 2018

Pathogenesis studies have suggested a similarly complex disease progression using animal models as observed when *M. tuberculosis* infects a human host. Infection by a strain of the *M. tuberculosis* complex can result in an active disease state, elimination through an acquired or innate immune response, or the bacteria can adapt to the hostile environment (Drain *et al.*, 2018; de Martino *et al.*, 2019). Variation in host immune response, inter-host variation to treatment responses, genetic variation among strains, and possibly heterogeneity of mycobacterial populations upon initial infection has been responsible for these inconsistent outcomes, which could be explained by the yin-yang paradigm.

Under different laboratory designs, the yin-yang paradigm relates to both patients and pathogens. The paradigm theorizes that upon infection the overall bacterial populations consists of growing (yang-orange), slow-growing and non-growing (yin-grey) sub-populations with different metabolic statuses in a consortium which can interconvert at the level of the bacteria (expressed by the gradient from light to dark) (Fig 1.2) (Zhang, Yew and Barer, 2012; Zhang, 2014a). The grey dot in Yang is connected and rooted in the Yin half (grey), and the orange dot in Yin, reverters/persisters, is connected to the Yang half (red). In the case of TB treatment, TB antibiotics kills growing bacteria (Yang), while leaving reverters (orange dot) untouched. Antibiotic-tolerant persisters/reverters can regress to a replicating state, causing TB disease regression. The yin-yang paradigm could explain LTBI in humans.

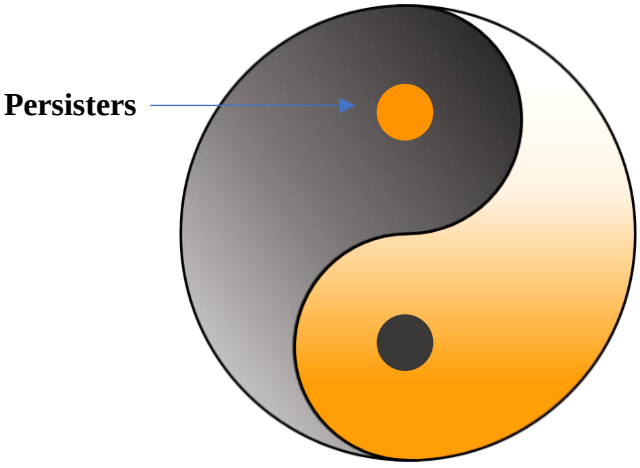


Figure 1.2 The yin-yang paradigm adapted from Zang., 2014

M. tuberculosis genomes are about 4.4 Mbp in length, GC rich, and comprise of ~ 4000 genes with ~99% similarity between *M. tuberculosis* complexes (Fig 1.3). However, specific strains assigned to a specific *M. tuberculosis* complex species and *M. tuberculosis* lineages depict notable differences in their virulence, pathogenesis and phenotypes, which have been reported

to have an impact on clinical appearance (Coscolla and Gagneux, 2010). Increased virulence has been observed especially in modern lineages including the Euro-American strains (Lineage 4) and the Beijing family (Lineage 2) compared to other ancient *M. tuberculosis* lineages for example the *M. africanum* strains (Lineage 5 and 6) and East-African-Indian (Lineage 1) (Merker *et al.*, 2015).

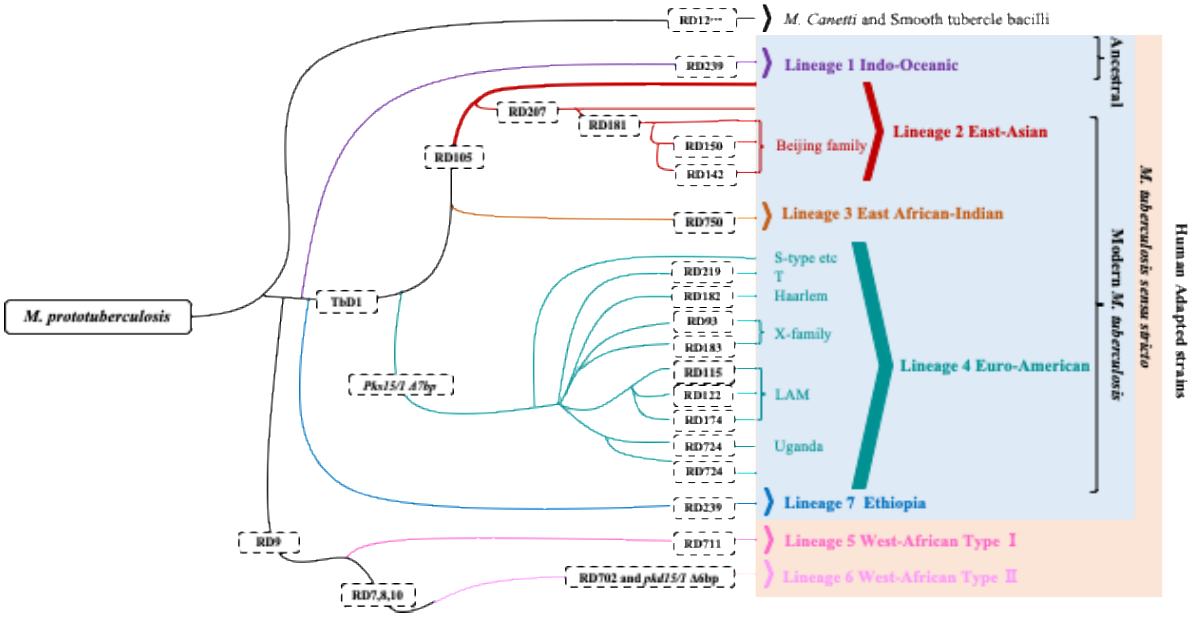


Figure 1.3: *M. tuberculosis* complexes adapted from (Tientcheu *et al.*, 2017).

Phenotypic and genotypic heterogeneity of mycobacterial populations may contribute to the variation in patient outcome as described by Figure 1.1. Various studies have shown that killing curves of *M. tuberculosis* under drug stress are biphasic, indicative of heterogeneous populations consisting of a mix of rapidly killed bacteria and those tolerant to antibiotic stress (Ahmad *et al.*, 2009). During antibiotic treatment of heterogeneous mycobacterial populations, antibiotic resistant and antibiotic tolerant (termed persisters) bacterial populations could arise. Antibiotic resistance differs from antibiotic tolerance. Antibiotic resistance is driven by non-reversible genetic mutations, and is either antibiotic-specific or drug class-specific; antibiotic resistance alters the lowest antibiotic concentration required to inhibit bacterial replication, known as the minimum inhibitory concentration (MIC). However antibiotic tolerance can occur in strains with no resistance-conferring mutations, can be observed across antibiotic classes (known as multidrug tolerance), and does not affect the MIC. Various bacterial states which contribute to survival under antibiotic pressure have previously been observed such as tolerance, persistence and dormancy.

1.1.3. Persister formation and the mechanisms of antibiotic tolerance.

Innate immune responses coupled with the adaptability of *M. tuberculosis* strains to these innate immune responses have been suggested as a mechanism that allows individuals with LTBI to maintain a dynamic relationship with the bacterium (de Martino *et al.*, 2019). Innate immune components include macrophages (Mϕ), dendritic cells, neutrophils, mast cells, airway epithelial cells and natural killer cells (de Martino *et al.*, 2019).

Mϕs are suggested to be the first line of defense against *M. tuberculosis*. Mϕs subject the bacteria to stressors such as hypoxia, low pH, reactive oxygen species, and reactive nitrogen species (Flynn and Chan, 2001). Exposure to Mϕs has been shown to slow or halt the replication of *M. tuberculosis* (Levitte *et al.*, 2016). The bacilli adapt to macrophage uptake conditions by entering a persistent state, where the persister bacilli are slowly or non-replicating and phenotypically drug-tolerant and can resume growth upon removal of the stressor to re-activate infection (Keren, Mulcahy and Lewis, 2012; Balaban *et al.*, 2019). Phagosomes inside Mϕs has acted as a safeguard against *M. tuberculosis* during the latent infection phase where in some Mϕ's *M. tuberculosis* growth is partially restricted and in others *M. tuberculosis* is actively growing. (Flynn and Chan., 2001; Orme., 1988; Russell., 2019). *M. tuberculosis* can neutralize strategies of Mϕs to suppress the pathogen. These strategies include intracellular trafficking, neutralization of toxic components such as reactive oxygen species and toxic metals, the acquisition of cytosol access, inhibition of autophagy, and the induction of host cell death, (Xu *et al.*, 1994; Vergne *et al.*, 2004; van der Wel *et al.*, 2007; Simeone *et al.*, 2012; de Martino *et al.*, 2019; Chen *et al.*, 2020).

To date little is known regarding the impact of *M. tuberculosis* persisters on disease progression. However, the persister phenomenon has been identified in numerous bacteria, including *Escherichia coli, Pseudomonas aeruginosa* and *Salmonella spp*. which have been utilized as model organisms for identifying mechanisms of persister formation and drug-induced tolerance (Möker, Dean and Tao, 2010; Hill and Helaine, 2019). Mechanisms of tolerance which have been identified across bacterial species and include, but are not constrained to, translational and post-transcriptional gene regulation, lowering the metabolic activity of a sub-population of bacteria, metabolic shifting (shifting between energy generating pathways), cell wall thickening, and various genetic adaptations conferring tolerance (Goossens, Sampson and Rie, 2021). Undoubtedly, a variety of factors affect the course of *M. tuberculosis* infections within individuals. It is suggested that persisters play a role in the

dynamic continuum of the TB disease as one of the factors influencing disease outcome and recurrence.

1.2. Problem Statement

Despite extensive research on TB, several aspects of the disease and its causative agent, *M. tuberculosis* are still poorly understood. A major knowledge gap surrounds the physiological state of the bacteria involved in LTBI. LTBI is in part attributable to the phenomenon of bacterial persistence. Persister bacteria are defined here as non- or slowly replicating, antibiotic-tolerant bacteria, where antibiotic tolerance is reversible and not genetically resistant. High treatment failure rates highlight that these persister populations pose a major problem for effective TB treatment.

Until recently, identification and isolation of persister bacteria has been extremely difficult. This is attributed to the low bacterial numbers (as only 1% of bacterial cultures comprise persisters) and lack of replication. Unfortunately, factors that trigger the entry into, survival in, and exit from, a persistent state are largely unknown. Determining the formation of persisters in clinical isolates from South African TB patients who have remaining lesion activity in the lung based on positron emission tomography PET-CT imaging could point to strains that are more likely to form persisters. This could provide valuable information about the underlying cause for unfavourable clinical outcome after treatment.

In this study we will be exploiting a novel replication reporter plasmid (pTiGc) and next-generation sequencing (NGS) data on sequence variants in *M. tuberculosis* for each patient to determine a correlation between patient outcome and bacterial data.

1.3. Hypothesis

We hypothesise that clinical isolates (taken at baseline) of individuals that show remaining lesion activity on PET-CT imaging and the presence of *M. tuberculosis* mRNA post TB treatment are predisposed to the formation of *M. tuberculosis* persisters.

1.4. Aim and Objectives

Aim:

To determine whether *M. tuberculosis* strains from TB patients who were considered cured, but have relapsed, or failed treatment, are more likely to be predisposed to persister formation than those who remained "cured".

Objectives:

Phenotypic

i. Assessing persister proportions in all clinical isolates (taken at baseline) from cured, recurrent/failed patient groups using fluorescence dilution (FD) and flow cytometry.
ii. Determine the correlation of persister formation with PET-CT scan classifications from all patient groups.

Genotypic

iii. Perform comparative next generation sequencing analysis of the isolates from all patient groups.
iv. Investigate strain evolution during treatment.
v. Determine whether sequence variation predisposes persister formation in clinical isolates (taken at baseline) from patient groups.

1.5. Thesis Overview

Chapter 2 **Literature review: Persisters and the genetic contributors to their formation.**

Chapter 3 **Assessment of persister proportions in clinical *M. tuberculosis* isolates (taken at baseline) from cured, recurrent/failed patient groups.**

Chapter 4 **Comparative next generation sequence analysis of clinical isolates (taken at baseline) from cured and recurrent/failed patient groups.**

Chapter 5 **Discussion and conclusions**

Chapter 6 **Reference list**

Supplementary Material

Appendices

Chapter 2

Literature review: Persisters and genetic contributors to their formation

Antibiotic resistance is undoubtedly one of the high-profile challenges human health currently faces. However, in the last decade, antibiotic tolerance has also come to the fore. This is as a result of identification of hard-to-treat bacterial infections, despite the lack of genetically encoded resistance, with the tendency to cause relapse. This chapter therefore provides an overview of the antibiotic tolerant subpopulation termed "persisters", the tools utilized to study them, triggers of persister formation and highlight the mechanisms persisters use to overcome environmental stress.

2.1. What is persistence?

2.1.1. Persistent infections vs antibiotic persistence

The term "persistent infections" is generally used when a pathogen resides in a host for prolonged time periods, independent of the host immune response (Balaban *et al.*, 2019). However, antibiotic persistence refers to a bacterial sub-population that is tolerant of prolonged antibiotic treatment, these bacterial populations are also referred to as 'persisters' (Lewis, 2010; Zhang, 2014a; Gollan *et al.*, 2019). Persistent infections are thought to be partly attributed to resistance or poor pharmacokinetics of the infecting populations as well as antibiotic persistence, which will be referred to as 'persistence' throughout (Levison, Matthew; Levison, 2013; Cicchese *et al.*, 2020).

2.1.2. Antibiotic resistance vs antibiotic persistence

Bacteria are defined as antibiotic resistant when the bacteria proliferate at lethal antibiotic concentrations. Resistance to a single drug or family of drugs is largely due to a genetic alteration of a non-resistant parent strain [de novo mutations or mutations acquired by horizontal gene transfer (HGT)](Figure 2.2.2)(Papavinasasundaram *et al.*, 2005). Once resistance has been acquired, the mutation is typically passed through the generations (Davies, J.; Davies, 2010; Gollan *et al.*, 2019). The level of resistance is usually measured utilizing a

minimum inhibitory concentration (MIC) assay either by culturing microorganisms in liquid media or on solid growth medium plates (Wiegand, Hilpert and Hancock, 2008). This refers to the lowest antibiotic concentration required to inhibit bacterial replication. Thus, the relation between MICs and antibiotic resistance is directly proportional (Balaban *et al.*, 2019).

'Persistence' is a phenomenon that occurs at a population level depicted by a biphasic killing curve indicative of a heterogeneous population (Fig 2.1) (Gold and Nathan, 2017; Balaban *et al.*, 2019). This heterogeneous population comprises of cells that are susceptible to lethal doses of antibiotics and a sub-population of antibiotic tolerant cells, where antibiotic tolerance is reversible and not genetically encoded (Helaine and Kugelberg, 2014). The progeny of persisters is drug-susceptible when regrown in the absence of antibiotics as depicted by Figure 2.1. Unlike resistant bacteria, persister cells are either slow or non-growing in the presence of antibiotics. Persisters are defined either as Type I ("triggered") persisters exit slowly from log to stationary phase, while type II ("spontaneous") persisters develop by phenotypic flipping in the absence of external stressors, which can switch back to a normal phenotype, and rise in numbers during the exponential growth phase (Levin-Reisman and Balaban, 2016; Gold and Nathan, 2017). However, it is worth noting, though, that persisters are much more complex than type I and type II, and are highly heterogeneous with variable metabolic activity.

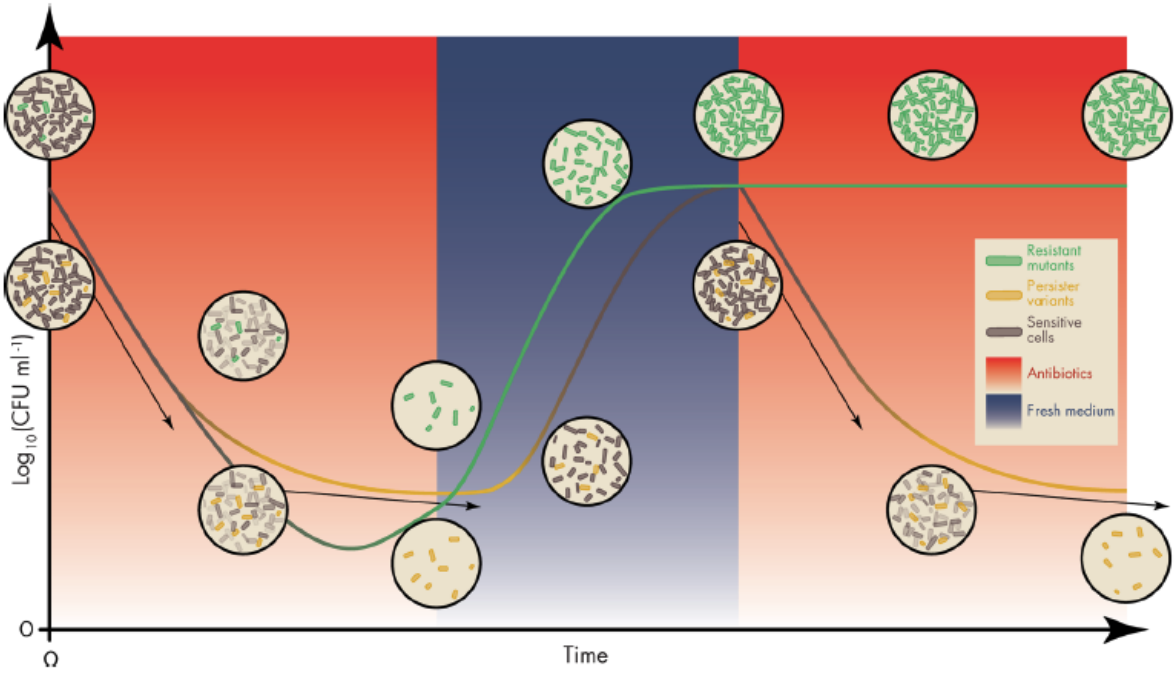


Figure 2.1: Bi-phasic killing curve and the difference between antibiotic resistance and antibiotic persistence. Review of drug treatment (red-plane) with persisters present (gold) or when antibiotic resistance occurs (green) within the bacterial population. Susceptible cells (grey) are destroyed in both cases after the initial procedure, as seen by the sharp decline arrow. However, resistant cells can

multiply during this process, while persisters can only survive (slowly falling arrow), resulting in a bi-phasic killing curve for persistence. Both populations of surviving cells multiply at natural rates until antibiotic therapy is stopped; persisters, however, can have a longer lag time. At the starting population (mid-blue plane), persisters will form a vulnerable population, while antibiotic-resistant cells would form a population made up entirely of resistant mutants. This distinction has an effect on a subsequent antibiotic therapy (second red plane)(adapted from van den Bergh, Fauvart and Michiels, 2017).

2.1.3. Antibiotic persistence vs tolerance and dormancy

Tolerance, persistence, and dormancy are all phenomena of survival to antibiotic treatment without an increase in the MIC (Meylan, Andrews and Collins, 2018). These three concepts are often used interchangeably, however, cells referred to as dormant are considered viable, yet do not replicate in optimal environmental conditions (Balaban *et al.*, 2019)(Fig 2.2). Dormant bacteria are usually tolerant of antibiotic treatment because of their growth arrest and decreased/inactive metabolism (Amato, Orman and Brynildsen, 2013). Literature suggests that the persistent state is an active stress response (Peyrusson *et al.*, 2020). Tolerance refers to the ability of an entire bacterial population to survive bactericidal activity due to having a lower killing rate, whereas persistence refers to the survival of a sub-population of non- or slowly growing drug-tolerant cells in response to antibiotic treatment as reflected by a biphasic killing curve. Mechanisms linked to dormancy and tolerance include reduced metabolic activity, which occurs in all three cell types (dormant, persistent and tolerant cells) (Greening, Grinter and Chiri, 2019). Two areas of interest in persister research are the molecular mechanisms of tolerance which enables the persister bacteria to survive stress conditions and the mechanisms that generate heterogeneity with the population. These mechanisms are studied utilizing various tools, which are further described below.

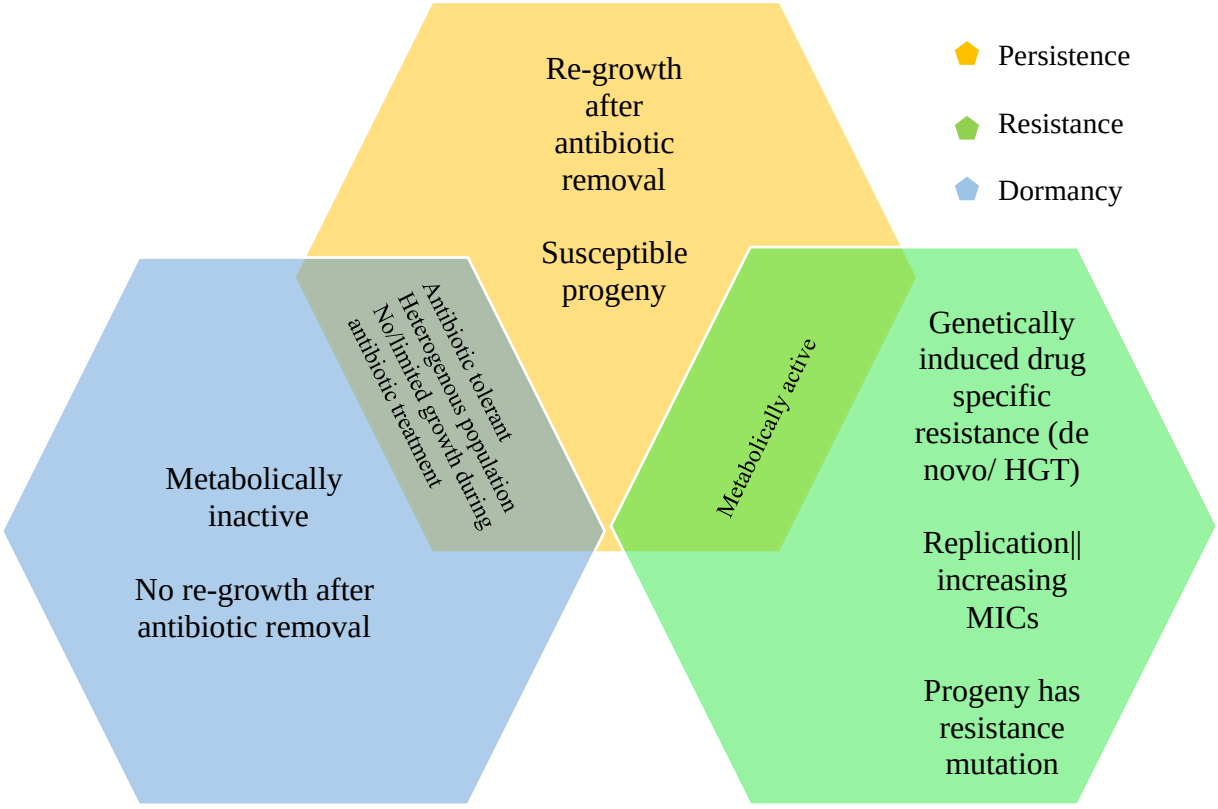


Figure 2.2: The overlapping characteristics of persistence, resistance, and dormancy. Created by JL Coetzee

2.2. Tools used to study persistence

The persistence phenomenon was discovered in the 1940s (Hobby, Meyer and Chaffee, 1942; Bigger, 1944). However, the transient nature of this subpopulation of bacteria, the limited evidence of their clinical impact and the lack of tools to study this hard-to-culture sub-population led to a major decline in interest. During the era of genetic engineering in the early 1980s, nearly 40 years after this phenomenon was first described, a breakthrough occurred (Moyed and Bertrand, 1983). The identification of three mutants in a gene termed *hipA* displaying the high persister (hip) phenotype reignited the interest in the field and subsequently tools were developed that could overcome challenges associated with studying persister bacteria (Moyed and Bertrand, 1983).

During the initial boom in renewed interest, bacterial populations were commonly environmentally or genetically manipulated to favor persister formation (McCune *et al.*, 1966; Wayne and Hayes, 1996). Although these techniques were essential in identifying pathways and stressors involved in mediating persister levels within a population, it became evident that

single-cell technology would be advantageous, as persisters are only a subgroup of bacterial populations (Verstraeten *et al.*, 2016). Thus, techniques such as microfluidics, flow cytometry, and fluorescent microscopy received preference (Fig 2.3.). Microfluidics is utilized to track behavior of single bacteria over time using time-lapse microscopy to analyze the history and fate of persister cells (Delincé *et al.*, 2016). This technique allowed researchers to show that persisters are pre-existing non-growers in *Escherichia coli* (Balaban *et al.*, 2004). Although microfluidics allows for the visualization and tracking of single cells, the number of cells that can be analyzed simultaneously is limited. Live cell microscopy in *M. tuberculosis* is time consuming as the doubling time of the organism is between 18 and 54 hours (Gill *et al.*, 2009). Therefore, flow cytometry provides an alternative that allows for high throughput assessment and sorting of single cells of interest, when combined with fluorescent reporters. The use of flow cytometry has been exploited to demonstrate that dormancy (lack of energy metabolism) is neither necessary nor adequate for persister formation (Orman and Brynildsen, 2013). Techniques that utilize flow cytometry in combination with a dual-fluorescence replication reporter that allows tracking of bacterial proliferation at the single cell level have provided new insights into persister populations (Roostalu *et al.*, 2008; Helaine *et al.*, 2010; Mouton *et al.*, 2016). Utilizing flow cytometry and fluorescence microscopy, Lerner *et al.*, 2017 found that necrotic macrophages provide a niche for replicating or non-replicating *M. tuberculosis* (Lerner *et al.*, 2017).

Omics techniques such as genomics, transcriptomics, and metabolomics coupled with improved bioinformatic analysis have led to the elucidation of changes in persister cells. An amalgamation of these techniques has been essential in identifying genes involved in persister formation, gene expression profiling, and persister evolution within a heterogeneous population (Cameron *et al.*, 2018; Stapels *et al.*, 2018; Choudhary *et al.*, 2019; Liu *et al.*, 2020). Furthermore, besides wet bench techniques utilized in tracking and analyzing persistence, mathematical modeling has been used for *in silico* persister behavior predictions and the design of molecular models (Lou, Li and Ouyang, 2008; Spalding *et al.*, 2018).

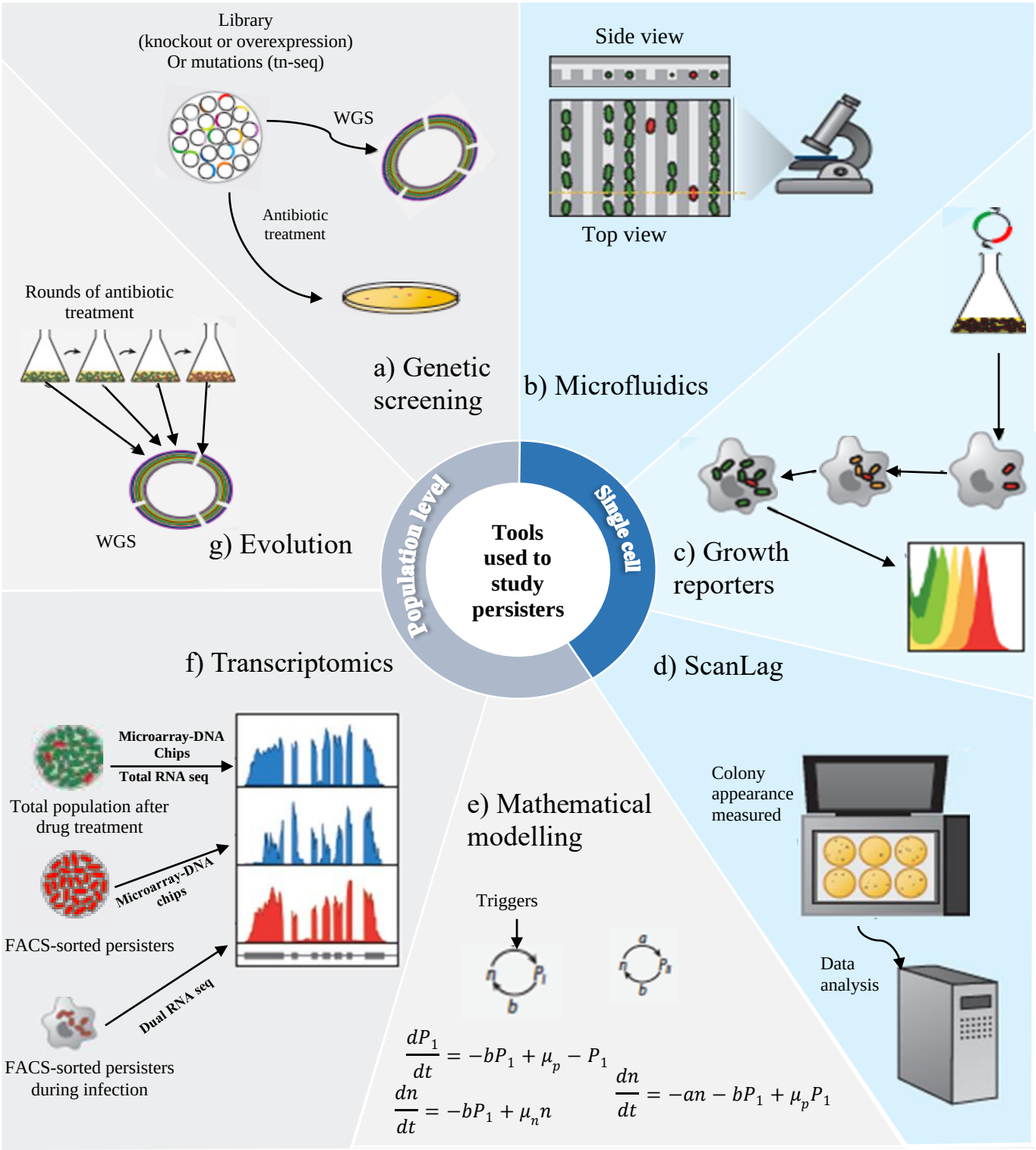


Figure 2.3: Tools utilized in understanding bacterial persistence. Non-single cell techniques are shown in grey. Single-cell techniques are shown in blue. (adapted from Gollan *et al.*, 2019)

2.3. Relevance of persisters in pathogenic diseases

The above-mentioned tools have been proven to be essential in the study of persisters in various pathogens including, but not restricted to, *Pseudomonas aeruginosa*, Salmonella spp., *Escherichia coli*, *Staphylococcus aureus* and *M. tuberculosis* as shown in Table 2.1.

Persister formation has generally been related to the inability of the host immune system to eradicate bacterial pathogens. This is due to host immune deficiency, bacterial evasion, and subversion (Monack, Mueller and Falkow, 2004; Grant and Hung, 2013; Fisher, Gollan and Helaine, 2017). However, more recently it has been shown that antibiotic treatment significantly increases the survival of a variety of pathogens in tissues (Agarwal *et al.*, 2016; Pham *et al.*, 2021). Whether these bacteria are persisters are unknown. Literature suggests that difficulty to distinguish between bacterial survival that is driven by poor pharmacokinetics (lack of penetration by antibiotics), phenotypic adaptation to stress or the induction of resistance genes *in vivo* gives rise to this ambiguity. Nonetheless, various studies show the role of persisters during infection (Table 2.1).

It has been speculated that antibiotic tolerance may be related to bacteria's capacity to survive during infection. Regardless of antibiotic therapy, chronic infections improve pathogen dissemination in host communities. The interaction between host immune systems and virulence factors of pathogens can bring about tolerant subpopulations (triggered or spontaneous) (Diard and Hardt, 2017; Bakkeren, Diard and Hardt, 2020). This is however difficult to test in clinical settings thus, macrophage, *in vitro* and *in vivo* models are used to analyse this hypothesis. *In vivo* models have observed long-term feacal shedding in persistent infections of *Salmonella enterica* subspecies within mice (Lawley *et al.*, 2008). Persister cells of *S. typhimurium* were recently found to occupy M2-like macrophages in the granuloma of murine spleens (Pham *et al.*, 2021). This observation correlates with *in vitro* studies that have shown that polarization of macrophages to a M2-like phenotype allows for increased pathogen survival within cells, including during antibiotic treatment (Stapels *et al.*, 2018; Thiriot *et al.*, 2020)

Cystic fibrosis leads to persistent lung infection and is connected to formation of biofilms containing heterogeneous populations of *P. aeruginosa* (Høiby, Ciofu and Bjarnsholt, 2010). To date, the clinical observations are unreproducible utilizing animal models, largely due to

the lack of standardized animal models (Moreau-Marquis, Stanton and O'Toole, 2008).This is however being addressed in recent studies (Geddes-McAlister, Kugadas and Gadjeva, 2019). Otherwise, *in vitro* biofilm systems have shown that large subpopulations of bacterial cells survived antibiotic treatment, hypoxia, and nutrient starvation although bacteria were genetically antibiotic susceptible. The rate of persister formation has been observed to be 1000-fold higher in biofilms compared to *in vitro* cultures (Spoering and Lewis, 2001). In a study by Ramsey *et al.*, 1999, patients receiving intermittent antibiotic treatment for cystic fibrosis displayed transient reductions of *P. aeruginosa* in sputum, however, reductions become less pronounced over time. Isolates obtained from patients who lacked genetic resistance are therefore indicative of evolution towards tolerance or persistence (Ramsey *et al.*, 1999). A follow-up study showed that high-persister mutants were isolated from these patients with cystic fibrosis (Mulcahy *et al.*, 2010).

S. aureus also forms persisters referred to as small-colony variants (SMVs), which are tolerant to host immune defense and antibiotic treatment (Sendi and Proctor, 2009). Infections thus result in persistent, relapsing infections such as osteomyelitis. After a long lag period before resuming growth after separation from mice or patient abscesses, and after growth under stresses such as low pH, SMVs have been observed (Vulin *et al.*, 2018). In accordance with previous literature, antibiotic tolerance has been observed in these SMVs, indicating a link between chronic infections, recurrence in a host, and antibiotic tolerance. (Vulin *et al.*, 2018).

Similar to the above-mentioned pathogens, *M. tuberculosis* causes recurrent infections that require lengthy antibiotic treatment. In patients who undergo antibiotic treatment, multiple subpopulations have been identified with variable molecular activity and a mosaic of resistance profiles (Wallis *et al.*, 1999; Fauvart, de Groote and Michiels, 2011; Liu *et al.*, 2016). Numerous *in vivo* and *ex vivo* studies have identified resistance-independent mycobacterial survival during antibiotic treatment (Dhar and J. D. McKinney, 2010). Studies utilizing animal models have shown that the caseum in lesions is a niche for drug-tolerant *M. tuberculosis* (Ramos *et al.*, 2017; Blanc *et al.*, 2018; Sarathy *et al.*, 2018; Sarathy and Dartois, 2020). A further example of *M. tuberculosis* persistence after anti-TB treatment demonstrated the presence of *M. tuberculosis* mRNA in culture-negative sputum samples and remaining lesion activity post pulmonary TB treatment (Malherbe *et al.*, 2016). *Hspx*, which has been related to long-term *M. tuberculosis* survival, was the most commonly identified transcript in the study (Yuan, Crane and Barry, 1996; Malherbe *et al.*, 2016). The detection of mRNA is suggestive

of ongoing transcription (based on the short-life of mRNA) (Pasipanodya *et al.*, 2007, 2010; Wejse *et al.*, 2008). Adaptation to various stress factors such as nutritional stress, oxidative stress, antibiotics and growth in mouse lungs have been found to increase persisters. The increased persister formation due to stresses have been shown to result from stress-induced noise in RNA expression (Leung and Lévesque, 2012). Infected mouse lungs have been shown to contain subpopulations of growing and non-growing, yet metabolically active, bacteria which are absent from mice lacking interferon gamma, a cytokine essential for anti-TB immunity (Manina, Dhar and McKinney, 2015). Survival of pathogens including *M. tuberculosis* is dependent on the bacterium's ability to adapt to stressors.

