Advancing the Understanding of the PE/PPE protein families through PPE_MPTR protein expression and analysis of *pe/ppe* Single Nucleotide Variants in differentially drug sensitive *Mycobacterium tuberculosis*

By

Tom Jack Bartizal

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

Supervisor: Prof. Samantha

Sampson Co-Supervisor: Dr.

Nastassja Kriel

December 2022

https://scholar.sun.ac.za

Declaration

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

Deur hierdie tesis elektronies in te lewer, verklaar ek dat die geheel van die werk hierin vervat, my eie, oorspronklike werk is, dat ek die alleenouteur daarvan is (behalwe in die mate uitdruklik anders aangedui), dat reproduksie en publikasie daarvan deur die Universiteit van Stellenbosch nie derdepartyregte sal skend nie en dat ek dit nie vantevore, of in die geheel of gedeeltelik, ter verkryging van enige kwalifikasie aangebied het nie.

Date: **28 August 2022**

Copyright © 2022 Stellenbosch University
All rights reserved

https://scholar.sun.ac.za

Abstract

Mycobacterium tuberculosis is one of the leading causes of infection and death by a single microorganism. The *M. tuberculosis* genome contains two prominent gene families, encoding proteins known as the PE (proline-glutamate) and PPE (proline-proline-glutamate) families. Several of these are essential for *M. tuberculosis* pathogenesis, however, the majority of the PE/PPE proteins, especially those within the PPE_MPTR subfamily are understudied. This is apparent regarding the sub-cellular localisation of PPE_MPTR proteins which may interact at the host-pathogen interface. Some functional aspects have also been overlooked, such as the potential for *pe/ppe* variants to contribute to drug resistance in *M. tuberculosis*. These genes are often overlooked due to their polymorphic nature, as well as the challenges associated with short-read sequencing technologies to reliably assemble their highly repetitive and GC-rich sequences. For the first part of our study, we aimed to identify the sub-cellular localisation of select PPE_MPTR proteins (PPE_MPTR10, -24, -40, 42, -53 and -62) by expressing them as fluorescently tagged proteins in *Mycobacterium smegmatis*. We successfully amplified all six genes from *M. tuberculosis* H37Rv DNA, and successfully cloned *ppe_mptr10* into the pCG expression vector. Unfortunately, no protein expression was detected and we were unable to determine the localisation of PPE_MPTR10 within *M. smegmatis*.

For the second aim of our study, we made use of a newly developed analytical pipeline to screen whole-genome sequencing (WGS) data and identify high-confidence *pe/ppe* single-nucleotide variants (SNVs) associated with drug resistance profiles. Nine SNVs were predicted to be unique to either drug susceptible, multi- or extensively-drug resistant (DS, MDR or XDR) classes of *M. tuberculosis*. We aimed to verify these findings by PCR and Sanger sequencing of targeted SNVs in an independent set of clinical *M. tuberculosis* isolates. The SNVs were determined to be true variants identified by the pipeline present in clinical isolates but absent from the reference *M. tuberculosis* H37Rv. None of the target variants were found to be unique to the drug resistance class for which they were originally predicted. The successful identification of *pe/ppe* variants by genomic analysis demonstrates the potential to screen these repetitive GC-rich regions for genetic variants. Future studies will be required to establish the role of PPE_MPTR proteins in *M. tuberculosis* pathogenesis.

https://scholar.sun.ac.za

Opsomming

Mycobacterium tuberculosis word beskou as een van die hoofoorsake van infeksie en dood deur 'n enkele mikro-organisme. Die *M. tuberculosis*-genoom bevat koderende streke vir twee prominente geenfamilies, wat kodeer vir proteïene bekend as die PE (prolien-glutamaat) en PPE (prolien-prolien-glutamaat) families. Daar is voorgestel dat verskeie van hierdie proteïene noodsaaklik is vir *M. tuberculosis* patogenese, maar die meerderheid van die PE/PPE proteïene, veral dié binne die PPE_MPTR subfamilie, word onderbestudeer. Dit is duidelik met betrekking tot die sub-sellulêre lokalisering van PPE_MPTR proteïene wat moontlik tot interaksies by die gasheer-patogeen koppelvlak betrokke mag wees. Sommige funksionele aspekte is ook oor die hoof gesien, soos die potensiaal vir *pe/ppe*-variante om by te dra tot ontwikkeling van middelweerstand in *M. tuberculosis*. Hierdie gene word dikwels geïgnoreer as gevolg van hul polimorfiese aard, sowel as die uitdagings met kortleesvolgordebepalingstegnologieë om hul hoogs herhalende en GC-ryke streke te onthul. Vir die eerste doel van ons studie het ons daarop gemik om die sub-sellulêre lokalisering van PPE_MPTR proteïene (PPE_MPTR10, -24, -40, 42, -53 en -62) te identifiseer deur hulle uit te druk as fluoresserende proteïene in *Mycobacterium smegmatis*. Al ses gene van *M. tuberculosis* H37Rv DNA is geamplifiseer en *ppe_mptr10* in die pCG uitdrukkingsvektor gekloneer. Ongelukkig kon geen proteïenuitdrukking bevestig word nie en kon ons nie die lokalisering van PPE_MPTR10 binne *M. smegmatis* bepaal nie.

Vir die tweede doel van ons studie, het ons gebruik gemaak van 'n nuwe analitiese pyplyn om heelgenoom data te gbruik om hoë-vertroue *pe/ppe* enkelnukleotied variante (SNVs) wat met geneesmiddelweerstandsprofiele geassosieer word, te identifiseer. Ons het nege SNVs geïdentifiseer wat uniek was aan óf dwelm vatbare, multi- of ekstensief-middel weerstandbiedende (DS, MDR of XDR) klasse van *M. tuberculosis*. Ons het ten doel gehad om hierdie bevindinge te verifieer deur PCR en Sanger-volgordebepaling van geteikende SNVs in 'n onafhanklike stel kliniese *M. tuberculosis*-isolate. Die SNVs is vasgestel as ware variante wat geïdentifiseer is deur die pyplyn teenwoordig in kliniese isolate maar afwesig van die *M. tuberculosis* H37Rv verwysing. Daar is gevind dat geen van die teikenvariante uniek is aan die geneesmiddelweerstandsklas waarvoor dit oorspronklik voorspel is nie. Die suksesvolle identifikasie van *pe/ppe*-variante deur genomiese analise demonstreer die potensiaal om hierdie herhalende GC-ryke streke vir genetiese variante te keur. Toekomstige studies sal vereis word om die rol van PPE_MPTR proteïene in *M. tuberculosis* patogenese vas te stel.

https://scholar.sun.ac.za

Acknowledgements

I would like to acknowledge the following for all their contributions and support throughout the duration of my MSc degree.

Firstly, I would like to thank Prof. Sampson for being an amazing supervisor. Even though she is incredibly busy and has many responsibilities, the care she shows her group and students is absolutely phenomenal. I must thank her for giving me the opportunity to join her group and complete this Master's degree, for which I will always be grateful. The road to the submission of my thesis was not always easy, but Prof. Sampson was a constant source of guidance and input, always supporting my personal goals for my future development. I see her as inspirational for the work she has done and still continues to do.

Secondly, I must thank my co-supervisor Dr. Nastassja Kriel, as she was an active presence throughout my whole project. I will always be especially grateful for her constant guidance and feedback in the first year of my degree, even at the final stages of her pregnancy. She made a point of trying to impart new skills and make me truly think about some of the work that I've done, why it is done and what the results can mean. Add onto this the care she often showed for my well-being and I don't really believe you could ask for a much better supervisor. She has served as a source of wisdom and I believe and hope I've learned a lot that I can carry with me going forward.

I also want to acknowledge Dr. Bahar Bagheri, on whose work the final chapters of my thesis are based. Her willingness to open her study for me to do the verification of her results is greatly appreciated, as well as her openness in sharing and discussing our results.

I would also like to thank all other members of the HPM group for their constant support either in the work I've done or as a source of morale when it was needed. This also extends to other members of the Division of MBHG who were willing to help and share their expertise for the work I've done during my degree.

Finally, I would like to acknowledge all the funding bodies who supported me as I dedicated my time and life to completing this degree, including the National Research Foundation, Harry Crossley Foundation and the Post-Graduate Office of Stellenbosch University.

https://scholar.sun.ac.za

List of Abbreviations

°C	Degrees Celsius
µg	microGram
µl	microlitre
aa	Amino acid
BCG vaccine	Bacille Calmette-Guérin vaccine
bp	Base-pair
BSA	Bovine serum albumin
CAF	Central Analytical Facility
CLEM	Correlative-light electron microscopy
DMSO	dimethyl sulfoxide
DR	Drug-resistant
DS	Drug susceptible
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
Erp	Exported, repetitive protein
ESAT-6	Early Secretory Antigenic Target 6
Esp	ESX Secretion-Associated Protein
et al.	*et alia* (and others)
FQ	Fluoroquinolone
GFP	Green Fluorescent Protein
HIV	Human Immunodeficiency Virus
HPM	Host-Pathogen Mycobactomics
IGV	Integrative Genomics Viewer
INH	isoniazid
iNOS	Inducible nitric oxide synthase
kb	kilobase
kDa	Kilo-Dalton
l	litre
LB	Luria-Bertani
MBHG	Molecular Biology and Human Genetics
MDR	Multi-drug resistant
ml	millilitre
MPTR	Major Polymorphic Tandem Repeat
MR	Mono-resistant
MTBC	*Mycobacterium tuberculosis* Complex
NF-kB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
ng	nanoGram
NO	Nitric oxide
NRP3	NLR family pyrin domain containing 3
OADC	oleic acid-albumin-dextrose-catalase
OD	Optical density
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PE/*pe*	Proline-Glutamate
PGRS	Polymorphic GC-Rich Repetitive Sequence

https://scholar.sun.ac.za

PPE/*ppe*	Proline-Proline-Glutamate
PPW	Proline-Proline-Tryptophan
PVDF	Polyvinylidene fluoride
RIF	rifampicin
SAMMtb	Severely Attenuated Mutant *Mycobacterium tuberculosis*
Sec	General Secretory
SLID	Second-line injectable drug
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SVP	Serine-Valine-Proline
TAE	Tris Acetate EDTA
TAG	Triacylglycerol
TAT	Twin-arginine translocation
TB	Tuberculosis
TBG	TB-Genomics
TBST	Tris-Buffered Saline – Supplemented with Tween
TF	Trigger Factor
Th2	T-Helper Cell Type 2
TLR	Toll-like Receptor
TNF-α	Tumour Necrosis Factor-α
TNT	Tuberculosis Necrotizing Toxin
WG	Whole-Genome
WGS	Whole-Genome Sequencing
WHO	World Health Organisation
XDR	Extensively-Drug Resistant

https://scholar.sun.ac.za

List of Figures

https://scholar.sun.ac.za

List of Tables

https://scholar.sun.ac.za

Table of Contents

https://scholar.sun.ac.za

https://scholar.sun.ac.za

https://scholar.sun.ac.za

Chapter 1

Thesis Overview

Chapter 1 **Introduction**

This chapter introduces the topic of the research project, discussing *Mycobacterium tuberculosis* and the knowledge gaps surrounding the fundamental molecular components, such as the PE/PPE proteins. The information gap that we aimed to address is discussed in the problem statement. A hypothesis is stated for the target research and the aims and objectives used to confirm the hypothesis is listed.

Chapter 2 **Literature Review: "Secretion and Localisation Signals Utilised by *Mycobacterium tuberculosis*"**

This review chapter provides an overview of current protein secretion literature in mycobacteria. This review will focus on the secretion signals and chaperones that are utilised within mycobacteria to target proteins to specific pathways, as well as current methods that have been utilised to investigate the specificity of protein secretion.

Chapter 3 **Localisation of PPE_MPTR Proteins in *Mycobacterium smegmatis*.**

This chapter describes the selection of PPE_MPTR proteins for sub-cellular localisation experiments. Targeted genes were PCR amplified and cloned into an episomal expression plasmid for expression in *M. smegmatis*. Unfortunately, only one expression construct was successfully generated, however, no recombinant protein expression could be detected.

https://scholar.sun.ac.za

Chapter 4 **Validating Single Nucleotide Genetic Variants Identified in PE/PPE Genes from Whole Genome Sequencing Analysis of *Mycobacterium tuberculosis*.**

This chapter aims to validate previous findings which predicted several *pe/ppe* single nucleotide variants (SNVs) within drug resistant strains of *M. tuberculosis* following the analysis of WGS data. The validation was carried out by targeted Sanger sequencing, which showed that *pe/ppe* SNVs were present within drug resistant clinical strains of *M. tuberculosis* but were not unique to a drug resistance class. The significance of this outcome is discussed, and suggestions made for further investigation and future approaches.

Chapter 5 **Conclusions**

This chapter elaborates on the results presented in Chapter 3 and 4 to make final conclusions on the relevance of the findings in this study.

Chapter 6 **References and Supplementary Materials**

This chapter includes all referenced studies, as well as supplementary data, procedures and reagents not discussed in the main body of the thesis.

https://scholar.sun.ac.za

Introduction

1.1 *Mycobacterium tuberculosis* – Unknowns on an Epidemic

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* or other members of the *M. tuberculosis* Complex (*MTBC*), has evolved and spread across the globe alongside humankind for approximately 70 000 years (Comas *et al.* 2013). TB remains the leading cause of death world-wide by a single infectious agent (Dheda *et al.* 2016; WHO 2021). The 2020 Covid-19 pandemic has significantly impacted progress to combating *M. tuberculosis* infections. A decrease in TB case reporting and efficient treatment in high-burden countries has been projected to cause an annual excess of 200 000 to 400 000 deaths from 2021-2025 (WHO 2021). This rate is likely to be exacerbated by the recent increases in infection with multi- (MDR) and extensively- (XDR) drug resistant strains, alongside the prevalence of co-infection with Human Immunodeficiency Virus (HIV) in high-burden countries in the African and South-East Asian regions (Shamu *et al.* 2019; WHO 2021). Improving our understanding of the biological components that underpin *M. tuberculosis* pathogenesis may reveal novel pathways for diagnosis or treatment. The roles that many proteins or secretory pathways play in the establishment of these infections are still unknown and under investigation (Ates *et al.* 2016).

M. tuberculosis translocation of proteins across the mycobacterial membrane to the bacterial surface or into host cells requires either the general secretion (Sec) pathway, twin-arginine translocation (TAT) system or the ESX secretion systems (McDonough *et al.* 2008; Daleke *et al.* 2012a; Korotkova *et al.* 2014; Miller *et al.* 2017). Five homologues of the ESX secretion system (ESX-1 to ESX-5) are encoded by *M. tuberculosis*, some of which are shared across mycobacterial species (Houben *et al.* 2014; Gröschel *et al.* 2016). The ESX-1, -3 and -5 pathways have all been shown to facilitate protein secretion and appear essential for virulence (Stanley *et al.* 2003; Abdallah *et al.* 2006; Siegrist *et al.* 2009). ESX-2 and -4 have only recently been implicated in the permeabilization of the phagocytic membrane and secretion of the tuberculosis necrotizing toxin (TNT) during *M. tuberculosis* infections (Pajuelo *et al.* 2021). Substrates of the ESX-1 and -5 pathways are often suggested to be essential components for mycobacterial virulence and infection (Bottai and Brosch 2009; Ekiert *et al.* 2014a; Gallant *et al.* 2021). PE (Proline – Glutamate) and PPE (Proline – Proline – Glutamate) proteins comprise protein families encoded by approximately 10% of the *M. tuberculosis* genome (Cole *et al.* 1998). Members of both protein families have been shown to be substrates of different ESX

https://scholar.sun.ac.za

secretion systems (Abdallah *et al.* 2006; Abdallah *et al.* 2009). Furthermore, the genetic expansion of the *pe/ppe* gene families in pathogenic *M. tuberculosis* alongside the *esx* gene clusters, suggested a role for some PE/PPE proteins in the survival and/or virulence of the bacterium during infection (Cole *et al.* 1998; Gey Van Pittius *et al.* 2006).

Members of both families have since been implicated in metabolic processes, growth and survival of *M. tuberculosis* extra-cellularly, or during infection of host cells (Fishbein *et al.* 2015; Mitra *et al.* 2017; Ates *et al.* 2018b; Ates 2019). Most focussed studies have revealed information regarding the PE family and its associated PE_PGRS (Polymorphic GC-Rich Repetitive Sequence) subfamily (Fishbein *et al.* 2015; Ates 2019). In contrast, the PPE subfamilies are relatively understudied, with significant knowledge gaps surrounding the functions of many member proteins. Recent studies have highlighted the biological relevance of specific PPE_MPTR (Major Polymorphic Tandem Repeat) proteins, which carry out roles in the metabolism and virulence of pathogenic *M. tuberculosis* (Ahmed *et al.* 2015; Brennan 2017; Mitra *et al.* 2017; Ahmed *et al.* 2018a; Gallant *et al.* 2021). Further investigation into both protein families could allow us to advance our understanding surrounding the biochemical and biological aspects, which underpin their function.

1.2 Origin of the PE/PPE Families

The *pe/ppe* genes were first defined after whole genome sequencing of *M. tuberculosis* H37Rv revealed these as two diverse protein coding families (Cole *et al.* 1998). The PE/PPE proteins, comprised of approximately 99 and 69 members respectively in *M. tuberculosis* H37Rv, are characterised by N-terminal proline-glutamate or proline-proline-glutamate amino acid residues (Bottai and Brosch 2009). Since their discovery, both families have been subdivided based on length and unique sequence components (**Figure 1.2.1**). PE proteins have been split into the PE and the PE_PGRS subfamilies. In turn, the PPE family has been divided into PPEs, PPE_PPW (Proline-Proline-Tryptophan), PPE_SVP (Serine-Valine-Proline) and, the recently evolved, PPE_MPTR subfamilies (Hermans *et al.* 1992; Poulet and Cole 1995; Gey Van Pittius *et al.* 2006; Ates *et al.* 2016). Long stretches of repetitive, GC-rich sequences, alongside numerous gene duplications and a lack of tryptic cleavage sites have complicated attempts to study either the gene families or the proteins they encode, hampering phylogenetic studies and leading to their exclusion from most bioinformatic pipelines (Schubert *et al.* 2013; Copin *et al.* 2014; Meehan *et al.* 2019). Despite this, it has been found that the evolution of both families could be linked to the diversification of the *esx*-1 to -5 gene clusters. The *pe/ppe* genes may have been inserted into the *esx*-1 region and subsequently co-duplicated with the expansion of

https://scholar.sun.ac.za

the *esx* gene clusters (Gey Van Pittius *et al.* 2006). Emergence of the PPE_MPTR subfamily (Sublineage V) is relatively recent and the evolution of this family is closely linked to the evolution of the *esx*-5 gene cluster and are restricted to members of the *MTBC* (Deng and Xie 2012). Secretion of MPTR subfamily proteins relies on the ESX-5 apparatus and requires a functional copy of PPE38 (Abdallah *et al.* 2009; Ates *et al.* 2018b; Gallant *et al.* 2021). The association of PE/PPE proteins with the virulence associated ESX secretion systems, alongside their genetic expansion in modern lineages of *M. tuberculosis*, suggests significant roles in the survival or virulence of pathogenic mycobacteria (Daleke *et al.* 2011; Brennan 2017; Delogu *et al.* 2017a; Kramarska *et al.* 2021).

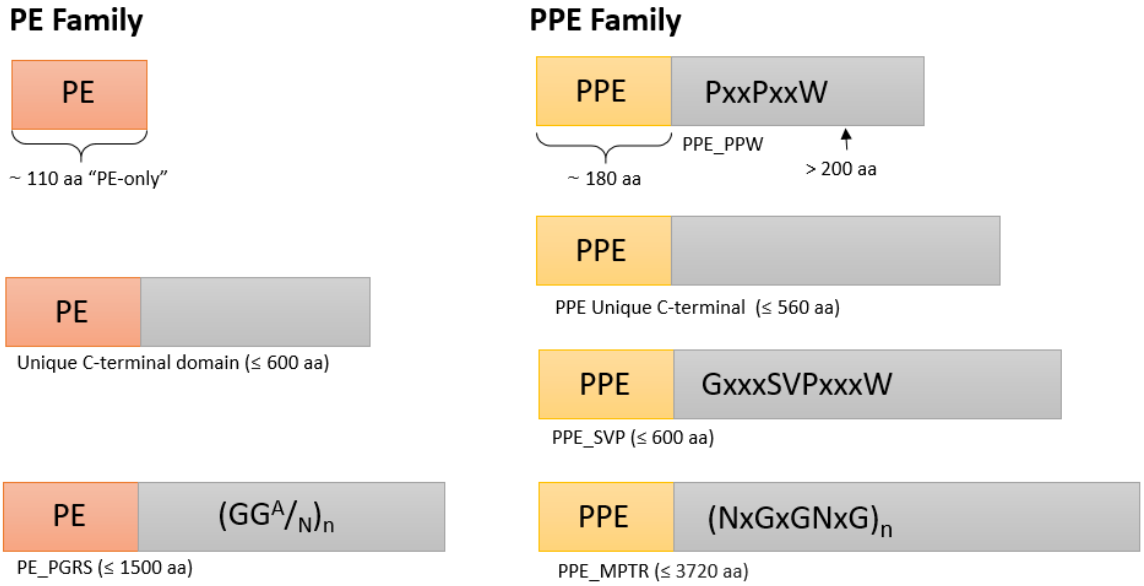


Figure 1.2.2: The PE/PPE protein families and their representative subfamilies. The PE family (left) is characterised by approximately 110 conserved amino acids at its N-terminal. This family is divided into the uncharacterised PE proteins with unique C-terminal regions and the PE_PGRS proteins, which are characterised by a C-terminal polymorphic domain, rich in glycine-glycine-alanine/glycine-glycine-asparagine amino acid (aa) repeats. The PPE family (right) is characterised by its approximately 180 conserved N-terminal aa. This family is divided into four subfamilies. The PPE_PPW subfamily is characterised by a conserved 44 aa repeat of glycine-phenylalanine-X-glycine-threonine and proline-X-X-proline-X-X-tryptophan motifs at its C-terminus. Proteins without any characteristic motifs are placed into the general PPE subfamily, which can range up to 560 aa in size. The largest subfamily, with up to 24 members, is the PPE_SVP, which is characterised by a glycine-X-X-serine-valine-proline-X-X-tryptophan motif between position 300 and 350 in its C-terminus. Finally, the PPE_MPTR subfamily is notoriously variable in size and characterised by multiple asparagine-X-glycine-X-glycine-asparagine-X-glycine repeats in its C-terminal region (Sampson, 2011; Deng and Xie, 2012; Ates, 2019).

https://scholar.sun.ac.za

1.3 Genetic Regulation of PE/PPE Proteins

Several *pe*/*ppe* genes have been shown to be differentially expressed under various stress conditions and during the establishment of bacterial persistence (Park *et al.* 2003; Schnappinger *et al.* 2003; Voskuil *et al.* 2004; Rachman *et al.* 2006). Differential regulation of *ppe_mptr* genes have also been shown by several research studies. Ten *ppe_mptr* genes (*ppe_mptr10, -12, -40, -6, -30, -16, -34, -35, -53* and *-62*) were differentially regulated in *M. tuberculosis* H37Rv within the lung and spleen tissue of mice (Soldini *et al.* 2011). Expression of *ppe_mptr6, -16, -34* and *-53*, were also significantly upregulated when mice were treated with sub-inhibitory concentrations of isoniazid and ethambutol (Soldini *et al.* 2011). Earlier work also highlighted *ppe_mptr35* and *-53* that were significantly upregulated in clinical samples taken from the lung granulomas of TB patients (Rachman *et al.* 2006). Sigma factors, cAMP dependent regulators and two component systems, associated with *M. tuberculosis* virulence, have been proposed as possible regulators of *pe* and *ppe* gene expression (Li *et al.* 2019a; Li *et al.* 2019c).