Table 2.1: Diseases known to be influenced by bacterial persisters (adapted from Van den Bergh *et al*.,2016 and Zhang.,2014).

Disease	Pathogen	References
Tuberculosis	*Mycobacterium tuberculosis*	Keren *et al*., 2011; Sarathy and Dartois., 2020; Liu *et al.*, 2016; Malherbe *et al*., 2016
Lyme Disease	*Borrelia burgdorferi*	Sharm *et al*., 2015; Feng, Auwaerter and Zhang., 2015; Feng J *et al*., 2020
Urinary tract infections	*Escherichia coli, Enterococcus, Pseudomonas aeruginosa, Chlamydia, Mycoplasma genitalium*	Keren *et al*., 2004 Michiels *et al*., 2016 Liebes *et al*., 2014; Zou and Shnaker., 2014; McAuliffe *et al*., 2006
Peptic ulcer	*Helicobacter pylori*	Fisher, Gollan and Helaine., 2017; Hathrobi *et al*., 2018
Bacteremia/sepsis	*Staphylococcus aureus*, Group B *Streptococcus*	Bigger.,1944; Lechner, Lewis and Bertram.,2012; Johnson and Levin.,2013
Endocarditis	*Streptococcus, Staphylococcus, Enterococcus*	Lueng and Lévesque.,2012; Elgharably *et al*.,2016
Brucellosis	*Brucella arbortus*	Amraei *et al*., 2020
Salmonellosis	*Salmonella enertica*	Vega *et al*., 2013; Arnoldini *et al*., 2014
Biofilms infections, periodontitis, and Prosthetic device infections	Multiple pathogens	Colon *et al*., 2014; Van Geelen *et al.*, 2020, Lewis., 2008

2.4. Triggers of persisters

The main triggers of persister formation are linked to environmental stressors like starvation, oxidative or extracellular metabolite signals (Harms, Maisonneuve and Gerdes, 2016).

2.4.1. Nutritional stresses

Various models have utilized nutrient starvation as an inducer of persister formation (Betts *et al.*, 2010; Grant *et al.*, 2013). The rationale behind this is that the antibiotic killing rate is dependent on the bacterial growth rate that is directed by carbon source availability and multiple other nutrient sources such as glycerol. Amino acid and nitrogen starvation examples have been found to increase persisters in *P. aeruginosa*, *S. aureus* and *E. coli* (Leung and Lévesque, 2012; Brown, 2019; Nguyen *et al.*, 2020). A carbon source shift has been associated with the stringent response and the observation of elevated levels of guanosine tetra- or pentaphosphate [(p)ppGpp] (Que *et al.*, 2013).

2.4.2. Oxidative, acidic and antibiotics

In addition to nutritional shifts, changes in acidic and oxidative stressors have been identified as triggers of persister formation. For example, treatment of *E.coli* cultures with reactive oxygen species (ROS) through addition of salicylate to the culture media led to an increase in persister formation (Vega *et al.*, 2012). Similarly, increased oxidative stress in *E.coli,* utilizing indole, was shown to promote persister formation (Vega *et al.*, 2012, 2013). Combination treatment with antibiotics and oxidative stress has been shown to cause DNA damage, increasing persisters (Wu *et al.*, 2012). Alterations in pH have been observed to promote persister formation, for example, Helaine *et al*. showed that pre-exposure of *Salmonella* to acidic environments (>4.5) significantly increased persister formation (Helaine *et al.*, 2014). Additionally, antibiotic treatment could increase persister formation. Specifically, *S. aureus*, and *E. coli* pretreated with sub-MIC concentrations of multiple antibiotics significantly increased persister formation (Dörr, Lewis and Vulić, 2009; Johnson and Levin, 2013; Kwan *et al.*, 2013; Gollan *et al.*, 2019). A similar observation was made when *M. tuberculosis* was treated with isoniazid (INH) (Walter *et al.*, 2015). Importantly, clinical strains of *M. tuberculosis* which were exposed to INH showed strong red fluorescence when using a dual reporter bacteriophage system indicating increased persister levels, as well as the subsequent emergence of resistant variants (Jain *et al.*, 2016). The dual reporter bacteriophage system consists of a green fluorescent reporter (GFP) and a red fluorescent protein (RFP). The bacteriophages follow a similar principle to FD where dilution of RFP is observed to determine persister proportions, those which retained RFP intensity were suggested to be persisters (Jain *et al.*, 2016). A similar occurrence was observed in a clinical strain of *Klebsiella pneumoniae*, where multi-drug tolerant persisters were identified utilizing killing experiments (Ren *et al.*, 2015). The impact of antibiotic treatment on persister formation is relevant to clinical settings

since antibiotic delivery and pharmacokinetics result in variable antibiotic concentrations being present in blood (Levison, Matthew; Levison, 2013). The low antibiotic concentration in blood could thus increase the persister population size and the pool from which resistant mutants could arise.

2.4.3. Host

During infection, bacteria experience a medley of stressors (mentioned above) in various degrees of intensity, which could favour persister formation. It is therefore suggested that persisters that form inside an infected host, unlike persisters formed under *in vitro* stress conditions, are uniquely adapted to host-specific triggers in response to which they are first formed. In the case of *M. tuberculosis*, a subpopulation of non-growing, but metabolically active mycobacteria was found to survive INH treatment and a combination of stressors in macrophages and after being transplanted from lungs of infected mice (Manina, Dhar and McKinney, 2015; Mouton *et al.*, 2016). These studies demonstrate that persisters formed in host environment are different than *in vitro* persisters as these are generally induced by a single stress factor.

2.5. Pathways and genes involved in persister formation in *M. tuberculosis*

Literature highlights that multiple interconnected pathways are responsible for the activation and formation of a persister state in response to the stress imposed (Helaine and Kugelberg, 2014; Amato and Brynildsen, 2015; Gollan *et al.*, 2019). Several of these have been identified with single-gene mutation studies, identifying these genes as possible drivers of persister formation (Glickman, Cox and Jacobs, 2000; Bryk *et al.*, 2008; Dhiman *et al.*, 2009). However, the majority of mechanisms exploited by persisters are not stressor-specific and occur in response to multiple stressors (Michiels *et al.*, 2016).

2.5.1. The Stringent Response

Persister research has identified starvation as an important trigger of persister formation (Potrykus and Cashel, 2008). The stringent response is a conserved stress response in all bacteria involving the construction of the hyperphosphorylated guanosine pentaphosphate or tretraphosphate (p)ppGpp which relocates cellular resources allowing the development of a VBNR state aiding in survival of cells to environmental stress (Gaca, Colomer-Winter and Lemos, 2015). The metabolism of (p)ppGpp is mediated by Rel/SpoT homolog (RSHs)

proteins (Atkinson, Tenson and Hauryliuk, 2011). Here RelA functions as a (p)ppGpp synthetase and SpoT is a bi-functional enzyme that is active in (p)ppGpp hydrolysis activity and weak (p)ppGpp synthetase. *M. tuberculosis* however, encodes a singular long RSH termed Rel_{Mtb}, that is conserved in all Mycobacterium species (Avarbock *et al.*, 1999; Prusa, Zhu and Stallings, 2018). Rel_{Mtb} was shown to complement a RelA *E .coli* mutant for growth in minimal media, confirming its ability to induce the stringent response (Avarbock *et al.*, 1999). The deletion of Rel_{Mtb} has been shown to produce a Rel_{Mtb} null mutant, suggesting its importance as the only functional (p)ppGpp synthetase in *M. tuberculosis*. Rel_{Mtb} has been found to be important in chronic infection of mice (Weiss and Stallings, 2013).

2.5.2. SOS response

M. tuberculosis has two DNA damage response pathways that are utilized during exposure to oxidative and antibiotic stress; the LexA/RecA-dependent SOS response and a LexA/RecA-independent pathway (Müller, Imkamp and Weber-Ban, 2018). Stressors generally cause DNA damage resulting in single stranded DNA (Dörr, Lewis and Vulić, 2009). RecA activates the LexA repressor, which leads to bacterial suppression of transcription (Dörr, Lewis and Vulić, 2009; Müller, Imkamp and Weber-Ban, 2018). Interestingly suppression/depletion of DNA gyrase in *M. tuberculosis* results in the activation of LexA/RecA-mediated SOS response and subsequently drug tolerance through formation of a persister subpopulation (Choudhary *et al.*, 2019).

2.5.3. Metabolic slowdown/shifting

Metabolic slowdown and shifting has been observed in persisters from *in vitro* models. This is because antibiotics and stressors such as Rifampicin target metabolically active bacteria (Hu *et al.*, 2000). For example, rifampicin kills metabolically active bacteria that are in log-phase of growth in comparison to the stationary phase bacteria that have shown to utilize metabolic slowdown via decreased replication is a successful mechanism for evading antibiotic killing and inducing antibiotic tolerance (Hu *et al.*, 2000). Similarly, Keijzer *et al*. observed a 10-fold reduction in ATP levels 7 days post rifampicin treatment of an *M. tuberculosis* Beijing strain (De Keijzer *et al.*, 2016). Genes involved in aerobic respiration, the TCA cycle and ATP synthesis are major pathways involved in energy metabolism which have been found to be down-regulated in response to stress (Walter *et al.*, 2015). Walter *et al* found that after 14 days post TB treatment initiation, Rv1304 (*atpB*), Rv1308 (*atpA*), which have been shown to be essential for *in vitro* growth (Dejesus *et al.*, 2017) were significantly downregulated. Cells in

exponential growth phase that were treated with arsenate, which decreases ATP concentrations, exhibited increased persister formation (Mohiuddin, Kavousi and Orman, 2020). Energy-related pathways include *menA*, required in menoquinone biosynthesis essential for ATP production, which inhibits survival of non-replicating persisters (Dhiman *et al.*, 2009). *SucB* (dihydrolipoamide acyltransferase), a subunit of the pyruvate dehydrogenase complex, under hypoxic and during *in vivo* infections was determined to be a drug target in persister baccili (Bryk *et al.*, 2008). Mutant *M. tuberculosis* with disrupted *cydC* that encodes a putative ATP-binding transporter system increased persister formation (Dhar and J. McKinney, 2010). When under stress, *M. tuberculosis* shifts from carbon metabolism, by upregulation of *tgs1* (triacylglycerol synthetase 1), to carbon storage of fatty acids (Tudó *et al.*, 2010; Baek, Li and Sassetti, 2011). Downregulation of various genes involved in mRNA synthesis and protein synthesis such as *tuf, gyrA, gyrB* has been associated with metabolic slowdown under antibiotic stress (Walter *et al.*, 2015).

2.5.4. Transcriptional and post-transcriptional gene regulation

The central dogma of molecular biology essentially explains that "DNA makes RNA, and RNA makes proteins". Based on this logic, changes at the transcriptional and post-transcriptional level during a stress response are an essential mechanism of the formation of non-growing or slowly growing *M. tuberculosis* persisters. Rv1152, which forms part of the Gnt transcriptional regulator protein family, WhiB, transcriptional factors and sigma factors function as stress regulators (Francez-Charlot *et al.*, 2009; Casonato *et al.*, 2012; Zeng *et al.*, 2016; Goossens, Sampson and Rie, 2021). Post-transcriptionally, toxin-antitoxin (TA) modules and small RNAs are thought to induce persistence. Specifically, *M. tuberculosis* Rv1152 has been shown to be involved in various persister models, namely antibiotic tress, hypoxia and nutrient starvation (Keren *et al.*, 2011; Iacobino *et al.*, 2021), overexpression of Rv1152 has been shown to increase tolerance of mutant *M. smegmatis* to vancomycin *in vitro,* cell surface and acid stress (Zeng *et al.*, 2016), while WhiB was involved in pathogenesis (Steyn *et al.*, 2002), cell division (Gomez and Bishai, 2000; Rybniker *et al.*, 2010)and stress response (Geiman *et al.*, 2006).

2.5.4.1.WhiB-like family genes

The WhiB-like (Wbl) family of genes is only found within Actinobacteria. WhiB was shown to be essential in *M. tuberculosis* virulence and antibiotic tolerance. Wbl proteins were found to be O_2-and NO-sensitive [4Fe-4S] (Smith *et al.*, 2010; Kudhair *et al.*, 2017). Recent developments shed light on Wbls as transcriptional regulators and sensors of O_2 or nitric oxide

(NO) (Kudhair *et al.*, 2017). In *M. tuberculosis* WhiB1 has been shown to be essential (Rv3219) (Smith *et al.*, 2010), WhiB3 (Rv3416) and WhiB4 (Rv3681c) were suggested to play a role in regulating virulence, WhiB3 and WhiB7 were found to play a role in antibiotic resistance while WhiB2 (Rv3260c) has been suggested to be essential and play a role in the regulation of cell division (Bush, 2018). WhiB1 was shown to be increasingly sensitive to NO, which is important for *M. tuberculosis* as NO is produced by lung macrophages (Kudhair *et al.*, 2017). Studies have shown that WhiB7 null mutants have increased drug susceptibility *in vitro* (Morris *et al.*, 2015). Studies have suggested that Whib7 may result in a drug-tolerant state by upregulation of drug efflux pumps (tap or Rv1258c) and through rearrangement of cellular processes that compensate for metabolic shifting, induced under antibiotic stress (Morris *et al.*, 2005). Mice infected with WhiB3 mutants were found to live longer when exposed to antibiotic stress, compared to mice infected with wild-type H37Rv. Considering WhiB3 encodes for a 4Fe-4S redox sensor protein, it is suggested to modulate drug sensitivity through regulating the balance between redox and bioenergetic homeostasis, both of which are affected during drug response. WhiB3 has been observed to regulate essential metabolic pathways such as glycolysis, the TCA cycle, the pentose phosphate pathway and amino acid biosynthesis (Saini *et al.*, 2016). The essential role the WhiB-like family of genes play in antibiotic resistance, NO sensitivity, antibiotic tolerance, metabolic shifting and increased survival upon chronic infections suggests their importance in *M. tuberculosis* persister formation as persister utilize these mechanisms to evade host defenses.

2.5.4.2.Sigma (σ) factors

Sigma factors are subunits of bacterial RNA polymerases that are responsible for binding RNA polymerases to form a holoenzyme that determines promoter specificity. These factors play an important part in post-transcriptional modifications such as acetylation and phosphorylation through protein kinases and anti-sigma factors. Post-translational modifications have been observed to increase persister formation (De Keijzer *et al.*, 2016). Generally, bacteria encode a single sigma factor regulating transcription of essential housekeeping genes, and a variety of sigma factors whose expression is stress-activated (Boldrin *et al.*, 2020). *M. tuberculosis* encodes for 13 sigma factors (SigA-M) (Rodrigue *et al.*, 2006). SigE was found to be imperative for response to acidic environments, human macrophage growth, detergent mediated surface stress, and oxidative stress (Manganelli *et al.*, 2001; Schnappinger *et al.*, 2003; Manganelli, 2014; Chauhan *et al.*, 2016). SigE is also responsible for transcription of SigB and the two component system MprAB (Dainese *et al.*, 2006). In response to antibiotic

treatment, expression of σ^B, σ^E, σ^F, σ^G, σ^H, σ^I and σ^J is increased and the expression of σ^A expression is decreased (Walter *et al.*, 2015; Miryala, Anbarasu and Ramaiah, 2019).

2.5.4.3.Toxin-antitoxin (TA) modules

TA modules are comprised of two genes: one encoding a toxin protein affecting bacterial growth and the other an antitoxin element (RNA, type I; or protein, type II) which nullifies the toxin under favorable conditions (Hall, Gollan and Helaine, 2017). A considerably large number of TAs (at least 88) has been identified in *M. tuberculosis* (Ramage, Connolly and Cox, 2009; Sala, Bordes and Genevaux, 2014). The majority of the TAs in *M. tuberculosis* are Type II TA systems. The high number of TA systems present in the *M. tuberculosis* genome suggests a highly important role in host-pathogen interactions (Yu *et al.*, 2020). Under specific environmental stressors, antitoxin degradation is induced, allowing the toxin to take effect on specific targets like the ribosome, specific transfer ribonucleic acid's (tRNAs) or messenger ribonucleic acids (mRNAs) (Slayden, Dawson and Cummings, 2018; Barth *et al.*, 2019; Barth and Woychik, 2020), which results in a slowdown of metabolism or dormancy. TA systems have been shown to increase the subpopulation of peristers stochastically in the presence or absence of stressors (Kim, Choi and Hwang, 2016; Yu *et al.*, 2020). TA modules have been assessed in response to antibiotic exposure where toxins such as MazF, Rv1577x, Rv2651c, and Rv0366c confer drug tolerance across multiple drug classes (Singh, Barry and Boshoff, 2010; Tiwari *et al.*, 2015; Tandon *et al.*, 2019). TA modules have also been detected in environmental stresses encountered during infection, directing bacteria toward an increased and constant dormant state during latent TB (Slayden, Dawson and Cummings, 2018).

2.5.4.4.Small RNAs (sRNAs)

To date sRNAs have been poorly studied in *M. tuberculosis* (Gerrick *et al.*, 2018). However, in non-mycobacterial species sRNAs regulate gene expression by binding mRNA, constricting mRNA translation increasing mRNA degradation (Storz, Vogel and Wassarman, 2011). sRNAs were found to play a role in regulation of genes associated with the efflux pumps, transport proteins, membrane proteins, metabolic enzymes and the mycobacterial cell wall (Chan *et al.*, 2017; Dersch *et al.*, 2017; Felden and Cattoira, 2018). In *E. coli* sRNAs were found to regulated persistence to multiple antimicrobials through the reduction of cellular metabolism (Zhang *et al.*, 2018).

2.5.4.5.Protein post-translation modifications

The role of protein post translation modifications as a mechanism of persister formation has been largely overlooked, as only one study to date addresses the topic. Keijzer *et al*. determined

that phosphorylation occurs in 132 unique proteins of *M. tuberculosis* and demonstrated that phosphorylation of Rv2986c/HupB upregulates protein expression thus increasing iron storage in a persister-like subpopulation (De Keijzer *et al.*, 2016). Post translation modification such as acetylation, pupylation (Barandun, Delley and Weber-ban, 2012)and phosphorylation (Sajid *et al.*, 2011) has recently been suggested to be a silent contributor to mycobacterial virulence, metabolism and pathogenesis. Phosphorylation has been observed to affect LexA binding to RecA, the inhibition has identified to increase DNA damage and subsequently Rif-resistance (Wipperman *et al.*, 2018). Subsequently this could suggest a possible mechanism for increased persister formation as RecA/LexA has been identified as a mechanism that increases persister antibiotic tolerance.

2.5.5. Genetic adaptations

Tolerance/persister formation is a phenotypic phenomenon, however, recent studies have suggested that specific variants in genes that are essential for *M. tuberculosis* increases the predisposition of a population to form persisters under stress conditions. Specifically, high persister "hip" mutants of genes are connected to carbon metabolism pathways and lipid biosynthesis, under selection by lethal doses of rifampicin and streptomycin (Torrey *et al.*, 2016). In hip mutants, *gltA1* (Rv1131) was upregulated which is associated with metabolic shifts in carbon away from the TCA cycle possibly by changing the propionyl-CoA metabolism. Similarly in hip mutants decreased activity of *fadE30*, a probable acyl-CoA dehydrogenase, is likely to prove the reduction in lipid catabolism, coherent with shifting from the TCA cycle to lipid synthesis (Torrey *et al*., 2016). Thus, FadE30 is observed to be essential for *M. tuberculosis* survival in macrophages (Rengarajan, Bloom and Rubin, 2005). Upregulation of *icl* and *tsg1* in hip mutants is suggestive of a redirection of carbon sources to lipid synthesis as well, as it supports glyoxylate bypass (Torrey *et al.*, 2016).

Glycerol was suggested to be an unimportant carbon source for *M. tuberculosis* as glycerol kinase (*glpK*) essential for glycerol catabolism is dispensable in an *in vivo* mouse model (Beste *et al.*, 2009; Pethe *et al.*, 2010). However, bacteria containing frameshift mutations in *glpK* gene have been shown to be multidrug tolerant *in vitro* and in clinical isolates (Bellerose *et al.*, 2019; Safi *et al.*, 2019). Interestingly variants accumulate during drug treatment suggesting their relevance in treatment failure and relapse.

2.6. Identifying genes associated with persister formation

The lack of knowledge regarding the impact genetic adaptions have on persister formation indicates the need to uncover the underlying genetic mutations which predisposes persister formation. Methods elucidating genes involved in persister formation includes inducing persisters with transposon sequences followed by next generation sequencing or RNA sequencing as a targeted approach determining genes involved in persisters. For example Carey *et al*., identified essential genes for *in vitro* growth of clinical isolates belong to the modern *M. tuberculosis* lineage, by utilizing a modified Himar transpon coupled with comparative WGS of H37Rv (Carey *et al.*, 2018). Although transposon sequencing (tn-seq) allows for the identification of functional genetic differences, the methodology is more technically challenging than that of WGS. WGS could provide an unbiased approach to identify genes involved in persister formation as it provides a comprehensive view of the genome (coding, non-coding and mitochondrial DNA). WGS could identify genes which could be included for targeted sequencing in a point of care setting (Nimmo *et al.*, 2019). Table 2.3. summarizes the advantages and disadvantages of utilizing tn-seq, RNA-seq and WGS for the identification of genes involved in persister formation.

Table 2.2 High throughput methods for identifying genes related to persister formation and their advantages and disadvantages

Method	Advantage	Disadvantages
Tn-seq	Study mutual freq of transposons	Limited bacterial studies
	Used to deduce fitness of genes within microorganisms	Errors during PCR amplification can lead to inaccurate sequence reads
	Robust, reproducible and sensitive	tn-seq analyses is not normalized
		Gene deletions results in false gene essentiality
		Gene duplications results in false gene non-essentialiy calls
		Sequence variants which changes transposon insertion sites creat subtle errors
		Time consuming
		Induces persistes by creating mutants
RNA-seq (whole exome sequencing)	Signal-to-noise ratio is low	RNA is easily contaminated
	Multiple publicly available databases	RNA is unstable
	Bacterial-to-human transcriptomic size ratio is greater	
WGS (DNA)	DNA is stable	Cannot deduce gene fitness
	High capacity to receive genetic material from archived samples	Identification of genes are coverage dependent
	Able to extract bacterial genomic DNA from sputum.	Certain regions/genes in the genome covered but not reliable in mapping
	Readily available	
	Low cost	
	Identify genes within a natural occurance	
	Robust, reproducible, sensitive	
	Identification of possible variants while sequencing can guide point of care	

2.7. Conclusion

In this review we highlighted that the subpopulation of bacteria termed persisters is extremely complex and is not completely understood. The large number of stressors that trigger persisters coupled with various genetic mutations linked to the multiple pathways which persisters utilize makes characterizing persisters in *M. tuberculosis* challenging. Their small population size makes investigating genes involved in persister formation in a non-targeted approach challenging. There is a lack of an efficient method to identify persister-related genes. WGS of hip mutants was thought to hold promise, but hip mutations are thought to create persisters in a way that does not reflect what occurs during infection. It has been speculated that if hip mutants are obtained from clinical isolates, then the mechanism of formation could be more productively investigated. This review highlights that future research should focus on

exploiting multi-omics approaches to elucidate mechanisms involved in persister formation, understanding the genetics of persisters and the role of epigenetic changes.

Chapter 3

Assessment of persister proportions in baseline clinical *M. tuberculosis* isolates from cured and recurrent/failed patient groups.

3.1. Introduction

It is thought that persistent mycobacteria arise in response to environmental stressors encountered in the host and adopt a slow or non-replicating state (North and Jung, 2004; Liu *et al.*, 2020). This small, viable, but non-replicating (VBNR) population is likely to be antibiotic-tolerant. Currently the majority of drug therapies target actively growing bacteria, however persister bacteria comprise an important subpopulation of bacteria that is recalcitrant to antibiotic treatment (Gill *et al.*, 2009). Importantly, VBNR bacterial populations are phenotypically drug tolerant, but not genetically resistant. Drug tolerant populations could contribute to the requirement for lengthy drug treatment and could themselves give rise to genetically resistant progeny.

Due to the difficulty of isolating persistent mycobacteria, little is known about them. Recent research has developed and used a technique known as Fluorescence Dilution (FD) to identify a VBNR Salmonella population in infected macrophages. (Helaine *et al.*, 2010), and to show for the first time that the internalization by macrophages induced the formation of VBNR populations (Helaine *et al.*, 2014) (Figure 3.1). Importantly, FD has recently been successfully adapted and optimized for use in mycobacteria (Mouton *et al.*, 2016, 2019). FD utilizes a dual-reporter plasmid containing GFP, a constitutively expressed green fluorescent protein and TurboFP635 that is under control of an inducible promoter (Seeliger *et al.*, 2012; Mouton *et al.*, 2016). Upon removal of the inducer a decrease in the inducible TurboFP635 fluorescence over time serves as a measure of mycobacterial replication within a population at a single cell level, bacteria that retains their maximum red fluorescence intensity represents non-growing *M. tuberculosis*. This technique has revealed considerable heterogeneity in intracellular mycobacterial replication (Mouton *et al.*, 2016).

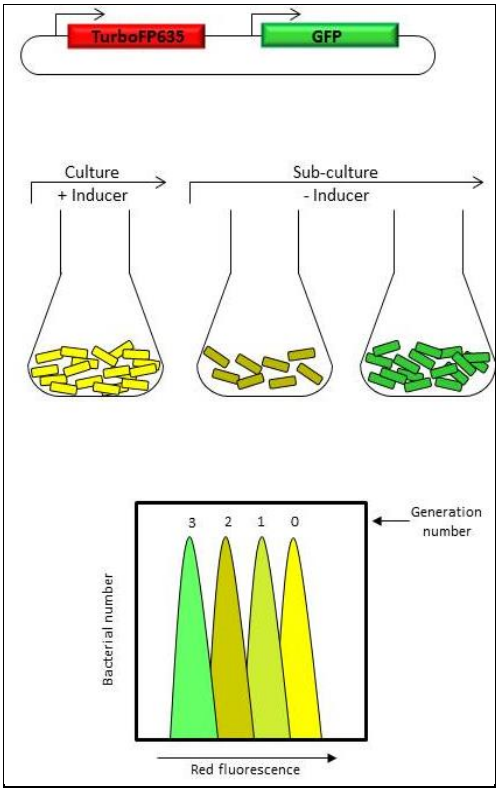


Figure 3.1. The principle of FD (Mouton *et al.*, 2016). A) pTiGc dual reporter plasmid schematic, with GFP being constitutively expressed providing a marker of viability, and TurboFP635 under control of a riboswitch promoter inducible by theophylline B) Utilization of FD assessing replication dynamics, where the fluorescence intensity of TurboFP635 is diluted as bacteria replicate. C) Flow cytometric detection of TurboFP635 intensity in bacterial population replication.

Here we aimed to apply FD in combination with flow cytometry to determine the underlying persister formation in strains from individuals previously infected with *M. tuberculosis* that have remaining lung lesion activity and the presence of *M. tuberculosis* mRNA. This could help to explain why apparently curative treatment for pulmonary tuberculosis (PTB) is not eradicating all of the *M. tuberculosis* bacteria in most patients (in the context of non-resolving and intensifying lesions on Positron Emission Tomography/Computed Tomography [PET/CT] images).

3.2. Methods and materials

All procedures involving the use of live cultures of *M. tuberculosis* have been performed in Biosafety Level 3 (BSL3) laboratories in the Division of Molecular Biology and Human Genetics, Stellenbosch University, Tygerberg, South Africa. Strict BSL3 safety precautions have been implemented as defined in the in-house BSL3 Departmental SOP. Care was taken to prevent any contamination of the bacterial cultures or surrounding environments to pathogenic mycobacterial cultures.

The work represented in the present study forms part of larger projects which received ethical approval from the Stellenbosch University Health Research Ethics Committee under the title: "*Mycobacterium tuberculosis* Biomarkers for diagnosis and cure", ethics number N10/01/013, and entitled "An investigation into the evolutionary history and biological characteristics of the members of genus Mycobacterium, with specific focus on the different strains of *M. tuberculosis*, other members of the *M. tuberculosis* complex and non-tuberculosis mycobacteria (NTM)", ethics number N14/03/022.

Bacterial samples were obtained in a parent study from patients at health clinics in Cape Town, Western Cape, South Africa during 2016. Patients received PET/CT scans at diagnosis (dx) and at later points during treatment to assess the outcome of the treatment. Sputum samples were subjected to sputum smear microscopy and cultured in a BD BACTEC™ MGIT™ 960 at Stellenbosch University. Confirmed *M. tuberculosis* isolates were subjected to standard genetic characterisation by RFLP analysis and spoligotyping (Van Embden *et al.*, 1993; Kamerbeek *et al.*, 1997).

Methods for preparing reagents and buffers used in this chapter are presented in Appendix A.

3.2.1. Plasmid constructs

Plasmids utilized in this study are listed in Table 3.2.1. Briefly, these were pST5552 (carrying hsp60(ribo)-gfp under the control of the theophylline-inducible riboswitch promoter) (Seeliger *et al.*, 2012),pSTCHARGE (encoding the inducible TurboFP635) and the pTiGc plasmid (carrying the inducible TurboFP635 and constitutive GFP) (Mouton *et al.*, 2016).

Table 3.2.1. Plasmids and strains adapted from Mouton *et al.*, 2016

Plasmid/Strain	Description	Source
pST5552	hsp60(ribo)-gfp (inducible GFP under control of theophylline-inducible riboswitch), $Kan^{R[1]}$, episomal	Seeliger et al. (2012), Addgene plasmid number 36255; Mouton *et al.* (2016)
pSTCHARGE	hsp60(ribo)-turboFP635 (inducible TurboFP635 under control of theophylline-inducible riboswitch), Kan^{R}, episomal	Mouton *et al.* (2016)
pTiGc	*leuD* and *panCD*-deficient attenuated strain of *M. tuberculosis* H37Rv carrying hsp60(ribo)-turboFP635 hsp60-gfp, Kan^{R}, episomal	Mouton *et al.* (2016)
H37Rv 102J23	Parent strain	Laboratory strain
SAMMtb	*leuD* and *panCD*-deficient attenuated strain of *M. tuberculosis* H37Rv, $Hyg^{R[2]}$	(Sampson *et al.*, 2004, 2011)

[1] Kanamycin resistant; [2] Hygromycin resistant

3.2.2. Bacterial strains and culturing

All reagents utilized in this study were purchased from Sigma-Aldrich, unless stated otherwise. Bacterial strains used in this study are listed in Table 1. The clinical *M. tuberculosis* isolates used in this study were received from the Catalysis TB-Biomarker Consortium (Malherbe *et al.*, 2016), obtained from smear positive sputum samples (Table 3.2). Dx isolates were utilized for downstream analysis. These isolates were classified into cured or failed/recurrent based on clinical outcome, Gene Xpert results, and PET/CT imaging. Results from Malherbe *et al*., 2016 showed that a fraction of individuals who has undergone the 6-month treatment for *M. tuberculosis* have remaining lesion activity based on PET/CT imaging of the lung and the presence of *M. tuberculosis* mRNA in sputum and bronchoalveolar lavage samples, suggestive of persister *M. tuberculosis*. PET/CT images were taken before treatment (Dx), 1 month after treatment was commenced (M1) and six months after the treatment commenced (M6). At M6, lesions were identified to show minimal or no activity (indicative of being clinically cured), showed moderate to high lesion intensity as compared to Dx scans (indicative of failed patients)

or showed new lesion activity representing re-current *M. tuberculosis* infection in patients who were re-diagnosed with pulmonary-TB 1 year after treatment. The isolates were assessed using the Ziehl Neelsen staining procedure for the identification of mycobacterium screening (Allen, 1992).

Liquid cultures of dx mycobacterial isolates were cultivated in 7H9 (Becton Dickinson, NJ, United States) complemented by 10% oleic acid-albumin-dextrose-catalase (OADC; Becton Dickinson, NJ, United States), 0.2% (v/v) glycerol (Sigma-Aldrich) and 0.05% (v/v) Tween 80 (Sigma-Aldrich) (7H9-OGT) and incubated at 37°C until OD_{600}=0.8-1. Electro-competent mycobacteria were prepared and transformed with pTiGc as described by Snapper *et al.* (1990). Two hundred microliters of the newly transformed mycobacterial isolates were plated onto 7H10 solid media (Becton Dickinson, NJ, United States) supplemented with 10% OADC, 1% (v/v) glycerol, and appropriate antibiotics, and cultured at 37°C for approximately 4 weeks.

To confirm the presence of the pTiGc plasmid, single colonies were picked into 96-well plates and cultured at 37°C for 6-10 days in 200 µl 7H9-OGT with appropriate antibiotics. The colonies were then duplicated into plates with and without 4 mM theophylline (Sigma-Aldrich) and cultured for a further 48 hours (h) before being read on a plate reader. The following settings were used for analyses using a FLUOstar Omega multi-mode 96 well microplate reader (BMG LABTECH, Offenburg, Germany); optic settings for GFP (green) detection were set at an excitation of 485 nm and an emission of 520 nm, the gain function was set at 1692. For TurboFP635 (Red), excitation was set at 584 nm and the emission set to 640 nm, the gain function was set at approximately 2800. The plates were shaken at a frequency of 500 rpm on a double orbital shaking mode for 2 seconds before the plate was read. To detect positive transformation a fold change of fluorescent intensity was calculated for both GFP and TurboFP635 of induced cells divided by the un-induced cells (data not shown). Fold-change calculations:

GFP:

Background -adjusted fluorescence intensity was initially determined by the background green intensity reading from wild-type (wt) from the raw fluorescent data per isolate. Fold induction was subsequently determined by dividing the background adjusted value by the background adjusted un-induced value per strain. GFP fold induction was expected to be -1.0 as GFP is constitutively expressed in all isolates at similar levels. Indicative of live bacterial populations.

TurboFP635:

Backgound-adjusted fluorescence and fold change for TurboFP635 was determined similarly to GFP. However, the fold change for TurboFP635 was expected to be ≥ 6 indicative that the TurboFP635 protein is induced by theophylline and can reliably track 6 generations of mycobacterial replication.

Table 3.2.2. Mycobacterial clinical isolates selected for this study (modified from Malherbe *et al.*, 2016)

Cured	Failed/Recurrent
C-S4dx	F/R-S43dx
C-S5dx	F/R-S43w24*
C-S29diag	F/R-S93dx
C-S41w4	F/R-S101dx
C-S105dx	F/R-S112dx
C-S105d2*	F/R-S130dx
C-S126dx	F/R-S137dx
C-S153dx	F/R-S152dx
C-S159dx	F/R-S163dx
C-S159x4	F/R-S163w24*
	F/R-S168dx
	F/R-S168w24*
	F/R-S169dx

*indicates isolates used for WGS only

3.2.3. Growth curve analysis of transformed clinical isolates

To assess whether carriage of the pTiGc plasmid imposes a fitness cost on the bacterial isolates, *in vitro* growth curves were used. Cultures were grown to an OD_{600nm}=0.6-0.8. On the day of setup, both the wild type (WT) clinical isolates and clinical isolates that were transformed with the pTiGc plasmid (number dx-pTiGc) were sonicated for 12 minutes at 37 kHz in a water bath (UC-1D; Zeus Automation) at room temperature and filtered through a 40 um cell strainer. The initial OD_{600nm} was adjusted to 1 (approx. $1X10^8$ CFU/ml). Thereafter, 100 µl bacteria (OD_{600nm}=0.1) was added to each well containing 100 µl 7H9-OGT containing the appropriate antibiotics. The growth curves were performed in a NUNC 96-well black, clear-bottomed plate and OD_{600nm} readings were taken every second day using the following parameters. Data was

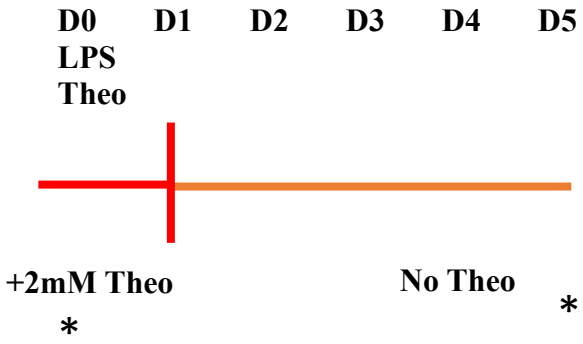
obtained utilizing a Polarstar Omega 96-well microplate reader (BMG Labtech, Ortenberg, Germany). The plate was read at an absorbance of 600 nm and shook at 200 rpm on the double orbital shaking setting for 5 seconds before the reading was done. This was done at 37°C with an absorbance path length of 120.

3.2.4. Infection of transformed isolates into THP-1

THP-1 cells (ATCC TIB-202) were cultured in Roswell Park Memorial Institute-1640 medium (RPMI), supplemented with 10% heat- inactivated fetal bovine serum (FBS) (R10) at 37°C in a 5% CO_2 atmosphere (both reagents were obtained from ThermoFisher). Cells were passaged every 2-4 days. For infections, cells were seeded at $1.25x10^5$ per well in 96 well plates and differentiated with 50 ng/ml of phorbol-12-myristate-13-acetate (PMA, Sigma-Aldrich) before incubation at 37°C with 5% CO_2 atmosphere for 3 days. Following the incubation, the R10 media containing PMA was replaced with fresh R10 and the macrophages were allowed to recover for 24h. One hour prior to infection THP-1 cells were stimulated with R10 containing 100 ng/ml lipopolysaccharide (LPS, Sigma-Aldrich) and incubated at 37°C, which induces macrophage activation. Before infection the R10 containing LPS was removed and replaced with 100 µl fresh R10 media.

Dx-pTiGc and SAMMtb-pTiGc (control) bacterial cultures were induced 7 days prior to infection with 4 mM theophylline in 7H9-OGT with the required antibiotic supplementation. On the day of infection mycobacteria were prepared for infection. Cultures were sonicated at 37 kHz for 12 minutes at room temperature in an ultrasonic bath to disperse clumps, and thereafter filtered through a 40 µm cell strainer. Bacterial OD_{600nm} was assessed with a spectrophotometer (Thomas Scientific) and adjusted to OD_{600nm}= 1 in R10 containing 4 mM theophylline. Thereafter, bacteria were added to the macrophages at a multiplicity of infection (MOI) of 10:1, and incubated at 37°C in a Thermo Ster-Cycle 5% CO_2 incubator (Marshall Scientific) for 3h in the presence of 2mM Theophylline. Following uptake, the cells were washed with 200 µl phosphate buffered saline (PBS, ThermoFisher) once before replacing the media with R10 media containing 100 U penicillin/streptomycin. This was followed by incubation at 37°C in 5% CO_2 for 1 hour to kill any non-phagocytosed, extracellular bacteria. Cells were washed three times with 200 µl PBS before adding fresh R10, containing 2 mM theophylline, to maintain expression of TurboFP635 for 24h after infection. After 24h, R10 media containing 2 mM theophylline was replaced with R10 without 2 mM theophylline (R10-

Theo) (Fig 3.2.1). To recover mycobacteria for flow cytometry analyses, the macrophages were lysed by the addition of 300 µl sterile distilled water. Intracellular bacteria were recovered from infected macrophages along with parallel *in vitro* bacterial cultures at day 0, and day 5.



Abbreviations: *= Sampling time points, D= day

Figure 3.2.1. Theophylline induction during macrophage infection

3.2.5. Flow cytometry sample preparation, acquisition and analysis.

In vitro cultured bacteria or intracellular bacteria (from lysed macrophages), were pelleted and fixed in 4% formaldehyde (Sigma-Aldrich) for 30 minutes, washed once by centrifugation at 10 000 rpm for 5 minutes and resuspended in PBS-Tween before storing at 4°C. On the day of flow cytometry analyses the samples were pelleted at 10 000 rpm for 5 minutes and resuspended in 300µl PBS. Samples were filtered through a 35 µM filter immediately prior to running on the flow cytometer. A volume of 5 µl microsphere standard beads (6.0 µm) from the LIVE/DEAD BacLight Bacterial Viability kit was added to samples after filtering (ThermoFisher,https://www.thermofisher.com/order/catalog/product/L7012#/L7012).

Samples were analysed on the BD FacsJazz™ flow cytometer (Becton Dickinson, United States) that is located in the BSL3 facility within the Division of Molecular Biology and Human Genetics, Stellenbosch University. The forward scatter (FSC) and side scatter (SSC) were investigated, as well as the fluorescent intensity of GFP which was captured by excitation at 488 nm, using a 530/30 filter and TurboFP635 fluorescence intensity was captured by excitation at 561 nm, using a 610/20 filter. Compensation was carried out using single color and unlabeled controls in each experiment. For samples from both *in vitro* cultures and bacterial samples recovered from macrophages, 20 000 events were captured.

FlowJo v10 software was used to analyze flow cytometric data. A primary gate was set according to FSC/SSC properties, followed by gating on the GFP-positive (live) population. The TurboFP635 fluorescence intensity of the population was then analysed (Fig 3.2.2).

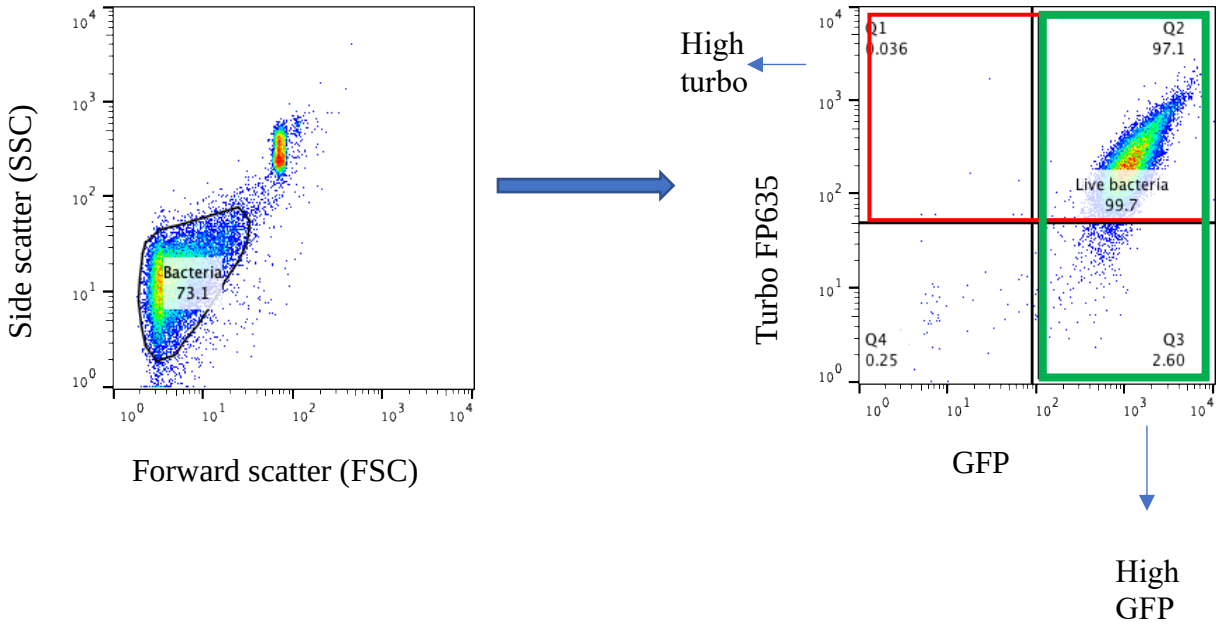


Figure 3.2.2: Gating strategy for flow cytometry created by JL Coetzee

3.2.6. Determination of bacterial uptake and survival within macrophages utilizing counting beads.

Enumeration of bacterial uptake by and survival in THP-1 cells post infection was determined by exploiting the LIVE/DEAD Backlight Bacterial Viability and Counting kit. The beads from the kit served as an alternative to using colony forming units (CFU's) for determining uptake of bacteria from inoculum, since it would provide a more rapid and accurate readout compared to CFU's. Fig 3.2.3 depicts the gating strategy for the bead and bacterial gating. Calculating the bacteria/ml was dependent on the bacterial events captured, the number of bead events captured, the number of beads added, sample volume and the dilution factor.

Bacteria/ml=

$$\frac{Bacterial\ events}{Bead\ events} \times \frac{number\ of\ beads\ added}{Sample\ volume} \times Bacterial\ dilution\ factor$$

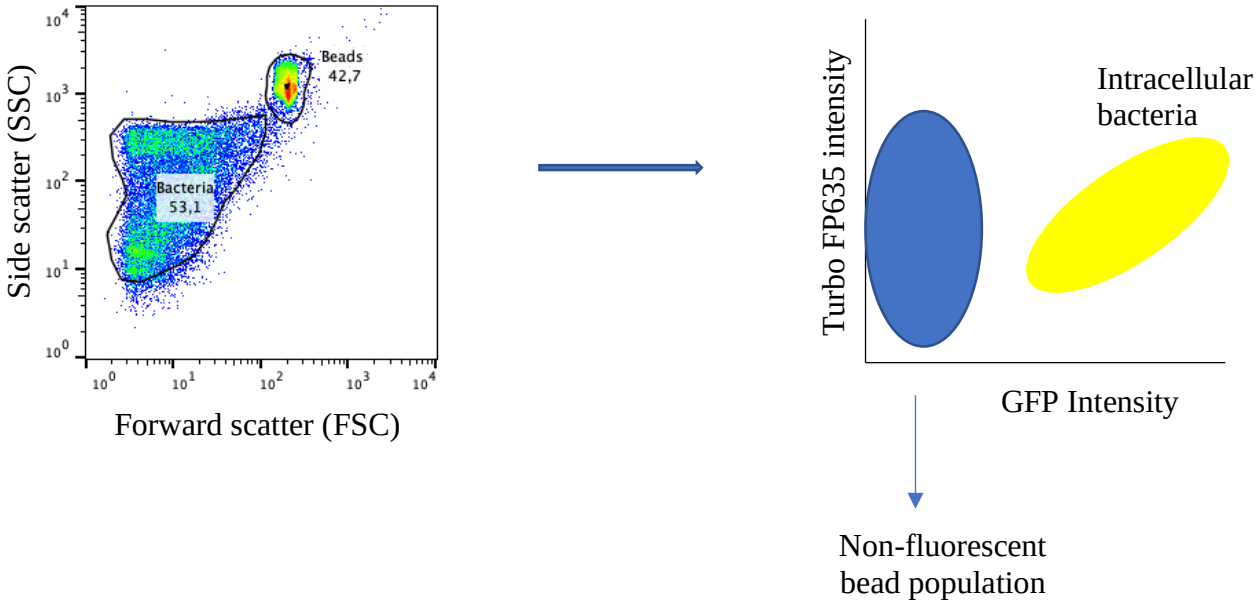


Figure 3.2.3. Schematic representation of the gating strategy for accurate bead population identification created by JL Coetzee

3.2.7. Statistical analysis

An OD_{600} growth curve data was utilized to calculate generation time per isolate at various time points and analysed using GraphPad Prism v9.01 and expressed as the mean ± standard deviation. Differences between mean generation times at 96h – 120h were analysed with a multiple t-test and unpaired pairwise comparison between wt and dx-pTiGc strains. If the measured p-value was <0.05, the variations were deemed statistically significant.

To calculate if a statistically significant difference in bacterial uptake percentage between isolates from cured and failed/recurrent treatment groups by THP-1 were observed, a pairwise comparison was performed for the mean ± standard deviation in GraphPad prism v9.01. If the measured p-value was <0.05, the variations were deemed statistically significant.

To determine differences in persister frequency between the cured treatment group and failed/recurrent patient treatment group, Differences between means were analysed with a grouped unpaired t test in GraphPad prism v9.01. If the measured p-value was <0.05, the variations were deemed statistically significant**.**

Results

3.3.1. Rationale

The objective of this section was to assess persister proportions in clinical isolates taken at baseline from both cured and failed/recurrent patient groups and to compare persister proportions to disease outcome. The assessment exploited the florescence dilution (FD) tool previously utilized in replication dynamics of *Mycolicibacterium smegmatis, M. tuberculosis a*nd similarly in *Salmonella* (Mouton *et al*., 2016; Helaine *et al.,* 2014). The tool uses a dual reporter plasmid, pTiGc, which encodes an inducible far red fluorescent protein (TurboFP635), enabling monitoring of bacterial replication over time and a constitutively expressed green florescent reporter (GFP) that allows assessment of bacterial viability.

3.3.2. Patient and isolate information

Clinical isolates selected for this study were obtained from a parent study conducted by the Catalysis TB Biomarker consortium (Malherbe *et al.*, 2016). The study followed a cohort of patients undergoing treatment for pulmonary TB (PTB) residing in the Western Cape, South Africa. Briefly, patients were grouped based on PET/CT scans at the end of treatment into cured (having no lesion activity), re-current (having new lesion activity), and failed (having intensified lesion activity). The presence of *M. tuberculosis* mRNA from culture-negative South African patients with PTB at the end of treatment is suggested to be indicative of viable, but non-culturable bacteria likely being persisters (Malherbe *et al.*, 2016).

For the current study 18 patients were selected of whom 8 (45%) were classified as cured and 10 (55%) as failed/recurrent after 6 months TB treatment. The cured patient group consisted of 7 (90%) males and 1 (10%) female with ages ranging from 19-42 years old. The failed/recurrent patient group consisted of 5 (50%) males and 5 (50%) females with ages ranging from 18-44 years. The prevalence of smokers overall was 17 (90%). Interestingly, in the cured patient group 50% either quit smoking/non-smoking during treatment, whereas the majority [8 (80%)] of the failed/recurrent patient group continued smoking during treatment.

Table 3.3.1. Mycobacterium catalysis clinical patient information

Sample id	Age	Sex	Smoking	Outcome
C-S4dx	40	Male	Quit smoking	Cured
C-S5dx	30	Male	Quit smoking	Cured
C-S29dx	42	Male	Smoking	Cured
C-S41w4	35	Male	Smoking	Cured
C-S105dx	21	Male	Smoking	Cured
C-S126dx	39	Male	Quit smoking	Cured
C-S153dx	25	Male	Smoking	Cured
C-S159dx	19	Female	Never smoked	Cured
F/R-S43dx	18	Male	Quit smoking	Failed
F/R-S93dx	30	Male	Smoking	Recurrent
F/R-S101dx	28	Female	Smoking	Recurrent
F/R-S112dx	52	Male	Smoking	Recurrent
F/R-S130dx	29	Male	Smoking	Recurrent
F/R-S137dx	44	Female	Smoking	Recurrent
F/R-S152dx	23	Female	Smoking	Recurrent
F/R-S163dx	25	Male	Smoking	Failed
F/R-S168dx	23	Female	Quit smoking	Failed
F/R-S169dx	30	Female	Smoking	Failed

3.3.3. Confirmation of clinical isolates, transformation with replication reporter plasmid and growth of transformed strains.

All bacterial samples from both cured and recurrent/failed groups were assessed utilizing the Ziehl Neelsen (ZN) staining procedure to confirm purity of mycobacterium cultures. No non-mycobacterial species were detected in any of the clinical isolates which was determined by Gene Xpert and whole genome sequencing (data not shown). Fig 3.3.1 represents a clean ZN stain. Streaking of cultures onto blood agar plates confirmed the absence of contamination in all clinical isolates (data not shown).

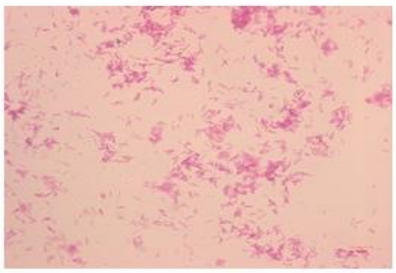


Figure 3.3.1. Ziehl-Neelsen (ZN) staining of H37Rv representing pure *M. tuberculosis* bacteria (scalebar =2µm)

To assess the persister proportions within both the cured and failed/recurrent patient groups all bacterial isolates were transformed with the dual reporter plasmid, pTiGc. Fluorescent plate reading results confirmed the expression of GFP and TurboFP635 in all isolates that were transformed with the dual reporter plasmid pTiGc (Table 3.3.2). A TurboFP635 fold change more than or equal to six were considered as an adequate induction, since it allows for identification of several generations within an actively growing bacterial population after induction has been removed. However, one isolate (C-S4dx-pTiGc) could not be sub-cultured into a larger volume after transformation for respective follow-up analysis and was excluded from all subsequent experiments.

Table 3.3.2. Transformation information of baseline clinical isolates.

Group	Sample id	TurboFP635 fold change
Cured	C-S4dx	10,05 *
	C-S5dx	77,51
	C-S29dx	75,47
	C-S41w4	43,89
	C-S126dx	88,39
	C-S105dx	234,39
	C-S153dx	71,15
	C-S159dx	33,41
Failed/recurrent	F/R-S43dx	6,15
	F/R-S93dx	28,00
	F/R-S101dx	275,23
	F/R-S112dx	64,94
	F/R-S130dx	168,96
	F/R-S137dx	139,32
	F/R-S152dx	9,64
	F/R-S163dx	8,76
	F/R-S168dx	1222 ,90
	F/R-S169dx	34,20

*Indicates isolates which did not grow following transformation.

To assess the fitness of bacterial isolates carrying the dual reporter, comparative *in vitro* growth curves were performed. Fig 3.3.2 and 3.3.3 depict the growth curves of 13 bacterial isolates from cured and failed/recurrent groups respectively. Four clinical isolates were excluded from further analyses (Samples F/R-S101dx, F/R-S130dx, F/R-S93dx, F/R-S168dx and their pTiGc counterparts) as starter cultures were difficult to initiate from freezer stocks inclusive of isolate C-S4-pTiGc. In an attempt to overcome this growth limitation, isolates were cultured from duplicate freezer stocks in 0.5x the initial volume of 7H9-OGT (containing double the concentration of glycerol). Still, no growth was observed for these clinical isolates.

Nonetheless, due to these difficult-to-culture bacteria only 13 of the initial 18 samples were selected for further analyses.

Visually, the wild type (wt) (e.g. F/R-S112dx) and the pTiGc transformed isolates (e.g. F/R-S112-pTiGc) show similar growth rates (Fig 3.3.2 and 3.3.3). This is explored in more quantitative detail in Fig 3.3.4, which represents the generation time of both the untransformed wild-type (wt) and the transformed (pTiGc) isolates at early time points (96 -120h) (Formula 3.3.1). The carriage of the pTiGc plasmid did not impose a fitness cost on bacterial growth of the clinical isolates (p-value > 0.05).

$$\text{Generation time} = \frac{time\ (minutes)}{number\ of\ generations}$$

$$Number\ of\ generations = \frac{\ln\ (fold\ change)}{\ln\ (2)}$$

$$Fold\ change = \frac{geomometric\ mean\ of\ time\ x}{geometric\ mean\ of\ time\ 0}$$

Formula 3.3.1. Generation time calculation

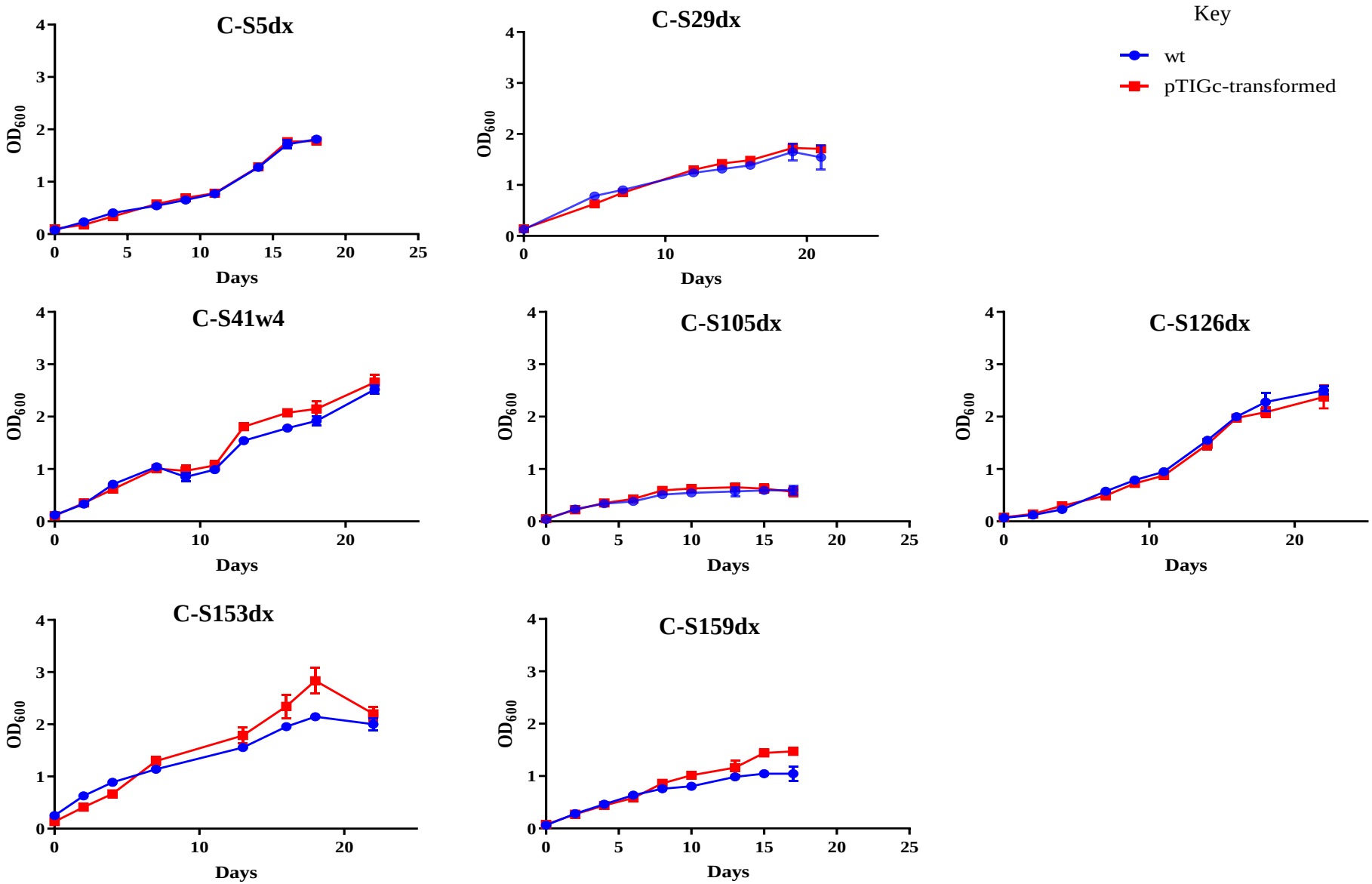


Figure 3.3.2. OD_{600}-based growth curve assessing fitness of bacterial isolates from the cured group carrying the pTiGc plasmid. All time points represent four technical replicates.

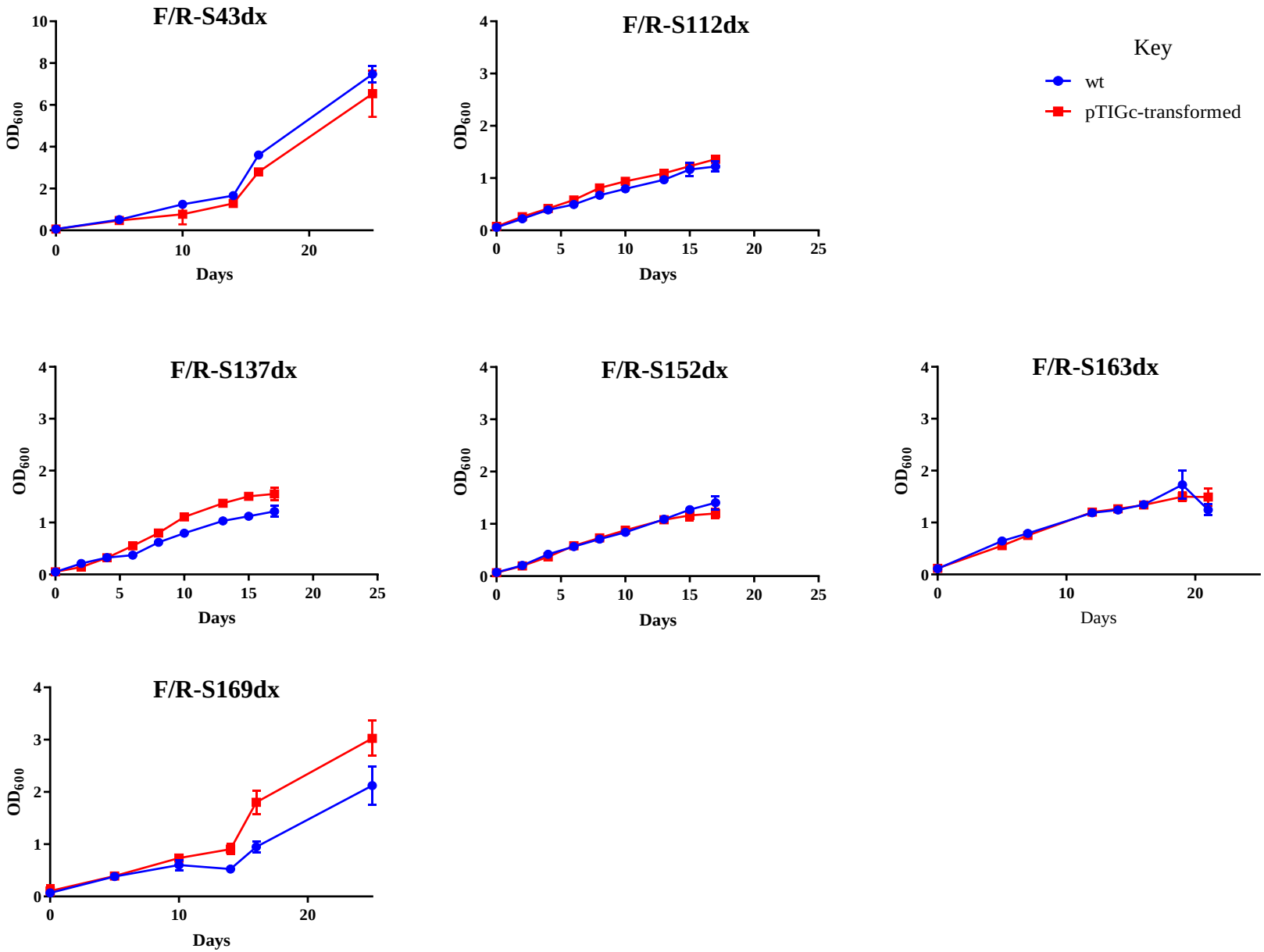


Figure 3.3.3. OD_{600}-based growth curve assessing fitness of bacterial isolates from the failed/recurrent group carrying the pTiGc plasmid. All time points represent four technical replicates.

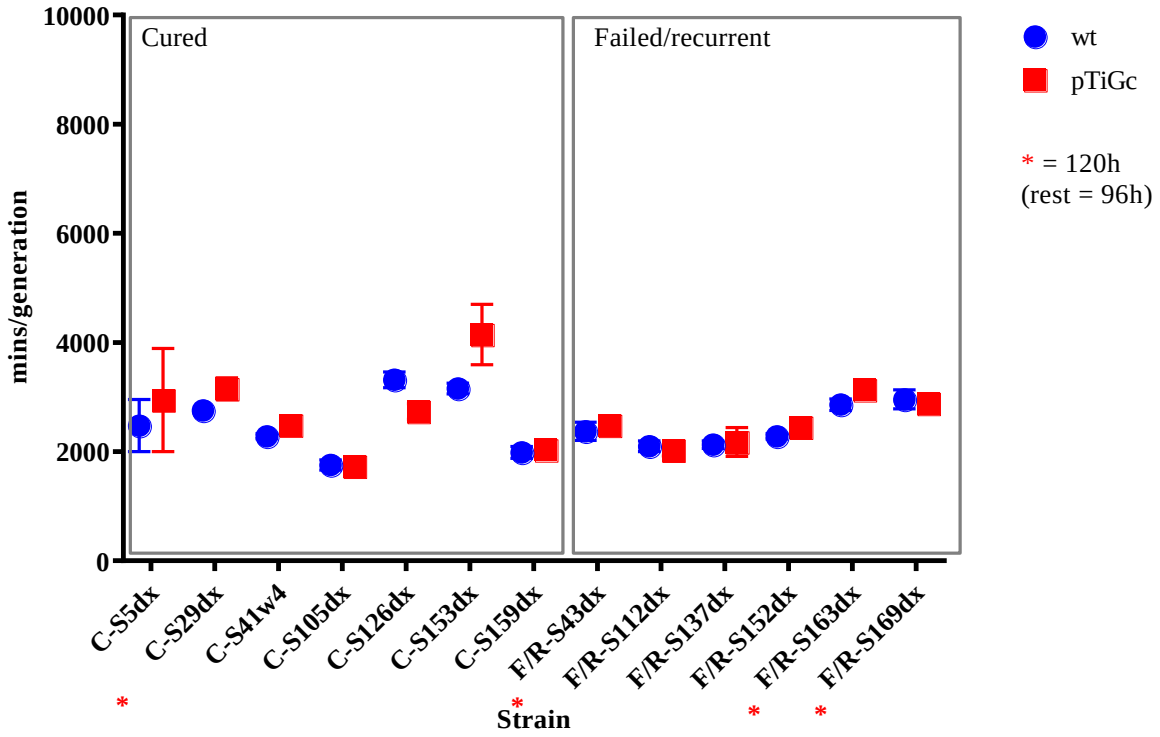


Figure 3.3.4. Generation time of all strains from cured and failed/recurrent treatment groups [wild-type(wt) vs pTiGc transformed isolates (pTiGc)] Data is representative of mean with SD values. The asterisk depicts the 120h time interval, while the other sample represent the 96h time interval. Multiple t-tests were run with no significance obtained between wt and pTiGc.

3.3.4. Intracellular mycobacterial uptake and survival following macrophage infection.

Uptake of the clinical isolates from both cured and failed/recurrent patient groups by THP-1 macrophages showed high variability between isolates and groups (Fig 3.3.5). Uptake of isolates from the cured patient group ranged from 1-400%, while uptake of isolates from the failed/recurrent patient groups had a lower uptake of 1-18% (Fig 3.3.5). Despite variability in the uptake between strains, all technical replicates within strains showed uniformity. Isolates that showed high bacterial uptake, e.g. C-S41w4 that showed an average uptake of 478%, suggests that 5 bacteria were taken up per macrophage. Mycobacterial survival after 24h in THP-1 macrophages (Fig 3.3.6) showed high variability between strains and triplicates.

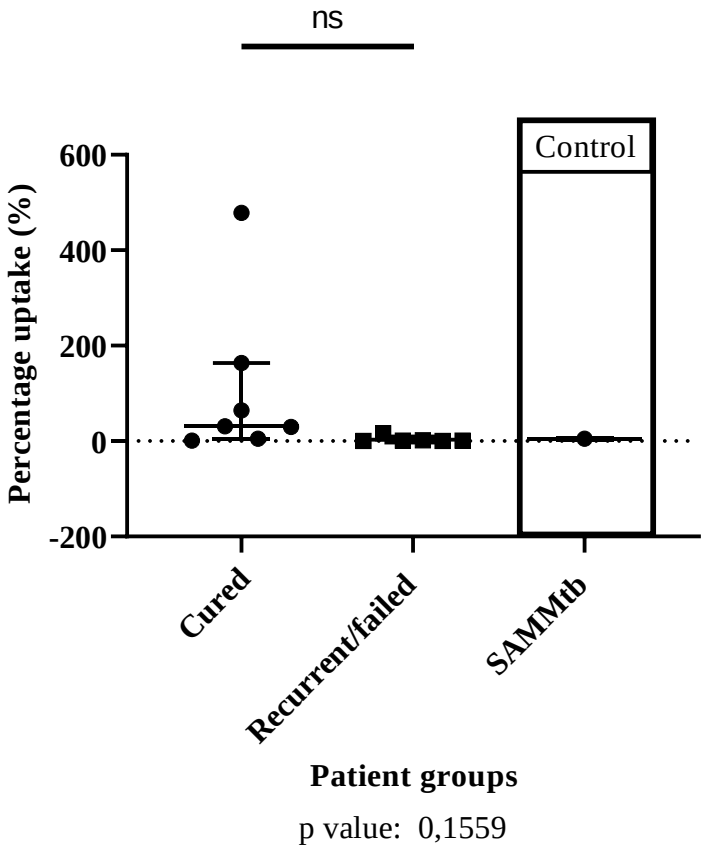


Figure 3.3.5 Uptake percentage of cured versus failed/recurrent patient groups' 0h post infection. Intracellular uptake of inoculum was assessed utilizing the by comparing bacterial uptake of isolates obtained from cured, failed/recurrent patient groups and the control SAMMtb reporting median and interquartile range. Data shown is representative of three technical replicates and 2 biological repeats.

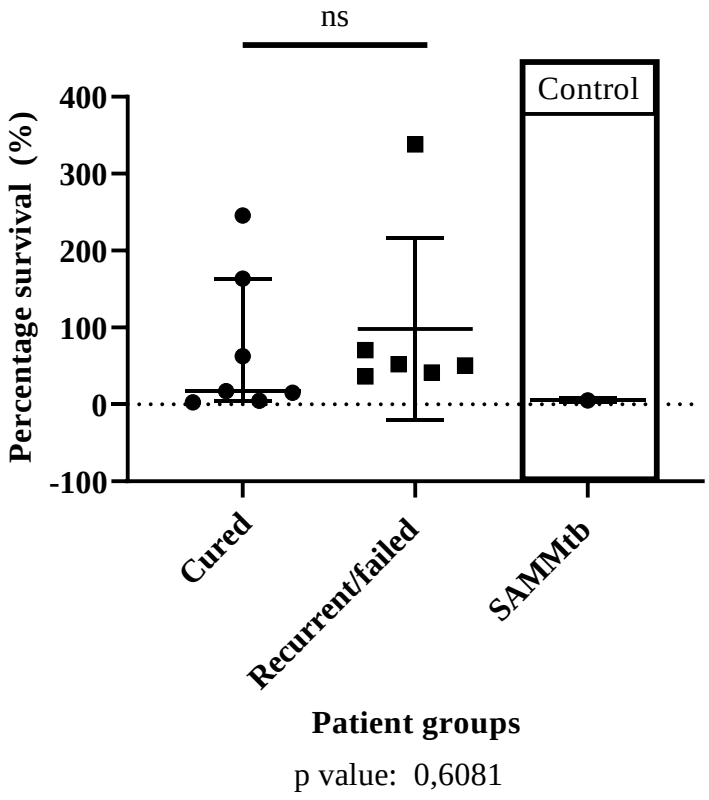


Figure 3.3.6 Percentage intracellular survival of isolates from cured and failed/recurrent patient groups 120h post infection. Intracellular survival of all patient groups and SAMMtb control reporting median and interquartile range (excluding obvious outliers). Pairwise group comparison indicating non-significance between group means. Data shown is representative of three technical replicates and 2 biological repeats.

3.3.5. Fluorescence dilution (FD) analysis assessing persister formation within patient groups at baseline.

To determine whether there is a difference in replicating populations between cured and failed/recurrent patient groups, changes in TurboFP635 fluorescent signal in response to THP-1 macrophage infection was assessed using flow cytometry. Fig 3.3.7 depicts representatives of both cured and failed/recurrent patient groups taken at baseline following infection.