1.4 Sub-cellular secretion and localisation

Pathogenic mycobacteria have been shown to utilise the Type VII secretion systems, divided into the ESX-1 to ESX-5, for aspects of bacterial virulence (Gröschel *et al.* 2016). The ESX-1 secretion pathway is involved with phagosomal rupture and bacterial escape into the host cell, with some PE/PPE proteins also associated with this pathway (Smith *et al.* 2008; Simeone *et al.* 2012; Ekiert *et al.* 2014a). Homologues of the ESX-5 secretion system are only present in slow growing mycobacteria, including *M. tuberculosis* (Newton-Foot *et al.* 2016).

Little is known about the ESX-dependent mechanisms for the secretion of PE_PGRS or PPE_MPTR substrates. However, ESX-5 is required for the translocation of PE_PGRS and PPE_MPTR proteins, which are isolated to modern lineages of *M. tuberculosis* (Abdallah *et al.* 2006; Abdallah *et al.* 2009). The characteristic PE/PPE N-terminal motif may act as a recognition sequence to facilitate the movement of PE/PPE proteins to the outer membrane (Daleke *et al.* 2011). Once translocated, the N-terminal is cleaved and the functional C-terminal remains, a process demonstrated for the PE protein, known to be a triacylglycerol lipase ($LipY_v$) (Daleke *et al.* 2011). C-terminal tails of PE proteins also contain a highly conserved tryptophan-xxx-aspartate/glutamate (YxxxD/E) motif, which is homologous in proteins associated with ESX secretion (Esp's and EsxB) (Daleke *et al.* 2012a). The YxxxD/E motif has been suggested as a requirement for secretion of PE substrates through the ESX systems.

https://scholar.sun.ac.za

However, this motif does not confer any ESX system specificity. When the YxxxD/E motif of LipYv (secreted via ESX-1) and PE25 (secreted via ESX-5) were exchanged, the proteins were shown to still secrete through their respective pathways (Daleke *et al.* 2012a).

It has been previously suggested that most PE/PPE substrates may be secreted as heterodimers, an assumption that is based on PE25/PPE41 crystallisation and the genomic localization of *pe/ppe* genes. The EspG chaperones, which can recognise a hydrophobic binding sequence on PPE proteins, have been suggested to be responsible for binding and conferring specificity to ESX systems (Phan *et al.* 2017; Damen *et al.* 2020). This was demonstrated by exchanging the $EspG_1$ sequence found in PPE68, of the PE35/PPE68 dimer, with the $EspG_5$ sequence from PPE18, which successfully redirected the secretion of the heterodimer to the ESX-5 system (Phan *et al.* 2017). Studies performed in *Mycobacterium marinum* have shown that a functional ESX-5 and PPE38 expression is required for the secretion of PE_PGRS and PPE_MPTR proteins (Ates *et al.* 2018b). PE/PPE proteins may not exclusively utilise ESX secretion pathways, as several proteins have been shown to contain recognition sequences for the general secretory (Sec) pathway (Fishbein *et al.* 2015).

Proteomic and computational studies have both suggested outer membrane localisation of PE/PPE proteins (Wolfe *et al.* 2010; Målen *et al.* 2011). This would allow them to interact at the host-pathogen interface on the cell surface (Sampson 2011). If confirmed, further study could provide more information regarding the functional significance of this localisation (Sampson 2011; Brennan 2017).

1.5 Functional Roles and Biological Significance of the PE/PPE Proteins

The majority of the PPE subfamily remains understudied compared to the PE and PE_PGRS subfamilies. However, some studies have suggested functions of various PPE_MPTR proteins which are described in **Table 1.5.1**.

https://scholar.sun.ac.za

Table 1.5.2: PPE_MPTR proteins of interest. Details provided on their associated biological functions and characteristics as described in available literature.

Gene	Protein	Functions and Characteristics from Literature
Rv0305c	PPE_MPTR6	Increased expression under cellular stress conditions (Soldini *et al.* 2011).
Rv0354c	PPE_MPTR7	Gene disruption led to *in vitro* growth advantage for *M. tuberculosis* H37Rv (Dejesus *et al.* 2017).
Rv0355c	PPE_MPTR8	Highly polymorphic C-terminal sequence observed from over 300 clinical *M. tuberculosis* isolates (Srivastava *et al.* 2006).
Rv0442c	PPE_MPTR10	Upregulated during *M. tuberculosis* infection of alveolar macrophages (Kruh *et al.* 2010). Potential role in evasion of phagosomal acidification during infection (Stewart *et al.* 2005). Linked to capsular integrity and cell viability during infection (Ates *et al.* 2016). Shown to increase cell survivability and attenuate host immune responses (Asaad *et al.* 2021).
Rv0755c	PPE_MPTR12	Significantly upregulated 24h post infection with *M. tuberculosis* (Kruh *et al.* 2010).
Rv0878c	PPE_MPTR13	Differentially expressed in clinical isolates from lung granulomas (Rachman *et al.* 2006). Directly activates the NLRP3 inflammasome cascade in macrophages (Yang *et al.* 2020b).
Rv1135c	PPE_MPTR16	Potential role in evasion of phagosomal acidification during infection (Stewart *et al.* 2005). Showed increased expression under stress conditions (Soldini *et al.* 2011).
Rv1548c	PPE_MPTR21	Potential role in evasion of phagosomal acidification during infection (Stewart *et al.* 2005).
Rv1753c	PPE24	Contains Gly-, Asn-rich regions interrupted by unique near perfect 26 aa repeats (Sampson *et al.* 2001a). Promotes bacterial survival within macrophages (Sassetti *et al.* 2003).
Rv1917c	PPE34	Contains Gly-, Asn-rich regions interrupted by unique near perfect 26 aa repeats (Sampson *et al.* 2001a). Reported to interact with immune cells to alter immunogenic reactions during infection, contributing to immune evasion (Bansal *et al.* 2010). Showed increased expression under stress conditions (Soldini *et al.* 2011).
Rv1918c	PPE_MPTR35	Contains Gly-, Asn-rich regions interrupted by unique near perfect 26 aa repeats (Sampson *et al.* 2001a). Differentially expressed in clinical isolates from lung granulomas (Rachman *et al.* 2006). Increased expression in immune-competent versus immune-deficient mice (Talaat *et al.* 2004).

https://scholar.sun.ac.za

Rv2353	PPE_MPTR39	Found completely or partially deleted in various clinical *M. tuberculosis* isolates (Tsolaki *et al.* 2004).
Rv2356c	PPE_MPTR40	Constitutively expressed during *in vivo* mouse infections with *M. tuberculosis* (Soldini *et al.* 2011).
Rv2608	PPE_MPTR42	Highly variable C-terminal sequence (McEvoy *et al.* 2012). Elicits powerful humoral response which has led to its inclusion in polyprotein *M. tuberculosis* vaccines (Bertholet *et al.* 2010). Reported as most immunogenic component of the ID93 vaccine, with highest level of population coverage (Ong *et al.* 2020).
Rv3159c	PPE_MPTR53	Expression downregulated after extended periods of starvation (Betts *et al.* 2002). Showed increased expression under stress conditions (Soldini *et al.* 2011).
Rv3343c	PPE_MPTR54	Promotes bacterial survival within macrophages (Sassetti *et al.* 2003). Potential role in evasion of phagosomal acidification during infection (Brodin *et al.* 2010). Highly variable C-terminal sequence (McEvoy *et al.* 2012). Differentially expressed in clinical isolates from lung granulomas (Rachman *et al.* 2006).
Rv3347c	PPE55	Stimulates a hyper-stimulatory immune response during macrophage infection (Singh *et al.* 2005a). Identified as potential *M. tuberculosis* vaccine candidate (Zvi *et al.* 2008).
Rv3350c	PPE_MPTR56	Exhibits higher expression in a *phoP*|Rv0757 mutant than in wildtype *M. tuberculosis* H37Rv strains (Walters *et al.* 2006).
Rv3533c	PPE_MPTR62	Surface exposed heme-binding protein (Mitra *et al.* 2017). Differentially expressed in clinical isolates from lung granulomas (Rachman *et al.* 2006). Constitutively expressed during *in vivo* mouse infections with *M. tuberculosis* (Soldini *et al.* 2011).

https://scholar.sun.ac.za

1.5.2 PPE Protein Family and Subfamilies

Proteins within the PPE subfamilies have also been shown to carry out essential roles during infections. PPE24, -53 and -54 have been reported as essential for bacterial survival in macrophages, while PPE31, -68, -25 and 26 are required for mycobacterial growth *in vivo* (Sassetti *et al.* 2003; Sassetti and Rubin 2003; Mi *et al.* 2017). Targeted investigation of some PPE proteins has revealed more specific roles for some member proteins, as described below.

Studies in *Mycobacterium marinum* have linked PPE_MPTR10 to capsular integrity, cell viability and survival of *M. tuberculosis* during growth, as well as infection (Ates *et al.* 2016). Similarly, expression of PPE_MPTR10 in *Mycobacterium smegmatis* also increased cell resistance to stressors that usually limit growth (Asaad *et al.* 2021). Much like the PE proteins mentioned previously, SNVs in certain *ppe* genes have been observed alongside canonical resistance-conferring mutations and have been associated with increased drug tolerance in *M. tuberculosis*. These include SNVs in *ppe*18, -19, -46 and -47, all of which have been detected in mycobacterial strains with increased resistance to isoniazid (Hang *et al.* 2019). Several genetic variations were also reported in *ppe_mptr*54 that appeared in bacteria with rifampicin, isoniazid and ethambutol resistance (Cui *et al.* 2016).

It has been suggested that both PE_PGRS and PPE_MPTR proteins are dependent on the functional expression of PPE38 for secretion (Ates *et al.* 2018b). A lack of PPE38, alongside subsequent abolition of associated PE/PPE secretion, was linked with a significant increase in bacterial virulence (Ates *et al.* 2018b). Several MPTR proteins, including PPE38, are also strongly associated with host modulation, leading to bacterial evasion of macrophage defence mechanisms during infections (Nair *et al.* 2009; Bansal *et al.* 2010; Brennan 2017; Asaad *et al.* 2021).

Loss of PPE38 abrogated the standard pro-inflammatory macrophage response upon infection (Gallant *et al.* 2021). Instead, infection with a PPE38-deficient mutant induced an alternative RelB/p50 NF-kB pathway, resulting in a dampened immune response in infected macrophages (Gallant *et al.* 2021). Expression of recombinant PPE_MPTR10 in *M. smegmatis* improved intracellular survival by suppressing the transcription of inflammatory cytokines and attenuating host-cell apoptosis (Asaad *et al.* 2021). Conversely, PPE_MPTR13 activates the NLRP3 inflammasome cascade, leading to the maturation and secretion of the pro-inflammatory Interleukin 1-β (IL-1β) (Yang *et al.* 2020b). Both the PPE18 and the MPTR

https://scholar.sun.ac.za

protein PPE34 have also been directly implicated in modulation of host cytokine responses during *M. tuberculosis* infections. PPE18 interacts with Toll-like receptor 2 (TLR2), leading to the inhibition of NF-kB (Nuclear Factor kappa-light-chain-enhancer of activated B cells), subverting the immune response of infected cells (Nair *et al.* 2011). PPE34 is implicated in promoting the secretion of anti-inflammatory cytokines by triggering dendritic cell maturation, shifting immune responses to an Th2 (T-Helper Cell Type 2) based response, which is less effective in clearing cellular infections (Bansal *et al.* 2010). Transposon mutants of PPE_MPTR10, -16, -21 and -54 were all heavily enriched in fully acidified phagosomes, which has suggested roles for these proteins in subverting phagosomal maturation (Stewart *et al.* 2005; Brodin *et al.* 2010).

PPE/PPE_MPTR proteins have also been reported as potential sources of antigenic variation within *M. tuberculosis* strains. Specific examples include the C-terminal sequences of PPE24, -34, -42 and -54, which have all been reported as highly variable and enriched in SNVs within clinical isolates (Chakhaiyar *et al.* 2004; McEvoy *et al.* 2012). This has contributed to the speculation regarding the role of these proteins as sources of antigenic variation that may contribute to immune evasion, although this has yet to receive experimental support. Recent studies also highlighted that restoring the secretion of up to 89 PE_PGRS/PPE_MPTR proteins in the *Mycobacterium bovis* Bacille Calmette-Guérin vaccine (BCG) seemed to have little effect on immune stimulation or protective efficacy (Ates *et al.* 2018c).

PPE/PPE_MPTR proteins have become linked to various critical roles during *M. tuberculosis* infections. Their association with the ESX-5 secretion system, alongside their suggested roles in *M. tuberculosis* metabolism and host modulation, has highlighted them as targets for diagnosis or treatment of mycobacterial infections.

1.5.3 Potential Applications

Previous studies have demonstrated that both PE and PPE proteins can elicit significant T-cell responses during *M. tuberculosis* infections (Delogu *et al.* 2017a). Some of these reactions were able to confer protective immunity against further infection, which lead to the inclusion of PPE18 and PPE_MPTR42 in the *M. tuberculosis*72F and ID93 polyprotein vaccines (Skeiky *et al.* 2004; Bertholet *et al.* 2010). Both vaccines have demonstrated significant levels of protection, with *M. tuberculosis*72F reported as safe in humans (Reed *et al.* 2009; Spertini *et al.* 2013; Baldwin *et al.* 2021). This highlights the value of the PPE/PPE_MPTR proteins in the potential future treatments for *M. tuberculosis* or other mycobacterial infections, suggesting

https://scholar.sun.ac.za

that further detailed studies would be beneficial (Brennan 2017; Delogu *et al.* 2017a; Ates 2019; Li *et al.* 2019b).

1.6 Problem Statement

The emergence of *M. tuberculosis* drug tolerant or resistant strains, capable of withstanding standard antibiotic regimens are of great concern (Ates *et al.* 2018a; Hang *et al.* 2019). Significant gaps also exist in our understanding of the biological components that underpin mycobacterial pathogenesis. Both PE and PPE proteins have been suggested to play essential roles in the development of drug resistant strains, while also fulfilling various biological functions (Delogu and Brennan 2001; Wang *et al.* 2020b; Asaad *et al.* 2021).

Molecular, genomic and high-throughput proteomic analyses have suggested that several PE/PPE proteins may surface-localise and contribute to interactions at the host-pathogen interface (Wolfe *et al.* 2010; Mazandu and Mulder 2012). The majority of PPE_MPTR proteins are however understudied in regards to their localisation and functional roles (Qian *et al.* 2020). Select PE/PPE proteins are also known to carry unique genetic variations in their coding sequences within drug-resistant strains (Kanji *et al.* 2015; Hang *et al.* 2019). Due to the highly repetitive GC-rich nature of the *pe/ppe* sequences, they are often excluded in bioinformatic pipelines that aim to analyse the whole genome sequence (WGS) of clinical *M. tuberculosis* (Copin *et al.* 2014). This could potentially lead to the exclusion of PE/PPEs of interest in studies looking to explore the role of these proteins in drug resistance.

For many members of the PPE family, specifically members of the MPTR subfamily, there is a significant gap in our understanding of their fundamental biology. Focussed studies are required to provide information about the sub-cellular localisation and potential partners of the proteins in this subfamily (Ates 2019; Qian *et al.* 2020; Asaad *et al.* 2021). Our study aims to identify the sub-cellular location of select PPE_MPTR proteins in *M. smegmatis*. Furthermore, recent work within our research group (Bagheri, B.) has also highlighted several *pe/ppe* genes, that would normally be excluded by analytical pipelines, which contain unique SNVs that appear isolated to drug-sensitive (DS), MDR and XDR *M. tuberculosis* (**Table 4.2.1**). We aim validate these findings in a selection of clinical isolates. These results may identify PE/PPE proteins of interest for future studies that may wish to explore the roles of specific PE/PPE proteins in drug resistance.

https://scholar.sun.ac.za

1.7 Hypotheses

1. Target PPE_MPTR proteins expressed in *M. smegmatis* will localise to the mycobacterial membrane.

2. Genetic variants in *pe/ppe* predicted to be associated with antibiotic resistance will be identified in a selection of clinical isolates using targeted sequencing.

1.8 Aim and Objectives

Aim 1: Determine the sub-cellular localisation of selected PPE_MPTR proteins in *Mycobacterium smegmatis*.

Objective 1: Clone pCG constructs for the expression of target PPE_MPTR-GFP fusion proteins.

Objective 2: Express fluorescently tagged PPE_MPTR proteins in *M. smegmatis* and mutant strains of *M. smegmatis* (ΔESX-1, ΔESX-3, ΔESX-4, ESX-5_{Xenopi}).

Objective 3: Identify the location of fluorescently tagged PPE_MPTR proteins *in vitro* using fluorescent microscopy.

Aim 2: Confirm the presence of identified SNVs in *pe* and *ppe* genes in clinical isolates of DS, MR, MDR and XDR strains of *M. tuberculosis*.

Objective 4: Select DNA isolates from 100 clinical strains, including DS, MDR and XDR strains. Screen available WGS data of these isolates to determine the presence or absence of previously identified SNVs of interest.

Objective 5: Amplify genes of interest from selected clinical isolates using polymerase chain reaction (PCR) to verify SNVs of interest using targeted Sanger sequencing.

Objective 6: Determine whether the identified SNVs are associated with their respective drug resistance class across all selected clinical *M. tuberculosis* isolates.

https://scholar.sun.ac.za

Chapter 2

Literature Review

Secretion and localisation signals Utilised by *Mycobacterium tuberculosis*

2.1 Introduction

Mycobacterium is a genus of Actinobacteria that are primarily represented by saprophytic soil bacteria, but a select few are the causative agents of clinically relevant diseases (Brennan 1995). These include members of the *Mycobacterium tuberculosis* (*M. tuberculosis*) *complex (MTBC)* and *Mycobacterium leprae*, respectively the causative agents of tuberculosis (TB) and leprosy (Cosma *et al.* 2003; Franco-Paredes *et al.* 2019). *M. tuberculosis* is ranked as the deadliest infectious agent globally, accounting for up to 10 million annual cases of infection (WHO 2021). Numerous mycobacterial proteins expressed by *M. tuberculosis*, as well as other infectious mycobacteria, have been demonstrated as essential for bacterial survival and growth during infection (Griffin *et al.* 2011; Dejesus *et al.* 2017). A large portion of these proteins include various members of the PE (Proline – Glutamate) and PPE (Proline – Proline – Glutamate) protein families, that interact with host cells to modulate the internal immune response or increase mycobacterial resistance to intracellular stressors during infections (Sassetti and Rubin 2003; Reddy *et al.* 2012; Ahmed *et al.* 2018a; Gupta *et al.* 2020; Wang *et al.* 2020a; Wang *et al.* 2020b). The localisation of these proteins to a specific sub-cellular domain may be essential for their functionality, but may also indicate the role these proteins play at the host pathogen interface (Fishbein *et al.* 2015; Delogu *et al.* 2017b).

Secretion systems are responsible for transporting the majority of proteins to the mycobacterial membrane or surface, where they may function and interact at the host-pathogen interface (Sampson 2011). Mycobacteria utilise several pathways for the translocation of proteins to their target location, including the general Secretory (Sec), Twin-Arginine Translocation (Tat) and Early Secretory Antigenic Target (ESX) secretion systems. The Sec pathway exports proteins in an unfolded state across the cytoplasmic membrane, but is also essential for integrating proteins into the membrane itself (Miller *et al.* 2017; Phan and Houben 2018). Proteins are targeted to the SecYEG channel by a mildly hydrophobic N-terminal signal sequence, which becomes cleaved upon translocation (Digiuseppe Champion and Cox 2007a). Folded proteins that also contain this N-terminal sequence, but carry a twin-arginine motif

https://scholar.sun.ac.za

within it, can become targeted to the general Tat pathway for export across the cytoplasmic membrane via the TatABC channels instead (Palmer and Berks 2012). Gram-negative bacteria also possess the conserved Type I to VI secretion pathways, which may be one-part systems (T1S, T3S, T4S, T6S) that translocate proteins across the inner and outer membrane, or two-part systems (T2S and T5S), which utilises the general Sec or Tat pathways for export across the inner cytoplasmic membrane and a second system for transport across the outer membrane (Phan and Houben 2018). Mycobacteria utilise the Type VII secretion system, which is essential for the export of numerous virulence factors. *M. tuberculosis* encodes five homologous T7S subsystems known as the ESX-1 to -5 secretion systems (Houben *et al.* 2014; Gröschel *et al.* 2016; Beckham *et al.* 2017). These ESX secretion systems have an array of substrates that are targeted to each, guided by signal peptides and corresponding chaperones to facilitate their transport and translocation. The exact signal involved in targeting the various homologues of these systems still remains unknown (Gröschel *et al.* 2016). Several studies have increased our understanding of the transport signals and chaperones involved in the secretion of substrates, showing how they connect to systems that allow for the translocation and sub-cellular localization of key mycobacterial proteins (Daleke *et al.* 2011; Korotkova *et al.* 2014; Ates and Brosch 2017; Miller *et al.* 2017; Phan *et al.* 2017).