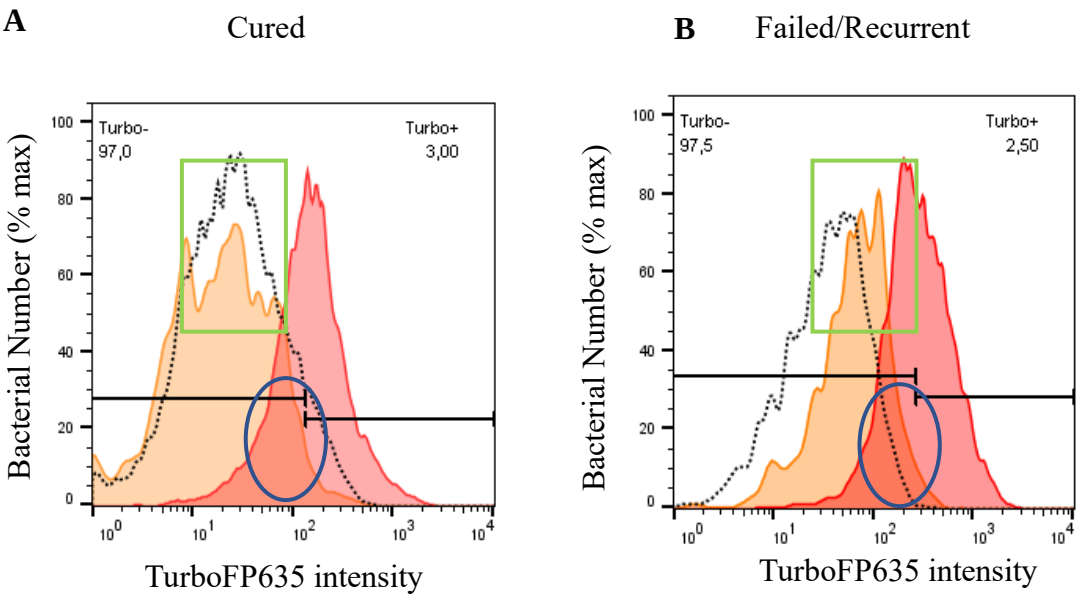


Figure 3.3.7. Population-wide replication dynamics from representatives of both cured and failed/recurrent patient groups upon macrophage infection. Intracellular bacteria lysed from macrophages 0h (red*), In vitro* bacteria 120h (dotted black), intracellular bacteria lysed from macrophages 120h, (orange) **A)** Flow cytometric identification of TurboFP635 fluorescence intensities intracellular and in vitro cultured mycobacteria for isolate C-S105dx, a representative of the cured patient group, at designated intervals. **B)** Flow cytometric identification of TurboFP635 fluorescence intensities of intracellular and *in vitro* mycobacteria at selected time points for isolate F/R-S43dx as a representative of the failed/recurrent. Turbo+ is indicative of the proportion intracellular (orange) bacterial population that remains high red (visible to the right of the black threshold line). Turbo – is indicative of the proportion of intracellular bacteria that are actively replicating (to the left of the black threshold line). Data shown are representative of three technical replicates, and two biological replicates.

Utilizing the gating strategy in Fig 3.2.2, high GFP fluorescent signal remained relatively unchanged throughout infections, indicative that the majority of bacteria that were selected for analysis were viable. The population dynamics of bacterial isolates from the two patient groups demonstrated a homogenous intracellular population at 0h, and heterogeneous intracellular bacterial replication inside the THP-1 macrophages and *in vitro* (black dashed line) at 120h in both patient groups (Fig 3.3.7).

At 120h post infection a small population of high-red bacteria overlaps with the intracellular bacterial population at 0h infection, indicating retarded growth that is suggestive of enrichment for a persister-like subpopulation (blue circle, Fig 3.3.7). In the majority (shown in supplementary, FigS3.2) of the failed/recurrent isolates (5/6) intracellular growth at 120h post infection vs *in vitro* growth 120h post infection was observed to be slower than that of the isolates obtained from cured patients (green square).

Interestingly, intracellular growth of S126dx 120h post infection was observed to be faster than *in vitro* growth 120h post infection (Fig 3.3.9). This was in contrast to previous observations of isolates obtained from cured patients, where the *in vitro* and intracellular replication rates were relatively similar (Fig 3.3.7). Various aspects needed to be considered such as whether intracellular is growth faster than other samples in the cured patient group, or whether the *in vitro* is growth slower than other samples. Comparison of Figure 3.3.8 to Figure 3.3.7(a) shows that the TurboFP635 intensity for the highest peak at D5 *in vitro* (black line) is similar to that of Figure 3.3.7(a), suggesting that the intracellular growth is indeed faster than the growth rate *in vitro*. This result is representative of 2 biological replicates.

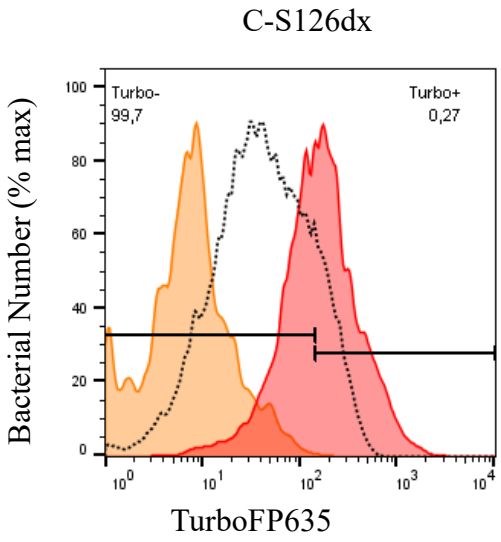


Figure 3.3.8. Population wide replication dynamics of C-S126dx upon macrophage infection. The histogram represents intracellular bacteria at 0h (red histogram), *in vitro* bacteria at 120h (black) and intracellular bacteria at 120h, (orange histogram). Data shown are representative of three technical replicates and two biological replicates.

To ascertain whether there was a statistically significant variation between the persister-like subpopulation in the cured and failed/recurrent patient group (at baseline) we applied the gating strategy outlined in Fig 3.2.6.1. Following the selection of viable bacteria (high GFP), TurboFP635 florescence was assessed in a histogram plot (Fig 3.3.7 and 3.3.8). A threshold gate was set based on the median TurboFP635 florescence intensity (MFI) of intracellular bacteria at 0h for each isolate (Fig 3.3.19a). To determine the frequency of a persister-like subpopulation, the top 50th percentile of TurboFP635 signal was selected and termed “high red”. This gate was used to determine the frequency of “high-red” persister bacteria in the *in vitro* bacteria and intracellular bacteria at 120h post infection (Fig 3.3.9b, c). The frequency of the “true” macrophage-induced persister-like subpopulation was determined by subtracting the frequency of high-red *in vitro* cultured bacteria from intracellular bacteria at 120h post

infection (equation below). FlowJo V10.7.1 was used to select the flow cytometry gates and GraphPad Prism V9.0.1 was utilized to determine statistical significance.

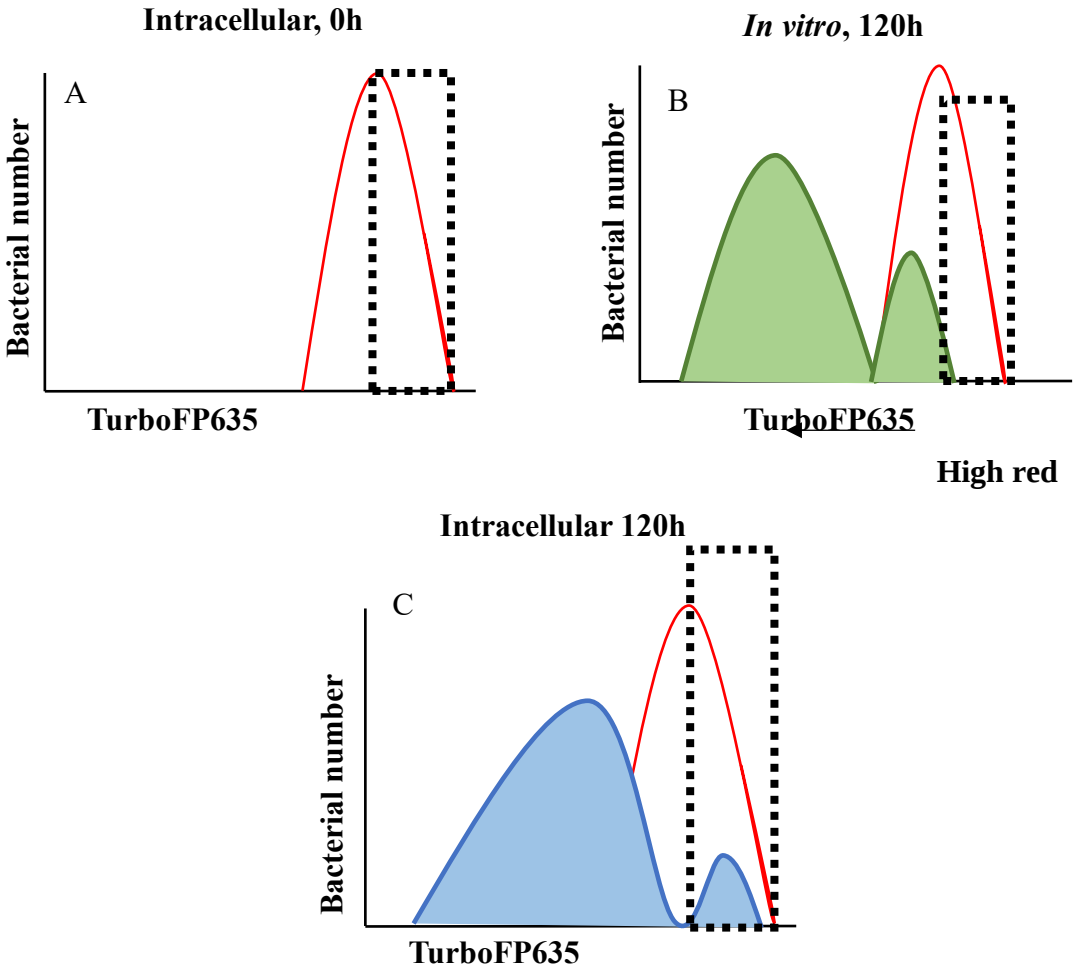


Figure 3.3.9. Persister analysis gating strategy. A) Initial gate determined by MFI of 0h intracellular bacteria select the top 50 percentile of MFI and is termed "high red". **B)** High red gate overlaid on *in vitro* bacteria 120h post infection. **C)** High red gate overlaid on intracellular bacteria 120h post infection. "True" persister = % intracellular bacteria 120h − % *in vitro* bacteria 120h

Significantly more persister-like proportions were observed in isolates from the failed/recurrent group compared to the cured group (Fig 3.3.10). This is an indication that the proportion of persister-like subpopulations at baseline could impact patient outcome during PTB treatment.

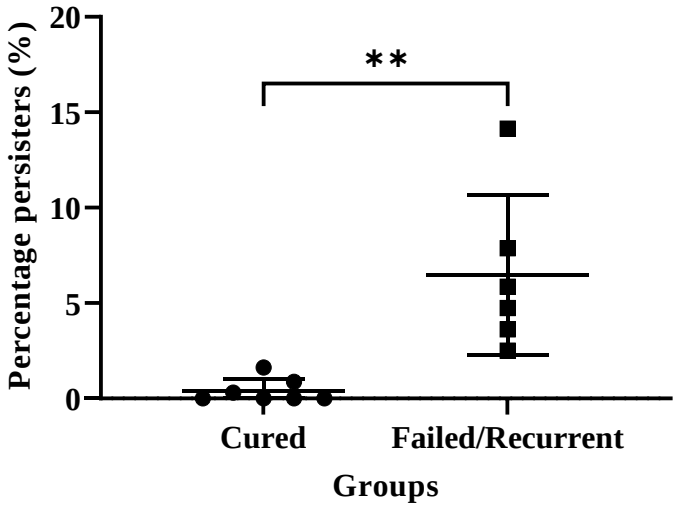


Figure 3.3.10. Persister frequency in isolates obtained from the cured and failed/recurrent patient groups following macrophage infections. Plots showing mean and SD values of **p<0.05

3.4. Discussion

To determine whether isolates from different patient groups (cured vs failed/recurrent) show variable growth dynamics upon macrophage infection, we utilized a THP-1 infection model in combination with the FD tool. The macrophage (THP-1) model was utilized to mimic aspects of the environment that *M. tuberculosis* bacteria are exposed to during PTB. We aimed to determine if clinical isolates from failed/recurrent groups are more likely to form a persister-like subpopulation that contributes to the poor treatment outcome by exposing all the isolates to the same environment.

3.4.1. Patient information and PTB outcome

Table 3.3.1. represents patient outcome of individuals undergoing PTB treatment and the general demographics of the patient groups. It was observed that the male: female ratio is similar between the cured and failed/recurrent patient groups. Khan *et al.*, 2020 has shown that smoking has a significant impact on TB treatment outcome, where a halt in smoking during treatment showed a~0.70% success in TB treatment (Khan *et al.*, 2020). This supports that a halt in smoking is an effective way to decrease treatment failure and subsequently drug resistance. In our study, the majority of patients in the failed/ recurrent group were smoking during treatment. However, results from the cured patient outcome group suggest that smoking had an irregular impact on patient outcome (Table 3.3.1). Gene Xpert inclusive of WGS results were utilized as a tool for differentiation between M. tuberculosis and non-tuberculosis mycobacterium (data not shown). Lian et al, 2020 assessed current diagnostic algorithms for detection of mixed infections by comparing Gene Xpert results and mycobacterial culture plus DNA sequencing. Gene Xpert only identified m. tuberculosis DNA presence while DNA sequencing identified NTMS as they comprised majority of the culture.

3.4.2. *In vitro* mycobacterial growth

Isolate C-S4dx-pTiGc failed re-growth after sub culturing (Table 3.3.2, indicated with an *). Factors which could affect bacterial growth of C-S4dx-pTiGc include the culture volume. An increased culture volume increases the nutrient-to-bacterium ratio. However, a decrease in proximity has been shown to decrease growth in some bacterial species. This could be due to the lack of growth enhancing stimulus secreted by neighboring bacteria as highly dilute

cultures cannot grow with limited resuscitation promoting factors (Rpfs) (Mukamolova *et al.*, 2002).

Re-growth from freezer stocks was observed to be problematic for some samples, as depicted by the inability to obtain reliable growth for isolates F/R-S93dx, F/R-S101dx, F/R-S130dx and F/R-S168dx (section 3.3.3.). One possibility is that bacterial stocks underwent freeze-thawing cycles during storage. Multiple freeze-thawing cycles have been observed to be detrimental to *Mycobacterium lepraemurium in vivo*, where a loss of viable bacteria of 60-97% was observed (Portaels *et al.*, 1988). Other factors which could affect bacterial growth include the need for multiple carbon/nitrogen sources, the change in environmental conditions and, storage time of isolates prior to usage for the current study. Storage time prior to usage was 2 years. Furthermore, Kim and Kibuca et al., 1972 found that storage of H37Rv at -70°C maintained its' viability however, the experiment was only implemented for a duration of 1 year (Kim and Kubica, 1972). However, previous literature has supported the hypothesis that a prolonged lag phase in bacterial growth is indicative of high persister proportions within the overall population (Şimşek and Kim, 2019). This is supported by the yin-yang model which suggests that the bacterial population consists of both growing and non/slow growing bacteria in a consortium which interconverts at various stages (Zhang, Yew and Barer, 2012; Zhang, 2014b). Thus, it can be suggested that these isolates which belonged to the failed/recurrent group consisted of a high proportion of persisters compared to the isolates from the cured patient group.

OD-based growth curves and generation calculations showed a strong correlation in *in vitro* growth between the untransformed wt and transformed pTiGc isolates (Fig 3.3.2, 3.3.3 and 3.3.4). In the majority of failed/recurrent isolates there is a longer stationary phase. The prolonged stationary phase could increase *in vitro* heterogeneity of these isolates and in turn increase the bacteria's adaptability to host environments. Jôers and Tenson have showed that wild type strains of *Escherichia coli* displayed increased heterogeneity correlating to a longer lag phase (Jõers, Kaldalu and Tenson, 2010; Jõers and Tenson, 2016). Thus, showing that despite favorable conditions, delaying growth is potentially advantageous to surviving stressors.

3.4.3. Uptake percentage and survival of mycobacterial strains within THP-1 macrophages

Literature suggests that the expected uptake of mycobacteria within macrophages after an infection at MOI 10:1 is approximately 10% (Li, Petrofsky and Bermudez, 2002). Results from the uptake percentages for the controls strain SAMMtb confirmed uptake of approximately 10% in THP-1 macrophages (Fig 3.3.6). However, within strains variability in uptake percentages were observed, however, no significant differences between the cured group and the failed/recurrent group were observed. Variability in uptake % could be the result of variability in inoculum (bacteria/ml). However, bacteria were all in exponential growth phase prior to infection initiation. Variability could also be due to a lack in host binding factors that are present on bacteria for phagocytosis (Ernst, 1998). Additionally, PMA stimulation could have affected THP-1 macrophages as PMA was found to induce a significant tumor necrosis factor-α production in resting macrophages where increased concentrations and prolonged treatment has led to rapid macrophage death (Mendoza-Coronel and Castañón-Arreola, 2016; Starr *et al.*, 2018). Macrophage uptake has also been observed to be strain dependent (Chakraborty *et al.*, 2013), where ~60% of infected THP-1 showed >10 bacilli per cell from both Beijing (lineage 2) and Latin-American-Mediterranean (LAM-6) lineage 4 strains, compared to ~40% of THP-1 showed to be infected with >10 bacilli per cell H37Rv. Reiling *et al* found that clinical strains belonging to lineage 2 (East-Asian) had low uptake in human monocyte derived macrophages (Reiling *et al.*, 2013). In our results, we observed a similar relationship since strains F/R-S152dx, F/R-S112dx and F/R-S163dx, belonging to lineage 2, showed low uptake by THP-1 macrophages (Fig 3.3.5). These strains have been isolated from the failed/recurrent patient group, which could suggest that uptake has an impact on treatment outcome. However, these observations need to be followed-up, as the sample numbers are too low to make a definitive conclusion.

Fig 3.3.6 depicts high inter strain variability in intracellular bacterial survival, which is presumably due to variable initial mycobacterial uptake or virulence effectors that are produced following macrophage infections. Literature suggests that mycobacteria from lineage 2 (East-Asian) has an increased ability to survive upon macrophage phagocytosis (Chakraborty *et al.*, 2013; Reiling *et al.*, 2013). However, this trend was not observed in the current study.

3.4.4. Replication dynamics of intracellular *M. tuberculosis* clinical isolates reveals population heterogeneity on multiple levels

At 0h after infection, *in vitro* and intracellular cultures of isolates from cured and failed/recurrent groups demonstrated a homogenous population, similar to previous studies (Fig 3.3.7) (Mouton *et al.*, 2016; Helaine *et al.*, 2014). However, at 120h post infection, heterogeneity within isolates both *in vitro* and intracellularly were observed. This showed that the majority of clinical mycobacterial populations (taken at baseline), regardless of treatment outcome, are inherently heterogeneous 120h after infection. Corresponding with these results, Cohen *et al.*, 2016 showed that upon initiation of treatment 21.1% of patients demonstrated *M. tuberculosis* bacterial heterogeneity at baseline based on mycobacterial interspersed repetitive units-variable tandem repeat (MIRU-VNTR) typing. This indicates that in 21.1% of patients had >1 strain of *M. tuberculosis* upon initial infection. Cohen *et al.*, found that bacterial heterogeneity at baseline is associated with a 2-fold increase in the odds of persistent culture positivity (Cohen *et al.*, 2016). Post infection, it was observed that intracellular growth is slowed compared to the *in vitro* growth in isolates from the failed/recurrent group compared to the cured group (indicated by the blue circle in Figure 3.3.7a-b). An increased subpopulation of bacteria that retain high TurboFP365 intensity suggests an increased frequency of bacteria with slow/no growth, suggestive of a persister-like subpopulation.

The growth of isolate C-S126dx-pTiGc was observed to be faster intracellularly compared to *in vitro* growth (Fig 3.3.8). Various aspects could explain the increased intracellular bacterial growth, such as increased macrophage lysis during infection, the percentage uptake of bacteria into macrophages, or more importantly the strain specific adaptability to the host environment. However, data validating increased macrophage lysis and strain specific adaptability would need to be explored in future.

3.4.5. Persister-like cell formation between cured and failed/recurrent patient groups

Previous work suggested that remaining lesion activity post PTB treatment is a result of persister-like formation (Malherbe *et al.*, 2016). Thus, we hypothesized that isolates obtained from these patients at baseline had a higher propensity to form persisters. This study provides preliminary support for these suggestions.

From our results, we observed persister-like subpopulations in the failed/recurrent group (Fig 3.3.10). A significant increase in persister-like frequency in isolates from the failed/recurrent patient group compared to the cured patient group was observed ($p<0.005$, Fig 3.3.10). Within the failed/recurrent patient group there was heterogeneity in persister-like frequency between isolates. This heterogeneity could be lineage dependent. The isolates with an average high persister frequency were determined to be S112dx, S163dx and S169dx. Isolates S112dx and S163dx belong to the (East-Asian) lineage that has previously shown to be hyper virulent and exhibit various adaptations to the host environment (Reiling *et al.*,2013). However, a definitive correlation between persister formation and mycobacteria lineages could not be made due to a limited sample size, and restriction of isolates to modern *M. tuberculosis* lineages.

Heterogeneity within persister subpopulations has previously been observed where persister subpopulations where found to respond diversely to stressors to achieve a persister or drug tolerant phenotype (Vilchèze *et al.*, 2013; Berney, Hartman and Jacobs, 2014; Amato and Brynildsen, 2015; Jain *et al.*, 2016). Nguyen *et al*. found that within *Staphylococcus aureus* clinical isolates from 36 patients, presenting with unresolved or reactive infections that were susceptible to moxifloxacin, showed high persister formation in 17% of isolates after macrophage infections and moxifloxacin treatment (Nguyen *et al.*, 2020). These findings are suggestive that high persister formation in antibiotic susceptible bacteria plays a role in clinical outcome (Nguyen *et al.*, 2020). This study has found that isolates consisting of large persister proportions in stationary phase that are phagocytized by THP-1 monocytes gave rise to larger persister proportions that remained unaffected by moxifloxacin intracellularly (Nguyen *et al.*, 2020). In our current study we focused on persister-like populations in exponential growth phase; in future studies, focusing on persister-like population dynamics in stationary phase could provide a better representation of the effect of persisters in LTBI. The conclusions from the study depicts the impact of high persister proportions on clinical outcome in *S. aureus* infections, these conclusions could thus be extended to *M. tuberculosis* as a macrophage-induced persister-like subpopulation was observed predominantly in the failed/recurrent patient group (38%).

3.4.6. Adaptation to host environment

It has previously been reported that upon uptake, macrophages kill the majority of intracellular bacteria when combined with antibiotic treatment post infection (Anes *et al.*, 2006). Generally, intracellular bacterial growth commences an undetermined period of time after uptake.

Heterogeneity observed in bacterial replication in all isolates in macrophages 120h post infection suggests that mycobacteria adapt to survive the host environment. In this study we have observed that there is variable replication within macrophages by utilizing FD. The frequency of persister-like cells of *M. tuberculosis* in the failed/recurrent patient group suggests it allows progression of infection by evading host immune responses. Slow mycobacterial growth has been suggested to be an adaptive mechanism, allowing for in-host persistence that results in no or limited clinical symptoms (Barry *et al.*, 2009). Decreased growth has also been associated with a decrease in drug efficacy as mechanisms involving replication serve as antibiotic targets. Persisters are known to have an increased expression of efflux pumps that actively export first line TB drugs, rifampicin and isoniazid thus increasing antibiotic tolerance (Adams *et al.*, 2011). This could suggest future experiments utilizing combination stresses (macrophage in addition to a first line drug) to determine the likelihood of increased persister-like proportions post treatment.

During antibiotic persister enrichment in sputum, differential gene expression was observed in genes involved in ATP synthase, stress responses, growth and division, NADH dehydrogenase and the DosR regulon (Jain *et al.*, 2016). Similarly, Walter *et al*. found that within sputum, upregulation in genes involved in drug efflux and stress responses was observed, while a decrease in replication, ribosomal protein production, expression of DNA gyrase and topoisomerase occurred, enabling survival (Walter *et al.*, 2015). In sputa, WGS revealed resistance-associated variants, depicted as heterozygous alleles, and in some provided a genotypic explanation for phenotypic resistance (Nimmo *et al.*, 2019). *M. tuberculosis* adaptations were previously identified by mouse and real-time *in vitro* models utilizing time lapsed fluorescence microscopy and microfluidics during a variety of stressors (Manina *et al.*, 2015). *M. tuberculosis* has been observed to prevent maturation once within the phagosome compartment by inhibition of phagosomal acidification and fusion with lysosomes (Armstrong and Hart, 1971; VanderVen, Brian C. Huang, Lu; Rohde, K; Russell, 2016). Liu *et al*. observed that NapM a nucleoid associated protein binds to DnaA both in vitro and in vivo inhibiting DNA replication as well as ATP hydrolysis activity enhancing *M. tuberculosis* survival (Liu *et al.*, 2019). Mechanisms underlying *M. tuberculosis* adaptation to survive within host environments, that subsequently cause recurrence in patients, are complex and need to be further explored. However, bacterial genotyping could shed some light on the subject.

3.5. Limitations

A limited number of clinical isolates were included in the current study. The reason for this was that criteria for isolate selection needed to include drug susceptible isolates that failed treatment or resulted in recurrent PTB. This experiment would thus need to be repeated with a larger cohort of separated patient groups cured, failed and recurrent as failed and recurrent are two different clinical phenotypes. This would allow for a clear understanding of persister proportions and their relevance in clinical outcome. Although the utilization of counting beads allowed for a more rapid and robust enumeration of bacterial uptake and survival percentage, confirmatory experiments would need to be completed with solid agar plate-based CFU counts.

3.6. Future work

- To assess differences in macrophage lysis in response to different strains, macrophage viability could be assessed with the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide MTT assay pre and post infection.
- An aspect which could be interesting to address is the level of persisters in the follow up samples. However, antibiotic treatment could increase the prevalence of phenotypic changes within a population as a mechanism to overcome the antibiotic stress.
- To determine antibiotic tolerance of the heterogeneous population's bacterial survival in response to a first line TB drug such as isoniazid could be assessed.
- Primary macrophage cells derived from TB patients could be used for macrophage infections, since immunity in these cells might be different from that of the THP-1 cell line.
- Following macrophage infections, bacterial cell sorting could be done to determine transcriptomic changes of the macrophages harboring persister vs actively replicating populations.
- Isolates determined to have increased persister proportions could be subjected to rounds of antibiotic treatment to determine if persister cells have a propensity to cause antibiotic resistance.
- Murine models could be used to assess persister formation in an *in vivo* setting.

3.7. Conclusion

Little is known about the impact persisters have on TB disease outcome. FD in combination with the THP-1 macrophage infection model allowed for the assessment of heterogeneous

mycobacterial populations both *in vitro* and intracellularly at a single cell level. Additionally, persister-like cells were more abundant within the failed/recurrent group indicative of their importance in TB disease treatment outcome. Suggesting their relevance in recurrence and failed treatment outcome. However, further studies would need to assess the relation between persister proportions and clinical outcome with a larger sample size.

Chapter 4

Whole genome sequencing analyses of clinical isolates.

4.1. Introduction

The utilization of WGS has become widely popular in recent years due to the swift advances in next generation sequencing techniques and the decline in costs. To date various strains of *M tuberculosis* including other mycobacterial species have been subjected to WGS, providing genetic information with a greater power than previously used methods. WGS has been applied to a variety of topics namely transmission investigations, studies of bacterial evolution, as well as those examining host-pathogen co-evolution, and in combination with transposon sequencing for genetic identification of VBNR populations known as persisters (Walker *et al.*, 2013; Dippenaar *et al.*, 2015; Copin *et al.*, 2016; Jajou *et al.*, 2018).

In this section we aim to use WGS as a non-targeted approach to identify if there is a genetic component which predisposes persister formation within isolates obtained from patients grouped into cured and failed/recurrent based on PET/CT post pulmonary TB treatment.

4.2. Materials and Methods

4.2.1 Genomic DNA extraction

Clinical isolates of *M. tuberculosis* from all patient groups selected by preliminary screening were initially cultured into 5 ml 7H9-OGT (7H9 supplemented with OADC, glycerol and Tween80) from mycobacterium growth indicator tube (MGIT) stocks and incubated at 37°C for 7-14 days in T25 vented flasks. Upon identification of growth the isolates were sub-cultured into 15 ml 7H9-OGT media and incubated for 7-14 days in T75 vented flasks to increase bacterial numbers for an increased DNA yield.

The liquid culture was centrifuged at 4000 rpm with an Eppendorf 5804 Benchtop centrifuge (Marshall Scientific) for 10 minutes at room temperature (~21°C) and the resulting bacterial pellet was re-suspended in 400 µl TE Buffer (0.01M Tris-HCl, 0.001 M EDTA [pH 8]) obtained from Sigma-Aldrich and transferred into a 2 ml cryogenic storage tube. The 2 ml tubes containing the TE buffer cell suspension were subjected to heat-killing at 80°C for 30 minutes. Thereafter, 50 µl lysozyme (10 mg/ml) was added. The cell suspensions were incubated at 37°C overnight to ensure cell wall degradation. Following incubation, 70 µl (10%) sodium dodecyl sulphate (SDS, Sigma-Aldrich) in combination with 5 µl proteinase K (Sigma-Aldrich) (10 mg/ml) were added and the suspensions were incubated for 10 minutes at 65°C for the digestion of bacterial proteins. Next, 100 µl (5 M) sodium chloride (NaCl) and 100 µl pre-warmed (65°C) hexadecetyltrimethylammonium bromide (CTAB)/NaCl (0.1/0.041 g/ml) (Sigma-Aldrich) was added, aiding in the separation of polysaccharides. The suspensions were incubated for an additional 10 minutes at 65°C. Seven hundred and fifty microliters chloroform: isopropanol (Sigma-Aldrich) (24:1 v/v) was added to the cell suspensions. After centrifugation (Marshal Scientific) at 11,000 x g for 8 minutes, the aqueous phase was transferred to a fresh tube and 0.6 volumes of isopropanol was added to precipitate the DNA followed by incubation at 20°C for 1 hour. Thereafter, the precipitated DNA was centrifuged (Marshal Scientific) at 11,000 g for 15 minutes and washed with an equal volume of cold 70% ethanol and centrifuged for 5 minutes at 11,000 g (Marshall Scientific). The DNA pellet was re-suspended in 35 µl TE buffer. A NanoDrop spectrophotometer was used to calculate the concentration of extracted DNA (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

4.2.2. Next-generation sequence analysis

WGS was performed on the DNA collected from the M. tuberculosis isolates chosen for this project initially using the Illumina NextSeq 550 instrument (Illumina, California, USA) and

thereafter using the Illumina MiSeq platform (Illumina, California, USA) as sequencing data from 3 isolates were not obtained from initial sequencing. With approximately 600 base fragment sizes and a read length of 101 bp (base pairs), a paired-end technique was used, resulting in insert sizes between 350 and 550 bases. One microgram of DNA was used to prepare libraries for sequencing per the manufacturer's instructions using the Illumina NEBNext sample preparation kit (Illumina, Inc, San Diego, CA). The theoretical depth of coverage of all isolates were estimated to be at least above 100x based on the predicted data output. This, in combination with the sequence data quality, warranted a high confidence level for the variations identified in the genomes.

4.2.3. FASTQ file format

The FASTQ format is a text-based format for the storing of biological sequences and corresponding quality scores in one file. It is concise and compact, and was originally utilized in Sanger sequencing, although it is now used as a standard format for transporting next-generation sequencing data.

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

Figure 4.2.1. A typical read from a FASTQ file generated by the Illumina sequencing platform version 1.5.

The data within the FASTQ file is commenced with a '@' symbol, followed by the sequence identifier. The second line consists of the sequence information where the end is indicated by a new line and the third line is represented by an optional '+' symbol. The '+' symbol is occasionally followed by the same sequence identifier, which is followed by the sequence quality information in the fourth line.

4.2.4. Phred-scaled quality values

Phred quality scores are automatically assigned during the sequencing run produced by WGS technologies. The phred score is a value representing the probability that a base is called incorrectly by the sequencer utilized:

$$Q = -10\, log_{10}\, P$$

Where:

Q = phred score

P = error probability

Sequencing quality scores are portrayed as ASCII (American standard code for informational change) with characters having an offset of +33. This system connects a character with a number. For example ") " represents a Phred score of 8 which correlates to an error probability of 0.15849. Table 4.2.1 shows some of the quality scores and their correlating ASCII characters found in the datasets analyzed in the current study.

Table 4.2.1. Examples of phred scores and the correlating ASCII characters

Phred score (Q)	Error probability	ASCII
2	0.63096	#
5	0.39811	&
10	0.10000	+
20	0.01000	5
30	0.00100	?
34	0.00040	67

4.2.5. Automated WGS Data Analysis Pipeline (USAP)

Worldwide there is a variety of free specialized software packages available for the analysis of high throughput next-generation sequencing. However, this software is highly generic and publicly available pipelines do not take a specific organism's genome into account. More specifically, traits of *M. tuberculosis* such as high GC content, genome size, and high percentage of repetitive regions are not accounted for. Therefore, various software packages together with in-house developed scripts were optimized (by members of the TB Genomics group, Stellenbosch University) for the analysis of mycobacterial genomes. An overview of the pipeline is represented by Fig 4.2.2.

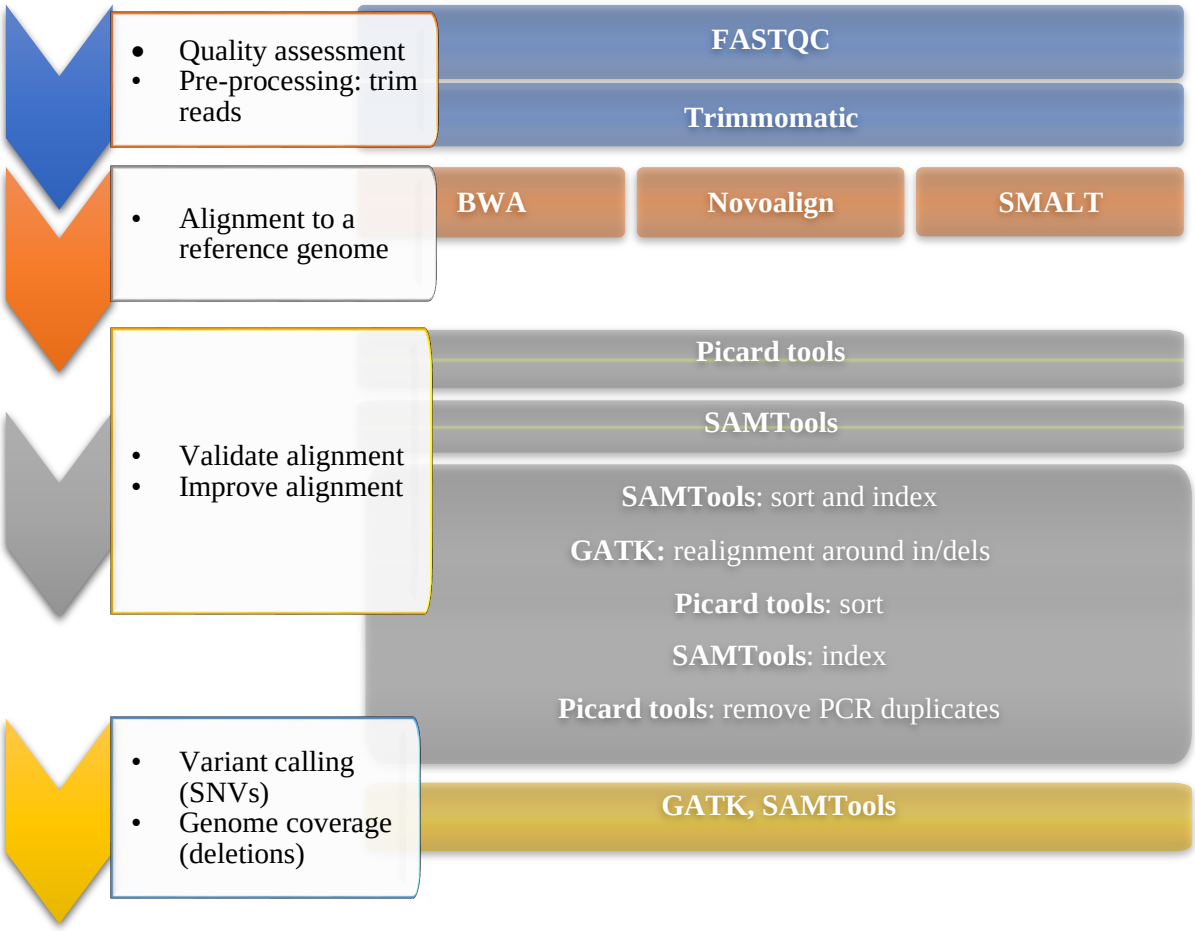


Figure 4.2.2 Workflow of the computational analysis of WGS data.

4.2.6. Quality control

Raw sequence data per isolate were subjected to quality checks. FastQC was used to check the quality of the Illumina reads to see whether any features in the data could affect downstream analysis. FastQC is a Java-based tool which utilizes FASTQ files as inputs and the results are produced in a HTML format; the data is evaluated using a seven-step package that includes the following steps:

- Basic statistics includes information describing the platform used, input file, sequence length, amount of reads processed, and percentage of the GC content.
- Calculating the per base sequence quality.
- Calculating the per base sequence content ascertaining the distribution of the four nucleotides throughout the reads.
- Calculating the GC content throughout the reads and comparing them to the theoretical value.
- Calculating the probability of read contamination.

- Calculating the number of uncalled bases throughout the reads – number of "Ns".
- Calculating the amount of duplicate sequences.

4.2.7. Trimming of sequences

Trimmomatic, a fast, multi-threaded command-line tool, was used for accurate trimming of the 3'end of the sequences. The tool was run in the paired-end mode, which maintained correspondence of the read pairs as well as the additional information to better determine adapters. While it is paramount that focus be placed on the quality of the reads (Phred~30), a balance had to be upheld between quality and read length. The FASTQ files were trimmed according to the command to produce high quality FASTQ files, which are used for further analysis. The command consists of the following (Appendix B):

- Removal of adapters
- Removal of leading low quality or N bases (below quality 20) (leading 20)
- Removal of trailing low quality or N bases (below quality 20) (trailing 20)
- Scanning the read with a 4-base wide sliding window trimming when the average quality per base drops below 20.
- Reads are dropped which were below 36 bases long. (MINLEN 36)

4.2.8. Alignment and mapping

Three distinct alignment software programs for alignment and mapping of WGSs were used. These mappers use different algorithms for mapping short sequencing reads to a reference genome, which was *M. tuberculosis* H37Rv (NC_000962.3). The mapping software included Novoalign 2.07.18 relies on a Needleman-Wunsch algorithm (Novocraft Technologies http://novocraft.com), BWA, which uses a Burrows-Wheeler transform Algorithm (Li and Durbin, 2009, 2010), and SMALT, which employs a hash table-based algorithm (Sanger Institute https://www.sanger.ac.uk). A collection of free software packages were used for the downstream analysis of the alignment file and quality control procedures.

Employing three different alignment tools minimized the identification of false positive variants since the aligners use distinctive algorithms. All alignment tools produced an output in the Sequence Alignment/Map (SAM) format. This format is generic for the storage of large nucleotide sequence alignments up to 128 Mb, it permits for the majority of procedures on the alignment to work on a stream without loading the entirety of the alignment to memory (Li *et al.*, 2009).

4.2.8.1. Novoalign

Novoalign is an alignment tool aligning short sequences against an indexed referenced sequence. The aligner aims to accurately identify variants in FASTQ format to a reference genome in fasta format. Indexing of the reference genome is completed using the "Novoindex" command. The indexed genome is saved to a corresponding file and thus can be reused if necessary. The indexing strategy utilizes a k-mer indexing size of 13 and an indexing step size of 1. Novoalign takes the input sequences and employs a Needleman-Wunch algorithm to find the best alignments. The aligner does a gapped global alignment and for this analysis the default of 6 was utilized, therefore allowing for six matches per alignment, producing a SAM file as an output.

4.2.8.2. Burrows-Wheeler Aligner (BWA)

BWA is a software package that aligns relatively short reads to a reference genome by executing three algorithms: BWA-backtrack (for reads up to 100 bp), BWA-SW and BWA-MEM (for longer reads 70 base pairs to 1 Mega base pairs). BWA-backtrack was utilized for analysis. The aligner requires the reference genome to be indexed, therefore the "faidx" command was used to index H37Rv from SAMTools and the "index" command was used in BWA home directory. BWA takes FASTQ reads as inputs and utilizes the "bwa-aln" command followed by the command "bwa sampe" to align the forward and reverse reads in combination to the reference genome (H37Rv) and produced a SAM file as an output. Default command line parameters were used for the alignment procedure.

4.2.8.3. SMALT

SMALT is a pairwise sequence alignment tool mapping reads to a reference genome. It utilizes a short-word hashing algorithm. SMALT encompasses a two-step process. Firstly, an index of short reads needed to be built utilizing the reference genome (fasta format) with the command "smalt index". Secondly, the sequenced reads in FASTQ format were mapped to the reference genome called by the "smalt map" command. The aligner matches reads to the reference utilizing a k-mer hash index method. Based on potential matching, segments were selected for alignment by a Smith-Waterman algorithm. The aligner utilizes FASTQ as the input sequence file format, and the output is a SAM file format.

4.2.9. SAM File Validation

The SAM files generated were validated with the "ValidateSAMFile" command in Picard tools (http://picard.sourceforge.net) this was used to identify the authenticity of the SAM file, the program validates the existence of reads group within the SAM file. The program was run in "mode=summary", which summarized all errors and warnings. Prioritization was set on severe errors.

4.2.10. Converting the Sequencing Alignment Map (SAM) File Format to Binary Alignment (BAM) File Format

SAMTools (http://samtools.sourceforge.net.) is a software package which invokes multiple utilities for post-processing and manipulation of alignments in SAM/BAM format (Li *et al.*, 2009). The software is able to index, sort, and merge SAM files. The SAMTools commands utilized for the conversion from the SAM alignment file to the binary alignment (BAM) format were "view" and "sort". The BAM format improves performance, due to the compression in size, while retaining all information from the SAM alignment format. The format can be indexed ensuring fast and efficient retrieval of all reads at a specific chromosomal locus.

4.2.11. Alignment Statistics

Qualimap was used to produce extensive alignment statistics for isolates analyzed in Chapter 4, section 4.2.8. The program inspects sequence alignments in an input SAM or BAM file format and provides a comprehensive report of the data concerning the depth of coverage of the reference genome, mean and median values of the insert size, and nucleotide distributions (García-Alcalde *et al.*, 2012; Okonechnikov, Conesa and García-Alcalde, 2016).

4.2.12. Post Alignment Processing of BAM Files

BAM files were subjected to processing for error corrections which were integrated during the alignment step.

4.2.12.1. Coordinate sorting and indexing of BAM files

The SAMTool commands "sort" and "index" were employed for the conversion of the BAM file into a format easily readable and manageable. The loading of extra alignments was avoided by sorting the BAM files by coordinate into computational memory.

4.2.12.2. Realignment focused on in/dels (insertions and deletions)

Insertions and deletions (in/dels) have the ability to affect the alignment of reads, which leads to the identification of false positive single nucleotide variants due to misalignment to a reference genome (Fletcher and Yang, 2010). This can occur due to bases mismatching to the reference sequence and can be misread as single nucleotide polymorphisms (SNPs). Minimization of such an occurrence across all reads was achieved by using the Genome Analysis Toolkit (GATK), which realigned misaligned sequencing reads (Mckenna *et al.*, 2009; Depristo *et al.*, 2011). The process consists of two steps. Firstly, small intervals that were misaligned were identified using the "RealignerTargetCreator" command. Secondly, questionable intervals were realigned using the "IndelRealigner" command, which realigned the intervals to the reference genome, thus amending the misaligned reads.

4.2.12.3. Coordinate sorting and indexing of realigned BAM files

Realigned BAM files were sorted with the command "sortsam" with Picard tools and indexed with "index" command using the SAMTools software package.

4.2.12.4. Removal of PCR duplicates

Polymerase chain reaction (PCR) amplification during the library construction may produce duplicate reads. The Picard command "MarkDuplicates" was utilized to locate the duplicate reads in BAM files, which were flagged in the BAM output files. This was used to decrease the bias established by PCR amplification.

4.2.13. Variant calling

Two autonomous variant callers were used for the identification of SNPs and short in/dels to the reference genome. The three different mapping files received from the prior stages were analyzed with two SNP callers – SAMTools and GATK (Li *et al.*, 2009; Mckenna *et al.*, 2009; Depristo *et al.*, 2011). The variants are kept in the variant call (vcf) file format. In addition to the identification of variants, GATK was used to determine small in/dels from each alignment, thus resulting in three vcf files comprising of potential in/dels for each isolate that was analyzed. The usage of 3 mapping alignment tools and two variant calling programs decreased the likelihood of identifying false positive variants.

4.2.13.1. GATK

The 'UnifiedGenotyper' tool from GATK was used for SNP and in/dels calling, and produced a file in the output call variant format (vcf). The value of stand-call-conf was set at 50, which

allowed for variants with a minimum phred-scaled threshold of 50 or greater to be reported as polymorphic sites. The stand-emit_cof value was set at 10. This allowed for variants with a phred-scaled confidence equal to or greater than 10, but less than the calling threshold of 50, to be reported and marked as filtered. The output vcf file contained information regarding the position, alternative sequence and the phred scale probability of the polymorphisms at that position. The vcf file also contains alternative base specific information, inclusive of the number of reads bridging that position and the number of reads containing the reference and alternative base at the position.

4.2.13.2. SAMTools

The command "mpileup" was used in SAMTools. This created a pileup of all reads relative to the reference genome and simultaneously identified SNPs relative to the reference genome. Default parameters were utilized. In/dels identified by SAMTools were excluded from downstream analysis. Vcf files generated by this software tool contained information relating to position, alternative sequence, and quality score in the phred scale for each variant. Furthermore, files include variant specific information such as the number of reads aligning to that position.

4.2.14. Annotation of variants obtained from the different aligners

In-house scripts written as part of the USAP pipeline were used to: 1) annotate the identified variants, 2) calculate the resultant amino acid changes created by SNPs located within genes and 3) annotate the identified in/dels. The genes in which the variants occur were classified based on its cellular function as reputable in the TubercuList knowledgebase and Mycobrowser (Cole *et al.*, 1998; Kapopoulou, Lew and Cole, 2011).

4.2.15. Comparison of annotations obtained from aligners

The annotated variants which were previously compiled from all three aligners were placed into a single variant file. An in-house Python script called dirCompare32 written by Dr. Ruben van der Merwe. The in-house script utilized two annotated variant input files from the isolate taken at diagnosis and at w24 to identify unique variants from both input files as well as overlapping variants. The output files produced were in text format, which were viewed and analysed in Microsoft Excel. These text files contained information such as the genomic position, the heterogeneity frequency, the number of reads, and functional category of all

unique variants identified in all three mappers with the two different variant callers as described in section 4.2.16. Information originated from six analyses strategies, namely BWA-GATK, BWA-SAMTools, Novoalign-GATK, Novoalign-SAMTools, SMALT-GATK and SMALT-SAMTools.

Similarly, annotated variants obtained from the three aligners were utilized for an inter-patient group comparison. The in-house script utilized all annotated variant input files from the cured patient group and compared to all variant input files from the failed/recurrent patient group to identify unique variants in the failed/recurrent patient group not present in the cured patient group.

4.2.16. Filtering of unique variants after pairwise comparison

Filtering of variants took place in Microsoft Excel and was used to: 1) Filter for variants that were identified in alignments of all three mappers, and both variant callers that are unique to a particular isolate when compared to the baseline/follow-up from the same patient. 2) Filter based on heterogeneity frequency $1 \leq x \geq 0.3$, which identified whether the variants were fixed in the population or not. A value of 0.7-1 represented a fixed variant, translating to 70% - 100% of reads supporting an identified variant at a specific position. A value of 0.3-0.7 was interpreted as a heterogeneous variant. A heterogenous value below 0.3 were excluded as variants below 0.3 are untrustworthy and considered false positives. 3) Filter based on number of reads at the position of the identified variant. This is defined by the average depth of coverage of each individual isolate, where if the unique variant is covered by less than 30% of the average depth of coverage of the entire genome of the specific isolate, the variant is excluded. 4) Filter based on functional category, where all variants identified in known repetitive regions were removed, such as insertion sequences, and phages. These genomic regions have a high rate of producing false positives with short read sequences, due to having high repetitive regions in *M. tuberculosis* (Treangen and Salzberg, 2012; Torrey *et al.*, 2016) 5) Filter based on position within 5 bases of each other. Variants positioned with a close proximity to each other often indicate an error in sequencing. Filtering allowed for a high confidence in unique variants.

4.2.17. Drug susceptibility and lineage prediction of isolates obtained from cured and failed/recurrent patient group

A pipeline, TB-profiler, similar to the automated in-house pipeline optimized for analysis of the *M. tuberculosis* genome was employed for drug resistance and lineage predictions (Coll *et al.*, 2015; Phelan *et al.*, 2019). The pipeline utilizes Trimmomatic to trim reads, for alignment to *M. tuberculosis* H37Rv and *BCFtools mpileup* and *BCFtools call* for variant calling. Command line usage of TB-profiler required Conda software package manager. Commands utilized are stipulated in appendix B.

4.2.18. Phylogenetic tree construction

High confidence SNPs of 15 sequences from the current study and 21 publicly available representatives of the *Mycobacterium tuberculosis* complex (MTBC) were included in the phylogenetic analysis (Comas *et al.*, 2010; Blouin *et al.*, 2012). Variants identified by both SAMTools and the GATK in three alignments were filtered to exclude variants in the *pe/ppe* family region, repeat regions, insertion sequences and phages, and only variants with an allele frequency of >0.95 were considered. A python script (Appendix C) written by Dr Ruben van der Merwe was used to generate a connected sequence of all high confidence SNPs recognized for each isolate. The principle is illustrated by the example below:

Reference strain partial genome sequence: ATGCAGTTGCGCACAGCTGCGGAT
Strain A partial genome sequence: ATCCAGTACCGCACCGCTGCGGAT
Strain B partial genome sequence: ACGCAGTTCCGCACAGGTGCGCTT

Concatenated SNP strings based on variable positions:
Reference: TGTGACGAT
Strain A: TCACCCGAT
Strain B: CGTCAGCTT

The connected sequences that contained variable sites were secured in multi-FASTA format. Sequences were converted to the Phylip format (.phy) and used for phylogenetic inference in IQ-TREE v 1.6.1.2 (Nguyen *et al.*, 2015). IQ-TREE uses an ultra-fast and automatic nucleotide substitution model selection method (Modelfinder) for phylogenetic analysis. IQ-TREE was run in the ultra-fast bootstrapping mode, using 1000 bootstrap iterations. Lineage and drug resistance annotation files produced by TB-profiler was used to annotate the resulting phylogenetic tree in the Interactive Tree of Life (iTOL) online phylogenetic tree visualization tool (Letunic and Bork, 2019).

4.3. Results

4.3.1 Introduction

WGS of *Mycobacterium tuberculosis* clinical isolates from Cape Town in South Africa was done in collaboration with the Centres for Disease Control and Prevention, Atlanta, GA, USA. The genomic DNA from 18 *M. tuberculosis* clinical isolates was subjected to WGS on an Illumina NextSeq 550 platform, 3 clinical isolates were sequenced on a Illumina Miseq at Inqaba Biotech. A customised in-house WGS data analysis pipeline was used to analyse the WGS data. Multiple measures were taken to ensure high quality sequencing and mapping of sequence reads for variant calling with high confidence. The reads obtained were aligned to the complete genome sequence of the *M. tuberculosis* H37Rv laboratory strain (Genbank accession number: NC000962.3). Table 4.1.1 depicts general information regarding the clinical isolates. All isolates selected for this study were predicted to be drug susceptible and belonged to either lineage 2 (Beijing, 37.5%) or lineage 4 (Euro-American, 62.5%). This was expected, since the majority of pulmonary TB in Africa is caused by *M. tuberculosis* from these strain families (Rutaihwa *et al.*, 2019).

Table 4.3.1. Information of *M. tuberculosis* clinical isolates

Sample ID	Age	Sex	Treatment outcome	Lineage	Drug susceptibility prediction
C-S5dx	30	Male	Cured	4.3.2.1 (LAM)	Susceptible
C-S29dx	42	Male	Cured	4.3.2.1 (LAM)	Susceptible
C-S41w4	35	Male	Cured	4.1.1.3 (X)	Susceptible
F/R-S43dx	18	Male	Failed	2.2 (Beijing)	Susceptible
F/R-S43w24				2.2.1.1 (Beijing)	
C-S105dx	21	Male	Cured	4.2 (Ural)	Susceptible
C-S105d2					
F/R-S112dx	52	Male	Recurrent	2.2 (Beijing)	Susceptible
S126dx	39	Male	Cured	4.1.2.1(X)	Susceptible
F/R-S137dx	44	Female	Recurrent	2.2.1.1 (Beijing)	Susceptible
F/R-S152	23	Female	Recurrent	2.2 (Beijing)	Susceptible
C-S153dx	25	Male	Cured	4.1.1.3 (X)	Susceptible
C-S153w8					
C-S159dx	19	Female	Cured	4.9 (T1)	Susceptible
C-S159w4					
F/R-S163dx	25	Male	Failed	2.2 (Beijing)	Susceptible
F/R-S163w24					
F/R-S169dx	30	Female	Failed	4.1.2.1 (X)	Susceptible

Abbreviations: d=day, w=week

4.3.2. Read Assessment and Trimming

The quality of all raw sequences in both the forward and reverse orientation was assessed using open access quality control tools. FastQC was used to assess the quality of the reads and the subsequent mapping strategies (Andrews, 2010). The quality of the raw sequences was considered, trimming of the reads were done accordingly using Trimmomatic (Bolger, Lohse and Usadel, 2014). Reads were on average 101 bases long and the per base quality scores decrease towards the ends of the reads similar to what was previously reported (Patel and Jain, 2012). However, reads from samples subjected to Miseq sequencing were on average 35-301 bases long with a low per base sequence quality for 2/3 sequences. The tapering of quality towards the end of reads is generally attributed to the Illumina sequencing technology, which relies on the synthesis procedure. However, Illumina sequencing frequently produces sequences of high quality regardless. Trimming of reads was completed utilizing the sliding window approach in Trimmomatic. This approach considers 4 bases at a time, determines the average quality score, and once parameters are not met, one base is trimmed. An average of 20 bases was used with a minimum read length of 36. Trimming produced reads with an average Phred scaled quality greater than 33, translating to a sequence error probability of 0.00050 (99,93% accuracy). Following trimming, samples subjected to Miseq sequencing had a minimum Phred of 16, translating to a sequence error probability of 0.025 (97,48% accuracy). Fig 4.3.1represents an example of the quality of the trimming of raw sequences from NextSeq; (a-b) shows raw sequences of isolate S105dx before trimming, showing larger errors that relate to having large variation between bases that were called, while (c-d) depicts an example of the quality of the trimmed reads in the forward and reverse orientation, respectively. Similarly, Fig 4.3.2 depicts a representative of the quality trimming of raw sequences subjected to the Illumina Miseq; (a-b) shows raw sequences of isolate S5 before trimming, showing larger errors that relate to having large variation between bases that were called, while (c-d) depicts an example of the quality of the trimmed reads in the forward and reverse orientation, respectively.

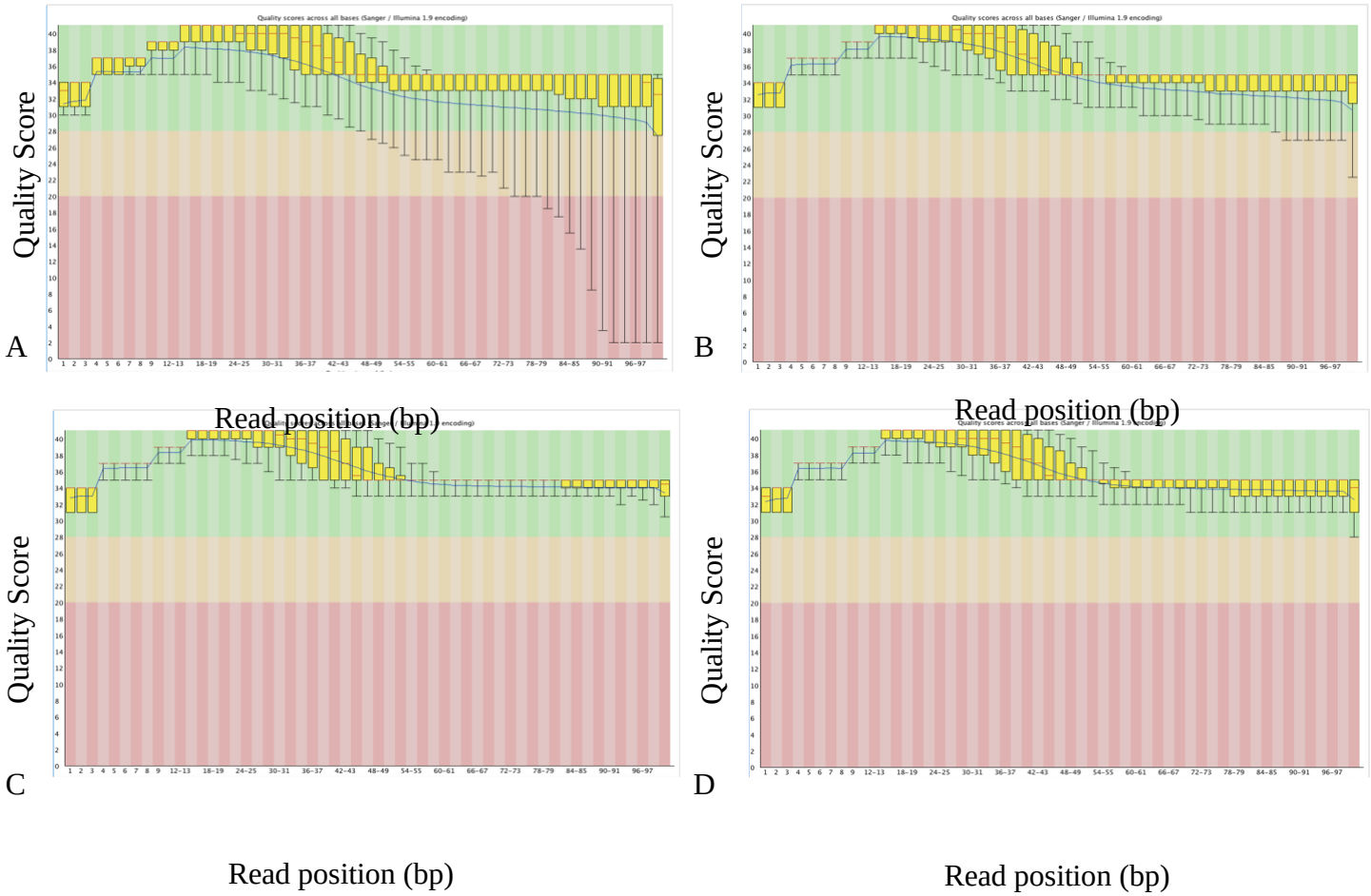


Figure 4.3.1. The per base quality of the sequencing reads of a representative strain C-S105dx from NextSeq platform. A) A per base quality graph of the forward read (R1) pre-trimming. B) A per base quality graph of the reverse read (R2) pre-trimming. C) A per base quality graph of the forward read (R1) post-trimming. D) A per base quality graph of the reverse read (R2) post-trimming.

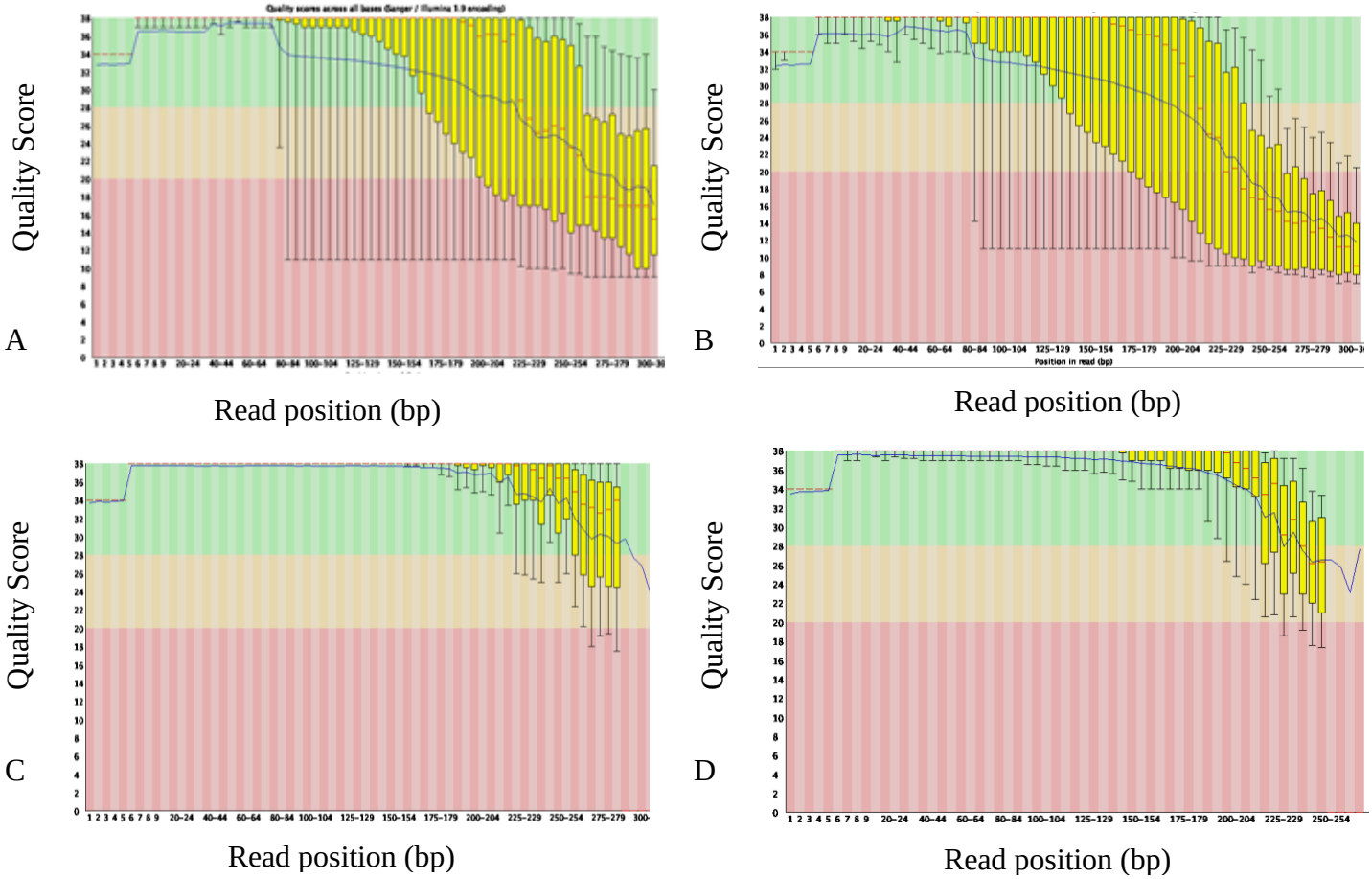


Figure 4.3.2. The per base quality of the sequencing reads of a representative strain C-S5 from Miseq platform. A) A per base quality graph of the forward read (R1) pre-trimming. B) A per base quality graph of the reverse read (R2) pre-trimming. C) A per base quality graph of the forward read (R1) post-trimming. D) A per base quality graph of the reverse read (R2) post-trimming.

4.3.3. Read Alignment and Mapping Statistics

Three independent mapping software packages (BWA, NovoAlign and SMALT) were utilised. These employ various algorithms for mapping short sequencing reads to the reference genome, *M. tuberculosis* H37Rv (Genbank accession number: NC000962.3). Qualimap2, an independent Java and R application which examines sequence alignments and produces graphical and statistical evaluation of the data from BWA, Novoalign and SMALT, was used to obtain mapping statistics for sequenced *M. tuberculosis* genomes (Okonechnikov, Conesa and García-Alcalde, 2016). Selected mapping statistics viewed from Qualimap2 and determined by SAMTools flagstat for the alignments of all clinical isolates to the reference genome *M. tuberculosis* H37Rv, are summarised in Table 4.3.2. The range of values obtained from each of the parameters listed in Table 4.3.2 reflects the diverse algorithms utilized by the independent mapping software packages. For the majority of isolates, >90% of the reads were

mapped to the H37Rv reference genome, suggesting no contamination within the samples, and that high sequence quality was obtained from using the NextSeq platform. However, the Miseq platform produced a majority of sequences with low sequence quality. Eleven of eighteen isolates were sequenced to a depth of coverage between 87,96 x – 140,06 x; two isolates had an average depth of coverage of ~0 and were excluded from further analysis (C-S5 and F/R-S137); while 5 isolates (C-S105d2, C-S126dx, C-S153w8, C-159w4 and F/R-S43dx) had an average depth of coverage of 17,72-58,85x.

Table 4.3.2. Average percentage of mapped reads and depth of coverage calculated based on Qualimap results from alignments produced by BWA, Novoalign and SMALT

Treatment Outcome	Sample ID	Mapping Aligners	Percentage mapped reads	Average depth of coverage (X)	Average % mapped reads (3 mappers)	Average depth of coverage (3 mappers)
Cured	C-S5dx	BWA	98,40	0,67 (+/- 1,12)	98,80	0,67 (+/- 1,13)
		Novoalign	98,23	0,67 (+/- 1,13)		
		SMALT	99,76	0,68 (+/- 1,13)		
	C-S29dx	BWA	99,18	140,4183 (+/- 30,20)	98,97	140,06 (+/-31,45)
		Novoalign	98,27	139,081 (+/- 32,24)		
		SMALT	99,45	140,6721 (+/- 31,75)		
	C-S41w4	BWA	99,34	143,6368 (+/- 31,74)	97,33	143,42 (+/-32,35)
		Novoalign	98,74	142,724 (+/- 33,36)		
		SMALT	99,63	143,9106 (+/-31,96)		
	C-S105dx	BWA	99,1	88,09 (+/- 21,078)	97,43	87,96 (+/-21,44)
		Novoalign	98,48	87.52 (+/- 22.01)		
		SMALT	99,43	87,96 (+/- 21,45)		
	C-S105d2	BWA	95,77	17,74 (+/- 5,89)	95,70	17,72 (+/- 5,95)
		Novoalign	96,19	17,79 (+/- 5,93)		
		SMALT	95,15	17,62 (+/- 6,03)		
	C-S126dx	BWA	98,65	13,28 (+/- 5,37)	99,04	21,86 (+/- 15,06)

		Novoalign	97,80	38,90 (+/- 34,40)		
		SMALT	98,70	13,39 (+/- 5,40)		
	C-S153dx	BWA	99,34	182,72 (+/- 42,37)	97,07	182,41 (+/-43,15)
		Novoalign	98,69	181,47 (+/- 44,46)		
		SMALT	99,61	183,04 (+/- 42,63)		
	C-S153w8	BWA	99,31	58,95 (+/- 15,92)	99,60	58,85 (+/- 16,14)
		Novoalign	98,58	58,52 (+/-16,49)		
		SMALT	99,60	59,07 (1+/-6,02)		
	C-S159dx	BWA	99,25	92,07 (+/- 23,26)	97,61	91,90 (+/-23,77)
		Novoalign	98,42	91,29 (+/- 24,54)		
		SMALT	99,64	92,34 (+/- 23,52)		
	C-S159w4	BWA	98,93	18,70 (+/- 6,15)	98,78	18,67 (+/- 6,20)
		Novoalign	98,10	18,54 (+/-6,34)		
		SMALT	99,31	18,75 (+/-6,20)		
Failed/recurrent	F/R-S43dx	BWA	98,87	40,66 (+/- 12,22)	98,86	40,77 (+/-15,90)
		Novoalign	98,86	40,36 (+/- 14,85)		
		SMALT	98,86	41,28 (+/- 20,61)		
	F/R-S43w24	BWA	98,85	87,79 (+/- 21,44)	98,57	87,52 (+/- 21,90)
		Novoalign	97,72	86,77 (+/- 22,37)		
		SMALT	99,15	87,99 (+/- 21,89)		
	F/R-S112dx	BWA	99,36	158,48 (+/- 36,45)	97,22	157,96 (+/-37,31)
		Novoalign	98,2	156,56 (+/- 38,22)		
		SMALT	99,66	158,84 (+/- 37,27)		
	F/R-S137	BWA	97,55	0,87 (+/- 1,32)	98,26	0,88 (+/- 1,32)
		Novoalign	97,59	0,87 (+/-1,32)		
		SMALT	99,65	0,89 (+/- 1,33		

	F/R-S152	BWA	98,12	89,00 (+/- 24,33)	98,26	88,97 (+/- 24,73)
		Novoalign	97,13	88,13 (+/- 25,55)		
		SMALT	99,53	89,78 (+/- 24,30)		
	F/R-S163dx	BWA	99,31	149,18 (+/- 33,65)	97	148,71 (+/-34,53)
		Novoalign	98,19	147,44 (+/- 35,53)		
		SMALT	99,61	149,52 (+/- 34,41)		
	F/R-S163w24	BWA	99,29	107,86 (+/- 25,66)	99,01	107 (+/- 26,28)
		Novoalign	98,16	106,59 (+/- 26,97)		
		SMALT	99,57	108,08 (+/- 26,22)		
	F/R-S169dx	BWA	83,71	90,70 (+/-21,90)	83,80	90,58 (+/-22,43)
		Novoalign	83,12	90,00 (+/- 23,18)		
		SMALT	84,58	91,04 (+/-22,21)		

4.3.4. High Confidence Variants

4.3.4.1. Comparison of variants identified in the cured patient group vs failed/recurrent patient group

In an attempt to identify common features in the genome of isolates from the failed/recurrent patient group that could explain the increased persister proportions observed in chapter 3 (section 3.3.5), we examined variants that were shared (or unique) in the failed/recurrent patient group, but not identified in the cured patient isolates. However, the data did not reveal an obvious genetic contributor to the differential persister phenotypes or clinical outcome observed.

4.3.4.2. Pairwise comparisons of isolates obtained at diagnosis (Dx) vs later time points (d2, w4, w8 and w24)

To determine if there was any strain evolution during TB treatment, variants identified at diagnosis (dx) and at later time points were compared utilizing the script mentioned in section 4.2.16. Importantly, only 5/13 patients have follow-up samples, with 3 belonging to the cured group (C-S105d2, C-S153w8 and C-S159w4) and 2 belonging to the failed/recurrent patient group (F/R-S43w24 and F/R-S163w24). Unique SNPs were only identified in two patients (F/R-S43 and F/R-S163) from the failed/recurrent patient groups (data not shown).