2.2 General Protein Trafficking

2.2.1 Chaperones

Prokaryotes utilise DnaK, GroEL and Trigger Factor (TF) as three conserved general chaperones to guide and maintain proteins in an ideal state for translocation. Both DnaK and GroEL are highly conserved across several species, and act as chaperones that can keep proteins in a translocation-competent state by binding to extended hydrophobic sequences on proteins, as demonstrated in *Escherichia coli* (Rüdiger *et al.* 1997; Houry 2001). There are two paralogs of GroEL in *M. tuberculosis* and other mycobacteria, namely GroEL1 and GroEL2 (Cpn60.1 and Cpn60.2). These have been shown to function at a lower oligomer state and have several key variations in hinge regions in comparison to the similar proteins in *E. coli* (Qamra *et al.* 2004; Chilukoti *et al.* 2016). Within *M. tuberculosis*, they can be found in relatively unstable, double ringed tetradecameric structures similar to GroEL in *E. coli*, instead of their base monomeric states (Kumar *et al.* 2009; Fan *et al.* 2012). GroEL2 is essential to mycobacterial survival, likely acting as the primary chaperone between the two homologs, as it was shown to rescue the activity of native GroEL, when expressed at higher levels in *E. coli* (Fan *et al.* 2012). In contrast, GroEL1 is dispensable in several mycobacterial species, but may

https://scholar.sun.ac.za

still have the potential to function as a chaperone when not contributing to other metabolic processes involved in the bacterial stress response (Kumar *et al.* 2009; Ansari *et al.* 2020). It has been shown to be involved in copper, nickel and iron binding in *M. tuberculosis* and *M. smegmatis* (Kumar *et al.* 2009; Ansari *et al.* 2020). This has suggested a role for GroEL1 in maintaining homeostasis of metal ions that could cause oxidative damage to bacteria during infection or biofilm formation when present in excess (Ojha *et al.* 2005; Yang *et al.* 2020a). Otherwise, it also has potential as a nucleotide associated protein (NAP), as it binds chromosomal DNA with high affinity, forming stable structures that protect DNA from digestion (Basu *et al.* 2009). Although DnaK is functionally redundant in *E. coli* and not essential as a chaperone, studies in *M. smegmatis* have demonstrated an essential role in ensuring the solubility and function of larger peptides in conjunction with ClpB (Deuerling *et al.* 1999; Fay and Glickman 2014; Yu *et al.* 2018). This includes acting as chaperone to the Fatty Acid Synthase I (FASI) protein, which is a function that cannot be supplemented by the TF chaperone as in *E. coli* (Harnagel *et al.* 2020). Both GroEL2 and DnaK localise to the capsule and are presented on the surface of *M. tuberculosis*, which could allow them to interact with the largely hydrophobic phospholipids and other lipid-rich molecules that make up the mycobacterial cell wall (Hickey *et al.* 2009). TF has, in contrast to GroEL and Dnak, targets short amino acid motifs enriched in basic and aromatic residues in *E. coli*, although major differences in TF ribosomal binding and structure have been observed in *M. tuberculosis* in comparison to *E. coli* models (Patzelt *et al.* 2001; Ferbitz *et al.* 2004; Li *et al.* 2019c). In addition to these, more specific chaperones can be found in mycobacteria, as well as various other organisms, that also contribute to the internal trafficking and targeting of protein substrates to their specific secretion pathways (Gröschel *et al.* 2016). Beyond essential chaperones that assist in translocation, specific and general target sequences play a role in the transport of proteins through the various available pathways (Daleke *et al.* 2012a; Lou *et al.* 2017; Phan *et al.* 2017).

2.2.2 General Secretion Signals and Systems

The general Sec pathway is composed of a SecYEG channel, with additional SecD and –F membrane components, as well as a SecA ATPase, which is responsible for recognising the Sec signal sequence, demonstrated in **Figure 2.2.1** (Digiuseppe Champion and Cox 2007a). This N-terminal recognition domain is composed of a positively charged N-terminal sequence, followed by a core hydrophobic region and ends on an uncharged polar region that contains

https://scholar.sun.ac.za

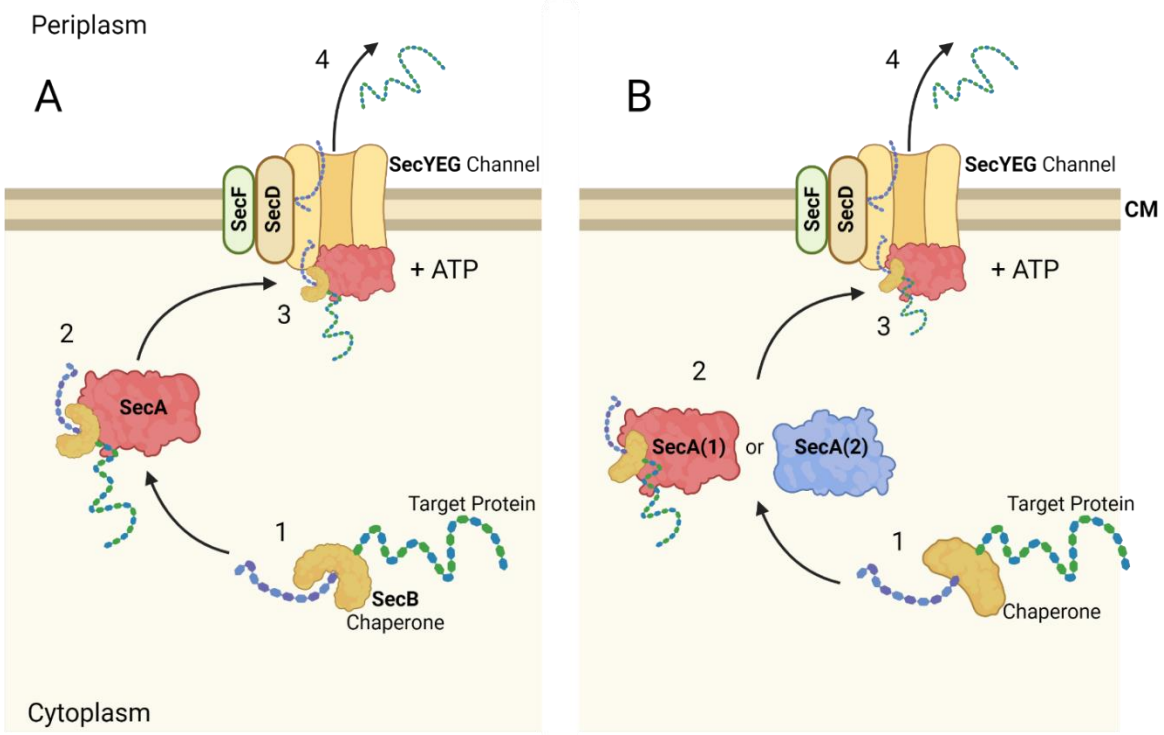


Figure 2.2.3: General SecYEG secretion system. (**A**) Canonical pathway of the Sec system. The SecB chaperone recognizes and binds the N-terminal secretion signal of a target protein (1). Target proteins are guided through SecB to the SecA, ATP-ase (2), which accompanies the protein to the SecYEG channel, where it binds to a high-affinity SecA binding site (3). The protein is translocated through the membrane, a process which requires ATP hydrolysis, becomes cleaved at the recognition signal and moved into the periplasmic space for further processing (4). (**B**) A non-canonical Sec system is present in some Gram-positive bacteria and mycobacteria. These systems may utilize another SecB-like chaperone to bind and guide proteins (1). These bacteria have also been shown to contain two SecA paralogs (2), which can carry out independent functions, although only SecA1 is essential in mycobacteria, both proteins have been associated with secretion in a SecYEG dependent manner (Gerlach and Hensel, 2007; Prabudiansyah *et al.*, 2015). CM, Cytoplasmic membrane. Figure was produced using BioRender (https://biorender.com/).

a cleavage site, which in mycobacteria most closely resembles the sequences found in Gram-positive bacteria (Wiker *et al.* 2000). Several PE/PPE proteins have also been found to contain an N-terminal Sec recognition sequence, which suggests that some may be able to utilise this more conventional system instead of the ESX pathways (Fishbein *et al.* 2015). Mycobacteria lack the canonical SecB chaperone, usually found in Gram-negative bacteria, which as shown in **Figure 2.2.1 (A)**, is responsible for guiding unfolded substrates from the ribosome to SecA. *M. tuberculosis* does however encode a SecB-like protein (Rv1957), which

https://scholar.sun.ac.za

is able to functionally replace the SecB chaperone in *E. coli* (Scott and Barnett 2006; Bordes *et al.* 2011). The Sec pathway is associated with the secretion of the majority of proteins in bacteria, which is a concept carried over to mycobacteria. The exact substrates and virulence factors that rely on the Sec pathway for secretion is difficult to study in mycobacteria, due to mutations in SecA1 mostly being lethal (Digiuseppe Champion and Cox 2007a). However, some virulence factors from *Listeria monocytogenes* and *Staphylococcus aureus* that rely on the pathway have been identified through mutations in the non-essential components of the Sec system (Quiblier *et al.* 2011; Burg-Golani *et al.* 2013). Bioinformatic analysis has suggested that up to 593 proteins may be exported in a SecA1 dependent manner in *M. tuberculosis* (Miller *et al.* 2017). Several of these potential substrates are essential to the viability of *M. tuberculosis* and other mycobacteria. These include cell wall synthesis and remodelling factors, such as the peptidoglycan hydrolase RipA and all eight of the *M. tuberculosis* annotated Penicillin Binding Proteins (PBPs), which are essential to cell wall modifications (Hett *et al.* 2007; Machowski *et al.* 2014). Lipo-proteins exported via the Sec pathway have also been shown to contain a lipobox motif at the C-terminal end of the Sec signal sequence, with an invariant cysteine, which acts as the site of lipid attachment (Nakayama *et al.* 2012). These proteins notably include the Type II signal peptidase LspA, as well as the Sec signal containing virulence factors LprG and LpqH, which are both exported to the cell wall and lead to severe *M. tuberculosis* attenuation when supressed (Henao-Tamayo *et al.* 2007; Gaur *et al.* 2014). Numerous factors speculated to be involved in the dormancy and resuscitation of *M. tuberculosis* also fall within this potential secretory group, including three L,D-transpeptidases (LdtMt2, PBPA, Rv1433) and five resuscitation-promoting factors (RpfA-E), among various other potential virulence factors, which have yet to be assigned a defined function, such as Erp (Exported, repetitive protein) and Rv0888 (an extracellular nuclease) (Berthet *et al.* 1998; Tufariello *et al.* 2006; Dutta *et al.* 2010; Dang *et al.* 2016).

Mycobacteria also encode for a secondary SecA2 secretion system, which in contrast to SecA1, is not essential to mycobacterial survival, but has been shown to be required for the full virulence of *M. tuberculosis* (Kurtz *et al.* 2006). SecA2-dependent export requires the core Sec accessory proteins of the SecYEG channel (Ligon *et al.* 2012). Furthermore, although it has been recognised to bind and export proteins carrying the standard Sec signal peptides, other substrates, such as the SodA (superoxide dismutase) and KatG (catalase) that both lack the standard signal sequence, have also been shown to be dependent on SecA2 for export in *M. tuberculosis* (Braunstein *et al.* 2003; Hinchey *et al.* 2007). Beyond direct export, the

https://scholar.sun.ac.za

standard SecA pathway, alongside the insertase YidC, is also essential to the integration of proteins into the inner cytoplasmic membrane (Luirink and Sinning 2004). This integral membrane localisation of proteins through the Sec pathway is a co-translational process that involves the ribonucleoprotein signal recognition particle (SRP) (Luirink and Sinning 2004). SRP generally identifies the transmembrane domains of target proteins by a signal sequence with a core 8-12 hydrophobic amino acids that adopt an α-helical structure (Valent *et al.* 1997; Akopian *et al.* 2013). This leads SRP to deliver the protein to the FtsY inner membrane receptor, which then delivers it to SecYEG and allows it to pass into the cytoplasmic membrane through a lateral channel in SecY (Egea and Stroud 2010).

Other than the general Sec pathway, bacteria can also possess the previously mentioned Twin-Arginine Translocation (Tat) pathway. Although it is not a universally available system for protein export, it is often associated with the export of virulence factors in bacterial pathogens (Palmer and Berks 2012). Studies in *E. coli* have shown the Tat system to be composed of a TatA channel, complemented by TatB and TatC membrane components, that act as the docking region for target proteins (Goosens *et al.* 2014). Little is known about the exact composition and function of the Tat system in the majority of mycobacteria. It is believed to also be composed of components like the TatA, -B and –C, found in Gram-negative bacteria, based on findings from *M. smegmatis*, demonstrated in **Figure 2.2.2** (Posey *et al.* 2006). Substrates are transported to the periplasm in a folded state, however, how they are translocated to the outer membrane or secreted after moving through the Tat system, is currently unknown in mycobacteria (Goosens *et al.* 2014). Tat recognition sequences consist of a positively charged N-terminal sequence, followed by a core hydrophobic region and ending on an uncharged polar region containing a cleavage site. The key difference is found in the twin-arginine motif found between the charged and hydrophobic domains of the signal sequence, generally denoted as R-R-X-Φ-Φ, where Φ represents a hydrophobic residue (Berks *et al.* 2000). It has however been shown that the twin-arginine motif is not completely essential, with several Tat substrates only containing a single arginine residue, leading to speculation that the degree of hydrophobicity in the domain following the Tat signal, or the presence of a positively charged residue at the end of the signal, may also target the substrate to the pathway (Cristóbal *et al.* 1999; Hinsley *et al.* 2001). Tat dependent substrates have been demonstrated for several mycobacteria, including *M. smegmatis* and *M. tuberculosis* (McDonough *et al.* 2005; McDonough *et al.* 2008). Tat secretion contributes to virulence and viability of *M. tuberculosis*, facilitating the secretion of active β-lactamases and the phospholipase C protein (McDonough

https://scholar.sun.ac.za

et al. 2005; McDonough *et al.* 2008). *M. tuberculosis* Tat substrates have been shown to carry a translocation signal denoted as Φ-Σ-R-R-X-Φ-Φ, where Σ corresponds to a small amino acid (McDonough *et al.* 2008; Ligon *et al.* 2012).

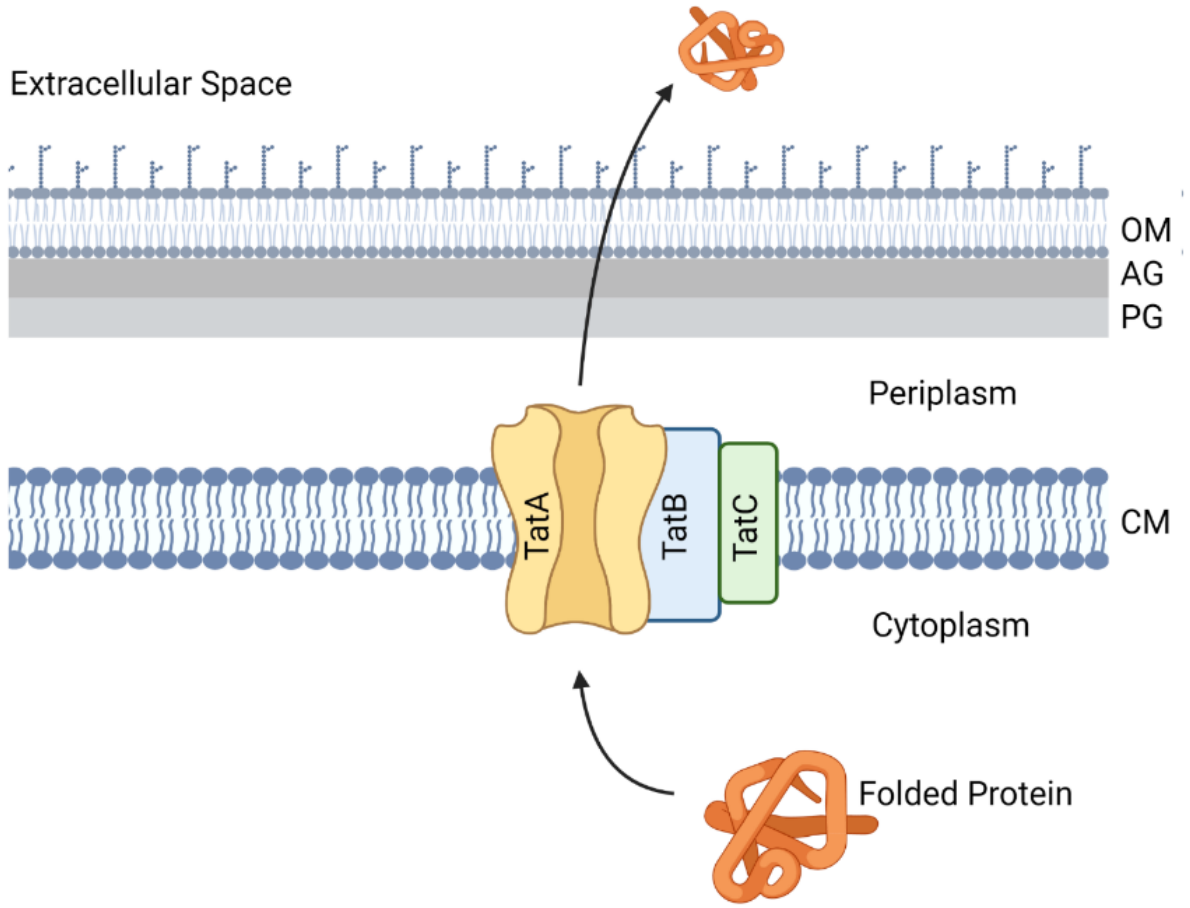


Figure 2.2.4: General Tat secretion system theorised for diderm mycobacteria. The TatA, TatB and TatC-like complex is embedded into the inner membrane. Folded proteins are translocated through the TatA-like channel, after binding to the docking region composed of the TatB and TatC. Through the Tat system proteins are translocated into the periplasm, but the system used for total secretion following transloaction is currently poorly understood in mycobacteria (Posey *et al.* 2006; Goosens *et al.* 2014). OM, Outer membrane; AG, Arabinogalactan; PG, Peptidoglycan; CM, Cytoplasmic membrane. Figure was produced using BioRender (https://biorender.com/).

https://scholar.sun.ac.za

2.3 Type VII Secretion: ESX pathways in Mycobacteria

Mycobacteria, especially pathogenic species such as *M. tuberculosis*, have been shown to utilise the Type VII (ESX) secretion systems in addition to the general secretory pathways (Beckham *et al.* 2017). The ESX systems are present in mycobacteria, as well as various Actinobacteria and Firmicutes (Bitter *et al.* 2009). These are primarily represented by ESX-4 systems, which have been suggested as the genetic origin of later ESX secretion systems (Bitter *et al.* 2009; Punta *et al.* 2012). ESX secretion systems are composed of a core complex, which contains the EccB, EccC, EccD and EccE protein channel, as well as a Mycosin Component (MycP), demonstrated in **Figure 2.3.1** (Houben *et al.* 2012). MycP is essential to the function of most ESX systems. While its complete deletion inhibits secretion in ESX systems, its proteolytic activity appears dispensable (Ohol *et al.* 2010). Another core component associated with ESX secretion, is the ATPase, EccA, which is essential to the function of ESX-1, -3 and -5 systems, although its exact role is currently still poorly understood (Sassetti *et al.* 2003; Bottai *et al.* 2012).

ESX secretion, as shown in **Figure 2.3.1**, has been modelled as both a two-step and one-step process. In two step processes, proteins are translocated into the periplasm, where they can be processed before being secreted through a second cognate system. In contrast, the one-step process is supposed to utilise an already existing component of the ESX system to directly translocate proteins to the bacterial surface. Potential channels or components that would facilitate both these models have been proposed, but none confirmed (Houben *et al.* 2012; Roy *et al.* 2020). However, ESX-1 associated EspC proteins polymerise and form filaments that are visible on the mycobacterial surface. These potentially act as an ESX-1 secretion needle, bridging the ESX channel through the mycobacterial outer membrane (Ates and Brosch 2017; Lou *et al.* 2017).

Selected members of the Type VII secretion systems are essential to the virulence of slow growing, pathogenic strains of mycobacteria, such as *M. tuberculosis*, which encodes all five ESX homologues (ESX-1 to ESX-5) (Gröschel *et al.* 2016). The ESX-1, -3 and -5 systems all facilitate the secretion of protein substrates, whereas ESX-2 and -4 have not been directly implicated in secretion (Stanley *et al.* 2003; Abdallah *et al.* 2006; Siegrist *et al.* 2009). Although the role of the ESX systems in protein translocation has been well established, some have also been implicated in essential metabolic processes. This includes ESX-1 in *M. smegmatis*, which appears involved in DNA conjugation, while ESX-3 in *M. tuberculosis*

https://scholar.sun.ac.za

and *Mycobacterium bovis* is essential for the uptake of iron and bacterial growth (Coros *et al.* 2008; Siegrist *et al.* 2009).

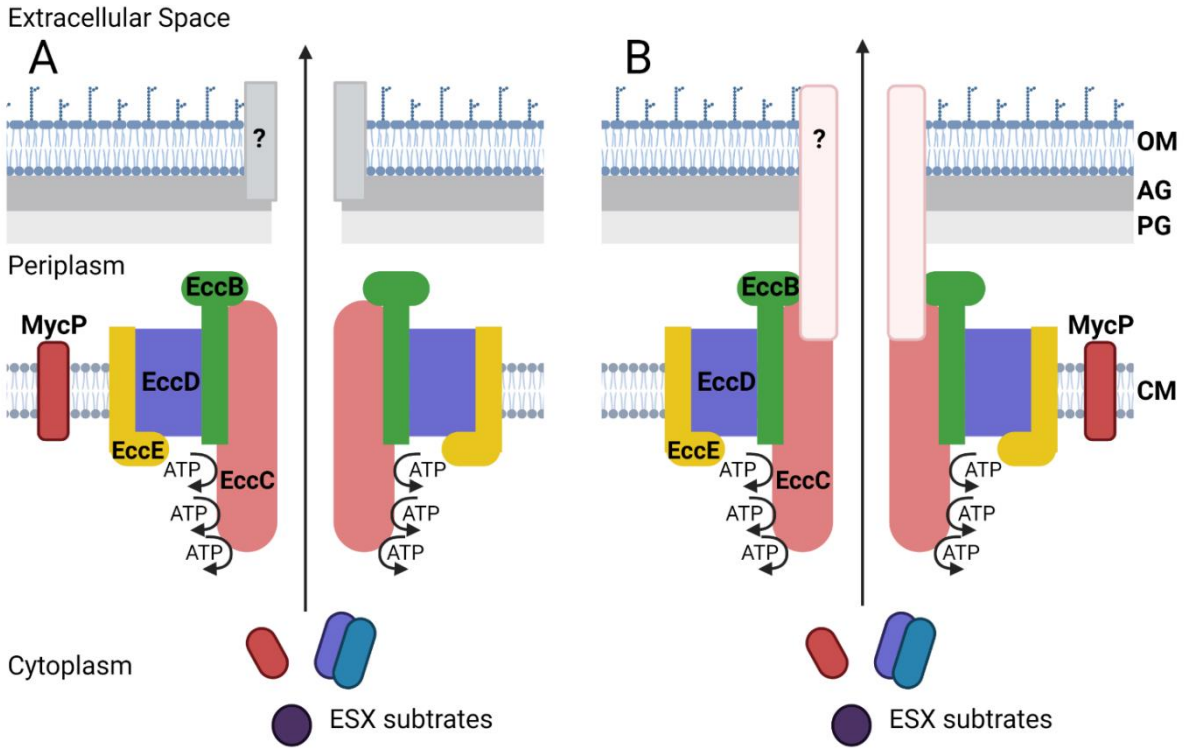


Figure 2.3.3: General composition of the Type VII ESX secretion systems. (**A**) A two-step ESX translocation system. The core EccC, EccB, EccD, EccE and Mycosin (MycP) components are embedded into the inner membrane, forming the core ESX channel and complex. Substrates are translocated through the channel by an energenic process, likely catalysed by EccC, which contains three nucleotide binding domains. In this model, substrates are moved into the periplasm, processed and then secreted out of the cell by an currently unknown pore/channel. (**B**) A one-step ESX secretion system has also been suggested, where a transmembrane component allows substrates to be directly secreted, without making contact in the periplasm (Houben *et al.* 2012; Houben *et al.* 2014; Gröschel *et al.* 2016). OM, Outer membrane; AG, Arabinogalactin; PG, Peptidoglycan; CM, Cytoplasmic membrane. Figure was produced using BioRender (https://biorender.com/).

https://scholar.sun.ac.za

2.3.1 ESX Secretion and Target Sequences

Several studies over the past two decades have expanded our understanding of the molecular mechanisms that underpin the function of the ESX secretion systems in mycobacteria. Early research identified the EsxA/ESAT-6 (Early Secreted Antigenic Target 6 kDa) and EsxB/CFP-10 (Culture Filtrate Protein 10 kDa) proteins as substrates of the ESX-1 secretion system in *M. tuberculosis* (Sorensen *et al.* 1995; Wards *et al.* 2000). EsxA and EsxB were shown to be interdependent for secretion through ESX-1, also relying on Rv3870 and Rv3871, believed to act as ATPases to facilitate the transport of the substrates across the membrane (Renshaw *et al.* 2002; Stanley *et al.* 2003). Seven amino acids at the C-terminal end of EsxB are required for the secretion of both EsxB and EsxA, acting as a likely secretion signal (Champion *et al.* 2006).

Some ESX substrates, such as EsxA/B, EspA/C and specific PE/PPE proteins, have been shown to secrete as heterodimers (**Figure 2.3.2**). These form anti-parallel, α-helical bundles containing a flexible WxG (Tryptophan-x-Glycine, where "x" represents a small amino acid) motif separating the two helices, acting as a general secretion signal, with a C-terminal ESX secretion signal in one of the protein partners (Renshaw *et al.* 2005; Ilghari *et al.* 2011). Further alignment of PE protein C-terminal tails revealed a highly conserved C-terminal tyrosine residue separated from a negative amino acid (D/E) by three other residues (YxxxD/E), which is homologous in both EsxB and Esp (ESX secretion-associated proteins) substrates (Daleke *et al.* 2012a). Mutagenesis showed that this signal motif, including the exact spacing between the conserved residues, is essential for the secretion of substrates through the ESX system. However, this signal was shown to not confer any system specificity when the sequences from the LipY (ESX-1 dependent) and PE25 (ESX-5 dependent) substrates where exchanged, as secretion continued down their respective pathways (Daleke *et al.* 2012a). Further study has however shown that system specificity may be linked to pathway specific chaperones, at least for some ESX dependent proteins (Damen *et al.* 2020; Williamson *et al.* 2020).

EspG proteins are ESX specific chaperones that assist in maintaining the solubility of their substrates, for both $EspG_1$ and $EspG_5$, specific to ESX-1 and -5 respectively (Daleke *et al.* 2012b; Korotkova *et al.* 2014). Exchanging the hydrophobic binding domain of $EspG_1$ in the ESX-1 dependent PPE68 with the $EspG_5$ domain from PPE18, successfully redirected the PE35/PPE68 heterodimer from ESX-1 to ESX-5 dependent secretion (Phan *et al.* 2017). This method of determining system specificity of proteins secreted through the ESX pathways may however not be universal, especially for ESX-1 substrates (Phan *et al.* 2017). Some Esp heterodimers themselves rely on Esp chaperones to guide them to the ESX-1 secretion system,

https://scholar.sun.ac.za

the primary example of this being the EspA/C dimer requiring EspD as a chaperone for secretion (Chen *et al.* 2012).

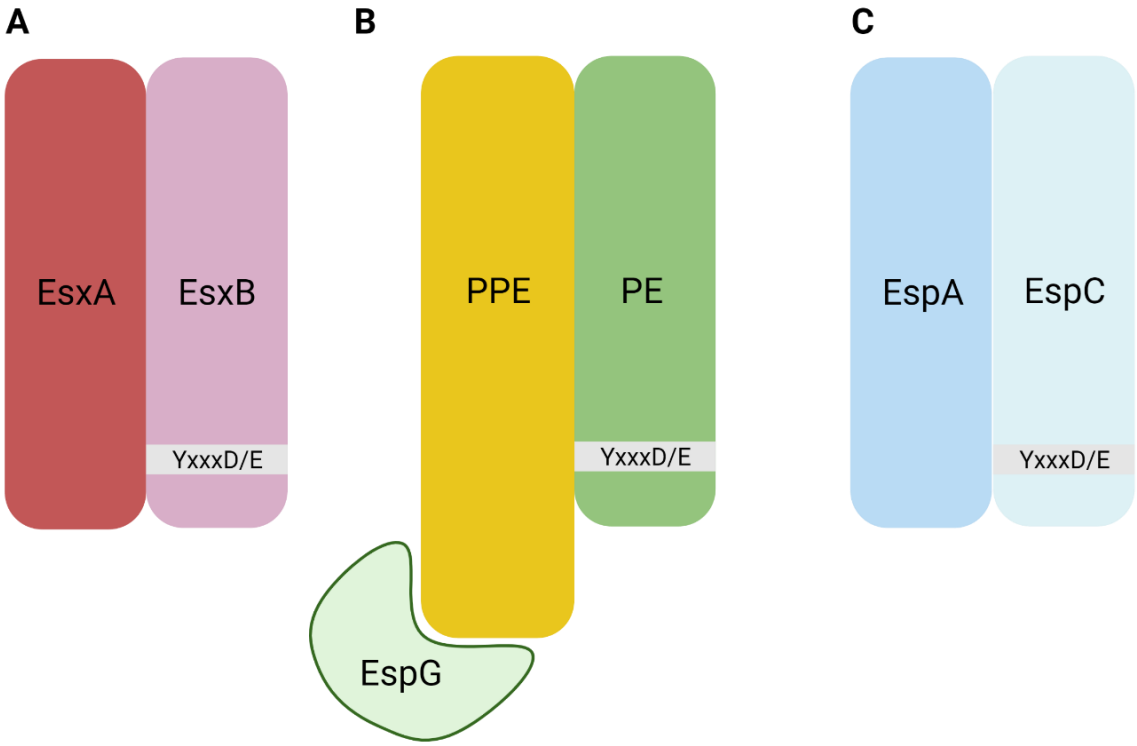


Figure 2.3.8: ESX substrate pairs. EsxA/B **(A)**, EspA/C **(C)** and some PE/PPE substrates **(B)** such as PE25/PPE41; have all been shown to be exported in an Esx dependent manner as semi-folded heterodimers. The General YxxxD/E secretion motif is always present in one of the substrate pairs, such as EsxB in EsxA/B, EspC in EspA/C and all PE proteins in PE/PPE pairs. PPE proteins have also been shown to likely confer Esx system specificity during secretion, due to containing a hydrophobic EspG binding domain, which escorts substrate pairs to their cognate Esx system (Ekiert *et al.* 2014; Phan and Houben 2018). Figure was produced using BioRender (https://biorender.com/).

https://scholar.sun.ac.za

2.4 Methods to Identify Secretion Pathway Components, Signals and Substrates

Numerous approaches have been applied to study the interaction of proteins and the components that allow them to be secreted and localised to specific regions in mycobacteria. Several of these approaches, including protein crystallography, mutagenesis combined with protein precipitation and mass-spectrometry, as well as more bioinformatic methods, have all become standard procedures (McDonough *et al.* 2005; Yang *et al.* 2016; Beckham *et al.* 2017; Li *et al.* 2019c). These methods are rarely used in isolation, more commonly being combined in tandem to produce more complete images of the interactions between proteins and what effects alterations in specific components may carry for the whole organism. This section will discuss how these methods have been used previously, while highlighting their benefits and use, alongside their limitations.

2.4.1 Structural Analysis of Recognition Sequences

X-ray crystallography has been utilised since the 1980s to view molecules crystallised from solution and visualise their molecular structure (Su *et al.* 2015). It has seen use in protein studies that aim to capture the folded state of proteins, where the images could help identify key amino acid domains that act as binding sites or interaction points between proteins. Nuclear Magnetic Resonance (NMR) spectroscopy alongside other electron microscopic (EM) techniques have also seen increasing use in studies that aim to expose the structure of protein molecules and identify sites that could be promising for protein-substrate interactions (Miao *et al.* 2015; Su *et al.* 2015; Zheng *et al.* 2015). General secretion chaperones such as GroEL, DnaK and TF have been extensively studied utilizing crystallographic methods in *E. coli* (Zhu *et al.* 1996; Rüdiger *et al.* 1997; Xu *et al.* 1997; Fujiwara *et al.* 2010). Similar approaches have been used in the homologues of these general chaperones specific to mycobacteria, but to a lesser extent (Qamra and Mande 2004).

Previously a Multiwavelength Anomalous Diffraction (MAD) approach was used to identify the hydrophobic binding pocket of DnaK in *E. coli*, which appeared to have a high affinity for leucine (Zhu *et al.* 1996). Crystallography, allowed the identification of the apical domain of GroEL in *E. coli* as the entryway into its cavity, where substrates enriched in α/β domains that have highly hydrophobic surfaces can become bound and enter the chaperone cavity to undergo folding (Xu *et al.* 1997; Fujiwara *et al.* 2010). In *M. tuberculosis*, crystallography was able to demonstrate the structure and binding domains of the two distinct GroEL paralogs (GroEL1

https://scholar.sun.ac.za

and GroEL2), that each appeared to function at lower oligomeric states during expression (Qamra *et al.* 2004; Qamra and Mande 2004; Sielaff *et al.* 2011). Further study has identified that they are able to oligomerize into relatively unstable tetradecameric structures when expressed in native cellular conditions (Fan *et al.* 2012). The structure of individual components that make up the general secretory pathways found in *M. tuberculosis* have also been explored with both crystallography and NMR (Lu *et al.* 2016; Swanson *et al.* 2016).

SecA1 and SecA2, SecA homologues within *M. tuberculosis*, have been compared by crystallography to identify structural similarities and deviations that effect the way the components interact with the SecYEG channel, while attempts have also been made to crystallise the SecB-like chaperone (Rv1957) of *M. tuberculosis* (Sharma *et al.* 2003; Lu *et al.* 2016; Swanson *et al.* 2016). The structure of the TatABC complex has yet to be explored in mycobacteria, although crystallisation and NMR in *E. coli* have revealed the core conformation and plasticity of the complex, which allows it to bind to the Tat-recognition sequence and interact with substrates (Rollauer *et al.* 2012; Ramasamy *et al.* 2013; Zhang *et al.* 2014). Further structural analysis has however been carried out on the secretion systems specific to mycobacteria and the Type VII secretion systems found in *M. tuberculosis* to determine specific sites of interaction that may confer system specificity for ESX associated substrates (Beckham *et al.* 2017; Lou *et al.* 2017; Phan *et al.* 2017). Crystallography has previously been utilised to explore the higher order structures and binding of protein substrates into their cognate ESX secretory complexes, such as the case for the EspB proteins associated with the ESX-1 pathway (Solomonson *et al.* 2015). It has also been used to visualise the structures of more specific ESX-5 associated substrates within *M. tuberculosis*, showing the broad conserved region of interaction between the EspG chaperone and its target PE-PPE protein heterodimer (Ekiert *et al.* 2014a). This acted as a recent clue to the suggested role of EspG as the primary guide responsible for directing specific PE-PPE (or single PPE proteins) to their cognate ESX secretion system (Ekiert *et al.* 2014a).

High resolution electron microscopy has also been utilised to explore the binding and polymerisation of the ESX-1 associated EspC substrate into a more complex filamentous structure, which contributes to the secretion of other associated substrates (Lou *et al.* 2017). Furthermore, single-particle, transmission electron microscopy (TEM), alongside more conventional crystallography, has been utilised to explore the overall ESX-5 secretion complex in *Mycobacterium xenopi*, which revealed a model structure of the complex and protein binding domains, as well as secretion channels (Zhang *et al.* 2015; Beckham *et al.* 2017).

https://scholar.sun.ac.za

In general, crystallography and electron microscopy are reliable methods that have been extensively utilised to expand our understanding of the base structure and dynamics involved during protein binding and interactions (Zhu *et al.* 1996; Rüdiger *et al.* 1997; Qamra *et al.* 2004; Solomonson *et al.* 2015; Swanson *et al.* 2016). Despite its usefulness, the upper limits of classical crystallography are being reached as more complex and smaller scale protein structures are targeted for research (Martin-Garcia *et al.* 2016; Srivastava *et al.* 2018). Proteins that are naturally disordered or membrane associated have also been notoriously difficult to visualise, leading to significant gaps in information on available reference databases (Oldfield and Dunker 2014; Djinovic and Carugo 2015). In response to this, several new supplementary approaches have arisen to expand the capabilities of, or even replace, crystallography. Development of higher energy X-ray diffraction in combination with more sophisticated computational modelling, has reduced sample damage while improving the quality of low-resolution data (DiMaio *et al.* 2013; Su *et al.* 2015). Cryo-cooling during crystallography has been established as a standard approach to minimise the damage or changes brought on by extensive exposure of protein structures to radiation during the imaging process (Garman and Owen 2006). However, extensive sub-zero cooling of protein crystals can cause significant conformational shifts away from the native fold and shape that some proteins adopt at physiological temperatures (Tilton *et al.* 1992; Halle 2004; Sierra *et al.* 2016). The development of X-ray Free-Electron Laser (XFEL) sources, which can be utilised to generate electron density maps by quick and high-energy blasts on protein structures in femtosecond crystallography, provide the opportunity to carry out these experiments at various temperatures (Martin-Garcia *et al.* 2016). This method has also been explored for use on un-crystallised proteins in their native states (Miyashita and Joti 2017). Beyond these enhancements, several complementary methods may be used instead of standard crystallography. Both NMR and cryo-electron transmission microscopy are valuable tools to supplement the shortcomings of low-resolution crystallography, with multiple data fitting models being developed to overlay results and capture atomic scale images of protein structures (Kudryashev *et al.* 2015; Kim and Sanbonmatsu 2017).

2.4.2 Gene Cloning and Mutagenesis Based Approaches

The genetic manipulation of genes to alter specific protein regions, and the expression of proteins in non-native systems, have been used to study recognition and translocation signals. These approaches refined several concepts surrounding the general secretory pathways (Sec, Tat) observed in prokaryotes (Wandersman 1993; Stanley *et al.* 2000). Similar approaches have

https://scholar.sun.ac.za

been utilised to determine the components of these secretory systems important to the translocation of proteins in mycobacteria (Braunstein *et al.* 2003; McDonough *et al.* 2008; Chilukoti *et al.* 2016) and more specifically in trafficking proteins through the Type VII, ESX secretion systems of *M. tuberculosis*.

Mutations were utilised to determine the putative C-terminal recognition sequence for the ESX-1 secretion system (Champion *et al.* 2006). This motif however did not confer system specificity, but as previously discussed, the exchange of the EspG recognition sequences between substrates of the ESX-1 and ESX-5 systems rerouted secretion (Phan *et al.* 2017). It has also highlighted that the EspG chaperone itself is not able to interact with the ESX complex, but the exchange of the recognition sequences in the PE35/PPE68 heterodimer, was able to reroute the secretion of the EsxA/EsxB pair from ESX-1 to ESX-5 secretion (Damen *et al.* 2020). Truncation of the ESX-1-associated EspC, C-terminal domain identified it as an essential portion of the protein. Without the full domain, the protein was unable to form filamentous structures that allow for the function of the ESX-1 system, leading to the attenuation of *M. tuberculosis* (Lou *et al.* 2017). Few direct limitations exist to utilising these approaches when studying biological systems. However, it has been noted that in some instances, altering or removing certain secretory pathways or partners may result in a lethal phenotype for bacteria being studied, although this may also serve to highlight the importance of specific proteins and domains (Luirink and Sinning 2004; Burg-Golani *et al.* 2013; Ates *et al.* 2016). The expression of mutant genes in non-native hosts may also produce conflicting results, where the observed functions or protein structures cannot be directly compared to those in the original organism, which must be considered when drawing conclusions (Ates *et al.* 2016).

2.4.3 Sequence Identification by Bioinformatics

Numerous protocols and web-services have been developed and refined to identify proteins that are likely secreted and thus valuable targets for further focussed study, based on the presence of select sequence features (Leversen *et al.* 2009; Zhu *et al.* 2015; Yang *et al.* 2016). Various models have been designed to represent protein sequence information that may be targeted for classification, including amino acid composition (AAC), functional domain composition and wavelet analysis models (Cai *et al.* 2003; Yang *et al.* 2007; Rezaei *et al.* 2008). This has also been applied in the identification of mycobacterial membrane associated or soluble proteins, by utilising algorithms that exploit over-represented tripeptides, unbiased di-peptide compositions (Unb-DPC) or evolutionary position matrixes (Bi-PSSM), among

https://scholar.sun.ac.za

many others, to classify proteins (Ding *et al.* 2012; Khan *et al.* 2017a; Khan *et al.* 2017b). These approaches have shown value in their ability to narrow lists of proteins that could likely be secreted, or otherwise be soluble or membrane associated (Ahmed *et al.* 2018b).

This can be especially useful for suggesting potential roles of target proteins in *M. tuberculosis*, and by extension their likely sub-cellular localization (Lamichhane *et al.* 2003; Mazandu and Mulder 2012). Ultimately, however, to confirm the secretion of any protein and the pathway that may be utilised based on these motifs, hands-on experimental approaches will likely still be required in the foreseeable future. Previous studies have shown that many widely available secretion pathway detection proGrams may not reliably detect proteins that are secreted (McDonough *et al.* 2008). Diverse proteins such as the PE/PPE families found in *M. tuberculosis* have commonly been excluded from bioinformatic pipelines due to sequence repetition, high GC-content, gene duplications and limited trypsin cleavage sites (Schubert *et al.* 2013; Copin *et al.* 2014; Meehan *et al.* 2019). These barriers can limit the reliability of predictive software, but many of these techniques are under constant improvement and the development of more robust reference genomes may allow for more consistent prediction of secretion signals in the near future (Maciuca *et al.* 2016; Yang *et al.* 2016).

https://scholar.sun.ac.za

Chapter 3

Localisation of PPE_MPTR Proteins in *Mycobacterium smegmatis*

3.1 Introduction

The PE/PPE protein families have undergone significant genetic expansion in modern lineages of *Mycobacterium tuberculosis* (Cole *et al.* 1998). Studies have implicated members of both families in the survival and virulence of infectious mycobacteria (Daim *et al.* 2011; Ates *et al.* 2018b). Several family members have not yet undergone targeted studies, leaving their potential biological roles unknown. Fundamental questions surrounding numerous PPE_MPTR proteins remain, especially with regards to their mechanism of secretion and final localisation (Abdallah *et al.* 2009; Ekiert *et al.* 2014b).

Sub-cellular localisation can have direct correlation with the role of a protein (Sampson 2011). PPE_MPTR proteins predicted to contribute to host-pathogen interactions or immune evasion may be associated with the outer membrane, or otherwise secreted into host cells (Sampson *et al.* 2001; Wolfe *et al.* 2010; Damen *et al.* 2020). PE_PGRS/PPE_MPTR proteins are also associated with the ESX secretion systems (Abdallah *et al.* 2009; Ates *et al.* 2016). Modes of PE/PPE targeting and specificity towards distinct ESX systems are still under investigation (Damen *et al.* 2020; Williamson *et al.* 2020). Studying the localisation of the PPE_MPTR subfamily will advance our understanding of the functional roles of this protein sub-family.

We aimed to identify the sub-cellular localisation of targeted PPE_MPTR proteins labelled with a green fluorescent protein (GFP) tag. Fluorescently tagged proteins would be expressed in various strains of *Mycobacterium smegmatis*, including mutants lacking native ESX secretion systems, as well as a strain encoding the non-native ESX-5 from *Mycobacterium xenopi* (ΔESX-1, ΔESX-3, ΔESX-4, ESX-5_{Xenopi}). The goal was to apply a combination of fluorescent imaging and correlative-light electron microscopy (CLEM) to determine the sub-cellular localisation of the PPE_MPTR proteins in these bacterial strains. Localisation of these PPE_MPTR proteins to the cell surface may suggest that these proteins might be active at the host-pathogen interface.

https://scholar.sun.ac.za

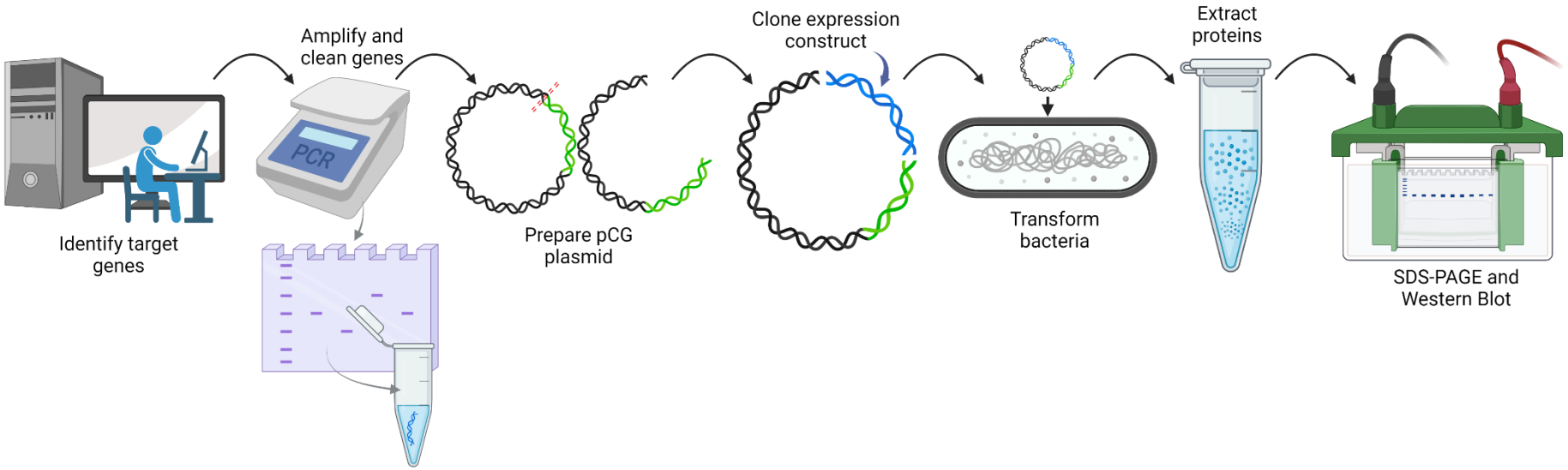


Figure 3.1.2: Simplified workflow for localisation of PPE_MPTR proteins. The steps followed to clone and express GFP tagged PPE_MPTR proteins included: Identify genes of interest and design primers to amplify them from *M. tuberculosis* H37Rv DNA; Separate amplicons through gel electrophoresis, excise and purify; Linearise the pCG plasmid (the backbone for expression constructs) by BamHI and MluI double digestion; Ligate inserts into the linearised plasmid and cloned in *Escherichia coli*; Screen clones to confirm plasmid integrity, then transform into *M. smegmatis* for constitutive protein expression; Prepare whole-cell lysate protein extracts from cultures and separated by SDS-PAGE; Stain or Western blot resulting gels to confirm GFP tagged protein expression. Figure was produced using BioRender (https://biorender.com/).

https://scholar.sun.ac.za

3.2 Methods and Materials

3.2.1 Candidate PPE_MPTR Protein Selection

A total of 69 PPE proteins have been previously identified, 22 of which were PPE_MPTR proteins that were considered for this study. Six candidate proteins were selected from this list to be cloned and expressed (**Table 3.2.1**). Candidate proteins were selected using available literature and unpublished data to identify proteins that may perform crucial functions during *M. tuberculosis* infections. Preference was given to proteins that had been previously studied within the Host-Pathogen Mycobactomics (HPM) research group, either in the context of protein secretion (PPE_MPTR10, -40 and -62) or host immune stimulation (PPE_MPTR42).