Unsurprisingly, patient F/R-S43 showed 130 unique variants in both the diagnosis bacterial isolate (dx) and follow-up isolate (w24), as this patient samples revealed reinfection with a different mycobacterial strain. Patient S163 was identified to have a non-synonymous SNP in dx bacterial isolate in *Rv3795*. A separate non-synonymous SNP was identified in the later time point of patient F/R-S163, F/R-S163w24, in gene *Rv2142A*. The pairwise comparison identified little strain evolution during treatment at a genomic level. However, a factor which contributed to the lack of evolution observed in the cured groups is the lack of later time point isolates i.e w24.

4.3.5. Phylogenetic Tree Construction

To assess the relationship of the included *M. tuberculosis* isolates to members of the *Mycobacterium tuberculosis* complex (MTBC), a comprehensive phylogenetic tree was constructed (Fig 4.3.3). The strings of the concatenated variants from all isolates used in this study were considered in fasta format. These strings were matched to each other and members of the MTBC. Previous phylogenetic analyses of *M. tuberculosis* evolution have used a related methodology (Jones *et al.*, 2019). Fig 4.3.3 depicts that all isolates were drug susceptible with majority (62.5%) belonging to lineage 4 (Euro-American) and (37, 5%) belonging to lineage 2 (East-Asian). Due to F/R-S43, F/R-S112, F/R-S163, F/R-S152 and F/R-S137 being part of the same sub-lineage. Fig 4.3.3 depicts a diverse collection of Euro-American sub-lineages, which were not dominated by a single sub-lineage, as isolates of the LAM family and X family predominantly made up the portion of lineage 4 data set.

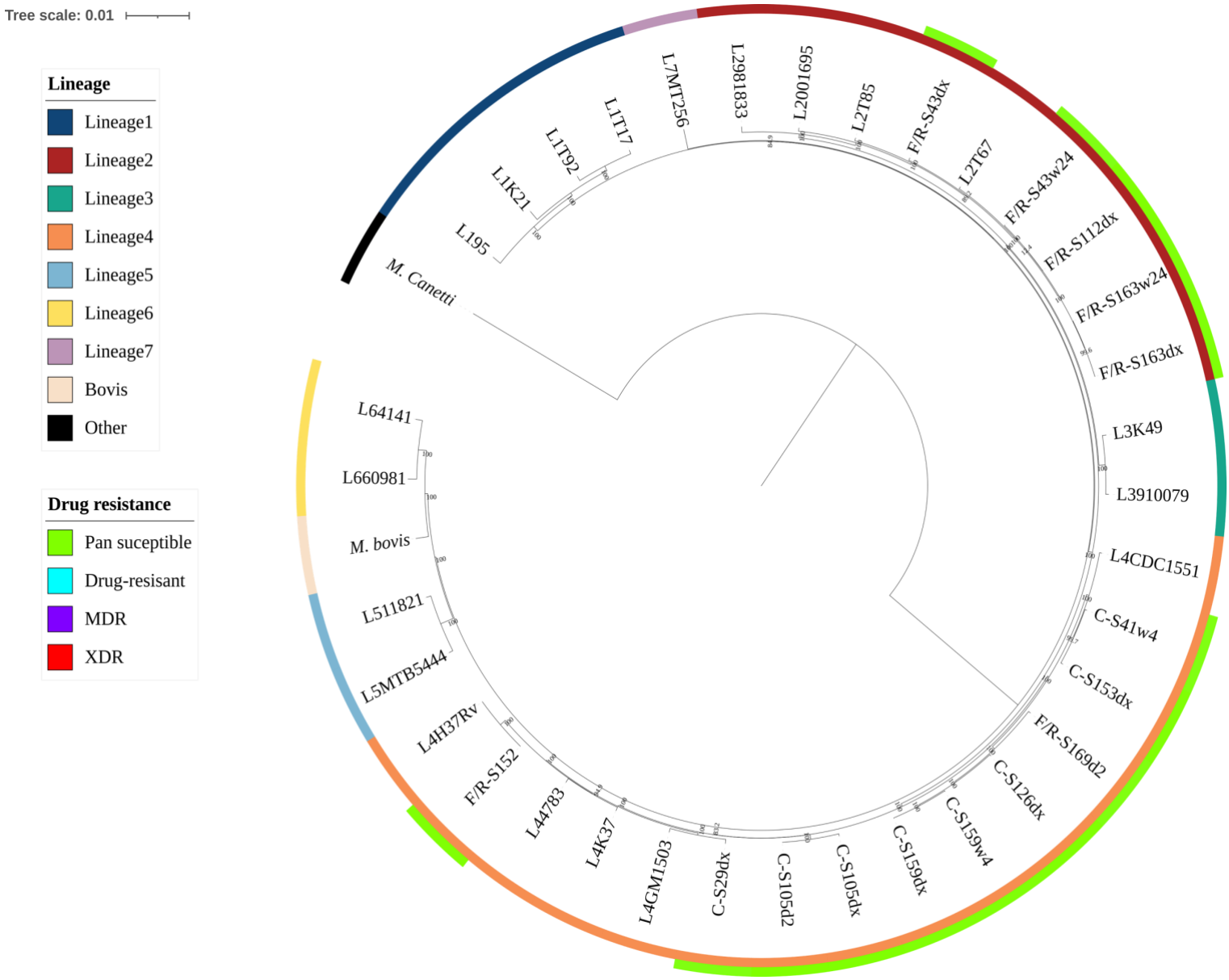


Figure 4.3.3. Molecular phylogenetic analysis by maximum likelihood method with 1000 bootstrap replicates showing the relationship of the included *M. tuberculosis* isolates to other members of the *Mycobacterium tuberculosis* complex. The bootstrap support values are shown next to the nodes. The phylogenetic tree was produced by IQ-TREE which was based on variable sites identified when compared to the *M. tuberculosis* H37Rv reference sequence (Letuni and Bork., 2019).

4.4. Discussion

In this component of the study, we wished to determine whether there were genetic contributors to the differential persister formation observed in the two patient groups with different clinical outcomes (section 3.3.5). We further wished to rule out that the clinical outcomes are linked to drug resistance-conferring mutations. This was aided by WGS analysis of selected clinical isolates. Genome sequencing was carried out on 2 different Illumina platforms, namely the NextSeq and Miseq platforms. For data analysis, we utilized a reference genome sequencing approach to *M. tuberculosis* H37Rv in combination with a *M. tuberculosis* customised University Stellenbosch automated pipeline (USAP) which utilized a multi-software approach to determine unique variants per isolate.

4.4.1. Data clean-up and quality control

A major challenge for WGS data analyses of *M. tuberculosis* is that the genome has several highly repetitive regions, which complicates bioinformatics analysis of the data. This property may cause alignment algorithms to map a read to the wrong location in the reference genome, predominantly around repetitive areas of low-complexity regions. Utilizing multiple aligners minimises the identification of false positive variants. Thus, regions which are annotated as *pks*, insertions, *esx*, repeat, phages, polyketide, or transposase in gene name or gene product description in the annotated variant file were removed prior to variant calling analysis.

The USAP pipeline includes several data clean-up and quality control steps. One of the steps to ensure good sequence quality is the use of the Trimmomatic tool which is a pre-processing tool that removes and poor-quality sections allowing downstream analysis on good quality reads. Fig 4.3.1. depicts a representative of the per base quality as observed in FastQC post sequencing on the Nextseq platform, showing good per base sequence quality. In contrast, Fig 4.3.2. depicts a representative of Miseq per base quality, showing the majority of per base quality scores being in the (orange-red) y-axis background indicative of reasonable to poor quality.

Following sequence trimming, data was subjected to the 3 alignment algorithms (BWA, Novoalign and SMALT). Using various software packages read alignment to the reference genome, including depth of coverage of each sequence was assessed. Overall, >90% of reads from clinical isolates mapped to *M. tuberculo*sis H37Rv, indicative of having no contamination (Table 4.3.2). Depth of coverage provided a further measure of data quality; a high depth of

coverage of ≥ 80x is recommended as this increases confidence in variants observed. However, a minimum depth of coverage of 30x has been observed to provide accurate variants in *M. tuberculosis* studies (Colman *et al.*, 2019).When depth of coverage was assessed for our sample set, it was revealed that isolates C-S5 and C-S137 showed a depth of coverage of <1, indicating that variants discovered in these isolates could not be assessed with any degree of confidence. These were therefore excluded from further analyses.

4.4.2. Antibiotic susceptibility and lineage specificity

To rule out genetically-encoded antibiotic resistance of mycobacterial isolates obtained from cured and failed/recurrent patient groups as a contributing factor to the different clinical outcomes (as previously observed by Malherbe *et al.*, 2016) we subjected WGS data to a publicly available profiling tool, TB-profiler. TB-profiler was specifically created for the identification of known antibiotic resistance mutations as well as for strain identification of *M. tuberculosis* (Coll *et al.*, 2015; Phelan *et al.*, 2019). Fig 4.3.3 showed that all isolates are drug susceptible. Thus, it can be reasonably concluded that drug resistance had no impact on the phenotypic persister-like population identified from macrophage infections.

Lineage identification was executed with the open source TB-profiler pipeline and confirmed by creating a phylogenetic tree with scripts and IQ-Tree. The pipeline utilizes similar tools such as USAP for drug susceptibility and lineage profiling. Strains identified were observed to belong to the modern MTBC lineages, with ~60% belonging to the Euro-American lineage 4 and ~40% belonging to the East-Asian lineage 2 (Fig 4.3.3). The observed lineage distribution is similar to what is expected in the Western Cape (Nicol *et al.*, 2005; Rutaihwa *et al.*, 2019). Strain evolution during treatment is inconclusive as absolute distance of SNP difference between baseline and follow-up could not be determined. Given the limited number of samples included in the phylogeny, two strains could cluster together on the tree but be more than 5-12 SNPs apart.

4.4.3. Unique variant recognition

Recent literature has provided evidence of genetic components which could predispose cells to enter a persister-like state (Torrey *et al.*, 2016; Safi *et al.*, 2019). Thus, we attempted to identify unique variants in isolates obtained from failed/recurrent patient group compared to the cured patient group. The lack of unique variants suggests that treatment outcome groups are not associated with genetic contributors, although current sample numbers would need to be

expanded for a definitive conclusion. It is noted that phenotypic antibiotic tolerance as well as epistasis could be responsible for the increased persister formation observed. However, regarding epistasis data suggests that non-synonmous mutation had limited or no contribution to the phenotype observed. Furthermore, phenotypic antibiotic tolerance has been shown to be directly proportional to the rate of antibiotic resistance emergence in clinical isolates of *E. coli* and *P. aeruginosa* (Vogwill *et al.*, 2016; Windels *et al.*, 2019). This could be a possible avenue of research in future.

Initially it was speculated that genetic variants could lead to the increased intracellular replication observed in isolate C-S126 (Figure 3.3.8). However, upon analysis of variants following WGS no variations in genomic data is suggested to explain the observed phenomenon.

Following a pairwise comparison, we assessed if unique variants could be identified between dx samples and follow-up samples at a later time point. Unsurprisingly, unique variants were only identified in one patient sample, F/R-S163. Variants identified in both dx and w24 isolates were synonymous SNPs in *Rv3795* and *Rv2142A* respectively, which do not cause amino acid changes affecting protein products. However, synonymous mutations have been shown to change translation initiation, mRNA stability as well as protein folding (Kristofich *et al.*, 2018). While the majority of synonymous mutations are neutral, their effects may be magnified under strong selection processes. Synonymous mutations could lead to codon bias. Codon biases could lead to either the overexpression or decreased expression of certain products which could impact persister formation (Walsh *et al.*, 2020). For example; in *E. coli*, a synonymous mutation in a gene upstream of *inhA*, encoding the target of isoniazid (used in TB treatment), generates a new promoter and increases *inhA* expression by 3-4 fold (Ando *et al.*, 2014). Mutations in *Rv3795* have been linked to low or moderate resistance as it is an essential gene involved in the biosynthesis of the mycobacterial cell wall arabinan, which is the drug target of ethambutol (Phelan *et al.*, 2016). Disruption of *Rv2142A* has been shown to increase growth of *in vitro* H37Rv, by analysing of saturated *HimaR* transposon libraries (Dejesus *et al.*, 2017). The use of RNA sequencing would be beneficial in determining whether codon bias has an effect on persister formation.

Additionally, isolates obtained from patient F/R-S43 showed approximately 130 unique variants (more than threshold of 5 or 12 SNPs) to the baseline (F/R-S43dx) and follow-up (F/R-

S43w24) isolates (data not shown). This is an indication that a reinfection with a closely related strain also belonging to lineage 2 has likely occurred. This explains why the two isolates appear adjacent to one another in the phylogenetic tree (Fig 4.3.3). Thus, it can be concluded that treatment failure was due to reinfection and likely not due to the unique genetic make-up of baseline vs follow-up isolates. As F/R-S43 and F/R-S43w24 is suggested to be an infection with a different strain of M. tuberculosis from the same lineage. It could be speculated that F/R-S43w24 would have an increased persister frequency as it was exposed to antibiotic treatment upon initial infection.

4.5. Limitations

Ideally, this portion of the study would need to be repeated in a larger sample size to make adequate conclusions of whether there is a genetic component that predisposes persister formation. Additionally, low quality sequences posed another limitation as samples had to be removed from sequencing analyses that reduced the sample size.

4.6. Future work:

- Following variant calling if unique variants were determined within the failed/recurrent group. Tools such as polyphen-2, swift and provean to assess the impact of SNVs and in/dels on the biological function of the protein. If mutations were observed it would be interesting to see in which pathways these mutations could affect. pathway analyses KEGG analyses would be good. You would be able to look at genotypic info and see whether pathway analyses highlight involvement in specific genes/functions.
- Resequencing of isolates with poor sequence quality would need to be done with the Illumina NextSeq platform.
- WGS would need to be performed on a larger cohort to increase accuracy of relationship between genomic data and phenotypic occurrence of persister-like cells. A larger cohort could be determined utilising power calculations.
- Deep WGS could be implemented to assess underlying minority populations in serial sputum samples. Similar to Liu et al., 2015.
- Epigenetic studies on persister-like cells could be performed.

4.7. Conclusion

Overall this section of this study has contributed new WGS data from strains with different *in vitro* phenotype and clinical outcomes. Unfortunately, the question of whether genetic composition predisposes persister of strains was not answered, due to the low sample number for what is a complex phenotype. A possibility to consider is that the persister phenotype observed in section 3.3.5 might be epigenetic rather than genetic.

Chapter 5

General conclusion

Approximately one third of the world's population is asymptomatically infected with *Mycobacterium tuberculosis,* (Gill *et al.*, 2009) the causative agent of tuberculosis (TB). In most *M. tuberculosis*-infected individuals, the infection persists in a latent, asymptomatic state that can continue for decades with the potential to reactivate later in life (Stewart, Robertson and Young, 2003). Therefore, therapies that aim to eliminate TB should target dormant organisms, since these could resume replication to cause active disease.

Previously it has been shown that persistent mycobacteria arise in response to environmental stressors encountered in the host and adopt a slow or non-replicating state (North and Jung, 2004; Liu *et al.*, 2019). This small, viable, but non-replicating (VBNR) population is likely to be antibiotic-tolerant (Balaban *et al.*, 2019). Currently the majority of drug therapies target actively growing bacteria, however persister bacteria comprise an important subpopulation of bacteria that is recalcitrant to antibiotic treatment (Gill *et al.*, 2009). Importantly, VBNR bacterial populations are phenotypically drug tolerant, but not genetically resistant. Drug tolerant populations have been determined to be a contributing factor to the requirement for lengthy drug treatment and give rise to genetically resistant progeny. However, little is known about mycobacterial persisters, since they comprise of only 1% of the bacterial population and are non- or slowly growing, making them difficult to isolate. A major knowledge gap exists regarding the genetic contributors to persisters, likely to be involved in recurrent/failed TB treatment outcome.

In a study by Malherbe *et al* lesion activity in lungs and the presence of *M. tuberculosis* mRNA were identified post TB treatment, which is suggestive of unculturable bacteria likely being persisters (Malherbe *et al.*, 2016). Lesions were found to have variable fluorodeoxyglucose F 18 (FDG) uptake, suggestive of a heterogeneous mycobacterial population in the failed/recurrent patient group.

Therefore, the aim of this study was to determine if *M. tuberculosis* isolates from failed/recurrent TB patients are more likely to be predisposed to persister formation than those who were cured in response to treatment. For assessing persister proportions in all clinical isolates from cured and recurrent/failed patient groups, isolates were transformed with a FD dual-reporter plasmid and a THP-1 infection model was used to mimic host environmental

conditions that *M. tuberculosis* encounter during PTB infections. The THP-1 infection model has been found to be similar in bacterial uptake, viability and host response as that of primary human monocyte-derived macrophages (MDMs), which are considered to be the first line defence against mycobacterial infection (Madhvi *et al.*, 2019). In addition, next generation sequencing analyses of the isolates were performed to characterise the isolates, investigate strain evolution during treatment and determine whether sequence variation predisposed persister formation in clinical isolates from cured and failed/recurrent patient groups.

Flow cytometric results demonstrated heterogeneous *in vitro* growth of *M. tuberculosis* clinical strains from both patient groups, as opposed to previous work (Mouton *et al.*, 2016). This suggests that clinical strains are phenotypically and genetically more heterogeneous prior to host environmental stress (Zhang, Yew and Barer, 2012; Khare and Tavazoie, 2020). Heterogeneity prior to infection has been linked to increased host adaptability of strains (Jain *et al.*, 2016), which could lead to uncleared/recurrent infection, even after antibiotic treatment. Interestingly, one isolate C-S126dx was observed to have an increased intracellular growth compared to *in vitro* growth. Initially it was speculated to be as a result of a genetic component. However, no obvious genetic variations were observed that could explain this observation. For example, deletion of the *pknH* gene was found to increase bacillary load during infections in BALB/c mice (Papavinasasundaram *et al.*, 2005). In addition to *in vitro* population heterogeneity, intracellular replication dynamics revealed the presence of a persister-like population in both cured and failed/recurrent groups 120 hours post infection. This suggests that host environments induce the formation of persister populations, which are inherently present prior to infection. This data which showed bacterial heterogeneity at baseline upon treatment initiation detected in replication dynamics have not been previously observed within clinical isolates. Furthermore, the frequency of persister-like cells was greater in the isolates from the failed/recurrent group compared to the cured group. It is tempting to speculate that these bacterial populations could possibly predispose patient treatment outcome. This however requires additional confirmation. Persister frequency within the failed/recurrent group was observed to be strain dependent as strains demonstrated variability in their propensity to form persisters in response to macrophage stress. The heterogeneity in persister proportions between strains could be attributed to various mechanisms bacterial persisters employ to survive host defense mechanisms, which could be lineage dependent. Isolates F/R-S112dx, F/R-S163dx, F/R-S169dx and F/R-S43dx are from lineage 2 and were found to have higher persister-like formation (FigureS3.2). Lineage 2 has been linked to increased virulence and improved

intracellular survival of *M. tuberculosis* in macrophages (Tram *et al.*, 2018). However, further analysis would be needed to determine the link between high virulence *M. tuberculosis* lineages and their predisposition for persister formation. Furthermore, it will be interesting to determine the effect of other models such as an *in vitro* granuloma model, or murine models on persister formation of failed/recurrent isolates, as this would more accurately represent host-pathogen interactions experienced during TB disease progression.

To determine whether antibiotic resistance and other genetic components predispose persister formation, a customized WGS analysis pipeline was applied. This allowed for the identification of genetic variation within clinical isolates between cured and failed/recurrent treatment groups at baseline and strain evolution during TB treatment. Based on the WGS analyses results, drug resistance was not a contributor to the persister formation identified in both cured and failed/recurrent patient groups as all isolates were identified to be drug susceptible. Following WGS analyses no unique variants were identified when treatment outcome groups were compared. Whether, the data suggests that there are no underlying genetic contributors to persister formation is inconclusive as the sample size is a limiting factor. Moreover, comparisons of variants of baseline and follow-up samples to determine strain evolution during treatment were inconclusive as limited isolates had follow-up samples from a later time point i.e w24. It can be speculated that persister formation that was observed during macrophage infections are as a result of epigenetic changes. An increased sample size would thus be beneficial to confirm that no genetic contributor is underlying the persister formation in these isolates. A study by Colengeli et al.,2018 used 1004 samples which was collected from 1995 till 2002 which determined a correlation between mics and relapse. It can therefore be speculated that a sample size similar to Colengeli et al., 2018 would be efficient.

This is one of the first studies to combine a macrophage infection model with WGS data to investigate the phenotypic and genotypic characteristics of persister formation in *M. tuberculosis* clinical isolates. In summary, the results from this study emphasize the heterogeneity of *M. tuberculosis* clinical isolates both *in vitro* and in response the host environment stress during infection, which could contribute to the adaptability of clinical isolates to stress environments and their ability to survive TB treatment. The study suggests that increased persister-like formation in isolates from the recurrent/failed group are likely not

as a result of genetic variation. Yet, data suggests persisters play a role in unfavourable TB treatment outcome.

Chapter 6

References

Adams, K. N. *et al.* (2011) 'Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism', *Cell*. Elsevier Inc., 145(1), pp. 39–53. doi: 10.1016/j.cell.2011.02.022.
Agarwal, P. *et al.* (2016) 'Mycobacterium tuberculosis can gain access to adipose depots of mice infected via the intra-nasal route and to lungs of mice with an infected subcutaneous fat implant', *Microbial Pathogenesis*. Elsevier Ltd, 93(January 2016), pp. 32–37. doi: 10.1016/j.micpath.2016.01.004.
Ahmad, Z. *et al.* (2009) 'Biphasic kill curve of isoniazid reveals the presence of drug-tolerant, not drug-resistant, mycobacterium tuberculosis in the guinea pig', *Journal of Infectious Diseases*. United States, 200(7), pp. 1136–1143. doi: 10.1086/605605.
Allen, J. L. (1992) 'A modified Ziehl-Neelsen stain for mycobacteria.', *Medical laboratory sciences*. England, 49(2), pp. 99–102.
Amato, S. M. and Brynildsen, M. P. (2015) 'Persister Heterogeneity Arising from a Single Metabolic Stress.', *Current biology : CB*. England, 25(16), pp. 2090–2098. doi: 10.1016/j.cub.2015.06.034.
Amato, S. M., Orman, M. A. and Brynildsen, M. P. (2013) 'Metabolic Control of Persister Formation in Escherichia coli', *Molecular Cell*. Elsevier Inc., 50(4), pp. 475–487. doi: 10.1016/j.molcel.2013.04.002.
Ando, H. *et al.* (2014) 'A silent mutation in mabA confers isoniazid resistance on Mycobacterium tuberculosis', *Molecular Microbiology*, 91(3), pp. 538–547. doi: 10.1111/mmi.12476.
Andrews, S. (2010) 'FastQC'. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
Anes, E. *et al.* (2006) 'Dynamic life and death interactions between mycobacterium smegmatis and J774 macrophages', *Cellular Microbiology*, 8(6), pp. 939–960. doi: 10.1111/j.1462-5822.2005.00675.x.
Armstrong, J. A. and Hart, D. (1971) 'Response of cultured macrophages to Mycobacterium Tuberculosis, with observations on fusion of lysosomes with phagosomes', *Journal of Experimental Medicine*, 134(3), pp. 713–740. doi: 10.1084/jem.134.3.713.
Atkinson, G. C., Tenson, T. and Hauryliuk, V. (2011) 'The RelA/SpoT Homolog (RSH) superfamily: Distribution and functional evolution of ppgpp synthetases and hydrolases across the tree of life', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0023479.
Avarbock, D. *et al.* (1999) 'Cloning and characterization of a bifunctional RelA/SpoT homologue from Mycobacterium tuberculosis', *Gene*, 233(1–2), pp. 261–269. doi: 10.1016/S0378-1119(99)00114-6.
Baek, S. H., Li, A. H. and Sassetti, C. M. (2011) 'Metabolic regulation of mycobacterial growth and antibiotic sensitivity', *PLoS Biology*, 9(5). doi: 10.1371/journal.pbio.1001065.
Bakkeren, E., Diard, M. and Hardt, W. D. (2020) 'Evolutionary causes and consequences of bacterial antibiotic persistence', *Nature Reviews Microbiology*. Springer US, 18(9), pp. 479–490. doi: 10.1038/s41579-020-0378-z.
Balaban, N. *et al.* (2004) 'Bacterial Persistence as a Phenotypic Switch', *American Association for the Advancement of Science*, 305(5690), p. 55. Available at: http://eprints.uanl.mx/5481/1/1020149995.PDF.
Balaban, N. Q. *et al.* (2019) 'Definitions and guidelines for research on antibiotic

persistence', *Nature Reviews Microbiology*. Springer US, 17(7), pp. 441–448. doi: 10.1038/s41579-019-0196-3.
Barandun, J., Delley, C. L. and Weber-ban, E. (2012) 'The pupylation pathway and its role in mycobacteria', *biomedcentral Journal of Biology*, 10(95).
Barberis, I. *et al.* (2017) 'The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus', *Journal of Preventive Medicine and Hygiene*, 58(1), pp. E9–E12. doi: 10.15167/2421-4248/jpmh2017.58.1.728.
Barry, C. E. *et al.* (2009) 'The spectrum of latent tuberculosis: rethinking the goals of prophylaxis', *Nat Rev Microbiol*, 7(12), pp. 845–855. doi: 10.1038/nrmicro2236.The.
Barth, V. C. *et al.* (2019) 'Toxin-mediated ribosome stalling reprograms the Mycobacterium tuberculosis proteome', *Nature Communications*. Springer US, 10(1). doi: 10.1038/s41467-019-10869-8.
Barth, V. C. and Woychik, N. A. (2020) 'The Sole Mycobacterium smegmatis MazF Toxin Targets tRNALys to Impart Highly Selective, Codon-Dependent Proteome Reprogramming', *Frontiers in Genetics*, 10(February), pp. 1–11. doi: 10.3389/fgene.2019.01356.
Bellerose, M. M. *et al.* (2019) 'Common variants in the glycerol kinase gene reduce tuberculosis drug efficacy', *mBio*, 10(4), pp. 1–15. doi: 10.1128/mBio.00663-19.
van den Bergh, B., Fauvart, M. and Michiels, J. (2017) 'Formation, physiology, ecology, evolution and clinical importance of bacterial persisters', *FEMS Microbiology Reviews*, 41(3), pp. 219–251. doi: 10.1093/femsre/fux001.
Berney, M., Hartman, T. E. and Jacobs, W. R. (2014) 'A Mycobacterium tuberculosis cytochrome bd oxidase mutant is hypersensitive to bedaquiline', *mBio*, 5(4), pp. 5–7. doi: 10.1128/mBio.01275-14.
Beste, D. J. V. *et al.* (2009) 'The genetic requirements for fast and slow growth in mycobacteria', *PLoS ONE*, 4(4). doi: 10.1371/journal.pone.0005349.
Bigger, J. (1944) 'TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION', *The Lancet*, 244(6320), pp. 497–500. doi: https://doi.org/10.1016/S0140-6736(00)74210-3.
Blanc, L. *et al.* (2018) 'Impact of immunopathology on the antituberculous activity of pyrazinamide', *Journal of Experimental Medicine*, 215(8), pp. 1975–1986. doi: 10.1084/jem.20180518.
Blouin, Y. *et al.* (2012) 'Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0052841.
Boldrin, F. *et al.* (2020) 'Tolerance and Persistence to Drugs: A Main Challenge in the Fight Against Mycobacterium tuberculosis', *Frontiers in Microbiology*, 11(August). doi: 10.3389/fmicb.2020.01924.
Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.
Brown, D. R. (2019) 'Nitrogen Starvation Induces Persister Cell Formation in Escherichia coli', *Journal of bacteriology*. United States, 201(3).
Bryk, R. *et al.* (2008) 'Selective Killing of Nonreplicating Mycobacteria', *Cell Host and Microbe*, 3(3), pp. 137–145.
Bush, M. J. (2018) 'The actinobacterial WhiB-like (Wbl) family of transcription factors', *Molecular Microbiology*, 110(5), pp. 663–676. doi: 10.1111/mmi.14117.
Cameron, D. R. *et al.* (2018) 'A genetic determinant of persister cell formation in bacterial pathogens', *Journal of Bacteriology*, 200(17), pp. 1–11. doi: 10.1128/JB.00303-18.
Carey, A. *et al.* (2018) 'TnSeq of Mycobacterium tuberculosis clinical isolates reveals strain-specific antibiotic liabilities', *Bulletin of the Korean Chemical Society*, 14(3), pp. 1729–1732.

doi: 10.1002/bkcs.10316.
Casonato, S. *et al.* (2012) 'WhiB5, a transcriptional regulator that contributes to mycobacterium tuberculosis virulence and reactivation', *Infection and Immunity*, 80(9), pp. 3132–3144. doi: 10.1128/IAI.06328-11.
Chakraborty, P. *et al.* (2013) 'Drug Resistant Clinical Isolates of Mycobacterium tuberculosis from Different Genotypes Exhibit Differential Host Responses in THP-1 Cells', *PLoS ONE*, 8(5), pp. 1–9. doi: 10.1371/journal.pone.0062966.
Chan, H. *et al.* (2017) 'Potential and use of bacterial small RNAs to combat drug resistance: A systematic review', *Infectiofile:///C:/Users/JLCOETZEE/Downloads/fmicb-08-00803.pdfn and Drug Resistance*, 10, pp. 521–532. doi: 10.2147/IDR.S148444.
Chauhan, R. *et al.* (2016) 'Reconstruction and topological characterization of the sigma factor regulatory network of Mycobacterium tuberculosis', *Nature Communications*, 7(May 2015). doi: 10.1038/ncomms11062.
Chen, D. Y. *et al.* (2020) 'MicroRNA-889 inhibits autophagy to maintain mycobacterial survival in patients with latent tuberculosis infection by targeting TWEAK', *mBio*, 11(1), pp. 1–17. doi: 10.1128/mBio.03045-19.
Choudhary, E. *et al.* (2019) 'Conditional silencing by CRISPRi reveals the role of DNA gyrase in formation of drug-tolerant persister population in Mycobacterium tuberculosis', *Frontiers in Cellular and Infection Microbiology*, 9(MAR), pp. 1–13. doi: 10.3389/fcimb.2019.00070.
Cicchese, J. M. *et al.* (2020) 'Both Pharmacokinetic Variability and Granuloma Heterogeneity Impact the Ability of the First-Line Antibiotics to Sterilize Tuberculosis Granulomas', *Frontiers in Pharmacology*, 11(March), pp. 1–15. doi: 10.3389/fphar.2020.00333.
Cohen, T. *et al.* (2016) 'Within-host heterogeneity of mycobacterium tuberculosis infection is associated with poor early treatment response: A prospective cohort study', *Journal of Infectious Diseases*, 213(11), pp. 1796–1799. doi: 10.1093/infdis/jiw014.
Cole, S. *et al.* (1998) 'Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence', *Nature*, 393(NOVEMBER), pp. 537–544.
Coll, F. *et al.* (2015) 'Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences', *Genome Medicine*. ???, 7(1), pp. 1–10. doi: 10.1186/s13073-015-0164-0.
Colman, R. E. *et al.* (2019) 'Whole-genome and targeted sequencing of drug-resistant Mycobacterium tuberculosis on the iSeq100 and MiSeq: A performance, ease-of-use, and cost evaluation', *PLoS Medicine*, 16(4), pp. 1–13. doi: 10.1371/journal.pmed.1002794.
Comas, Ĩ. *et al.* (2010) 'Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved', *Nature Genetics*, 42(6), pp. 498–503. doi: 10.1038/ng.590.
Copin, R. *et al.* (2016) 'Within Host Evolution Selects for a Dominant Genotype of Mycobacterium tuberculosis while T Cells Increase Pathogen Genetic Diversity', *PLoS Pathogens*, 12(12), pp. 1–16. doi: 10.1371/journal.ppat.1006111.
Coscolla, M. and Gagneux, S. (2010) 'Does M. tuberculosis genomic diversity explain disease diversity?', *Drug Discovery Today: Disease Mechanisms*, 7(1), pp. 1–26. doi: 10.1016/j.ddmec.2010.09.004.
Dainese, E. *et al.* (2006) 'Posttranslational regulation of Mycobacterium tuberculosis extracytoplasmic-function sigma factor σL and roles in virulence and in global regulation of gene expression', *Infection and Immunity*, 74(4), pp. 2457–2461. doi: 10.1128/IAI.74.4.2457-2461.2006.
Davies, J.; Davies, D. (2010) 'Origins and evolution of antibiotic resistance.', *Microbiología (Madrid, Spain)*, 12(1), pp. 9–16. doi: 10.1128/mmbr.00016-10.
Dejesus, M. A. *et al.* (2017) 'Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis', *mBio*, 8(1), pp. 1–17. doi:

10.1128/mBio.02133-16.
Delincé, M. J. *et al.* (2016) 'A microfluidic cell-trapping device for single-cell tracking of host-microbe interactions', *Lab on a Chip*. Royal Society of Chemistry, 16(17), pp. 3276–3285. doi: 10.1039/c6lc00649c.
Depristo, M. A. *et al.* (2011) 'HHS Public Access', 43(5), pp. 491–498. doi: 10.1038/ng.806.A.
Dersch, P. *et al.* (2017) 'Roles of regulatory RNAs for antibiotic resistance in bacteria and their potential value as novel drug targets', *Frontiers in Microbiology*, 8(MAY), pp. 1–12. doi: 10.3389/fmicb.2017.00803.
Dhar, N. and McKinney, J. (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*, 107(27), pp. 12275–12280. doi: 10.1073/pnas.1003219107.
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*. National Academy of Sciences, 107(27), pp. 12275–80. doi: 10.1073/pnas.1003219107.
Dhiman, R. K. *et al.* (2009) 'Menaquinone Synthesis is Critical for Maintaining Mycobacterial Viability During Exponential Growth and Recovery from Non- Replicating Persistence', *Molecular microbiology*, 72(1), pp. 85–97. doi: 10.1111/j.1365-2958.2009.06625.x.Menaquinone.
Diard, M. and Hardt, W. D. (2017) 'Evolution of bacterial virulence', *FEMS microbiology reviews*, 41(5), pp. 679–697. doi: 10.1093/femsre/fux023.
Dippenaar, A. *et al.* (2015) 'Whole genome sequence analysis of Mycobacterium suricattae', *Tuberculosis*. Elsevier Ltd, 95(6), pp. 682–688. doi: 10.1016/j.tube.2015.10.001.
Dörr, T., Lewis, K. and Vulić, M. (2009) 'SOS Response Induces Persistence to Fluoroquinolones in Escherichia coli', *PLoS Genetics*, 5(12), pp. 60–66.
Drain, P. K. *et al.* (2018) 'Incipient and subclinical tuberculosis: A clinical review of early stages and progression of infection', *Clinical Microbiology Reviews*, 31(4), pp. 1–24. doi: 10.1128/CMR.00021-18.
Van Embden, J. D. A. *et al.* (1993) 'Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology', *Journal of Clinical Microbiology*, 31(2), pp. 406–409. doi: 10.1128/jcm.31.2.406-409.1993.
Ernst, J. D. (1998) 'Macrophage Receptors for Mycobacterium tuberculosis', *Infection and Immunity*, 66(4), pp. 1277–1281. doi: 10.1128/iai.66.4.1277-1281.1998.
Fauvart, M., de Groote, V. N. N. and Michiels, J. (2011) 'Role of persister cells in chronic infections: Clinical relevance and perspectives on anti-persister therapies', *Journal of Medical Microbiology*, 60(6), pp. 699–709. doi: 10.1099/jmm.0.030932-0.
Felden, B. and Cattoira, V. (2018) 'Bacterial Adaptation to Antibiotics through Regulatory RNAs', *Antimicrobial Agents and Chemotherapy*, 62(5), pp. 1–11. doi: 10.1128/AAC.02503-17.
Fisher, R. A., Gollan, B. and Helaine, S. (2017) 'Persistent bacterial infections and persister cells', *Nature Reviews Microbiology*. Nature Publishing Group, 15(8), pp. 453–464. doi: 10.1038/nrmicro.2017.42.
Fletcher, W. and Yang, Z. (2010) 'The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection', *Molecular Biology and Evolution*, 27(10), pp. 2257–2267. doi: 10.1093/molbev/msq115.
Flynn, J. L. and Chan, J. (2001) 'Tuberculosis : Latency and Reactivation MINIREVIEW Tuberculosis : Latency and Reactivation', *Infection and immunity*, 69(7), pp. 4195–4201. doi: 10.1128/IAI.69.7.4195.

Francez-Charlot, A. *et al.* (2009) 'Sigma factor mimicry involved in regulation of general stress response', *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), pp. 3467–3472. doi: 10.1073/pnas.0810291106.
Gaca, A. O., Colomer-Winter, C. and Lemos, J. A. (2015) 'Many means to a common end: The intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis', *Journal of Bacteriology*, 197(7), pp. 1146–1156. doi: 10.1128/JB.02577-14.
García-Alcalde, F. *et al.* (2012) 'Qualimap: Evaluating next-generation sequencing alignment data', *Bioinformatics*, 28(20), pp. 2678–2679. doi: 10.1093/bioinformatics/bts503.
Geddes-McAlister, J., Kugadas, A. and Gadjeva, M. (2019) 'Tasked with a challenging objective: why do neutrophils fail to battle', *Pathogens*.
Geiman, D. E. *et al.* (2006) 'Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven Mycobacterium tuberculosis whiB-like genes', *Antimicrobial Agents and Chemotherapy*, 50(8), pp. 2836–2841. doi: 10.1128/AAC.00295-06.
Gerrick, E. R. *et al.* (2018) 'Small RNA profiling in mycobacterium tuberculosis identifies mrsi as necessary for an anticipatory iron sparing response', *Proceedings of the National Academy of Sciences of the United States of America*, 115(25), pp. 6464–6469. doi: 10.1073/pnas.1718003115.
Gill, W. P. *et al.* (2009) 'A replication clock for Mycobacterium tuberculosis', *Nature Medicine*, 15(2), pp. 211–214. doi: 10.1038/nm.1915.
Glickman, M. S., Cox, J. S. and Jacobs, W. R. (2000) 'A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis', *Molecular Cell*, 5(4), pp. 717–727. doi: 10.1016/S1097-2765(00)80250-6.
Gold, B. and Nathan, C. (2017) ' Targeting Phenotypically Tolerant Mycobacterium tuberculosis ', *Tuberculosis and the Tubercle Bacillus*, 5(1), pp. 317–360. doi: 10.1128/9781555819569.ch15.
Gollan, B. *et al.* (2019) 'Bacterial persisters and infection: Past, present, and progressing', *Annual Review of Microbiology*, 73, pp. 359–385. doi: 10.1146/annurev-micro-020518-115650.
Gomez, J. E. and Bishai, W. R. (2000) 'whmD is an essential mycobacterial gene required for proper septation and cell division', *Proceedings of the National Academy of Sciences of the United States of America*, 97(15), pp. 8554–8559. doi: 10.1073/pnas.140225297.
Goossens, S. N., Sampson, S. L. and Rie, A. Van (2021) 'Mechanisms of Drug-Induced Tolerance in Mycobacterium tuberculosis', (October 2020), pp. 1–21.
Grant, S. S. and Hung, D. T. (2013) 'Persistent bacterial infections, antibiotic tolerance, and the oxidative stress response', *Virulence*, 4(4), pp. 273–283. doi: 10.4161/viru.23987.
Greening, C., Grinter, R. and Chiri, E. (2019) 'Uncovering the Metabolic Strategies of the Dormant Microbial Majority: towards Integrative Approaches', *mSystems*, 4(3), pp. 1–5. doi: 10.1128/msystems.00107-19.
Hall, A. M., Gollan, B. and Helaine, S. (2017) 'Toxin–antitoxin systems: reversible toxicity', *Current Opinion in Microbiology*, 36, pp. 102–110. doi: 10.1016/j.mib.2017.02.003.
Harms, A., Maisonneuve, E. and Gerdes, K. (2016) 'Mechanisms of bacterial persistence during stress and antibiotic exposure', *Science*, 354(6318). doi: 10.1126/science.aaf4268.
Helaine, S. *et al.* (2010) 'Dynamics of intracellular bacterial replication at the single cell level', *Proceedings of the National Academy of Sciences*, 107(8), pp. 3746–3751. doi: 10.1073/pnas.1000041107.
Helaine, S. *et al.* (2014) 'Internalization of salmonella by macrophages induces formation of nonreplicating persisters', *Science*, 343(6167), pp. 204–208. doi: 10.1126/science.1244705.
Helaine, S. and Kugelberg, E. (2014) 'Bacterial persisters : formation , eradication , and experimental systems', *Trends in Microbiology*. Elsevier Ltd, 22(7), pp. 417–424. doi:

10.1016/j.tim.2014.03.008.
Hill, P. W. S. and Helaine, S. (2019) 'Antibiotic persisters and relapsing salmonella enterica infections', in Lewis, K. (ed.) *Persister Cells and Infectious Disease*. Cham: Springer International Publishing, pp. 19–38. doi: 10.1007/978-3-030-25241-0_2.
Hobby, G. L., Meyer, K. and Chaffee, E. (1942) 'Observations on the Mechanism of Action of Penicillin.', *Proceedings of the Society for Experimental Biology and Medicine*. SAGE Publications, 50(2), pp. 281–285. doi: 10.3181/00379727-50-13773.
Høiby, N., Ciofu, O. and Bjarnsholt, T. (2010) 'Pseudomonas aeruginosa biofilms in cystic fibrosis', *Future Microbiology*. Future Medicine, 5(11), pp. 1663–1674. doi: 10.2217/fmb.10.125.
Hu, Y. *et al.* (2000) 'Detection of mRNA transcripts and active transcription in persistent Mycobacterium tuberculosis induced by exposure to rifampin or pyrazinamide', *Journal of Bacteriology*, 182(22), pp. 6358–6365. doi: 10.1128/JB.182.22.6358-6365.2000.
Iacobino, A. *et al.* (2021) 'Moxifloxacin activates the sos response in mycobacterium tuberculosis in a dose-and time-dependent manner', *Microorganisms*, 9(2), pp. 1–9. doi: 10.3390/microorganisms9020255.
Jain, P. *et al.* (2016) 'Mycobacterium tuberculosis Persistent Cells in Human Sputum', *mBio*, 7(5), pp. 1–13. doi: 10.1128/mBio.01023-16.Editor.
Jajou, R. *et al.* (2018) 'Correction: Epidemiological links between tuberculosis cases identified twice as efficiently by whole genome sequencing than conventional molecular typing: A population-based study(PLoS ONE (2018)13:4 (e0195413) DOI: 10.1371/journal.pone.0195413)', *PLoS ONE*, 13(5), pp. 1–11. doi: 10.1371/journal.pone.0197556.
Jeon, D. (2020) 'Latent tuberculosis infection: Recent progress and challenges in South Korea', *Korean Journal of Internal Medicine*, 35(2), pp. 269–275. doi: 10.3904/kjim.2020.029.
Jõers, A., Kaldalu, N. and Tenson, T. (2010) 'The frequency of persisters in Escherichia coli reflects the kinetics of awakening from dormancy', *Journal of Bacteriology*, 192(13), pp. 3379–3384. doi: 10.1128/JB.00056-10.
Jõers, A. and Tenson, T. (2016) 'Growth resumption from stationary phase reveals memory in Escherichia coli cultures', *Scientific Reports*. Nature Publishing Group, 6(March), pp. 1–11. doi: 10.1038/srep24055.
Johnson, P. J. T. and Levin, B. R. (2013) 'Pharmacodynamics, Population Dynamics, and the Evolution of Persistence in Staphylococcus aureus', *PLoS Genetics*, 9(1). doi: 10.1371/journal.pgen.1003123.
Jones, R. *et al.* (2019) 'Phylogenetic Analysis of Mycobacterium tuberculosis Strains in Wales by Use of Core Genome Multilocus Sequence Typing To Analyze Whole-Genome Sequencing Data', *Journal of Clinical Microbiology*, 57(6), pp. 1–11.
Kamerbeek, J. *et al.* (1997) 'Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology', *Journal of Clinical Microbiology*, 35(4), pp. 907–914. doi: 10.1128/jcm.35.4.907-914.1997.
Kapopoulou, A., Lew, J. M. and Cole, S. T. (2011) 'The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes', *Tuberculosis*. Elsevier Ltd, 91(1), pp. 8–13. doi: 10.1016/j.tube.2010.09.006.
De Keijzer, J. *et al.* (2016) 'Mechanisms of Phenotypic Rifampicin Tolerance in Mycobacterium tuberculosis Beijing Genotype Strain B0/W148 Revealed by Proteomics', *Journal of Proteome Research*, 15(4), pp. 1194–1204. doi: 10.1021/acs.jproteome.5b01073.
Keren, I. *et al.* (2011) 'Characterization and transcriptome analysis of mycobacterium tuberculosis persisters', *mBio*, 2(3), pp. 3–12. doi: 10.1128/mBio.00100-11.
Keren, I., Mulcahy, L. R. and Lewis, K. (2012) *Persister eradication: Lessons from the world*

of natural products. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc. doi: 10.1016/B978-0-12-404634-4.00019-X.
Khan, A. H. *et al.* (2020) 'Effect of smoking on treatment outcome among tuberculosis patients in Malaysia; A multicenter study', *BMC Public Health*. BMC Public Health, 20(1), pp. 1–8. doi: 10.1186/s12889-020-08856-6.
Khare, A. and Tavazoie, S. (2020) 'Extreme Antibiotic Persistence via Heterogeneity-Generating Mutations Targeting Translation', *mSystems*, 5(1), pp. 1–12. doi: 10.1128/msystems.00847-19.
Kiazyk, S. and Ball, T. (2017) 'Latent tuberculosis infection: An overview', *Canada Communicable Disease Report*, 43(3/4), pp. 62–66. doi: 10.14745/ccdr.v43i34a01.
Kim, T. H. and Kubica, G. P. (1972) 'Long-term preservation and storage of mycobacteria.', *Applied microbiology*, 24(3), pp. 311–317. doi: 10.1128/aem.24.3.311-317.1972.
Kim, Y., Choi, E. and Hwang, J. (2016) 'Functional studies of five toxin-antitoxin modules in Mycobacterium tuberculosis H37Rv', *Frontiers in Microbiology*, 7(DEC), pp. 1–14. doi: 10.3389/fmicb.2016.02071.
Kristofich, J. C. *et al.* (2018) 'Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme', *PLoS Genetics*, 14(8), pp. 1–25. doi: 10.1371/journal.pgen.1007615.
Kudhair, B. K. *et al.* (2017) 'Structure of a Wbl protein and implications for NO sensing by M. tuberculosis', *Nature Communications*. Springer US, 8(1), pp. 1–12. doi: 10.1038/s41467-017-02418-y.
Kwan, B. W. *et al.* (2013) 'Arrested protein synthesis increases persister-like cell formation', *Antimicrobial Agents and Chemotherapy*, 57(3), pp. 1468–1473. doi: 10.1128/AAC.02135-12.
Lawley, T. D. *et al.* (2008) 'Host transmission of Salmonella enterica serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota', *Infection and Immunity*, 76(1), pp. 403–416. doi: 10.1128/IAI.01189-07.
Lerner, T. R. *et al.* (2017) 'Jcb_201603040 583..594', 216(3). Available at: https://doi.org/10.1083/jcb.201603040.
Letunic, I. and Bork, P. (2019) 'Interactive Tree of Life (iTOL) v4: Recent updates and new developments', *Nucleic Acids Research*. Oxford University Press, 47(W1), pp. 256–259. doi: 10.1093/nar/gkz239.
Leung, V. and Lévesque, C. M. (2012) 'A stress-inducible quorum-sensing peptide mediates the formation of persister cells with noninherited multidrug tolerance', *Journal of Bacteriology*, 194(9), pp. 2265–2274. doi: 10.1128/JB.06707-11.
Levin-Reisman, I. and Balaban, N. Q. (2016) 'Quantitative Measurements of Type I and Type II Persisters Using ScanLag.', *Methods in molecular biology (Clifton, N.J.)*. United States, 1333, pp. 75–81. doi: 10.1007/978-1-4939-2854-5_7.
Levison, Matthew; Levison, J. (2013) 'Pharmacokinetics and Pharmacodynamics of Antibacterial Agents', *Early Human Development*, 83(1), pp. 1–11. doi: 10.1016/j.earlhumdev.2006.05.022.
Levitte, S. *et al.* (2016) 'Mycobacterial Acid Tolerance Enables Phagolysosomal Survival and Establishment of Tuberculous Infection In Vivo', *Cell Host and Microbe*. The Authors, 20(2), pp. 250–258. doi: 10.1016/j.chom.2016.07.007.
Lewis, K. (2010) 'Persister cells', *Annual Review of Microbiology*, 64, pp. 357–372. doi: 10.1146/annurev.micro.112408.134306.
Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.
Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi: 10.1093/bioinformatics/btp698.
Li, Y., Petrofsky, M. and Bermudez, L. E. (2002) 'Mycobacterium tuberculosis Uptake by Recipient Host Macrophages Is Influenced by Environmental Conditions in the Granuloma of the Infectious Individual and Is Associated with Impaired Production of Interleukin-12 and Tumor Necrosis Factor Alpha', *Infection and Immunity*, 70(11), pp. 6223–6230. doi: 10.1128/IAI.70.11.6223.
Liu, J. *et al.* (2020) 'Effect of tolerance on the evolution of antibiotic resistance under drug combinations', *Science*, 367(6474), pp. 200–204. doi: 10.1126/science.aay3041.
Liu, Y. *et al.* (2016) 'Immune activation of the host cell induces drug tolerance in Mycobacterium tuberculosis both in vitro and in vivo', *Journal of Experimental Medicine*, 213(5), pp. 809–825. doi: 10.1084/jem.20151248.
Liu, Y. *et al.* (2019) 'NapM enhances the survival of Mycobacterium tuberculosis under stress and in macrophages', *Communications Biology*. Springer US, 2(1), pp. 1–9. doi: 10.1038/s42003-019-0314-9.
Lou, C., Li, Z. and Ouyang, Q. (2008) 'A molecular model for persister in E. coli', *Journal of Theoretical Biology*, 255(2), pp. 205–209. doi: https://doi.org/10.1016/j.jtbi.2008.07.035.
Madhvi, A. *et al.* (2019) 'Comparison of human monocyte derived macrophages and THP1-like macrophages as in vitro models for M. tuberculosis infection', *Comparative Immunology, Microbiology and Infectious Diseases*. Elsevier, 67(July), p. 101355. doi: 10.1016/j.cimid.2019.101355.
Mahomed, H. *et al.* (2011) 'Predictive factors for latent tuberculosis infection among adolescents in a high-burden area in South Africa', *International Journal of Tuberculosis and Lung Disease*. France, 15(3), pp. 331–336.
Malherbe, S. T. *et al.* (2016) 'Persisting positron emission tomography lesion activity and Mycobacterium tuberculosis mRNA after tuberculosis cure', *Nature Medicine*, 22(10), pp. 1094–1100. doi: 10.1038/nm.4177.
Manganelli, R. *et al.* (2001) 'The Mycobacterium tuberculosis ECF sigma factor σE: Role in global gene expression and survival in macrophages', *Molecular Microbiology*, 41(2), pp. 423–437. doi: 10.1046/j.1365-2958.2001.02525.x.
Manganelli, R. (2014) 'Sigma Factors: Key Molecules in Mycobacterium tuberculosis Physiology and Virulence', *Microbiology Spectrum*, 2(1), pp. 1–23. doi: 10.1128/microbiolspec.mgm2-0007-2013.
Manina, G., Dhar, N. and McKinney, J. D. (2015) 'Stress and host immunity amplify mycobacterium tuberculosis phenotypic heterogeneity and induce nongrowing metabolically active forms', *Cell Host and Microbe*. Elsevier Inc., 17(1), pp. 32–46. doi: 10.1016/j.chom.2014.11.016.
de Martino, M. *et al.* (2019) 'Immune Response to Mycobacterium tuberculosis: A Narrative Review', *Frontiers in Pediatrics*, 7(August), pp. 1–8. doi: 10.3389/fped.2019.00350.
McCune, R. M. *et al.* (1966) 'Microbial persistence. I. The capacity of tubercle bacilli to survive sterilization in mouse tissues.', *The Journal of experimental medicine*, 123(3), pp. 445–468. doi: 10.1084/jem.123.3.445.
Mckenna, A. *et al.* (2009) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research*, pp. 1297–1303. doi: 10.1101/gr.107524.110.20.
Mendoza-Coronel, E. and Castañón-Arreola, M. (2016) 'Comparative evaluation of in vitro human macrophage models for mycobacterial infection study', *Pathogens and Disease*, 74(6), pp. 1–7. doi: 10.1093/femspd/ftw052.
Meylan, S., Andrews, I. W. and Collins, J. J. (2018) 'Targeting Antibiotic Tolerance, Pathogen by Pathogen', *Cell*, 172(6), pp. 1228–1238. doi:

https://doi.org/10.1016/j.cell.2018.01.037.
Michiels, J. E. *et al.* (2016) 'Molecular mechanisms and clinical implications of bacterial persistence', *Drug Resistance Updates*. Elsevier Ltd, 29, pp. 76–89. doi: 10.1016/j.drup.2016.10.002.
Millet, J. P. *et al.* (2013) 'Factors that influence current tuberculosis epidemiology', *European Spine Journal*, 22(SUPPL.4). doi: 10.1007/s00586-012-2334-8.
Miryala, S. K., Anbarasu, A. and Ramaiah, S. (2019) 'Impact of bedaquiline and capreomycin on the gene expression patterns of multidrug-resistant Mycobacterium tuberculosis H37Rv strain and understanding the molecular mechanism of antibiotic resistance', *Journal of Cellular Biochemistry*, 120(9), pp. 14499–14509. doi: 10.1002/jcb.28711.
Mohiuddin, S. G., Kavousi, P. and Orman, M. A. (2020) 'Flow-cytometry analysis reveals persister resuscitation characteristics', *BMC Microbiology*. BMC Microbiology, 20(1), pp. 1–13. doi: 10.1186/s12866-020-01888-3.
Möker, N., Dean, C. R. and Tao, J. (2010) 'Pseudomonas aeruginosa increases formation of multidrug-tolerant persister cells in response to quorum-sensing signaling molecules', *Journal of Bacteriology*, 192(7), pp. 1946–1955. doi: 10.1128/JB.01231-09.
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*. England, 2(9), pp. 747–765. doi: 10.1038/nrmicro955.
Moreau-Marquis, S., Stanton, B. A. and O'Toole, G. A. (2008) 'Pseudomonas aeruginosa biofilm formation in the cystic fibrosis airway', *Pulmonary Pharmacology and Therapeutics*, 21(4), pp. 595–599. doi: 10.1016/j.pupt.2007.12.001.
Morris, R. P. *et al.* (2005) 'Ancestral antibiotic resistance in Mycobacterium tuberculosis', *Proceedings of the National Academy of Sciences of the United States of America*, 102(34), pp. 12200–12205. doi: 10.1073/pnas.0505446102.
Mouton, J. M. *et al.* (2016) 'Elucidating population-wide mycobacterial replication dynamics at the single-cell level', *Microbiology (United Kingdom)*, 162(6), pp. 966–978. doi: 10.1099/mic.0.000288.
Mouton, J. M. *et al.* (2019) 'Comprehensive characterization of the attenuated double auxotroph mycobacterium tuberculosisΔleudΔpanCD as an alternative to h37Rv', *Frontiers in Microbiology*, 10(AUG), pp. 1–13. doi: 10.3389/fmicb.2019.01922.
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. doi: 10.1128/jb.155.2.768-775.1983.
Mukamolova, G. V. *et al.* (2002) 'A family of autocrine growth factors in Mycobacterium tuberculosis', *Molecular Microbiology*, 46(3), pp. 623–635. doi: 10.1046/j.1365-2958.2002.03184.x.
Mulcahy, L. R. *et al.* (2010) 'Emergence of Pseudomonas aeruginosa strains producing high levels of persister cells in patients with cystic fibrosis', *Journal of Bacteriology*, 192(23), pp. 6191–6199. doi: 10.1128/JB.01651-09.
Müller, A. U., Imkamp, F. and Weber-Ban, E. (2018) 'The Mycobacterial LexA/RecA-Independent DNA Damage Response Is Controlled by PafBC and the Pup-Proteasome System', *Cell Reports*, 23(12), pp. 3551–3564. doi: 10.1016/j.celrep.2018.05.073.
Ncayiyana, J. R. *et al.* (2016) 'Prevalence of latent tuberculosis infection and predictive factors in an urban informal settlement in Johannesburg, South Africa: A cross-sectional study', *BMC Infectious Diseases*. BMC Infectious Diseases, 16(1), pp. 1–10. doi: 10.1186/s12879-016-1989-x.
Nguyen, L. T. *et al.* (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Molecular Biology and Evolution*, 32(1), pp.

268–274. doi: 10.1093/molbev/msu300.
Nguyen, T. K. *et al.* (2020) 'The Persister Character of Clinical Isolates of Staphylococcus aureus Contributes to Faster Evolution to Resistance and Higher Survival in THP-1 Monocytes: A Study With Moxifloxacin', *Frontiers in Microbiology*, 11(November), pp. 1–12. doi: 10.3389/fmicb.2020.587364.
Nicol, M. P. *et al.* (2005) 'Distribution of strain families of Mycobacterium tuberculosis causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa', *Journal of Clinical Microbiology*, 43(11), pp. 5779–5781. doi: 10.1128/JCM.43.11.5779-5781.2005.
Nimmo, C. *et al.* (2019) 'Whole genome sequencing Mycobacterium tuberculosis directly from sputum identifies more genetic diversity than sequencing from culture', *BMC Genomics*. BMC Genomics, pp. 1–9. doi: 10.1101/446849.
North, R. J. and Jung, Y. J. (2004) 'Immunity to tuberculosis', *Annual Review of Immunology*, 22, pp. 599–623. doi: 10.1146/annurev.immunol.22.012703.104635.
Okonechnikov, K., Conesa, A. and García-Alcalde, F. (2016) 'Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data', *Bioinformatics*, 32(2), pp. 292–294. doi: 10.1093/bioinformatics/btv566.
Orman, M. A. and Brynildsen, M. P. (2013) 'Dormancy is not necessary or sufficient for bacterial persistence', *Antimicrobial Agents and Chemotherapy*, 57(7), pp. 3230–3239. doi: 10.1128/AAC.00243-13.
Papavinasasundaram, K. G. *et al.* (2005) 'Deletion of the Mycobacterium tuberculosis pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice', *Journal of Bacteriology*, 187(16), pp. 5751–5760. doi: 10.1128/JB.187.16.5751-5760.2005.
Pasipanodya, J. G. *et al.* (2007) 'Pulmonary impairment after tuberculosis', *Chest*. The American College of Chest Physicians, 131(6), pp. 1817–1824. doi: 10.1378/chest.06-2949.
Pasipanodya, J. G. *et al.* (2010) 'Pulmonary impairment after tuberculosis and its contribution to TB burden', *BMC Public Health*, 10. doi: 10.1186/1471-2458-10-259.
Patel, R. K. and Jain, M. (2012) 'NGS QC toolkit: A toolkit for quality control of next generation sequencing data', *PLoS ONE*, 7(2). doi: 10.1371/journal.pone.0030619.
Pethe, K. *et al.* (2010) 'A chemical genetic screen in Mycobacterium tuberculosis identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy', *Nature Communications*. Nature Publishing Group, 1(1), pp. 1–8. doi: 10.1038/ncomms1060.
Peyrusson, F. *et al.* (2020) 'Intracellular Staphylococcus aureus persisters upon antibiotic exposure', *Nature Communications*. Springer US, 11(1). doi: 10.1038/s41467-020-15966-7.
Pham, T. H. M. *et al.* (2021) 'HHS Public Access', 27(1), pp. 54–67. doi: 10.1016/j.chom.2019.11.011.Salmonella.
Phelan, J. *et al.* (2016) 'Mycobacterium tuberculosis whole genome sequencing and protein structure modelling provides insights into anti-tuberculosis drug resistance', *BMC Medicine*. BMC Medicine, 14(1), pp. 1–13. doi: 10.1186/s12916-016-0575-9.
Phelan, J. E. *et al.* (2019) 'Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs', *Genome Medicine*. Genome Medicine, 11(1), pp. 1–7. doi: 10.1186/s13073-019-0650-x.
Portaels, F. *et al.* (1988) 'Effects of freezing and thawing on the viability and the ultrastructure of in vivo grown mycobacteria', *International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association*, 56(4), p. 580—587. Available at: http://europepmc.org/abstract/MED/3065421.
Potrykus, K. and Cashel, M. (2008) '(p)ppGpp: Still magical?', *Annual Review of Microbiology*, 62, pp. 35–51. doi: 10.1146/annurev.micro.62.081307.162903.
Prusa, J., Zhu, D. X. and Stallings, C. L. (2018) 'The stringent response and Mycobacterium tuberculosis pathogenesis', *Pathogens and Disease*, 76(5), pp. 1–13. doi:

10.1093/femspd/fty054.
Que, Y. A. *et al.* (2013) 'A quorum sensing small volatile molecule promotes antibiotic tolerance in bacteria', *PLoS ONE*, 8(12), pp. 1–9. doi: 10.1371/journal.pone.0080140.
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). doi: 10.1371/journal.pgen.1000767.
Ramos, L. *et al.* (2017) 'The minipig as an animal model to study Mycobacterium tuberculosis infection and natural transmission', *Tuberculosis*. Elsevier Ltd, 106, pp. 91–98. doi: 10.1016/j.tube.2017.07.003.
Ramsey, B. *et al.* (1999) 'INTERMITTENT ADMINISTRATION OF INHALED TOBRAMYCIN IN PATIENTS WITH CYSTIC FIBROSIS', *The New England Journal of Medicine*, pp. 23–30.
Reiling, N. *et al.* (2013) 'Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice', *mBio*, 4(4), pp. 1–10. doi: 10.1128/mBio.00250-13.
Ren, H. *et al.* (2015) 'Gradual increase in antibiotic concentration affects persistence of Klebsiella pneumoniae', *The Journal of antimicrobial chemotherapy*, 70(12), pp. 3267–3272. doi: 10.1093/jac/dkv251.
Rengarajan, J., Bloom, B. R. and Rubin, E. J. (2005) 'Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages', *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), pp. 8327–8332. doi: 10.1073/pnas.0503272102.
Rodrigue, S. *et al.* (2006) 'The σ factors of Mycobacterium tuberculosis', *FEMS Microbiology Reviews*, 30(6), pp. 926–941. doi: 10.1111/j.1574-6976.2006.00040.x.
Roostalu, J. *et al.* (2008) 'Cell division in Escherichia coli cultures monitored at single cell resolution', *BMC Microbiology*, 8(February). doi: 10.1186/1471-2180-8-68.
Rutaihwa, L. K. *et al.* (2019) 'Multiple introductions of Mycobacterium tuberculosis Lineage 2-Beijing into Africa over centuries', *Frontiers in Ecology and Evolution*, 7(MAR). doi: 10.3389/fevo.2019.00112.
Rybniker, J. *et al.* (2010) 'Insights into the function of the WhiB-like protein of mycobacteriophage TM4 - A transcriptional inhibitor of WhiB2', *Molecular Microbiology*, 77(3), pp. 642–657. doi: 10.1111/j.1365-2958.2010.07235.x.
Safi, H. *et al.* (2019) 'Phase variation in Mycobacterium tuberculosis glpK produces transiently heritable drug tolerance', *Proceedings of the National Academy of Sciences of the United States of America*, 116(39), pp. 19665–19674. doi: 10.1073/pnas.1907631116.
Saini, V. *et al.* (2016) 'Ergothioneine Maintains Redox and Bioenergetic Homeostasis Essential for Drug Susceptibility and Virulence of Mycobacterium tuberculosis', *Cell Reports*. The Authors, 14(3), pp. 572–585. doi: 10.1016/j.celrep.2015.12.056.
Sajid, A. *et al.* (2011) 'Phosphorylation of Mycobacterium tuberculosis Ser/Thr phosphatase by PknA and PknB', *PLoS ONE*, 6(3). doi: 10.1371/journal.pone.0017871.
Sala, A., Bordes, P. and Genevaux, P. (2014) 'Multiple toxin-antitoxin systems in Mycobacterium tuberculosis', *Toxins*, 6(3), pp. 1002–1020. doi: 10.3390/toxins6031002.
Sampson, S. L. *et al.* (2004) 'Protection Elicited by a Double Leucine and Pantothenate Auxotroph of Mycobacterium tuberculosis in Guinea Pigs', 72(5), pp. 3031–3037. doi: 10.1128/IAI.72.5.3031.
Sampson, S. L. *et al.* (2011) 'Extended safety and efficacy studies of a live attenuated double leucine and pantothenate auxotroph of Mycobacterium tuberculosis as a vaccine candidate', *Vaccine*. Elsevier Ltd, 29(29–30), pp. 4839–4847. doi: 10.1016/j.vaccine.2011.04.066.
Sarathy, J. P. P. *et al.* (2018) 'Extreme drug tolerance of mycobacterium tuberculosis in Caseum', *Antimicrobial Agents and Chemotherapy*, 62(2), pp. 1–11. doi:

10.1128/AAC.02266-17.
Sarathy, J. P. P. and Dartois, V. (2020) 'Caseum: A niche for mycobacterium tuberculosis drug-tolerant persisters', *Clinical Microbiology Reviews*, 33(3), pp. 1–19. doi: 10.1128/CMR.00159-19.
Schnappinger, D. *et al.* (2003) 'Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment', *Journal of Experimental Medicine*, 198(5), pp. 693–704. doi: 10.1084/jem.20030846.
Seeliger, J. C. *et al.* (2012) 'A Riboswitch-Based inducible gene expression system for mycobacteria', *PLoS ONE*, 7(1), pp. 3–8. doi: 10.1371/journal.pone.0029266.
Sendi, P. and Proctor, R. A. A. (2009) 'Staphylococcus aureus as an intracellular pathogen: the role of small colony variants', *Trends in Microbiology*, 17(2), pp. 54–58. doi: 10.1016/j.tim.2008.11.004.
Seung, K. J., Keshavjee, S. and Rich, M. L. (2015) 'Drug-Resistant Tuberculosis', pp. 1–20.
Simeone, R. *et al.* (2012) 'Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death', *PLoS Pathogens*, 8(2). doi: 10.1371/journal.ppat.1002507.
Şimşek, E. and Kim, M. (2019) 'Power-law tail in lag time distribution underlies bacterial persistence', *Proceedings of the National Academy of Sciences of the United States of America*, 116(36), pp. 17635–17640. doi: 10.1073/pnas.1903836116.
Singh, A. *et al.* (2020) 'Drug-resistant tuberculosis and hiv infection: Current perspectives', *HIV/AIDS - Research and Palliative Care*, 12, pp. 9–31. doi: 10.2147/HIV.S193059.
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. doi: 10.1128/JB.01285-09.
Slayden, R. A., Dawson, C. C. and Cummings, J. E. (2018) 'Toxin-antitoxin systems and regulatory mechanisms in Mycobacterium tuberculosis', *Pathogens and Disease*, 76(4), pp. 1–12. doi: 10.1093/femspd/fty039.
Smith, L. J. *et al.* (2010) 'Mycobacterium tuberculosis WhiB1 is an essential DNA-binding protein with a nitric oxide-sensitive iron-sulfur cluster', *Biochemical Journal*, 432(3), pp. 417–427. doi: 10.1042/BJ20101440.
Spalding, C. *et al.* (2018) 'Mathematical modelling of the antibiotic-induced morphological transition of Pseudomonas aeruginosa', *PLoS Computational Biology*, 14(2), pp. 1–28. doi: 10.1371/journal.pcbi.1006012.
Spoering, A. L. L. and Lewis, K. (2001) 'Biofilms and planktonic cells of Pseudomonas aeruginosa have similar resistance to killing by antimicrobials', *Journal of Bacteriology*, 183(23), pp. 6746–6751. doi: 10.1128/JB.183.23.6746-6751.2001.
Stapels, D. A. C. *et al.* (2018) 'Salmonella persisters undermine host immune defenses during antibiotic treatment', *Science*, 362(6419), pp. 1156–1160. doi: 10.1126/science.aat7148.
Starr, T. *et al.* (2018) 'The phorbol 12-myristate-13-acetate differentiation protocol is critical to the interaction of THP-1 macrophages with Salmonella Typhimurium', *PLoS ONE*, 13(3), pp. 1–13. doi: 10.1371/journal.pone.0193601.
Stewart, G. R., Robertson, B. D. and Young, D. B. (2003) 'Tuberculosis: A problem with persistence', *Nature Reviews Microbiology*, 1(2), pp. 97–105. doi: 10.1038/nrmicro749.
Steyn, A. J. C. *et al.* (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. doi: 10.1073/pnas.052705399.
Storz, G., Vogel, J. and Wassarman, K. M. (2011) 'Regulation by Small RNAs in Bacteria: Expanding Frontiers', *Molecular Cell*, 43(6), pp. 880–891. doi: 10.1016/j.molcel.2011.08.022.
Tandon, H. *et al.* (2019) 'Mycobacterium tuberculosis Rv0366c-Rv0367c encodes a non-

canonical PezAT-like toxin-antitoxin pair', *Scientific Reports*, 9(1), pp. 1–14. doi: 10.1038/s41598-018-37473-y.
Thiriot, J. D. *et al.* (2020) 'Hacking the host: exploitation of macrophage polarization by intracellular bacterial pathogens', *Pathogens and disease*. Oxford University Press, 78(1), pp. 1–14. doi: 10.1093/femspd/ftaa009.
Tientcheu, L. D. *et al.* (2017) 'Immunological consequences of strain variation within the Mycobacterium tuberculosis complex', *European Journal of Immunology*, pp. 432–445. doi: 10.1002/eji.201646562.
Tiwari, P. *et al.* (2015) 'MazF ribonucleases promote Mycobacterium tuberculosis drug tolerance and virulence in guinea pigs', *Nature Communications*. Nature Publishing Group, 6. doi: 10.1038/ncomms7059.
Torrey, H. L. *et al.* (2016) 'High persister mutants in mycobacterium tuberculosis', *PLoS ONE*, 11(5), pp. 1–28. doi: 10.1371/journal.pone.0155127.
Tram, T. T. B. *et al.* (2018) 'Virulence of Mycobacterium tuberculosis Clinical Isolates Is Associated With Sputum Pre-treatment Bacterial Load, Lineage, Survival in Macrophages, and Cytokine Response', *Frontiers in cellular and infection microbiology*, 8(November), p. 417. doi: 10.3389/fcimb.2018.00417.
Treangen, T. J. and Salzberg, S. L. (2012) 'Repetitive DNA and next-generation sequencing: Computational challenges and solutions', *Nature Reviews Genetics*, 13(1), pp. 36–46. doi: 10.1038/nrg3117.
Tudó, G. *et al.* (2010) 'Examining the basis of isoniazid tolerance in nonreplicating Mycobacterium tuberculosis using transcriptional profiling', *Future Medicinal Chemistry*, 2(8), pp. 1371–1383. doi: 10.4155/fmc.10.219.
VanderVen, Brian C. Huang, Lu; Rohde, K; Russell, D. (2016) 'The minimal unit of infection: M. tuberculosis in the macrophage', *Physiology & behavior*, 176(1), pp. 100–106. doi: 10.1128/microbiolspec.TBTB2-0025-2016.The.
Vega, N. *et al.* (2012) 'Signaling-Mediated Bacterial Persister Formation Nicole', *Natural Chemistry Biology*, 176(1), pp. 100–106. doi: 10.1038/nchembio.915.Signaling-Mediated.
Vega, N. M. *et al.* (2013) 'Salmonella typhimurium intercepts Escherichia coli signaling to enhance antibiotic tolerance', *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), pp. 14420–14425. doi: 10.1073/pnas.1308085110.
Vergne, I. *et al.* (2004) 'Cell biology of Mycobacterium tuberculosis phagosome', *Annual Review of Cell and Developmental Biology*, 20, pp. 367–394. doi: 10.1146/annurev.cellbio.20.010403.114015.
Verstraeten, N. *et al.* (2016) 'A Historical Perspective on Bacterial Persistence BT - Bacterial Persistence: Methods and Protocols', in Michiels, J. and Fauvart, M. (eds). New York, NY: Springer New York, pp. 3–13. doi: 10.1007/978-1-4939-2854-5_1.
Vilchèze, C. *et al.* (2013) 'Mycobacterium tuberculosis is extraordinarily sensitive to killing by a vitamin C-induced Fenton reaction', *Nature Communications*, 4(May). doi: 10.1038/ncomms2898.
Vogwill, T. *et al.* (2016) 'Persistence and resistance as complementary bacterial adaptations to antibiotics', *Journal of evolutionary biology*, 29(6), pp. 1223–1233. doi: 10.1111/jeb.12864.
Vulin, C. *et al.* (2018) 'Prolonged bacterial lag time results in small colony variants that represent a sub-population of persisters', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-06527-0.
Walker, T. M. *et al.* (2013) 'Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study', *The Lancet Infectious Diseases*. Elsevier Ltd, 13(2), pp. 137–146. doi: 10.1016/S1473-3099(12)70277-3.
Wallis, R. S. S. *et al.* (1999) 'Drug tolerance in Mycobacterium tuberculosis', *Antimicrobial*