3.2.2 Bacterial Strains

E. coli XL-1 Blue (Agilent Technologies, Santa Clara, U.S.) was used for molecular cloning and plasmid propagation. *M smegmatis* mc^2155 (ATCC 700084) was used for the expression and localization of pCG_MPTR constructs. Various mutant *M. smegmatis* strains were also considered for the expression and localisation of pCG_MPTR proteins to determine the effect of the ESX secretion systems on sub-cellular localisation. All bacterial strains are listed in **Table 3.2.2**

3.2.3 Preparing Competent Bacteria for Transformations

Chemically competent *E. coli* XL-1 Blue was prepared by picking a single colony from Luria-Bertani (LB) agar plates, which were streaked with bacteria from available freezer stocks and incubated at 37°C overnight. The colony was inoculated into 10 ml LB broth and cultured overnight at 37°C with shaking at 180 rpm, before being sub-cultured (1:50) into 50 ml LB broth and grown until an OD_{600} of 0.6. This culture was centrifuged at 4000 rpm for 10 min at 4°C, supernatant discarded and the resulting bacterial pellet resuspended in 10 ml ice-cold $MgCl_2$ [100 mM], followed by incubation on ice for 30 minutes. The bacteria were pelleted as before, then resuspended in ice-cold $CaCl_2$ [100 mM], supplemented with 15% glycerol. From this resuspension, aliquots were prepared and stored at -80°C for further use.

https://scholar.sun.ac.za

Table 3.2.5: PPE_MPTR proteins selected in the study alongside available information on expression and biological relevance.

Gene	Protein	Gene Length (bp)	Protein Length (aa)	Molecular Weight (kDa)	Reported as Essential (Y/N)	Criteria for Selection
Rv0442c	PPE_MPTR10	1464	487	47.3	N[3]	• Previously expressed in *M. smegmatis*[5]. • Upregulated in *M. tuberculosis* infected alveolar macrophages (Rachman *et al.*, 2006). • Predicted outer membrane localisation (Gey Van Pittius *et al.*, 2006; Ates, 2019). • Potential role in arresting phagosomal acidification (Stewart *et al.*, 2005). • Significantly modulates host immune response during infections (Asaad, *et al.* 2021)
Rv1753c	PPE24	3162	1053	104.7	Y[2,3]	• Previously expressed in *M. smegmatis*[5]. • Reported to promote mycobacterial survival in macrophages (Sassetti *et al.*, 2003). • Carries Glycine-, Asparagine-rich regions interrupted by unique, near perfect 26 aa repeats (Sampson *et al.*, 2001).
Rv2356c	PPE_MPTR40	1848	615	58.4	Y[3]	• Constitutively expressed during infection in mice (Soldini *et al.*, 2011).
Rv2608	PPE_MPTR42	1743	580	59.7	N[3,4]	• Previously expressed in *M. smegmatis*[5]. • Elicits a hyper-stimulatory reaction during macrophage infection (Bertholet *et al.*, 2010). • PPE_MPTR candidate in polyprotein vaccines (Delogu *et al.*, 2017).
Rv3159c	PPE_MPTR53	1773	590	56.6	N[3,4]	• Previously expressed in *M. smegmatis*[5]. • Significantly increased expression under stress conditions (Soldini *et al.*, 2011). • Downregulated after extended periods of starvation (Betts *et al.*, 2002). • Predicted membrane localisation (Song *et al.*, 2008).
Rv3533c	PPE_MPTR62	1749	582	55.5	N[4]	• Previously expressed in *M. smegmatis*[5]. • Showed to be differentially expressed in alveolar macrophages infected in *M. tuberculosis* (Rachman *et al.*, 2006).

https://scholar.sun.ac.za

	• Constitutively expressed during mouse infection (Soldini *et al.*, 2011). • Reported to act as a heme-binding protein (Mitra *et al.*, 2017).

1: *Lamichhane et al., 2003*
2: *Sassetti et al., 2003*
3: *Griffin et al., 2011*
4: *Dejesus et al., 2017*
5: *Previous work from HPM Laboratory*

Table 3.2.6: Bacterial strains used to express and determine the localization of recombinant PPE_MPTR proteins.

Parent Strain	Plasmid/Strains	Strain Description	Described by
E. coli	XL-1 Blue	Host strain for routine cloning applications using plasmid or lambda vectors.	Agilent Technologies
M. smegmatis	mc^2155	Fast growing and non-pathogenic strain with efficient plasmid-transformation ability.	(Snapper, Melton, Mustafa, Kieser & Jr, 1990)
M. smegmatis	ΔESX-1	Lacks the ESX-1 secretion system.	N. Steyn (Hons); M. Newton-Foot (PhD)
M. smegmatis	ΔESX-3	Lacks the ESX-3 secretion system.	M. Newton-Foot (MSc)
M. smegmatis	ΔESX-4	Lacks the ESX-4 secretion system.	M. Smit (MSc)
M. smegmatis	ΔESX-1; ΔESX-3	Lacks both the ESX-1 and ESX-3 secretion systems.	M. Newton-Foot (PhD)
M. smegmatis	ΔESX-1; ΔESX-4	Lacks both the ESX-1 and ESX-4 secretion systems.	M. Newton-Foot (PhD)
M. smegmatis	ΔESX-3; ΔESX-4	Lacks both the ESX-3 and ESX-4 secretion systems.	N. Steyn (Hons), M. Newton-Foot (PhD)
M. smegmatis	ΔESX-1; ΔESX-3; ΔESX-4	Lacks the ESX-1, -3 and -4 secretion systems.	M. Newton-Foot (PhD)
M. smegmatis	$ESX\text{-}5_{xenopi}$	Encodes the ESX-5 secretion system, which does not occur naturally in the strain.	Beckam *et al.*, 2017

https://scholar.sun.ac.za

Electrocompetent *M. smegmatis* mc^2155 was prepared by streaking bacterial stocks on LB agar and incubating at 37°C for three days. A single colony was inoculated into 10 ml 7H9 (Merck, Darmstadt, Germany) supplemented with 10% OADC (oleic acid-albumin-dextrose-catalase) (BD, Franklin Lakes, U.S.), 0.05% Tween-80 and 0.2% glycerol (7H9-Complete). These cultures were incubated at 37°C with shaking at 180 rpm for three days, then sub-cultured into 50 ml 7H9-Complete and allowed to grow overnight at the same parameters, until an OD_{600} of 0.5-1.0. The culture was centrifuged at 4000 rpm for 15 min at 4°C and the bacterial pellet washed in an equal volume of ice-cold glycerol (10%). The resuspended bacteria were centrifuged as before and the resulting pellet resuspended in 3 ml of ice-cold glycerol (10%). Aliquots of the final resuspension were prepared and stored at -80°C for future use.

3.2.4 Amplification of Target Genes

The *ppe_mptr* genes were amplified from *M. tuberculosis* (H37Rv) genomic DNA using the primers in **Table 3.2.3**. Stocks of H37Rv DNA used during this project were prepared by the CTAB (cetyltrimethylammonium bromide) approach (Dr. Nastassja Kriel). Primers were designed utilizing the TakaRa Bio (Kusatsu, Shiga, Japan) In-Fusion Cloning Primer Design Tool (https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools). Primers included the restriction enzyme recognition sequences for BamHI (GGATCC) and MluI (ACGCGT). These restriction enzymes were also used to linearise the pCG plasmid vector, to ensure overlapping compatible cut sites for downstream cloning processes. All constructs were checked during the design process to ensure that the recombinant protein will be in-frame with the encoded green fluorescent protein (GFP) tag. PCR amplifications were done in 50 µl reactions, utilizing the Phusion Hot Start, Flex 2x Master Mix (New England Biolabs, Ipswich, U.S.) according to the manufacturer's instructions. Reactions were prepared with a final primer and master mix concentration of 0.5 µM and 1x, respectively, with the addition of 3% DMSO. The PCR reactions were carried out with initial denaturation at 98°C for 30 seconds before 25 cycles of amplification, with annealing at 64°C for 30 seconds, followed by extension at 72°C for 30 seconds, after which final elongation was carried out at 72°C for 10 min. Resulting samples were kept at 4°C, before being separated on agarose gels. Samples for which amplicons of the correct size were obtained were subsequently purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, U.S.), according to the manufacturer's instructions.

https://scholar.sun.ac.za

Table 3.2.7: Oligonucleotide primer sequences used in the PCR amplification of *ppe_mptr* genes from *M. tuberculosis* H37Rv DNA for cloning into the pCG vector.

Gene	Protein	Primers (3'-5')		Primer Tm (°C)	Amplicon Size (bp)
Rv0442c	PPE_MPTR10	**Forward**	TATACCATGGGGATCCGTGACAAGCCCGCATTTTGC	75.1	1467
		Reverse	AACCGGCGGAACGCGTCTCCGAACCGACCGGCTG	82.0	
Rv1753c	PPE24	**Forward**	TATACCATGGGGATCCATGAATTTTTCTGTACTGCCGCCG	74.2	3165
		Reverse	AACCGGCGGAACGCGTGGACGTGGTGCCTTGGAAG	80.9	
Rv2356c	PPE_MPTR40	**Forward**	TATACCATGGGGATCCGTGGTGAATTTTTCGGTGTTGCC	75.5	1857
		Reverse	AACCGGCGGAACGCGTCCTGAAGAATCCCGAAAGTCC	78.5	
Rv2608	PPE_MPTR42	**Forward**	TATACCATGGGGATCCATGAATTTCGCCGTTTTGCCG	74.2	1752
		Reverse	AACCGGCGGAACGCGTGAAAAGTCGGGGTAGCGCC	80.8	
Rv3159c	PPE_MPTR53	**Forward**	TATACCATGGGGATCCATGAATTATTCGGTGTTGCCGCC	75.2	1782
		Reverse	AACCGGCGGAACGCGTGGGCTGACCGAAGAAGCCC	82.0	
Rv3533c	PPE_MPTR62	**Forward**	TATACCATGGGGATCCATGAACTATGCGGTATTGCC	71.4	1758
		Reverse	AACCGGCGGAACGCGTGTTCCCGAACCCCGACCG	82.3	

https://scholar.sun.ac.za

3.2.5 Agarose Gel Electrophoresis

Gel electrophoresis was carried out to separate and visualize PCR products and enzymatic DNA digestions. Gels were composed of 1% agarose containing 0.3X EZ-Vision Bluelight 10 000X DNA dye (VWR Life Sciences, Radnor, U.S.) and were separated in 1x TAE [40 mM Tris base, 20 mM Glacial Acetic Acid, 1 mM EDTA pH 8.0] at 120V for 1.5-2 hours. DNA samples were run alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher, Waltham, U.S.) to determine fragment sizes and visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad, Hercules, U.S.).

3.2.6 Plasmid Preparation

The pCG plasmid was selected as expression vector for PPE_MPTR-GFP fusion proteins. This plasmid is a modification on the pSE100 vector, which has a backbone of 5538 bp, containing the pse100 promoter and encodes GFP, conferring hygromycin resistance (Dr. Nastassja Kriel, unpublished). The plasmid was transformed into *E. coli* XL-1 Blue by heat shock at 42°C for 90 seconds, followed by plating onto LB agar plates containing 150 µg/ml of hygromycin. Plates were incubated at 37°C overnight, after which single colonies were picked and inoculated into liquid LB media containing the appropriate antibiotic and incubated overnight at 37°C, with shaking at 180 rpm. Cells were harvested the next day, the plasmid DNA extracted and purified according to instructions using the Wizard DNA Miniprep kit (Promega). Plasmid DNA was quantified using the Nanodrop mass spectrometer (Thermo Fisher). Circular plasmids were linearised by enzymatic digest in 10 µl reactions containing 3 µg of plasmid DNA, 1x CutSmart Buffer, 20 U BamHI and 20 U MluI (NEB). The reactions were carried out at 37°C for a total of 4 hours. Digestions were confirmed by gel electrophoresis (**Section 3.2.5**). Visualized bands which corresponded to linear pCG were excised from the gel and purified utilizing the Wizard SV Gel and PCR Clean-Up System (Promega), according to manufacturer instructions.

https://scholar.sun.ac.za

3.2.7 Cloning Expression Constructs

Amplified *ppe_mptr* genes of interest (**Table 3.2.1**) were cloned into the pCG plasmid to generate expression constructs. Genes were cloned into pCG between the BamHI and MluI cut sites, upstream and in-frame of the GFP coding sequence separated by a linker. The linker was composed of eight amino acids (Ser-Ala-Gly-Ser-Ala-Gly-Ser-Ala). Insertion of the *ppe_mptr* genes into the linear plasmid was performed using the In-Fusion HD Cloning kit (Takara Bio) according to the manufacturer instructions. *E. coli* XL-1 Blue was transformed with 5 µl of the in-Fusion reaction. Transformants were plated onto LB agar plates containing 150 µg/ml of hygromycin and allowed to grow overnight at 37°C. Resulting colonies were inoculated into LB broth containing 150 µg/ml of hygromycin and incubated overnight at 37°C with shaking at 180 rpm. Recombinant plasmids were isolated and purified from the bacterial cultures using the Wizard Plus SV Miniprep DNA Purification System (Promega) according to manufacturer's instructions.

3.2.8 Confirming Recombinant Clones

PCR reactions were done on prospective clone samples utilising the cloning primers shown in **Table 3.2.3** (pCG_FP1 and pCG_RP1) to screen for the expected gene inserts. These PCR reactions were set up in a 25 µl volume, using the Phusion Hot Start, Flex 2x Master Mix (NEB), as described by the manufacturer. Reactions were carried out with initial denaturation at 98°C for 30 seconds followed by 20 cycles of amplification, with annealing at 64°C for 30 seconds, extension at 72°C for 30 seconds, with final elongation carried out at 72°C for 10 min. The PCR amplicons were separated by gel electrophoresis and positive samples selected. Selected samples, in which PCR confirmed the presence of the genes, were subjected to 10 µl BamHI and MluI enzymatic digests (**described in section 3.2.6)**. Digests were separated by gel electrophoresis (**section 3.2.5**) to confirm the presence of inserts within the vector. pCG_PPE-MPTR10 clones that showed the presence of the correctly sized DNA fragment in the correct position within the vector were sent for Sanger sequencing (Inqaba Biotechnical, Pretoria, South Africa). External backbone and internal sequencing primers (**Table 3.2.4)** were used to generate sequence reads to determine sequence integrity of the insert and vector. All forward sequencing primers were paired with the pCG_RP1, reverse primer, for pCG plasmid sequencing.

https://scholar.sun.ac.za

Table 3.2.8: Oligonucleotide primers used for the Sanger sequencing of pCG *ppe_mptr* gene inserts. The external pCG_FP1 backbone primer, as well as all internal sequencing primers were paired with the pCG_RP1 primers to produce sequence reads.

Target	Primers (5'-3')		Length (bp)	Tm (°C)
pCG	**pCG_RP1**	ATCCTGTTGACGAGGGTGTC	20	62
ppe_mptr10	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE10_FP2	GCTCATGGACGTCGAGGC GGCATA	24	62.5
	PPE10_FP3	GCATCGGGACCTCGGGGA CAAT	22	60.4
ppe24	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE24_FP2	GTCGGCGATGTCTGCCTACCAT	22	58.6
	PPE24_FP3	CCAGATCGGAATCGGCGGGCTCAA	24	62.5
	PPE24_FP4	CCCGCCTTCAGTCTGCCGGCAAT	23	62.4
	PPE24_FP5	GCTGAGTATTCCTTCCGTAGCCAT	24	57.4
	PPE24_FP6	CCCGGAGCTAACCATCAACTCGAT	24	59.1
ppe_mptr40	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE40_FP2	GATTGCCGCTGCTGAGGCCACCTA	24	62.5
	PPE40_FP3	GGCGGGAACACCGGCGACTTCAAC	24	64.2
	PPE40_FP4	CAAGCAGGTGCCAACTCGGGCTT	23	60.6
ppe_mptr42	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE42_FP1	CGTGTCTCCAGCGATGGTC	19	62
	PPE42_FP2	GGAGTAACCGCGTTGGTCAT	20	62
	PPE42_FP3	AACGACTTCCCCAAATACCC	20	60
ppe_mptr53	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE53_FP2	GTGGGCTATCACGGCGGGGCAT	22	62.3
	PPE53_FP3	CGGTTTCTTCAACTCTGGCAACAA	24	55.7
	PPE53_FP4	CCGGCGGCTTCAACGTCGGCTT	22	62.3
ppe_mptr62	**pCG_FP1**	GCGGGAGAACTCCCTATCAG	20	64
	PPE62_FP2	CGTTGAGGGCGCCTACGAACAGAT	24	60.8
	PPE62_FP3	CTGGGGCTTCGGCAACAACGGCAT	24	62.5
	PPE62_FP4	GGGTTCCAGAACGGGGGCAGCAA	23	62.4

https://scholar.sun.ac.za

3.2.9 Mycobacterial Transformation and Fluorescent Protein Expression

Electrocompetent *M. smegmatis* mc^2155 was transformed with pCG_PPE-MPTR constructs. *M. smegmatis* mc^2155 stocks were thawed on ice before adding 200 – 1000 ng of DNA and transferred to a pre-chilled 0.2 cm gap Electroporation Cuvette (Bio-Rad, Hercules, U.S). Cells were electroporated at 2.5 Kv, 25 µF and 1000 Ω (Gene Pulser II, Bio-Rad) then suspended in 1 ml of 7H9-Complete. The suspensions were allowed to recover at 37°C with shaking at 180 rpm for three hours. Transformants were plated onto LB agar (Sigma-Aldrich) containing 50 µg/ml of hygromycin and incubated at 37°C for 3 days. Single colonies were inoculated into 15 ml 7H9-Complete, containing 50 µg/ml hygromycin, and incubated at 37°C with shaking at 180 rpm for three days. Positive controls of *M. smegmatis* transformed with the backbone pCG plasmid and negative controls of untransformed *M. smegmatis* were also prepared and cultured as described.

3.2.10 Monitoring Growth of *M. smegmatis* Following Transformation

The fitness of *M. smegmatis* strains was tested by assessing growth of cultures following transformation of the pCG_PPE_MPTR plasmids. Transformation was carried out as described in **Section 3.2.9**, single colonies of untransformed *M. smegmatis* (WT, ΔESX-1, -3, -4 and -1,3,4) and pCG_PPE_MPTR transformed *M. smegmatis* (WT, ΔESX-1, -3, -4 and -1,3,4) were inoculated into 15 ml 7H9-Complete in triplicate and incubated at 37°C for three days, shaking at 180 rpm. The Optical Density at 600 nm (OD_{600}) of each 15 ml culture was measured to inoculate 25 ml 7H9-Complete at a starting OD_{600} of 0.05, for the start of the growth curve. Cultures were incubated at 37°C with shaking at 180 rpm. OD_{600} of each culture was measured at 4-hour intervals for a total of 32 hours. Growth curves were generated by taking the average OD_{600} reading of each triplicate culture set at each time point.

3.2.11 Protein Extraction

Bacterial stocks of all transformed *M. smegmatis* cultures were made by taking an aliquot of liquid cultures to a final volume of 1 ml, containing 25% glycerol, then frozen at -80°C for future culturing and protein expression. Otherwise, transformed cultures grown in 15 ml volumes were pelleted by centrifugation at 5000 rpm for 15 minutes at 4°C. Cell pellets were resuspended in 500 µl of 1x PBS (Phosphate-Buffered Saline) (Gibco, Thermo Fisher) supplemented with protease inhibitor (1 tablet per 10 ml 1x PBS) as directed by manufacturer (cOmplete Protease Inhibitor Cocktail, Roche). Two hundred microlitres of glass beads (0.1 mm, BioSpec, Thermo Fisher) were added to each resuspension, which were beaten for four 20 sec bursts, each followed by 5 min periods of rest on ice (FastPrep-24). The resulting whole-

https://scholar.sun.ac.za

cell lysate was removed from the tube once the glass beads were allowed to settle. Protein lysates were stored at -20°C between experiments.

3.2.12 Bradford Assay to Determine Protein Concentration

Protein concentrations were determined by Bradford Assay. Bovine Serum Albumin (BSA) was diluted in 1x PBS (Thermo Fisher) to create a standard curve ranging in concentration from 0 mg/ml to 2 mg/ml. One in four (1:4) dilutions of samples were prepared in 1x PBS. Samples and BSA standards were loaded in duplicate into clear, flat-bottom, 96 well plates. Two hundred microlitres of Bradford reagent concentrate (Bio-Rad) (diluted 1:5 in distilled H_2O and filtered) was added to each well and plates were incubated in the dark for 15 minutes. The absorbance of the standards and samples was then recorded at 595 nm using a FlUOstar Omega microplate reader (BMG Labtech, Ortenberg, Germany).

3.2.13 Confirming Protein Expression by Western Blot

SDS-PAGE and subsequent blotting were carried out using NuPAGE and XCell II Blotting modules (Invitrogen). NuPAGE, LDS Sample buffer [4x] (Thermo Fisher) was mixed with 10% β-Mercaptoethanol (Sigma-Aldrich) in a 1:1 ratio. From this, 5 µl dye mixture was added to 20 µg of protein lysate prior to boiling at 100°C for 10 min. The entire volume of boiled samples were loaded into the wells of NuPAGE, pre-cast, 4 – 12% Bis-Tris acrylamide gels (Invitrogen). Proteins were separated alongside a PAGE-Ruler, Pre-stain Protein ladder (ThermoFisher) at 200 V for one hour utilizing pre-formulated NuPAGE 1x Running Buffer. Resulting gels were stained using InstantBlue Protein Stain (Sigma-Aldrich). Otherwise, when proceeding to Western Blotting, proteins from the SDS-PAGE gels were transferred onto methanol activated PVDF (Polyvinylidene fluoride) membranes according to manufacturer instructions at 30 V for one hour, utilizing pre-formulated 1x NuPAGE Transfer Buffer (Invitrogen). Membranes were blocked overnight at 4°C in 5% skim milk solution (5% Skim milk powder in 1x TBST [2 mM Tris base, 15 mM NaCl, 0.1% Tween-80]). Once blocked, membranes were washed three times for 10 minutes in 1x TBST and incubated with primary mouse, anti-GFP antibodies (# MA5-15256, Thermo Fisher) (1:2000 in 1x TBST) for one hour at room temperature with shaking. Membranes were then washed three times and incubated with secondary HRP-conjugated, Goat anti-Mouse antibodies (# 62-6520, Thermo Fisher) (1:10000 in 1x TBST), followed by a final three washes. Membranes were sealed in plastic bags and incubated in a 1:1 ratio of Clarity Western Peroxide Reagent and Clarity Western ECL substrate (Bio-Rad), with mild agitation, before being viewed under a GelDoc Imager system (Universal Hood III, 731BR02413; Bio-Rad).

https://scholar.sun.ac.za

3.3 Results

3.3.1 PPE_MPTR Genes Amplified by PCR

Amplicons that matched the size of PPE_MPTR genes of interest (**Table 3.3.1**) were produced by PCR reactions (**Figure 3.3.1**), showing the successful amplification of genes. Non-specific banding patterns were seen in the majority of samples in both the first (**Figure 3.3.1, A**) and the second round of amplifications (**Figure 3.3.1, B**). No non-specific bands were present in the PPE_MPTR42 reaction (**Figure 3.3.1, A**) and amplicons were subsequently cleaned using the Wizard PCR Clean-Up System (Promega). Bands of interest, indicated by red arrows in **Figure 3.3.1 A**, were excised from the gel and cleaned according to the Wizard SV Gel Clean-Up procedure (Promega).

Table 3.3.2: Expected size of target PPE_MPTR genes amplified from *M. tuberculosis* H37Rv.

Gene	Protein	Expected Size (bp)
Rv0442c	PPE_MPTR10	1464
Rv1753c	PPE24	3162
Rv2356c	PPE_MPTR40	1848
Rv2608	PPE_MPTR42	1743
Rv3159c	PPE_MPTR53	1773
Rv3533c	PPE_MPTR62	1749

https://scholar.sun.ac.za

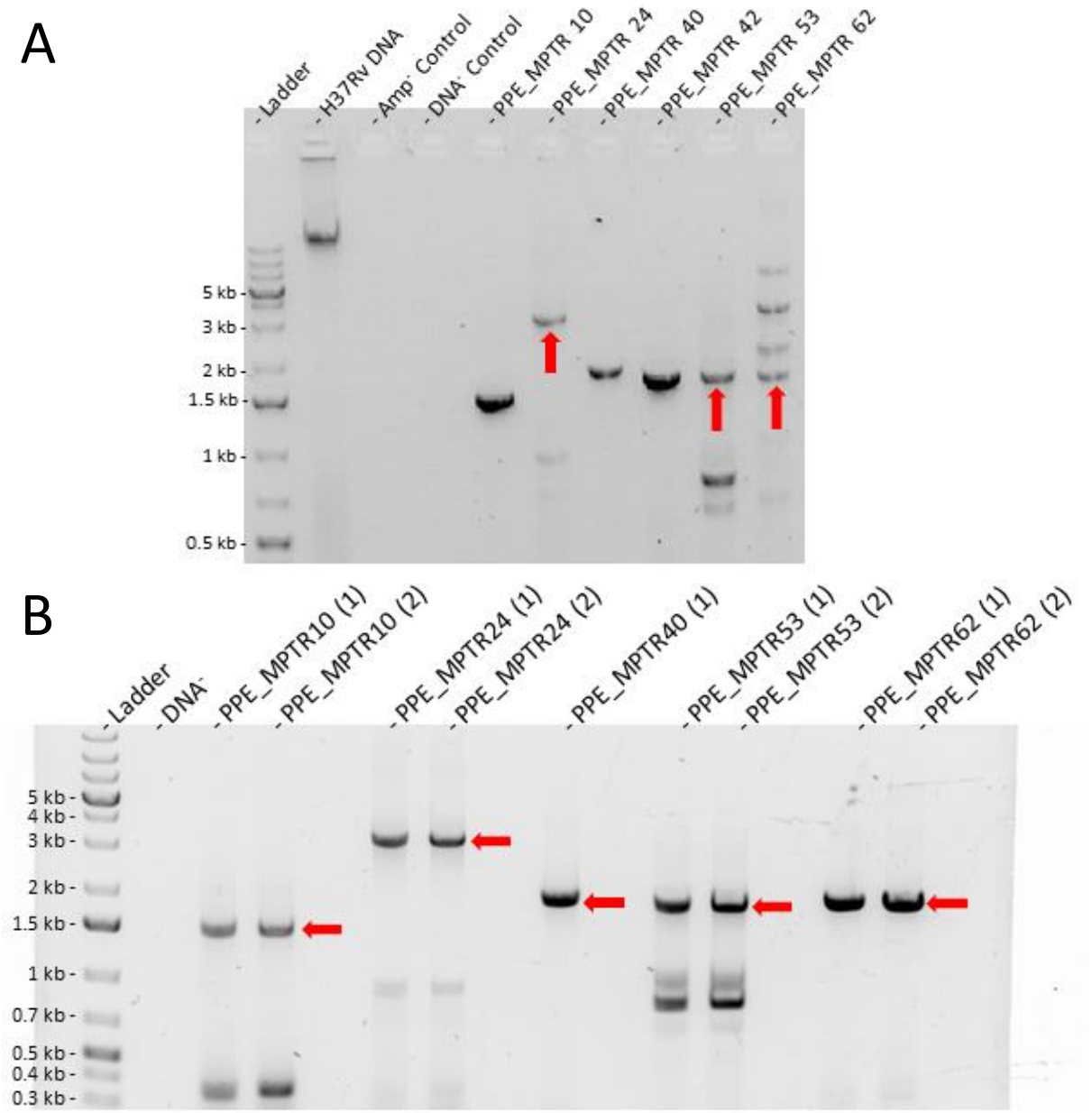


Figure 3.3.7: *ppe_mptr* genes amplified from *M. tuberculosis* H37Rv DNA. Bands corresponding to the expected genes were visible across all reactions. **A** represents the first round of amplifications, while **B** represents a new PCR amplification performed with the same primers following optimization using *M. tuberculosis* H37Rv DNA, where the melting temperature (Tm) was increased. The majority of the reactions had visible non-specific bands following PCR (**A and B**), in which case the expected bands are marked with red arrows. Expected sizes were as follows; PPE_MPTR10 = 1464 bp, PPE_MPTR24 = 3162 bp, PPE_MPTR40 = 1848 bp, PPE_MPTR42 = 1743 bp, PPE_MPTR53 = 1773 bp, PPE_MPTR62 = 1749. The Amp$^-$ control represents a blank reaction which contained 100 ng of H37Rv DNA, but no reagents to catalyse the PCR. DNA$^-$ controls represent PCR samples which contained all reagents and primers, but no H37Rv DNA. The first lane in **A** was loaded with 300 ng of H37Rv DNA as a reference, all other wells were loaded with PCR reactions, each of which only received 100ng of DNA in in 50 µl reactions. Gels were comprised of 1% agarose containing 0.3 EZ-Vision Bluelight DNA dye (VWR Life Sciences), and electrophoresis done in 1x TAE buffer at 120V. Samples were separated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

3.3.2 Screening Samples by PCR

Colonies were present on agar plates following the cloning reactions of pCG with all the amplified *ppe_mptr* genes. The positive control in-Phusion reaction also produced colonies on plates, indicating the competence of the *E. coli*. Negative controls comprised of untransformed *E. coli*, *E. coli* transformed only with linear plasmid, as well as only in-Phusion reagents, produced no colonies. A total of five colonies were selected from each plate for colony screening PCR, which aimed to amplify the target *ppe_mptr* genes from the plasmids to confirm the insert's presence. Following the colony PCR, only pCG_PPE10 amplifications produced any visible PCR products (**Figure 3.3.2**), which were at the expected size of 2325 bp. Plasmid DNA was prepped from all selected clones for another round of screening by enzymatic digest (**Figure 3.3.3**).

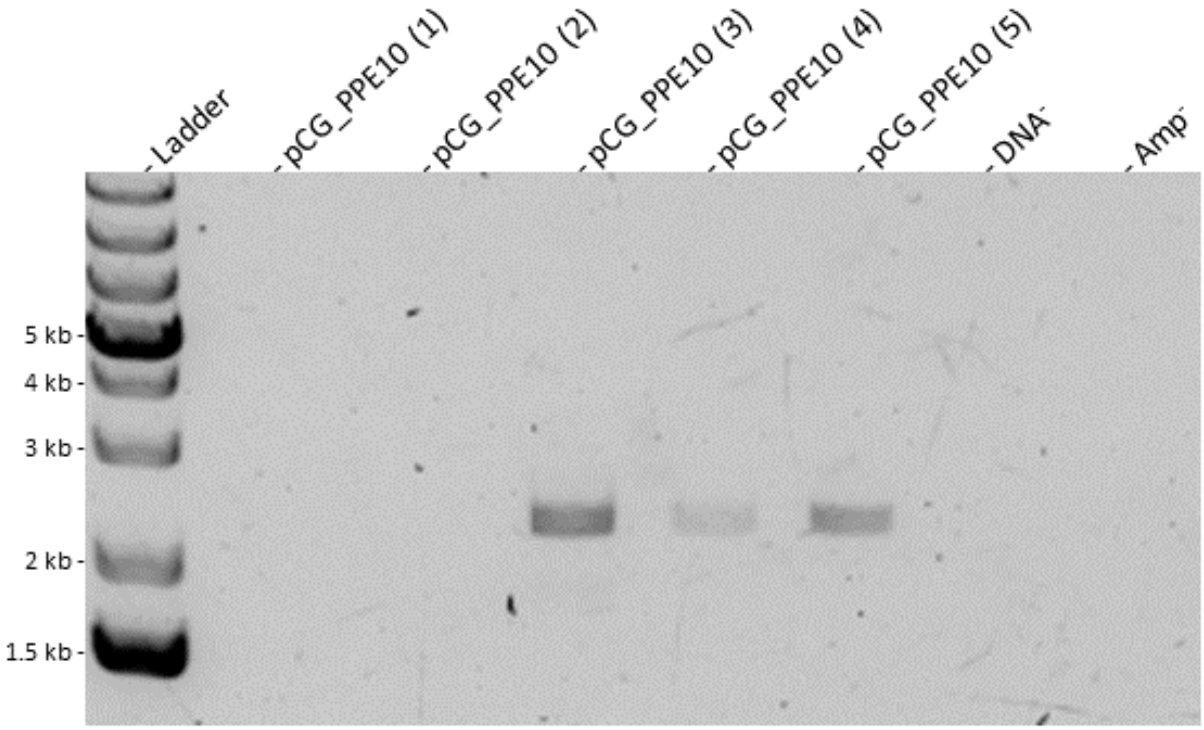
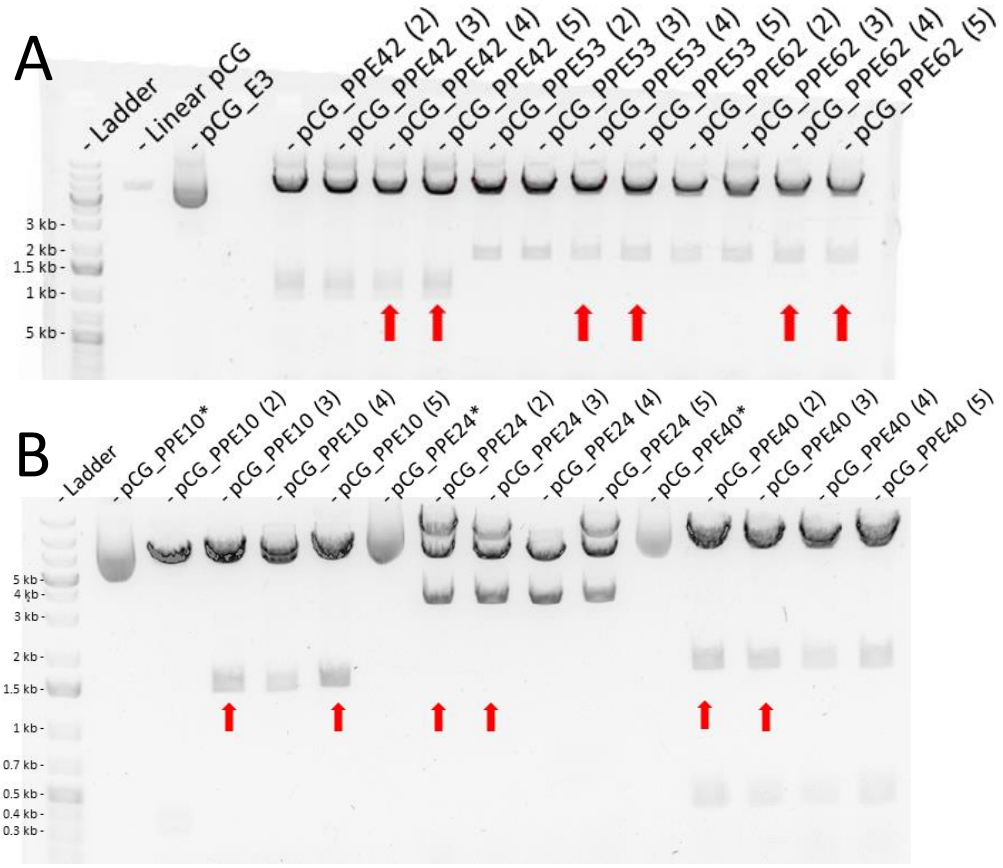


Figure 3.3.8: pCG_PPE10 Colony screening by PCR. Only colonies from the pCG_PPE10 clone plates showed any product following PCR screening reactions. Colonies (3 – 5) produced bands of expected sizes (2325 bp), which includes both the *ppe_mptr10* insert and the GFP coding sequence. Gels were comprised of 1% agarose containing 0.3X EZ-Vision Bluelight DNA dye (VWR Life Sciences) and electrophoresis done in 1x TAE buffer at 120V. Samples were seperated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

3.3.3 Fragments Produced by Screening Clones Using Enzymatic Double-Digests

DNA from clones selected in **section 3.3.2** were subjected to double digest with BamHI and MluI to determine whether the amplicon was inserted into the correct site on the expression vector before sequencing the samples. Fragments of expected size were excised from all samples, except pCG_PPE10(2) (**Figure 3.3.3, A and B**). Unexpected fragments were also generated by the digestion of digestion of pCG_PPE24 and -40 (**Figure 3.3.3, B**). Samples marked with red arrows were all sent for confirmation by Sanger sequencing.



***Undigested plasmid**

Figure 3.3.9: BamHI and MluI double digests of pCG_PPE constructs. All plasmid preparations were digested by BamHI and MluI. Fragments of expected sizes (**Table 3.3.1**) were produced across all samples, except pCG_PPE10 (2). Expected sizes were as follows; PPE_MPTR10 = 1464 bp, PPE_MPTR24 = 3162 bp, PPE_MPTR40 = 1848 bp, PPE_MPTR42 = 1743 bp, PPE_MPTR53 = 1773 bp, PPE_MPTR62 = 1749. Red arrows indicate samples that were sent for Sanger sequencing to confirm insert integrity. "*" indicates undigested plasmids which were loaded into wells. pCG_E3, represents a pCG_EccE3 construct used as a positive control for the double digest (Caitlyne Young). Gels were comprised of 1% agarose containing 0.3X EZ-Vision Bluelight DNA dye (VWR Life Sciences) and electrophoresis done in 1x TAE buffer at 120V. Samples were seperated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

3.3.4 Sequencing Results for pCG_PPE-MPTR Constructs

Sequencing results could only be obtained for the pCG_PPE10(3) plasmid clone following Sanger sequencing (**Figure 3.3.4**). All other isolates either did not produce usable sequencing results or alignments showed large errors in plasmid DNA in comparison to the pCG_PPE-MPTR reference sequences, rendering them unusable.

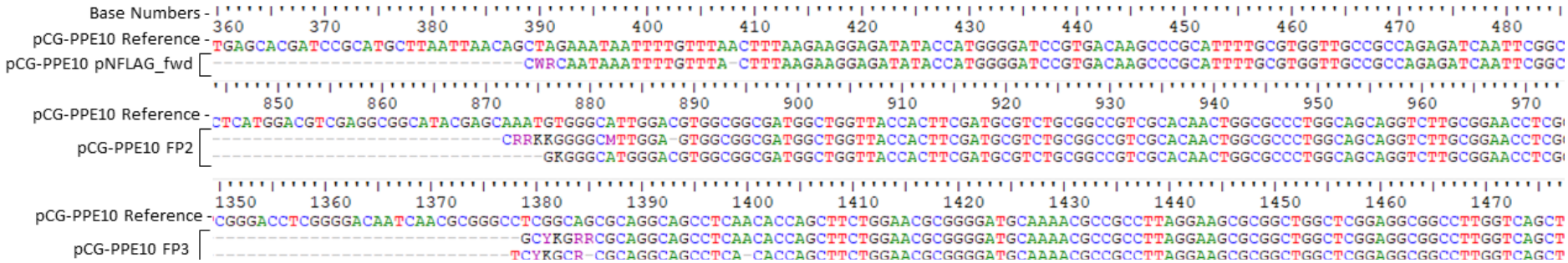


Figure 3.3.10: Alignment of pCG_PPE10 Sanger sequencing results. Cropped images showing the start of each Sanger sequencing fragment aligned to the pCG_PPE10 reference plasmid sequence, for two different clones. Sequencing primers were designed to produce fragments that could overlap to ensure PPE insert integrity. Each alignment starts with the reference sequence, followed by the respective sequencing results produced by sequencing primers.

https://scholar.sun.ac.za

3.3.5 *M. smegmatis* Growth following pCG_PPE10 Transformation

pCG_PPE10 plasmids were transformed into M. smegmatis as described in **section 3.2.10**. Growth of transformed cultures progressed at a similar rate to that of the untransformed cultures (**Figure 3.3.5**). This indicated that the transformation of wild-type (WT) *M. smegmatis*, as well as *M. smegmatis* mutants (ΔESX-1, -3, -4 and -1,3,4), with pCG_PPE10 did not impact the viability of the bacteria.

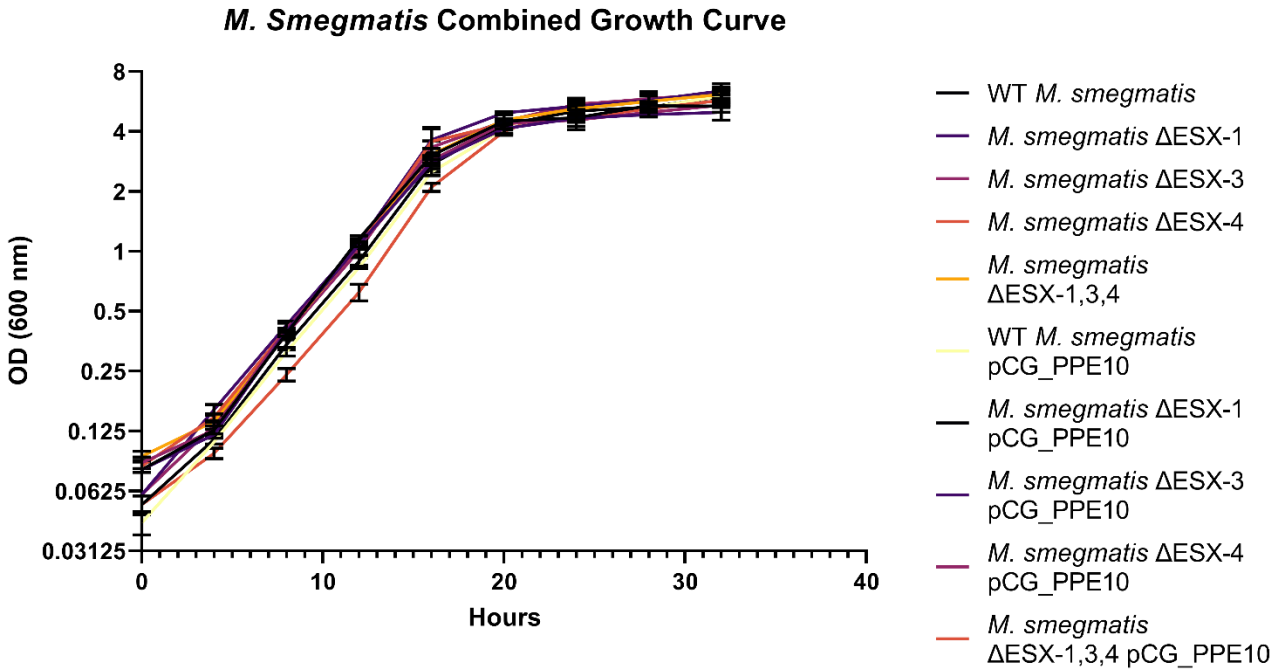


Figure 3.3.11: Growth curves of untransformed and pCG_PPE10 transformed *M. smegmatis*. Strains include WT *M. smegmatis* mc^2155, as well as the ΔESX-1, ΔESX-3, ΔESX-4 and ΔESX-1,3,4 knockout mutants. Each strain was grown individually and following transformation with pCG_PPE10. The growth rate and maximum of all strains, transformed or otherwise, was similar. Each datapoint is represented by the mean OD of triplicate cultures and error bars indicate standard deviation. The OD was measured at 600nm at 4-hour intervals for a total of 32 hours.

https://scholar.sun.ac.za

3.3.6 pCG_PPE10 Protein Expression

pCG_PPE10 plasmid DNA was repeatedly transformed into wild-type and ESX-5_{xenopi} *M. smegmatis* for constitutive protein expression. Whole-cell lysates were prepared as described in **Section 3.2.11** from cultured WT *M. smegmatis* transformed with an empty pCG plasmid, as well as pCG_PPE10 construct, untransformed WT *M. smegmatis* and ESX-5_{xenopi} transformed with the pCG_PPE10. Proteins were separated by SDS-PAGE electrophoresis and transferred to membrane for western blot using an anti-GFP antibody. No bands corresponding to the recombinant protein (75.2 kDa) were detected using SDS-PAGE (**Figure 3.3.6, A**) or western blotting (**Figure 3.3.6, B**) when compared to WT *M. smegmatis* or the pCG positive control.