Agents and Chemotherapy, 43(11), pp. 2600–2606. doi: 10.1128/aac.43.11.2600.
Walsh, I. M. *et al.* (2020) 'Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness', *Proceedings of the National Academy of Sciences of the United States of America*, 117(7), pp. 3528–3534. doi: 10.1073/pnas.1907126117.
Walter, N. D. *et al.* (2015) 'Transcriptional Adaptation of Drug-tolerant Mycobacterium tuberculosis During Treatment of Human Tuberculosis', *Journal of Infectious Diseases*, 212(6), pp. 990–998. doi: 10.1093/infdis/jiv149.
Wayne, L. G. and Hayes, L. G. (1996) 'An in vitro model for sequential study of shiftdown of Mycobacterium tuberculosis through two stages of nonreplicating persistence', *Infection and Immunity*, 64(6), pp. 2062–2069. doi: 10.1128/iai.64.6.2062-2069.1996.
Weiss, L. A. and Stallings, C. L. (2013) 'Essential roles for mycobacterium tuberculosis rel beyond the production of (p)ppGpp', *Journal of Bacteriology*, 195(24), pp. 5629–5638. doi: 10.1128/JB.00759-13.
Wejse, C. *et al.* (2008) 'TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course', *Scandinavian Journal of Infectious Diseases*, 40(2), pp. 111–120. doi: 10.1080/00365540701558698.
van der Wel, N. *et al.* (2007) 'M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells', *Cell*, 129(7), pp. 1287–1298. doi: 10.1016/j.cell.2007.05.059.
WHO (2020) *Global Tuberculosis report*. Geneva: World Health Organization.
WHO (2021) *WHO announces updated definitions of extensively drug-resistant tuberculosis*. Available at: https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis (Accessed: 22 April 2021).
Wiegand, I., Hilpert, K. and Hancock, R. E. W. (2008) 'Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances', *Nature Protocols*, 3(2), pp. 163–175. doi: 10.1038/nprot.2007.521.
Windels, E. M. *et al.* (2019) 'Bacterial persistence promotes the evolution of antibiotic resistance by increasing survival and mutation rates', *ISME Journal*. Springer US, 13(5), pp. 1239–1251. doi: 10.1038/s41396-019-0344-9.
Wipperman, M. *et al.* (2018) 'Mycobacterial Mutagenesis and Drug Resistance Are Controlled by Phosphorylation- and Cardiolipin-Mediated Inhibition of the RecA Coprotease', *Molecular Cell*, 72(1), pp. 100–106. doi: 10.1016/j.molcel.2018.07.037.Mycobacterial.
Wu, Y. *et al.* (2012) 'Role of oxidative stress in persister tolerance', *Antimicrobial Agents and Chemotherapy*, 56(9), pp. 4922–4926. doi: 10.1128/AAC.00921-12.
Xu, S. *et al.* (1994) 'Intracellular trafficking in Mycobacterium tuberculosis and Mycobacterium avium-infected macrophages.', *The Journal of Immunology*, 153(6), pp. 2568 LP – 2578. Available at: http://www.jimmunol.org/content/153/6/2568.abstract.
Yam, Y. K. *et al.* (2020) 'Extreme Drug Tolerance of Mycobacterium abscessus "Persisters"', *Frontiers in Microbiology*, 11(March), pp. 1–9. doi: 10.3389/fmicb.2020.00359.
Yu, X. *et al.* (2020) 'Characterization of a toxin-antitoxin system in Mycobacterium tuberculosis suggests neutralization by phosphorylation as the antitoxicity mechanism', *Communications Biology*. Springer US, 3(1), pp. 1–15. doi: 10.1038/s42003-020-0941-1.
Yuan, Y., Crane, D. D. D. and Barry, C. E. E. (1996) 'Stationary phase-associated protein expression in Mycobacterium tuberculosis: Function of the mycobacterial α-crystallin homolog', *Journal of Bacteriology*, 178(15), pp. 4484–4492. doi: 10.1128/jb.178.15.4484-4492.1996.
Zeng, J. *et al.* (2016) 'Mycobacterium tuberculosis Rv1152 is a Novel GntR Family

Transcriptional Regulator Involved in Intrinsic Vancomycin Resistance and is a Potential Vancomycin Adjuvant Target', *Scientific Reports*. Nature Publishing Group, 6(September 2015), pp. 1–12. doi: 10.1038/srep28002.
Zhang, S. *et al.* (2018) 'Small Non-coding RNA RyhB mediates persistence to multiple antibiotics and stresses in uropathogenic Escherichia coli by reducing cellular metabolism', *Frontiers in Microbiology*, 9(FEB), pp. 1–10. doi: 10.3389/fmicb.2018.00136.
Zhang, Y. (2014a) 'Persisters, persistent infections and the Yin-Yang model', *Emerging Microbes and Infections*, 3(October 2013), pp. 1–10. doi: 10.1038/emi.2014.3.
Zhang, Y. (2014b) 'Persisters, persistent infections and the Yin-Yang model', *Emerging Microbes and Infections*, 3. doi: 10.1038/emi.2014.3.
Zhang, Y., Yew, W. W. and Barer, M. R. (2012) 'Targeting persisters for tuberculosis control', *Antimicrobial Agents and Chemotherapy*, 56(5), pp. 2223–2230. doi: 10.1128/AAC.06288-11.

Supplementary material

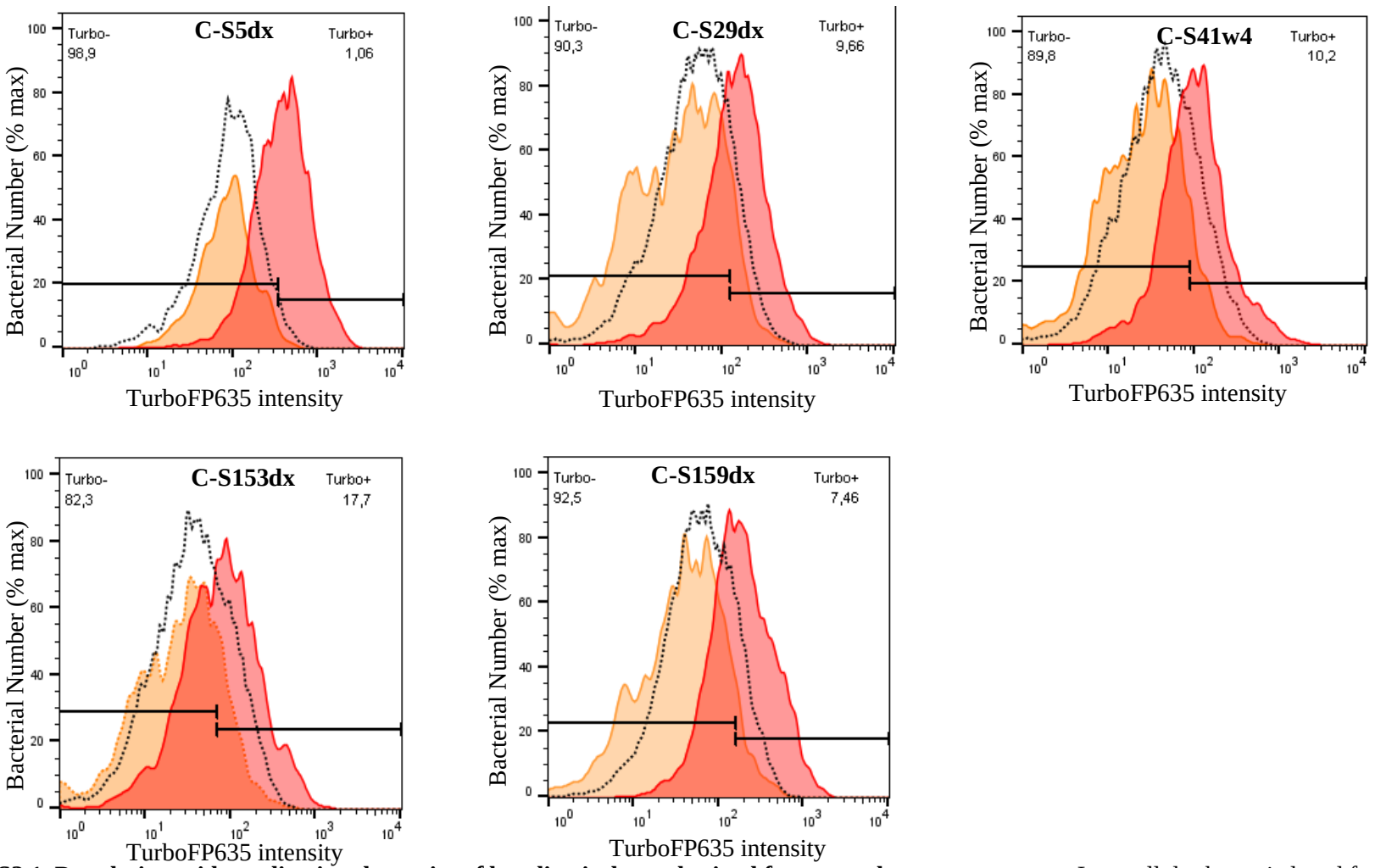


Figure S3.1. Population-wide replication dynamics of baseline isolates obtained from cured treatment group. Intracellular bacteria lysed from macrophages 0h (red), in vitro bacteria 120h (dotted black line), intracellular bacteria lysed from macrophages 120h (orange). Turbo+ indicative of proportion of bacterial population in "high-red persisters". Turbo – is indicative of intracellularly lysed bacteria from macrophages that are actively replication (left of black threshold). Data is representative of 3 technical replicates and 2 biological duplicates.

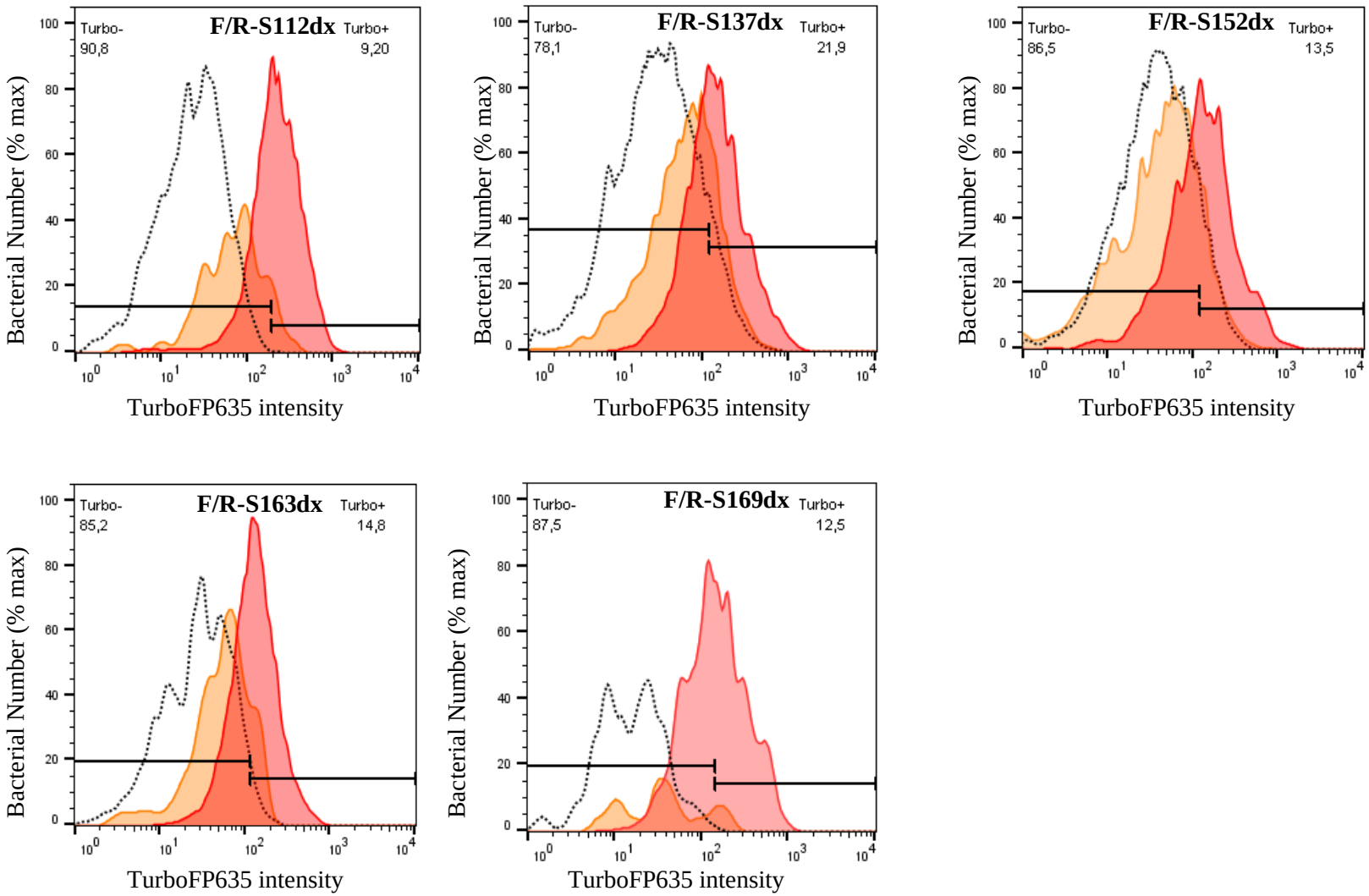


Figure S3.2. Population-wide replication dynamics of baseline isolates obtained from failed treatment group. Intracellular bacteria lysed from macrophages 0h (red), in vitro bacteria 120h (dotted black line), intracellular bacteria lysed from macrophages 120h (orange). Turbo+ indicative of proportion of bacterial population in "high-red persisters". Turbo – is indicative of intracellularly lysed bacteria from macrophages that are actively replication (left of black threshold). Data is representative of 3 technical replicates and 2 biological duplicates.

Appendices

Appendix A: Recipes and Protocols

7H9 liquid media

2.35g of 7H9 powder into 450ml ddH_2O

Autoclave at 121°C for 18 minutes

Aseptically add:

- 50ml OADC
- 1.25ml 20% Tween-80 (filter-sterilized)
- 2ml 50% glycerol (filter-sterilized)

7H10 solid media

19g 7H10 powder into 900ml MilliQ H_2O

Autoclaved at 121°C for 18 minutes.

Cool to approximately 50°C

Aseptically Add:

- 100 ml OADC (BD)
- 10 ml 50 % glycerol
- Required antibiotic

Measure out and pour 25-30ml per plate.

EDTA

MW 292.24

Weigh out 29.22 g for 200 ml of a 0.5M solution

pH as required - eg. to pH 8.0, with NaOH

Leucine
(200x stock is 10 mg/ml in water) Weigh out 2.5 g powder
Dissolve in 250 ml MilliQ H_2O,
Filter-sterilize (use cup filter)
Make 50 ml aliquots, cover with foil and store at 4˚C

Pantothenate
(1000x is 24 mg/ml in water)
Weigh out 2.4 g powder

Dissolve in 100 ml MilliQ H_2O,
Filter-sterilize
Cover with foil, store at -20˚C
Just prior to use, thaw tube, store at 4ºC for up to 1 week.

Theophylline (5x)
Since Theophylline is poorly soluble in aqueous solutions, can only make up at 5x or lower concentration, so make it up in 7H9 complete (or RPMI + 10% FCS for tissue culture) to avoid diluting media when adding to final culture.
Weigh out 90 mg into a 50 ml Falcon tube
Add 50 ml 7H9 complete, mix to dissolve (put on shaker at RT for 1 hour)
Filter-sterilize
Store at 4ºC for up to 1 week

Tris
MW: 121.14
Make up 1 M stock by dissolving
60.57 g in 500 ml
pH as required
CTAB/NaCl solution

Dissolve 4.1 g NaCl in 80 ml distilled water. While stirring, add 10 g CTAB.

Heat solution in 65°C incubator

Adjust volume to 100 ml with distilled water

Lysozyme (10 mg/ml)

Reconstitute lyophilised lysozyme (brought to room temperature) with distilled water to 10 mg/ml

- Add 100mg to 10ml of H2O (or 50mg to 5 ml of H2O)

Proteinase K (10 mg/ml)

Reconstitute vial of lyophilised Proteinase K with distilled water to 10 mg/ml. Freeze aliquots in 2 ml tubes at -20°C

10% SDS (500 ml; pH 7.8)

- 10 g SDS made up to 100 ml with distilled water

Dissolve by heating at 65°C for 20 min

NaCl (5M)

- 29.2 g NaCl made up to 100 ml with distilled water

Autoclave. Store at room temperature for up to 1 year

Chloroform/isoamyl alcohol (24:1)- store in fridge and bring to RT prior to use

- 384 ml Chloroform
- 16 ml Isoamyl alcohol

TE (Tris EDTA; pH 8; 1 litre) – store at RT

- 1.211 g Tris
- 0.372 g EDTA
- Adjust pH with HCl

Appendix B: Commands

Generic commands used in the analysis pipeline

1. Trimmomatic

```
java -jar trimmomatic-0.35.jar PE -phred33 input_forward.fq.gz input_reverse.fq.gz
output_forward_paired.fq.gz output_forward_unpaired.fq.gz output_reverse_paired.fq.gz
output_reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:20
TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36
```

```
bwa index ref.fa
bwa backtrack ref.fa read1.fq read2.fq > aln-pe.sam
bwa aln ref.fa short_read.fq > aln_sa.sam
bwa sampe ref.fa aln_sa1.sai aln_sa2.sai read1.fq read2.fq > aln-pe.sam
```

2. BWA

3. Novoalign (http://www.novocraft.com/documentation/novoalign-2/novoalign-ngs-quick-start-tutorial/basic-short-read-mapping/)

```
Indexing: Novoindex [-k -s] referencegenome.fasta
Alignment: Novoalign -d reference.nix -f input_forward.fq.gz input_reverse.fq.gz -
i200.50 -o SAM > alignment.sam >log.txt
```

4. SMALT

```
Indexing: Smalt index [Index Options] Index Refseq-file.fasta
Alignment: Smalt map [MAP-OPTIONS] Index Read-File. Read1.fq read2.fq_
output_aln.sam
```

5. SAM file validation in PicardTools (https://broadinstitute.github.io/picard/command-line-overview.html#CommandSyntax)

```
Java -jar picard.jar ValidateSamFile \\ input.sam/bam  \\ MODE=SUMMARY -o
output.sam
```

6. SAM to BAM conversion in SAMTools

```
Indexing: samtools faidx referencegenome.fasta
Sort: samtools sort alignedoutput.sam -o sorted.sam
View: samtools view -b sorted.sam -o viewed.bam
```

7. BAM processing in PicardTools

```
java -jar picard.jar CollectAlignmentSummaryMetrics \
        REFERENCE=my_data/reference.fasta \
        INPUT=my_data/input.bam \
        OUTPUT=results/output.txt
```

8. In/del realignment in GATK

(https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-

```
Java -ja GenomeAnalysisTK.jar -T IndelRealigner  -R reference. Fasta -I input.bam -
known indels.vcf  -targetIntervals intervalistFromRTC.intervals -o realignedBam.bam
```

0/org_broadinstitute_gatk_tools_walkers_indels_IndelRealigner.php)

9. Removal of PCR duplicates

```
Java -jar picard.jar MarkDuplicates input .bam -o marked_duplicates .bam -m
marked_dup_metrics.txt
```

10. Variant calling
    a. GATK

```
Java -jar GenomeAnalysisTK.jar -T UnifiedGenotyper -R reference. Fasta -I sample1.bam
[-I sample2.bam….] --dbsnp dbSNP.vcf -o snsps.raw.vcf --stand_call_cof [50.0] [-L
targets.interval_List]
```

    b. SAMTools

```
Mpileup: ./bcftools mpileup -f reference.fasta input.bam | call -vmO v -o variants raw.vcf
```

Generic command for TB-profiler for usage on Khaos server following activation:

```
tb-profiler profile -1 /path/to/reads/isolateID_R1.fastq.gz -2
/path/to/reads/isoalteID_R2.fastq.gz -p isolateID -t 8 –txt
```

The arguments provided here include:

- -1 : the absolute path to the forward reads of the isolate that you want to analyse with TB-profiler.
- -2 : the absolute path to the reverse reads of the isolate that you want to analyse with TB-profiler. If the isolate is sequenced single endedly and only one FATSQ file is available, leave out the -2 argument.
- -p : prefix, this is the name of the output file and should correspond to the isolate ID.
- -t : the number of threads to use on the server and translates to the computational resources that will be assigned to perform the task.
- --txt : include the results in a plain text format, as opposed to the default output in Jason format.

Appendix C: Scripts

1. SCRIPT USED TO TRIM SEQUENCES USING TRIMMOMATIC

```
#!/bin/bash

if [[ $1 == "" || $1 == "help" ]];
then
        echo "Hello, is me your looking for"
        echo "your params is as following"
        echo "1 = samples"
        echo "2 = output dir"
        echo "3 = raw file dir"
        echo "4 = ram (6)"
        echo "5 = cores (6)"
fi

samples="${1}"
output="${2}"
rawdir="${3}"
ram="${4}"
threads="${5}"

#masterdir
master_dir=/home/user/Desktop/final_script/
#myprograms
trim="${master_dir}/programs/Trimmomatic-0.36/trimmomatic-0.36.jar"
trim_PE="${master_dir}/programs/Trimmomatic-0.36/adapters/TruSeq2-PE.fa"

while read sample;
do

        echo "getting raw file names"
        raw_1="${rawdir}/${sample}"
```

```
        echo "Your file is: $raw_1"
        raw_2="${rawdir}/${sample}"
        echo "Your file is: $raw_2"

        echo "Trimming reads, Julian"
        java -Xmx"${ram}"g -jar $trim PE \
        -phred33 \
        -threads "$threads" \
        "$raw_1" "$raw_2" \
        "${output}/${sample}_R1_forward_paired.fq.gz"
"${output}/${sample}_R1_forward_unpaired.fq.gz" \
        "${output}/${sample}_R2_reverse_paired.fq.gz"
"${output}/${sample}_R2_reverse_unpaired.fq.gz" \
        ILLUMINACLIP:"${trim_PE}":2:30:15 LEADING:3 TRAILING:3
SLIDINGWINDOW:4:20 MINLEN:30

done< <(tr -d '\r' < "$samples")
```

2. SCRIPT USED TO ANNOTATE CONFIDENCE OF VARIANTS

```
#!/usr/bin/perl
# annotating SNPs for one strain
# Note that the H37RvAnno.txt and H37RvGeneSeq.fasta files used in this script were
downloaded from the Tuberculosis database (TBDB)
use strict;
my ($vcf)=@ARGV;
my @headers;
open(MUT, "$vcf") or die "Cannot open $vcf:$!\n";
while (<MUT>) {
chomp;
next if (/^##/);
if (/^#/) {
@headers = split(/\t/,$_);
```

```
print join ("\t", ("CHROM", "POS", "LOCUS", "SYMBOL", "REFBASE", "ALTBASE",
@headers[5,6,7,8,9], "CODONnr", "REFCODON", "REFAA", "MUTCODON", "MUTAA",
"CHANGE")), "\n";
next;
}
my ($CHROM, $POS, $ID, $REFBASE, $ALTBASE, $QUAL, $FILTER, $INFO,
$FORMAT, $STRAIN_1)=split(/\t/,$_);
my
$annofile="/home/adippenaar/Documents/Bioinformatics/Out_groups_output/H37RvAnno.tx
t";
my $line=0;
my $prevGene;
my $prevStrand;
my
$geneseqfile="/home/adippenaar/Documents/Bioinformatics/Out_groups_output/H37RvGen
eSeq.fasta";
my $codonsize = 3;
open(ANNO, "$annofile") or die "Cannot open $annofile:$!\n";
while (<ANNO>) {
if ($line==0) {
$line++;
next;
}
```

```
chomp;
my ($LOCUS, $SYMBOL, $SYNOYM, $LENGHT, $START, $STOP, $STRAND, $NAME)=split(/\t/,$_);
#print "'$STRAND'\n";
if ($POS>$START && $POS<$STOP) {
open(GENESEQ, "$geneseqfile") or die "Cannot open $geneseqfile:$!\n";
my $seq;
while (<GENESEQ>) {
next unless /^>$LOCUS/;
while (<GENESEQ>) {
last if /^>/;
chomp;
$seq.=$_;
}
last;
}
close GENESEQ;
my $posingene;
if ($STRAND eq "+") {
$posingene = ($POS - $START) + 1;
}
else {$posingene = ($STOP - $POS) + 1;
}
my $codonnr = int(($posingene - 1)/$codonsize + 1);
#print "'$codonnr'\n";
my $firstbase = (($codonnr - 1) * $codonsize) + 1 - 1;
#print "'$firstbase'\n";
my $lastbase = (($codonnr - 1) * $codonsize) + 3 - 1;
#print "'$lastbase'\n";
# -1: subst will start to count at 0
```

```
my $codon = substr($seq, $firstbase, $codonsize);
#print "'$seq'\n";
#print "'$codon'\n";
my $aa = &codon2aa($codon);
#print "aa is $aa\n";
my $offset;
if (($posingene % $codonsize) == 1) {
$offset = 1 - 1;
}
elsif (($posingene % $codonsize) == 2) {
$offset = 2 - 1;
}
elsif (($posingene % $codonsize) == 0) {
$offset = 3 - 1;
}
#print "offset value is $offset\n";
my $mutcodon = $codon;
my $mutaa;
my $change;
if ($STRAND eq "+") {
substr($mutcodon, $offset, 1) = $ALTBASE;
#print "altbase is $ALTBASE\n";
#print "my mutcodon is $mutcodon\n";
$mutaa = &codon2aa($mutcodon);
if ($mutaa eq $aa) {$change = "SYN";
}
elsif ($mutaa ne $aa) {$change = "NONSYN";
}
}
```

```
else {
if ($ALTBASE eq "T") {$ALTBASE = "B";
}
if ($ALTBASE eq "A") {$ALTBASE = "D";
}
if ($ALTBASE eq "C") {$ALTBASE = "E";
}
if ($ALTBASE eq "G") {$ALTBASE = "H";
}
if ($ALTBASE eq "B") {$ALTBASE = "A";
}
if ($ALTBASE eq "D") {$ALTBASE = "T";
}
if ($ALTBASE eq "E") {$ALTBASE = "G";
}
if ($ALTBASE eq "H") {$ALTBASE = "C";
}
substr($mutcodon, $offset, 1) = $ALTBASE;
#print "altbase is $ALTBASE\n";
#print "my mutcodon is $mutcodon\n";
$mutaa = &codon2aa($mutcodon);
if ($mutaa eq $aa) {$change = "SYN";
}
elsif ($mutaa ne $aa) {$change = "NONSYN";
}
}
```

```
# print join ("\t", ($CHROM, $POS, $LOCUS, $SYMBOL, $REFBASE, $ALTBASE, $QUAL, $FILTER, $INFO, $FORMAT, $STRAIN_1, $codonnr, $codon, $aa, $mutcodon, $NAME)), "\n";
print join ("\t", ($CHROM, $POS, $LOCUS, $SYMBOL, $REFBASE, $ALTBASE, $QUAL, $FILTER, $INFO, $FORMAT, $STRAIN_1, $codonnr, $codon, $aa, $mutcodon, $mutaa, $change, $NAME)), "\n";
last;
}
if ($POS<$START){
$SYMBOL = "-";
my $message="Intergenic";
#if($STRAND eq "+") {$message.=".Upstream of $LOCUS";
#}
#if($prevStrand eq "-") {$message.=".Upstream of $prevGene";
#}
print join ("\t", ($CHROM, $POS, $message, $SYMBOL, $REFBASE, $ALTBASE, $QUAL, $FILTER, $INFO, $FORMAT, $STRAIN_1)), "\n";
last;
}
$prevGene=$LOCUS;
$prevStrand=$STRAND;
}
close ANNO;
}
close MUT;
############################################################
sub codon2aa {
my $codon = uc shift;
if ( $codon =~ m/GC./ ) { return "A" } # Alanine
elsif ( $codon =~ m/TG[TC]/ ) { return "C" } # Cysteine
elsif ( $codon =~ m/GA[TC]/ ) { return "D" } # Aspartic Acid
elsif ( $codon =~ m/GA[AG]/ ) { return "E" } # Glutamic Acid
elsif ( $codon =~ m/TT[TC]/ ) { return "F" } # Phenylalanine
elsif ( $codon =~ m/GG./ ) { return "G" } # Glycine
```

```
elsif ( $codon =~ m/CA[TC]/ ) { return "H" } # Histidine
elsif ( $codon =~ m/AT[TCA]/ ) { return "I" } # Isoleucine
elsif ( $codon =~ m/AA[AG]/ ) { return "K" } # Lysine
elsif ( $codon =~ m/TT[AG]|CT./ ) { return "L" } # Leucine
```


3. SCRIPT USED TO COMPARE ANNOTATIONS

```
import dirCompare32

# EDIT THESE DIRECTORIES
dir1 = "/home/marisat/Dircompare/DirA_ECWCstrains/"
dir2 = "/home/marisat/Dircompare/DirB_KZNstrains/"

#dir1 = "C:/Ruben/WORK/[ALL_RESULTS]/2016/pnca_online/ALL_GATK/"
#dir2 = "C:/Ruben/WORK/[ALL_RESULTS]/2016/pnca_online/ALL_SAMTOOLS/"
outputDir = "/home/marisat/Dircompare/DirComparOut/"

dirCompare32.main(dir1,dir2,outputDir)
```

```
elsif ( $codon =~ m/ATG/ ) { return "M" } # Methionine
elsif ( $codon =~ m/AA[TC]/ ) { return "N" } # Asparagine
elsif ( $codon =~ m/CC./ ) { return "P" } # Proline
elsif ( $codon =~ m/CA[AG]/ ) { return "Q" } # Glutamine
elsif ( $codon =~ m/CG.|AG[AG]/ ) { return "R" } # Arginine
elsif ( $codon =~ m/TC.|AG[TC]/ ) { return "S" } # Serine
elsif ( $codon =~ m/AC./ ) { return "T" } # Threonine
elsif ( $codon =~ m/GT./ ) { return "V" } # Valine
elsif ( $codon =~ m/TGG/ ) { return "W" } # Tryptophan
elsif ( $codon =~ m/TA[TC]/ ) { return "Y" } # Tyrosine
elsif ( $codon =~ m/TA[AG]|TGA/ ) { return "_" } # Stop
else { die "Bad codon \"$codon\"!\n" }
}
```

4. SCRIPT USED TO CREATE SNP STRING FOR PHYLOGENETIC INFERENCE

```
#This python script assumes:
#1. A single chromosome
#2. Input files with the naming convention
# pos_alt_<sample_nr>_<list_identifier>.txt
# Entries in these files contain 2 columns: the position of a variant and its
# value
#3. A fasta file (reference sequence) with 1 header line and a column length of 60

#The output of this script is <sample_nr>.txt files, one file for each of the
#input files. The content of each file is a string of nucleotides, in order
#of position, for the set of all positions read from the input files. If a
#value for a certain position is not available for a sample, it contains the
#value of the the reference allele, as read from the FASTA file.

nr_fasta_header_lines = 1
col_len = 60
input_file_prefix = 'pos_alt_'

import sys, os
```

```
if len(sys.argv) != 4:
        print("Usage: python create_phylo_files.py <fasta_file> <input_dir> " + \
                "<output_dir>")
        sys.exit(-1)
else:
        fasta_file_name = sys.argv[1]
        input_dir = sys.argv[2]
        output_dir = sys.argv[3]


#Map containing a map of positions and variants, keyed on sample nr
sample_map = {}
#Map containing reference variants, keyed on position
ref_map = {}
#Set the column length to 60

input_dir_list = os.listdir(input_dir)
for file_name in input_dir_list:
        if file_name[0:8] == input_file_prefix:
                print('Processing ' + file_name + ' ...')
                #Determine the sample number
                sample_nr_length = file_name[9:].find('_')
                sample_nr = file_name[8:9+sample_nr_length]
                #Initialize the variant map for this sample, keyed on position
                var_map = {}
                #Read the file content and populate the maps
                in_file = open(os.path.join(input_dir, file_name))
                for line in in_file:
                data = line.strip().split()

                pos, var = int(data[0]), data[1]
                var_map[pos] = var
                if (pos in ref_map) == False:
                        #Get the reference allele from the fasta file
                        col_nr = str(pos % col_len)
                        row_nr = str((pos / col_len) + nr_fasta_header_lines + 1)
                        if col_nr == '0':
                        col_nr = str(col_len)
                        os.system('head -' + row_nr + ' ' + fasta_file_name + ' | ' + \
                        'tail -1 > tmp_fasta_line.txt')
```

```
                        ref = os.popen('cut -c' + col_nr + \
                        ' tmp_fasta_line.txt').read().strip().upper()
                        os.system('rm tmp_fasta_line.txt')
                        ref_map[pos] = ref
        in_file.close()
        sample_map[sample_nr] = var_map

#Sort the positions
positions = sorted(ref_map.keys())

for sample_nr in sample_map.keys():
        #Open the output file for this sample
        print('Writing ' + sample_nr + '.txt ....')
        out_file = open(os.path.join(output_dir, sample_nr) + '.txt', 'w')
        debug_file = open(os.path.join(output_dir, sample_nr) + '_debug.txt', 'w')
        debug_file.write('pos\tref\talt\n')

        #Get the variant map for this sample
        var_map = sample_map[sample_nr]

        #Write the output file
        for pos in positions:
                debug_file.write(str(pos) + '\t')
                if pos in var_map:
                        var = var_map[pos]
                        debug_file.write('*\t' + var + '\n')
        else:
                var = ref_map[pos]
                debug_file.write(var + '\t*\n')
        out_file.write(var)
out_file.close()
debug_file.close()
```