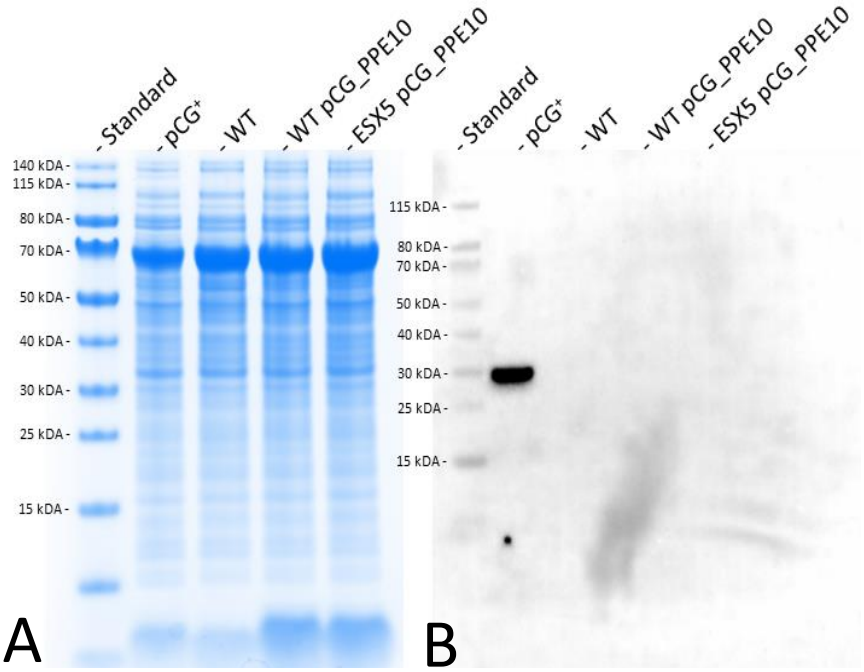


Figure 3.3.12: Example of pCG_PPE10 PAGE and Western blots. (A) Stained gel following SDS-PAGE gel electrophoresis separation of proteins in whole-cell lysates prepped from *M. smegmatis* cultures. This included pCG^+, transformed with an empty pCG plasmid (positive control), WT untransformed *M. smegmatis* (negative control), WT pCG_PPE10 transformed and $ESX5_{xenopi}$ mutant *M. smegmatis* transformed with pCG_PPE10. **(B)** Duplicate gels of SDS-PAGE were prepared for Western blot to confirm protein expression. pCG^+ positive produced bands, but no other experimental cultures showed the presence of PPE10-GFP protein expression. Gels were NuPAGE, pre-cast, 4 – 12% Bis-Tris acrylamide gels (Invitrogen) stained with InstantBlue Protein Stain (Sigma-Aldrich). Proteins were transferred to PVDF membranes utilizing pre-formulated 1x NuPAGE Transfer Buffer (Invitrogen). Band corresponding to positive GFP expression was seen after probing with monoclonal mouse, anti-GFP antibodies (# MA5-15256, Thermo Fisher) (1:2000 in 1x TBST) and detected by chemiluminescence using goat anti-mouse, HRP conjugates secondary antibody (# 62-6520, Thermo Fisher) (1:10000 in 1x TBST) using the GelDoc Imager system (Universal Hood III; Bio-Rad) (Bio-Rad). All lanes were loaded with 20 ng of protein as determined by Bradford protein concentration determination of whole-cell lysates. Proteins were run alongside a PAGE-Ruler, Pre-stain Protein ladder (ThermoFisher).

https://scholar.sun.ac.za

3.4 Discussion

Furthering our understanding of *M. tuberculosis* pathogenesis requires the targeted investigation of the molecular components that underpin the bacteria's function. Targeted studies show that mycobacterial PPE_MPTR proteins may have significant biological roles, being implicated in immune modulation, membrane stability and nutrient uptake (Ates *et al.* 2016; Mi *et al.* 2017; Wang *et al.* 2020b; Asaad *et al.* 2021; Gallant *et al.* 2021). PPE_MPTR proteins have been predicted as membrane associated based on proteomic analysis (Wolfe *et al.* 2010; Mazandu and Mulder 2012). Few targeted studies have yet to directly address the secretion and ultimate localisation of these proteins. We aimed to address this information gap by heterologous expression of fluorescently labelled PPE_MPTR proteins in *M. smegmatis* to identify their sub-cellular localisation.

All our target *ppe_mptr* genes were successfully amplified from *M. tuberculosis* H37Rv DNA. Of the six, we were able to confirm the presence of the insert within the pCG_PPE10 clone, utilising colony PCR. The accurate amplification of the ppe10 gene was confirmed by Sanger sequencing, which confirmed the integrity of our pCG_PPE10-GFP expression plasmid. Subsequent transformations showed that transformed bacteria were able to grow on hygromycin agar plates while untransformed competent cells did not grow, suggesting that the pCG-PPE10 plasmid was present within the cells. However, no expression of the PPE10-GFP fusion protein could be detected following the harvest of whole cell lysates (**Figure 3.3.6**). The PPE_MPTR10 protein has previously been predicted to be localised to the mycobacterial membrane (Gey Van Pittius *et al.* 2006; Ates 2019). Overexpression of membrane proteins can lead to toxicity in pro- and eukaryotic hosts, which may result in the death of cells and a significant reduction in protein yield (Wagner *et al.* 2007; Eguchi *et al.* 2018). This is unlikely to be the main contributor in our experiments, as the growth of *M. smegmatis* did not appear significantly impacted following transformation or during liquid culture (**Figure 3.3.5**). Solubility predictions for the PPE_MPTR10 protein do suggest it is highly insoluble in *Escherichia coli*, (SoluProt v1.0, https://loschmidt.chemi.muni.cz/soluprot/). SoluProt defines a solubility score above 0.5 as a protein being capable of soluble expression in *E. coli*. Running the amino acid sequence of PPE_MPTR10 through SoluProt produced a solubility score of 0.221, classed as insoluble expression. However, expression was done in *M. smegmatis*; previous studies have shown that using *M. smegmatis* as the expression host for *M. tuberculosis* proteins can significantly increase their chance of expressing as soluble proteins (Goldstone *et al.* 2008; Bashiri and Baker 2015).

https://scholar.sun.ac.za

Although heterologous expression of mycobacterial proteins in *M. smegmatis* is usually successful, membrane-associated proteins may deliver significantly lower yields (Grāve *et al.* 2022). Non-native proteins expressed in *M. smegmatis* that require translocation across the membrane may also become hindered, accumulating in the cell-wall and promoting stress responses that may result in their degradation (Lü *et al.* 2009). The availability of molecular pathways required for the folding and translocation of proteins may also be a limiting factor (Grāve *et al.* 2022). Wild-type *M. smegmatis* encodes the ESX-1, -3 and -4 secretion systems, but lacks the ESX-5 system, which has been suggested to translocate PPE_MPTR proteins (Digiuseppe Champion and Cox 2007b; Abdallah *et al.* 2009). We did attempt to address this issue by transforming a strain of *M. smegmatis* encoding the ESX-5 secretion system (ESX-5_{xenopi}), however this had no effect on target protein production (Beckham *et al.* 2017). Furthermore, even in the absence of ESX-5, PPE_MPTR10 has previously been expressed and detected by Western blot following expression in wild-type *M. smegmatis* mc^2155 (Asaad *et al.* 2021).

3.5 Limitations and Future Approaches

We successfully amplified all six selected *ppe_mptr* genes from *M. tuberculosis* H37Rv DNA. Subsequent cloning experiments that aimed to insert these genes into the pCG expression plasmid produced one successful clone (pCG_PPE10), as confirmed by Sanger sequencing. However, repeated attempts at heterogenous protein expression utilising this construct produced no product for analysis. Furthermore, the project also contained some notable limitations.

Cloning experiments were limited to the pCG plasmid, which may not be optimal for expression of PPE-GFP fusion proteins. Both 6X Histidine and FLAG (DYKDDDDK) tags have previously been used in constructs to successfully express and isolate PE/PPE proteins (Chen *et al.* 2017; Damen *et al.* 2020; Asaad *et al.* 2021). These tags are small enough to not significantly effect protein structure, solubility or localisation (Deng and Boxer 2018). 6X His and FLAG tags can also be utilised for Correlative-light electron microscopy (CLEM) instead of GFP. This would allow for high-resolution tracking of protein particles expressed inside bacteria, without the risk of disrupting the protein structure. Both 6x His and Flag tags have well-established protocols for the purification of expressed proteins (Lichty *et al.* 2005; Spriestersbach *et al.* 2015). Developing constructs to express target proteins with these tags could thus benefit future studies that require pure protein extracts. Previous studies have also

https://scholar.sun.ac.za

successfully utilised the pALACE shuttle vector for the stable expression of PPE_MPTR10 (Asaad *et al.* 2021). This vector allowed for the direct control of protein expression in *M. smegmatis* via acetamide based induction, which may allow us to regulate expression more tightly in future experiments, reducing the risk of toxicity or protein accumulation and degradation (Newton *et al.* 2003; Asaad *et al.* 2021).

PE/PPE proteins have also been shown to interact and translocate as heterodimeric partners (Phan *et al.* 2017; Damen *et al.* 2020). However, in this study target proteins were to be expressed in isolation. PPE10 has previously been shown to interact with the PE9 protein, although whether this would necessarily affect the localisation of PPE10 remains unclear (Tiwari *et al.* 2015). Co-expression of PE9/PPE10 could be worth future investigation to observe changes to PPE10 secretion or sub-cellular localisation.

All proteins initially targeted in this study are native to *M. tuberculosis*, but expression was carried out in the model *M. smegmatis* as an analogue to the pathogenic, slow growing *M. tuberculosis*. As previously mentioned, *M. smegmatis* lacks the ESX-5 secretion pathways, but to address this we did also transform into a mutant strain encoding ESX-5_{xenopi}. Future studies would still benefit from moving expression to *M. tuberculosis*, or models such as *SAMMtb* (*M. tuberculosis* Δ*leuD* Δ*panCD*), to track PPE_MPTR localisation in an environment closer to the natural setting (Mouton *et al.* 2019).

https://scholar.sun.ac.za

Chapter 4

Validating Single Nucleotide Genetic Variants Identified in *pe/ppe* Genes from Targeted Sequencing of *Mycobacterium tuberculosis*

4.1 Introduction

Incidence of Multi- and Extensively-drug resistant (MDR and XDR) strains of *M. tuberculosis* have steadily increased since the first reported cases of infection (Gandhi *et al.* 2006; WHO 2021). The resistance of these strains to commonly utilised first- or second-line antibiotics can result in prolonged infectiousness, risking outbreaks in vulnerable communities (WHO 2021). Whole genome sequencing (WGS) data from *M. tuberculosis* is a valuable tool in tracking and monitoring resistant strains (Jajou *et al.* 2019).

The advantage of utilising WGS data to predict *M. tuberculosis* resistomes and identify bio-markers associated with phenotypic drug resistance has already been shown (Miotto *et al.* 2017; Wu *et al.* 2022). However, highly repetitive, GC-rich regions are commonly excluded from bio-informatic pipelines that analyse the genome of *M. tuberculosis* (Bainomugisa *et al.* 2018). The reason for this is primarily the limitation of short read sequencing technologies in mapping and assembling reliable genomes from these regions (Phelan *et al.* 2016). This blind-spot also encompasses the sequences encoding the PE/PPE family of proteins, which have been implicated in *M. tuberculosis* pathogenesis and virulence (Bainomugisa *et al.* 2018; Ates 2019). These omissions have fostered a gap in the understanding of *pe/ppe* involvement in the development of drug resistance. More targeted studies have begun to reveal novel genetic variations in *pe/ppe* coding sequences within drug resistant *M. tuberculosis* strains (Cui *et al.* 2016; Bainomugisa *et al.* 2018; Hang *et al.* 2019). The relevance of these variations is still debated, as many have been found in strains already containing known resistance conferring mutations (Bainomugisa *et al.* 2018). However, expanding the pool of known PE/PPE genetic variations in drug susceptible and resistant populations of *M. tuberculosis* can aid in identifying specific *pe/ppe* genes of interest for targeted investigation.

A previous study in our laboratory developed a new bioinformatic pipeline which was applied to screen publicly available *M. tuberculosis* sequences and identify high confidence genetic variants in *pe/ppe* genes (Bagheri, unpublished). Several variants were identified that appeared isolated to either drug susceptible, MDR or XDR strains of *M. tuberculosis* (**Table 4.2.1**). We aimed to corroborate these findings by directly comparing the *in silico* results to targeted

https://scholar.sun.ac.za

sequencing results from an independent sample set. Samples included DNA extracted from clinical isolates that were phenotypically classified into four classes of drug susceptibility (DS, MR, MDR and XDR). Gene fragments containing the SNV sites were amplified from these isolates by PCR and Sanger sequenced to verify the presence of target variants.

Targeted sequencing and confirmation of these variants will provide insight into the genetic variants within *pe/ppe* genes, specifically those present in drug resistant strains of *M. tuberculosis*. These could act as promising targets for future and further investigation into the roles of PE/PPE proteins in bacterial resistance and virulence.

4.2 Methods and Materials

4.2.1 Single Nucleotide Variant Identification

The variants targeted for validation in this project were identified using a computational approach developed by a post-doctoral fellow within the Host-Pathogen Mycobactomics research group (Dr. B. Bagheri), illustrated in **Figure 4.2.1**. Briefly, Whole-genome (WG) sequences of 120 *M. tuberculosis* isolates were compiled from the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home). All analysis on the WG sequences was carried out within the Galaxy software package (Galaxy version 22.01.1.dev0, Galaxy Community). Lineage and drug resistance profiles of these sequences were determined with TB-Profiler and screened for contamination using Kraken-2 (v2.0.8-beta). Sequence quality was checked by combination of FASTQC-0.11.4 and Trimmomatic-0.35. Ninety-six sequences with a Phred score >30 and read length >100 bp were selected and processed for further analysis. Sequences that passed quality control were subjected to a stringent set of parameters for variant calling. Reads were aligned by Burrows-Wheeler Alignment (BWA-MEM) and duplicates eliminated by MarkDuplicates (GatK) (Li and Durbin 2009). A combination of three variant calling algorithms were used (FreeBayes-1.1.0.4, lofreq-2.1.5 and varscan-2.4.2). Subsequently, by using vcflib tool (v1.0.0 rc1), an intersection of single nucleotide variants (SNVs) called by all three different variant calling tools were identified for each drug category; variants were annotated by Snpeff-4.3.1. Furthermore, A Python script was written to exclude the variants that were located in regions of low mapping quality. This script was written based on the results obtained by a previous study, which found blind spots in the *M. tuberculosis* reference genome (Modlin *et al.* 2021). The presence of SNVs was manually confirmed using the Integrative Genomics Viewer (IGV, Broad Institute) on at least six selected sequences from each drug category. Lastly, a Venn diagram

https://scholar.sun.ac.za

(https://bioinformatics.psb.ugent.be/webtools/Venn/) was generated to illustrate the shared and unique SNVs found among different drug categories. Consequently, the variants present in all sequences of each drug category (at least 30 sequences) were identified as unique to that drug category. All variants that were identified as unique to specific drug resistant classes are listed in **Table 4.2.1**.

Table 4.2.3: *pe/ppe* encoding genes identified to contain unique SNVs in various drug resistant strains of *M. tuberculosis*.

Gene	Protein	Genome Position	Variant	Strain Class
Rv0159c	PE3	188800	T→C	MDR
Rv1088	PE_PGRS9	836658	A→G	DS
Rv1753c	PPE24	1983313	T→G	DS
Rv1803c	PE_PGRS32	2045310	A→G	MDR
Rv1809	PPE33	2051746	T→C	XDR
Rv1917c	PPE34	2165286	A→C	XDR
Rv2741	PE_PGRS47	3054321	A→G	MDR
Rv3347c	PPE55	3746409	A→G	XDR
Rv3347c	PPE55	3752207	A→G	XDR

https://scholar.sun.ac.za

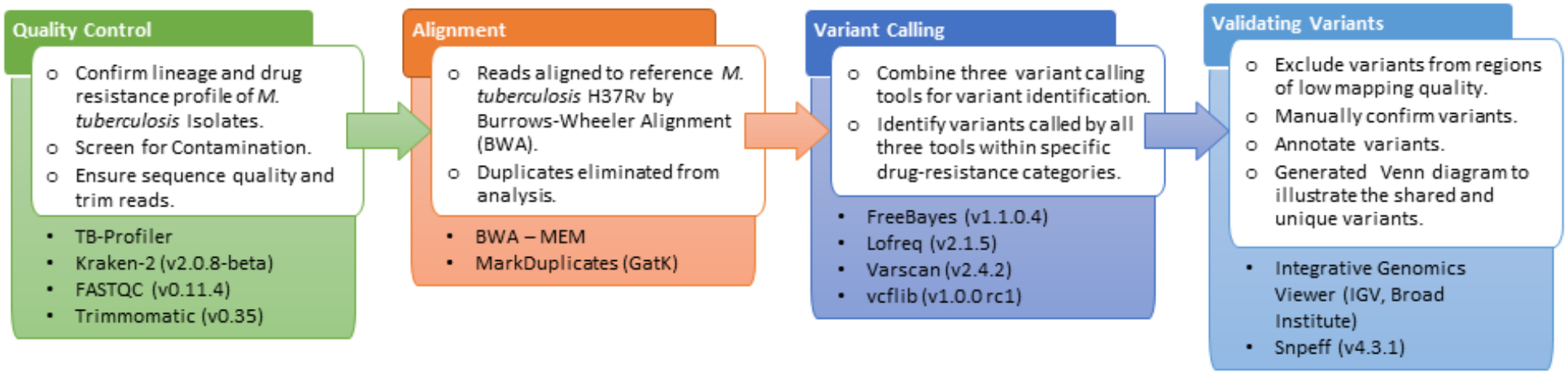


Figure 4.2.2: Workflow of the analytical pipeline developed for the identification of *pe/ppe* SNVs. All analysis was done within the Galaxy software package (Galaxy version 22.01.1.dev0). Isolate sequences were downloaded from the ENA. These isolates were passed through TB-Profiler to confirm their lineage and associated DR profile, as well as screened for contamination using Kraken-2 (v2.0.8-beta). Sequences were checked for quality using combination FASTQC (v0.11.4) and Trimmomatic (v0.35). The sequences that passed initial screening were aligned to *M. tuberculosis* H37Rv using BWA and duplicate sequences eliminated by MarkDuplicates. Three variant calling tools were used to identify SNVs (FreeBayes-v1.1.0.4, lofreq-v2.1.5 and varscan-v2.4.2) and variants that overlapped between all three listed using vcflib (v1.0.0 rc1). SNVs that fell within low mapping quality regions were excluded. The variants left were manually confirmed using the IGV and annotated by Snpeff (v4.3.1).

https://scholar.sun.ac.za

4.2.2 Selection of Clinical Isolates

Drug susceptibility testing (DST) is routinely performed by the National Health Laboratory Services (NHLS) of South Africa within the six healthcare districts of the Western Cape Province. Since 2006, clinical *M. tuberculosis* samples that were shown to carry resistance to antibiotics such as isoniazid (INH), rifampicin (RIF), fluoroquinolone (FQ) and second line injectables (SLIDs) by standard DST, were archived at Stellenbosch University (Oostvogels *et al.* 2022). These samples were de-identified from patient identifiers prior to being banked. A total of 100 *M. tuberculosis* isolates within this archive, listed in **Supplementary Table 6.1**, were selected for investigation. Both multi-drug resistant (MDR) and extensively-drug resistant (XDR) classes were represented by 40 isolates each. Furthermore, 12 mono-resistant (MR) and 8 drug susceptible (DS) isolates were also investigated. Each of these DR/DS isolate groups were also represented by strains classified as belonging to either *M. tuberculosis* Lineage 2 or 4, Beijing and Euro-American lineages, respectively. These lineages account for the majority of *M. tuberculosis* infections in the South-African setting.

4.2.3 Screening Isolates for Preliminary Results

All selected isolates had whole genome sequencing data available on the server housed within the Division of Molecular Biology and Human Genetics (@khaos.mb.sun.ac.za). Each entry was individually accessed to preliminarily determine the sequence read at the target SNV site within each gene. A minimum of 15 reads at the target SNV site was required for the isolate to be considered for the final selection.

4.2.4 Bioinformatic Profiling and Phylogenetic Analysis of Selected Isolates

To generate the phylogenetic tree of selected isolates, the annotated variant files of each isolate were recovered from the in-house Mycobactomics khaos server and filtered to remove low quality, heterogeneous variants and variants in hard-to-map regions. Only variants with allelic frequency >0.95 were considered. Filtered variant files were converted to a multi-FASTA format then transformed to Phylip format (.phy) for phylogenetic analysis in IQTree (v2.2.0) (Nguyen *et al.* 2015). ModelFinder was utilized alongside ultra-fast bootstrapping, using 1000 bootstrap iterations, during the IQTree run for phylogenetic analysis and tree generation (Thi Hoang *et al.* 2017; Kalyaanamoorthy *et al.* 2017). The phylogenetic tree was visualized using the Interactive Tree of Life (iTOL, v6.0) online phylogeny visualization software (Letunic and Bork 2021).

https://scholar.sun.ac.za

4.2.5 Amplification of Target Gene Regions

Gene fragments, encompassing the SNV sites of interest within the *pe/ppe* genes, were directly amplified from the selected DNA isolates (**Supplementary Table 6.1**) using the primers described in **Table 4.2.2**. PCR amplifications were performed in 25 µl reactions using the Phusion Hot Start, Flex 2x Master Mix (New England Biolabs, Ipswich, U.S.) according to the manufacturer's instructions. Reactions were prepared with a final primer and master mix concentration of 0.5 µM and 1x, respectively, with the addition of 3% DMSO. The PCR reactions were carried out with initial denaturation at 98°C for 30 seconds before 25 cycles of amplification, with annealing at optimal primer temperatures (**Table 4.2.2**) for 30 seconds, followed by extension at 72°C for 30 seconds, after which final elongation was carried out at 72°C for 5 min. Resulting samples were stored at 4°C, before being assessed using agarose gel electrophoresis (**Section 3.2.5**). Samples with PCR fragments of the correct amplification length were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, U.S.), according to the manufacturer's instructions.

4.2.6 Sanger Sequencing and Analysis

Gene amplicons were submitted for Sanger sequencing to verify identified SNVs, using the same primers as for the PCR reactions (**Table 4.2.2**). Sanger sequencing was performed at the Central Analytical Facility (CAF) at Stellenbosch University. Resulting sequencing reads were aligned using BioEdit (version 7.2.5) (Hall, 1999) to *M. tuberculosis* H37Rv reference genes to determine the presence or absence of the target SNV of interest.

https://scholar.sun.ac.za

Table 4.2.4: Identified gene variants alongside oligonucleotide primers utilised to amplify gene fragments containing the site of interest.

Gene	Location (bp)	Variant	Resistance Profile[a]	Primers (5'-3')		Primer Tm (°C)	Annealing Temperature in PCR (°C)	Size (bp)
Rv3347c (ppe55)	3752207	**A → G**	**XDR**	**Forward** **Reverse**	AATGCCGGTGGTTTCACTACG TAGATTCTGCCGACCGAGTACCA	59.8 62.8	64.0	605
Rv3347c (ppe55)	3746409	**A → G**	**XDR**	**Forward** **Reverse**	TGTCGGGATATTCAACCTGG CTGCCCGTGTTATAGCTA	57.3 52.6	62.0	497
Rv2741 (pe_pgrs47)	3054321	**A → G**	**MDR**	**Forward** **Reverse**	GATGAGGTGTCGATAGCCGT AAACAATCCGGCAGCTCCTCC	59.3 63.0	64.0	363
Rv1917c (ppe34)	2165286	**T → C**	**XDR**	**Forward** **Reverse**	CGCCATTTTTCCCCGGATTC GTAGCTACCGACGTTTCCGCTAC	59.3 63.0	64.0	860
Rv1809 (ppe33)	2051746	**A → G**	**XDR**	**Forward** **Reverse**	GGCGACCAACATTTTCGGACAAA ACACGTTCTGATAGGTCGTTGCG	62.7 63.0	64.0	395
Rv0159c (pe3)	188800	**T → C**	**MDR**	**Forward** **Reverse**	ATGTCCTACGTCATCGCGGC AGCGGTGTTGCTGGCTTCAG	62.0 63.0	64.0	288

a. Original interpretation of pipeline predicted the SNV to be associated with this resistance profile.

https://scholar.sun.ac.za

4.3 Results

4.3.1 Isolate Analysis and Phylogeny

Clinical isolates representing four drug resistance profiles were selected to screen for previously identified SNVs. These were selected to verify the variants identified within the previous pipeline within an independent set of isolates. Selected isolates were divided into DS, MR, MDR and XDR isolates, each represented by subsets of both *M. tuberculosis* Lineage 2 and 4. A maximum likelihood phylogenetic tree was generated using the available isolate data (**Figure 4.3.1**). Interestingly three samples did not cluster according to the expected lineages. Isolate R15732 clustered among Lineage 5 representatives, although it is listed in the database as an isolate belonging to Lineage 4.1. Both S43diag and S152 listed as Lineage 2, were placed among Lineage 4 representatives.

https://scholar.sun.ac.za

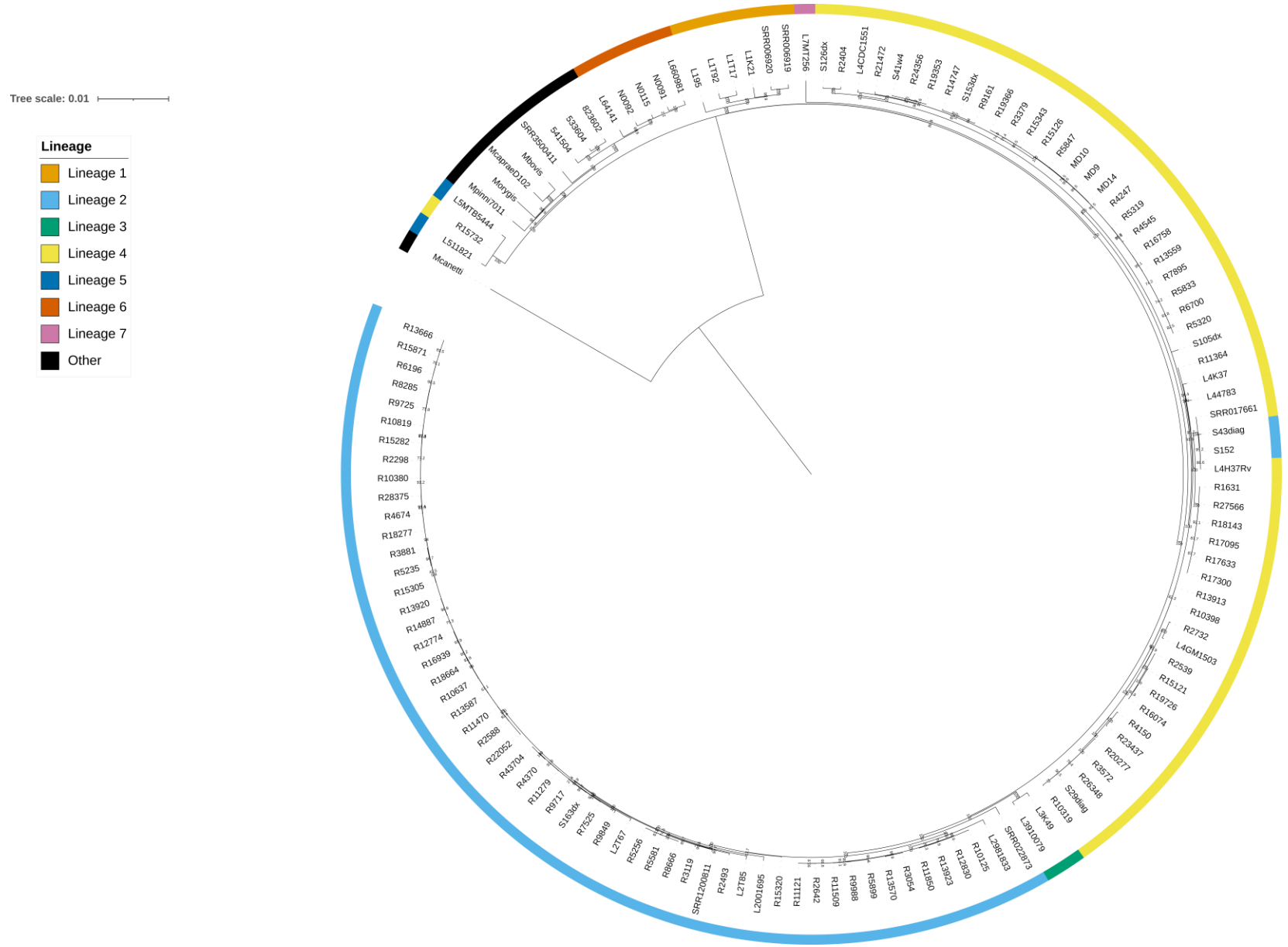


Figure 4.3.2: Phylogenetic analysis showing the relationship of selected isolates to other representatives of the *M. tuberculosis* complex and each other. Bootstrap support values are shown next to nodes. The tree was generated using IQTree, based on identified variable sites in comparison to the *M. tuberculosis* H37Rv reference genome (Letunic and Bork 2021).

https://scholar.sun.ac.za

4.3.2 Gene Amplifications

The target gene fragments were amplified from clinical *M. tuberculosis* DNA isolates (**Figure 4.3.2 – 4.3.4**). Gene fragments corresponding to *pe3* and *pe_pgrs47* were amplified from all forty MDR isolates (**Figure 4.3.2**), as well as all eight DS and twelve MR isolates (**Figure 4.3.3**). Gene fragments corresponding to *ppe33*, *ppe34* and *ppe55* were amplified from all forty XDR isolates (**Figure 4.3.2**), as well as from all eight DS and twelve MR isolates (**Figure 4.3.3**). Another subset was also amplified to cross-check variants between drug resistant classes. For these, *pe3* and *pe_pgrs47* fragments were amplified from six XDR isolates (**Figure 4.3.4**). Additionally, *ppe33*, *ppe34* and *ppe55* fragments were amplified from six MDR isolates (**Figure 4.3.4**). All amplifications produced single bands corresponding to the desired product, which were cleaned, quantified and sent for sequencing.

Bands produced following the amplification of *ppe34* fragments showed variable sizes between isolates, even within drug resistant classes (**Figure 4.3.3 – 4.3.5**). The *ppe34*, alongside *ppe24* and -*35*, is known to encode glycine-asparagine rich regions broken up by approximately 26 amino acid repeats (69 -75 bp each) (Sampson *et al*. 2001). Eight tandem repeats were in *ppe34* encoded by *M. tuberculosis* H37Rv. However, the number of repeats were polymorphic and varied between different isolates, which we have also observed in our fragment sizes (Sampson *et al*. 2001). No clear pattern of association to specific drug resistance classes or *M. tuberculosis* lineages was discerned. Although, fragments that appeared smaller than the H37Rv reference, appeared more commonly amplified from Lineage 4 isolates, regardless of drug resistance class. Regardless of their size discrepancies, the majority of *ppe34* fragments were successfully sequenced and the target SNV site aligned to the reference H37Rv genome.

https://scholar.sun.ac.za

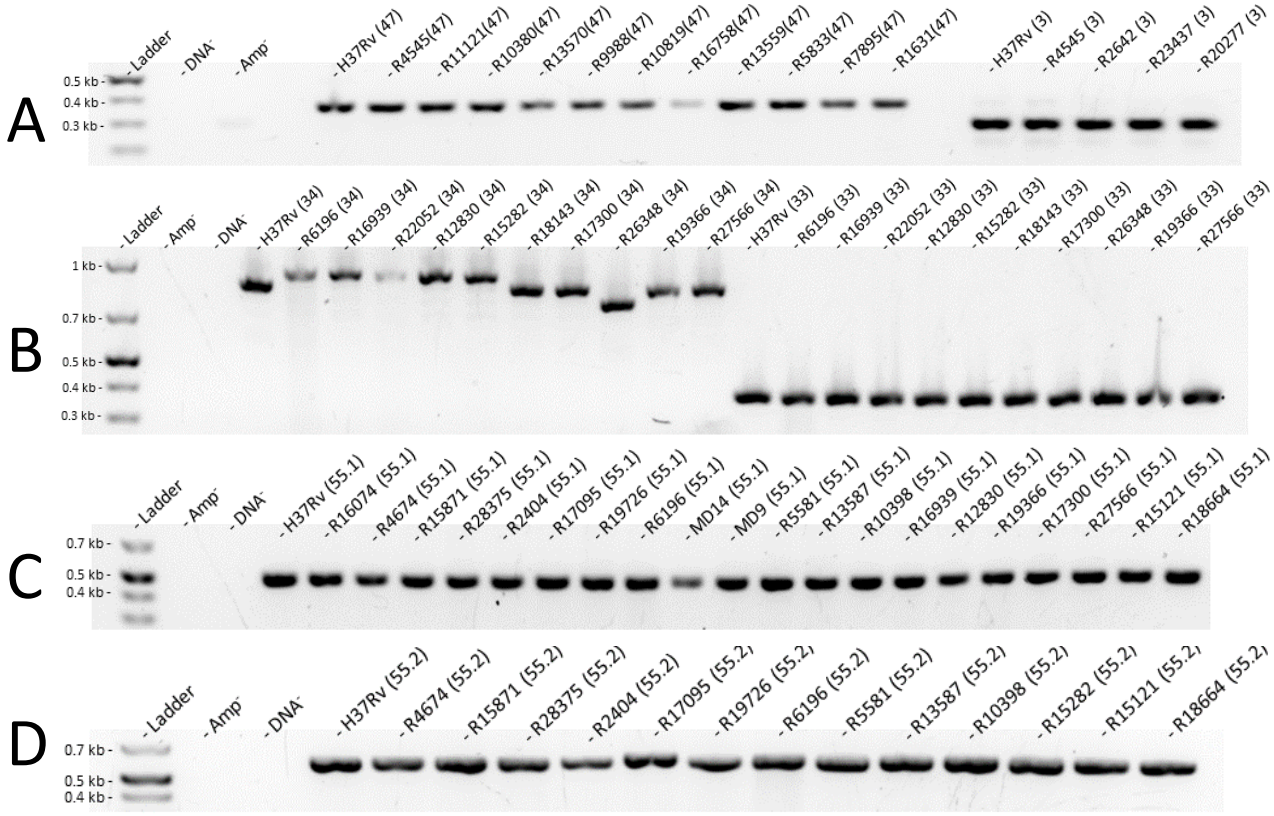


Figure 4.3.2: Amplicons produced following PCR amplification of target gene regions from MDR and XDR clinical *M. tuberculosis* isolates. **(A)** Bands for the amplified fragments targeting the *pe47* (363 bp) and *pe3* (288 bp) SNVs, both of which were predicted as unique to MDR *M. tuberculosis*. (**B, C and D**) Bands for the amplified fragments targeting the *ppe34* (860 bp), *ppe33* (395 bp) and two *ppe55* SNVs (498 bp and 605 bp), all predicted as unique to XDR *M. tuberculosis*. The Amp$^-$ control represents a blank reaction which contained 100 ng of H37Rv DNA, but no reagents to catalyze the PCR. DNA$^-$ PCR controls contained all required reagents except DNA. Gels were comprised of 1% agarose containing 0.3X EZ-Vision Bluelight DNA dye (VWR Life Sciences) seperated by electrophoresis in 1X TAE buffer at 120V. Samples were seperated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

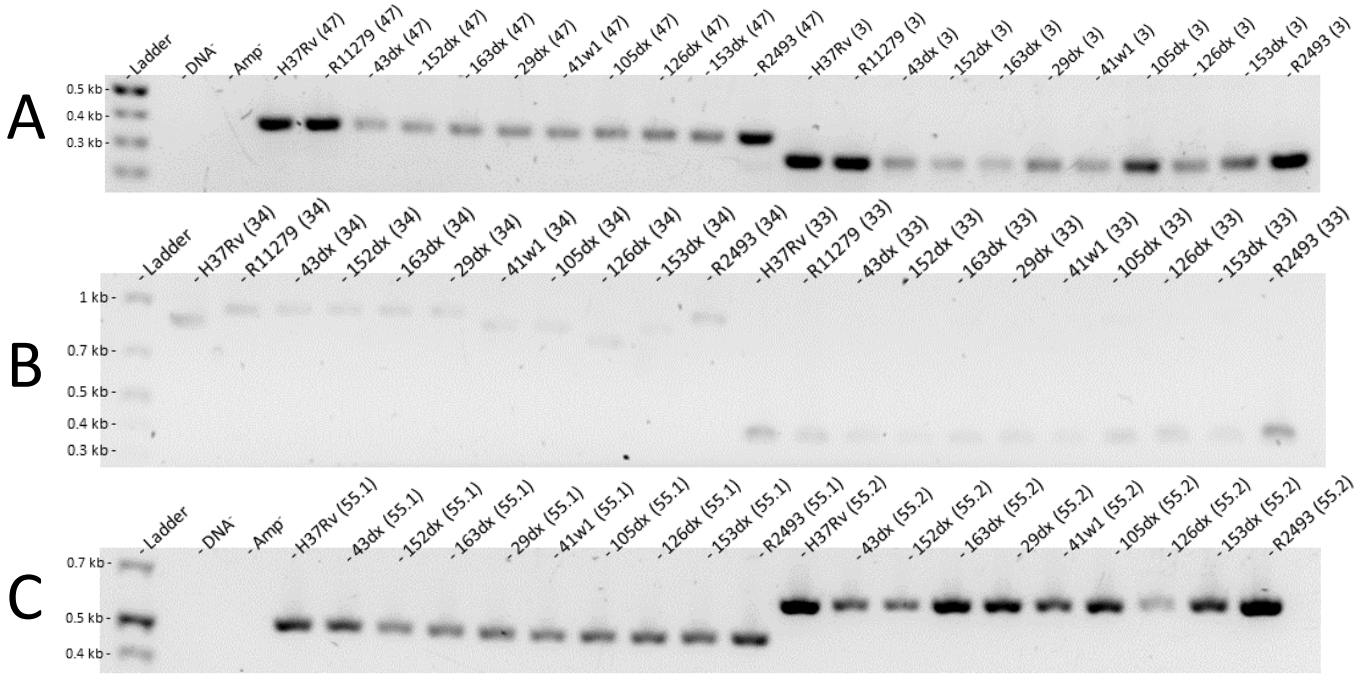


Figure 4.3.3: Amplicons produced following PCR amplification of target gene regions from DS *M. tuberculosis* DNA isolates. (A) Bands for the amplifed fragments targeting the *pe47* (363 bp) and *pe3* (288 bp) SNVs sites. **(B)** Bands for the amplified fragments targeting the *ppe34* (860 bp) and *ppe33* (395 bp) SNVs sites. **(C)** Bands for the amplified fragments targeting the *ppe55* (498 and 605 bp) SNVs sites. Each of these SNVs were predicted to be unique to MDR and XDR isolates respectively. The numbers in brackets represent the target gene from which the fragment was amplified (*pe47*, *pe3*, *ppe34*, *ppe33*, *ppe55*). The Amp$^-$ control represents a blank reaction which contained 100 ng of H37Rv DNA, but no reagents to catalyze the PCR. DNA$^-$ PCR controls contained all required reagents except DNA. Gels were comprised of 1% agarose containing 0.3X EZ-Vision Bluelight DNA dye (VWR Life Sciences) seperated by elecrophoresis in 1X TAE buffer at 120V. Samples were seperated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

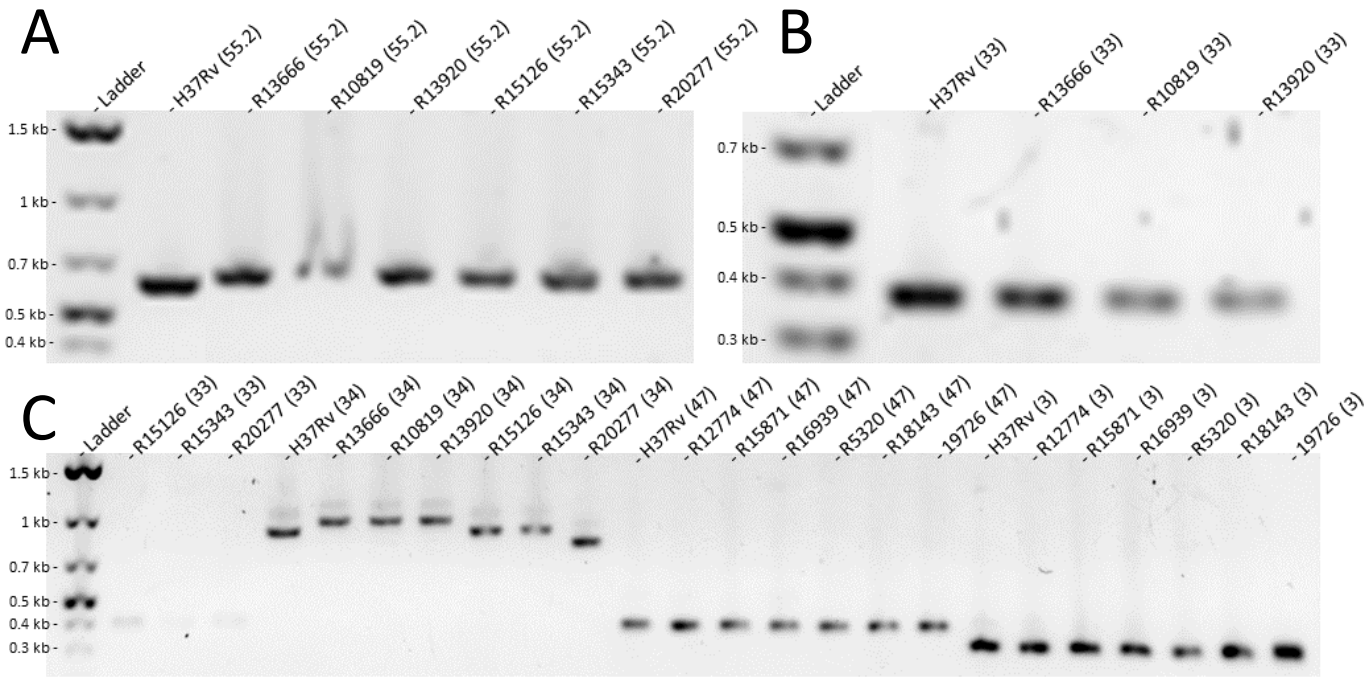


Figure 4.3.4: Amplicons produced from PCR amplification of gene fragments from clinical isolates to cross-evaluate SNVs. (A) Bands of *ppe55*, predicted as unique to XDR *M. tuberculosis*, these samples were amplified from MDR *M. tuberculosis*. **(B)** Half of the bands of *ppe33*, predicted as unique to XDR *M. tuberculosis* were amplified from MDR *M. tuberculosis*. **(C)** Faint bands corresponding to the next half of *ppe33* samples were seen in lane 2 - 4, alongside the bands of *ppe34*. Both were predicted as unique to XDR *M. tuberculosis*, but were amplified from MDR isolates. Bands of *pe_pgrs47* and *pe3* were also seen, these were predicted as unique to MDR isolates, but were amplified from XDR *M. tuberculosis* DNA. Gels were comprised of 1% agarose containing 0.3X EZ-Vision Bluelight DNA dye (VWR Life Sciences), separated by electrophoresis in 1X TAE buffer at 120V. Samples were seperated alongside a GeneRuler 1 kb plus DNA ladder (Thermo Fisher). Visualized using UV light in a GelDoc Imager system (Universal Hood III, Bio-Rad).

https://scholar.sun.ac.za

4.3.3 Sequencing Results

Sequencing data, including the *M. tuberculosis* H37Rv laboratory strain (ATCC 700084), was aligned to the *M. tuberculosis* H37Rv reference genome (https://www.ncbi.nlm.nih.gov/nuccore/NC_000962.3). **Figure 4.3.5** – **4.3.7** are examples of these alignments. These alignments helped to identify less reliable sequences, demonstrated conservation of sequence surrounding the target SNVs and confirmed SNVs in the majority of strains tested.

With regards to the reliability of sequences, for example the alignment of *pe3* gene fragments to the reference genomes showed that the site of interest was too close to the start of the sequence to be reliable (**Figure 4.3.6, B**). The primers for this gene would require redesign to allow more appropriate coverage of the *pe3* SNV site. As such, *pe3* was excluded from all further analysis over all isolates. The alignments also identified other examples where the sequencing results were of too low quality to generate reliable comparisons to the reference genomes. Some of these were relatively isolated to single gene fragments, such as the sequence alignments of *ppe34* from MD10 and *ppe55* from R10319 (**Table 4.3.1**). These differences in sequence quality could be attributed to the process of fragment amplification and subsequent purification of PCR products, which in some instances produced samples with low DNA concentrations and quality. However other isolates, such as R2539, R6196 and R2588, consistently produced low quality amplifications and sequencing reads and struggled to align to reference genomes (**Figure 4.3.7**, **Table 4.3.1**). These isolates were low in concentration and 260/280 ratios indicated poor quality DNA, indicating potential chemical contamination, which may have negatively impacted the Sanger sequencing workflow. In summary, five of the six target regions were successfully amplified, Sanger sequenced and aligned to the H37Rv reference genomes.

Besides the examples mentioned above, the majority of sequences (83/100) were successfully amplified and covered the entire sequencing region of interest to the H37Rv reference genome, from which variants could be identified within the gene fragments (**Table 4.3.1)**. These sequences also largely showed conservation of sequence relative to the reference genome. Surprisingly, given the interpretations from the initial bioinformatic pipeline, SNVs were observed across all samples where reliable sequence data was obtained. This included DS and MR isolates, that harboured SNVs which were predicted as unique to MDR and XDR strains. As a validation of the *M. tuberculosis* H37Rv reference genome sequence, we also sequenced

https://scholar.sun.ac.za

a laboratory strain of *M. tuberculosis* H37Rv. This demonstrated that all but one of the SNVs were present in the clinical isolates, but absent from the reference H37Rv genome. The exception to this was the SNV originally predicted to be unique to *ppe33* in XDR isolates; this was also found within the lab strain of *M. tuberculosis* H37Rv, alongside all other DS and MR isolates, but was absent from the online reference sequence (**Figure 4.3.5 and Figure 4.3.7**).

These unexpected findings prompted a revisit to the interpretation and results generated from the original bioinformatic pipeline. This revealed an error in the original analysis caused by an over-stringent filtering of the identified SNVs, this resulted in them being falsely flagged as unique to specific DR classes. Briefly, three separate variant calling software packages were utilised to identify variants in *pe/ppe* sequences across the original sample set collected from the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home). Once a variant was identified, all isolates were examined for this variant. If it was identified within all 30 sequences (100%) belonging to a specific DR class, the variant was called as unique to that class through the pipeline. A re-analysis revealed that the last step of the pipeline, where a Venn diaGram was generated to illustrate the shared and unique SNVs found among different drug categories, could have resulted in the false identification. Consequently, the results obtained from the last step of analysis (Venn diaGram) which resulted in identifying unique SNVs associated with MDR and XDR categories, were excluded from the previous study. To further evaluate the potential presence of SNVs in MDR and XDR sequences, the relative abundance of each SNV among all 96 sequences will be evaluated. Implementing this approach allows identifying the SNVs that may present high percentages (>80%) in MDR and XDR sequences.

https://scholar.sun.ac.za

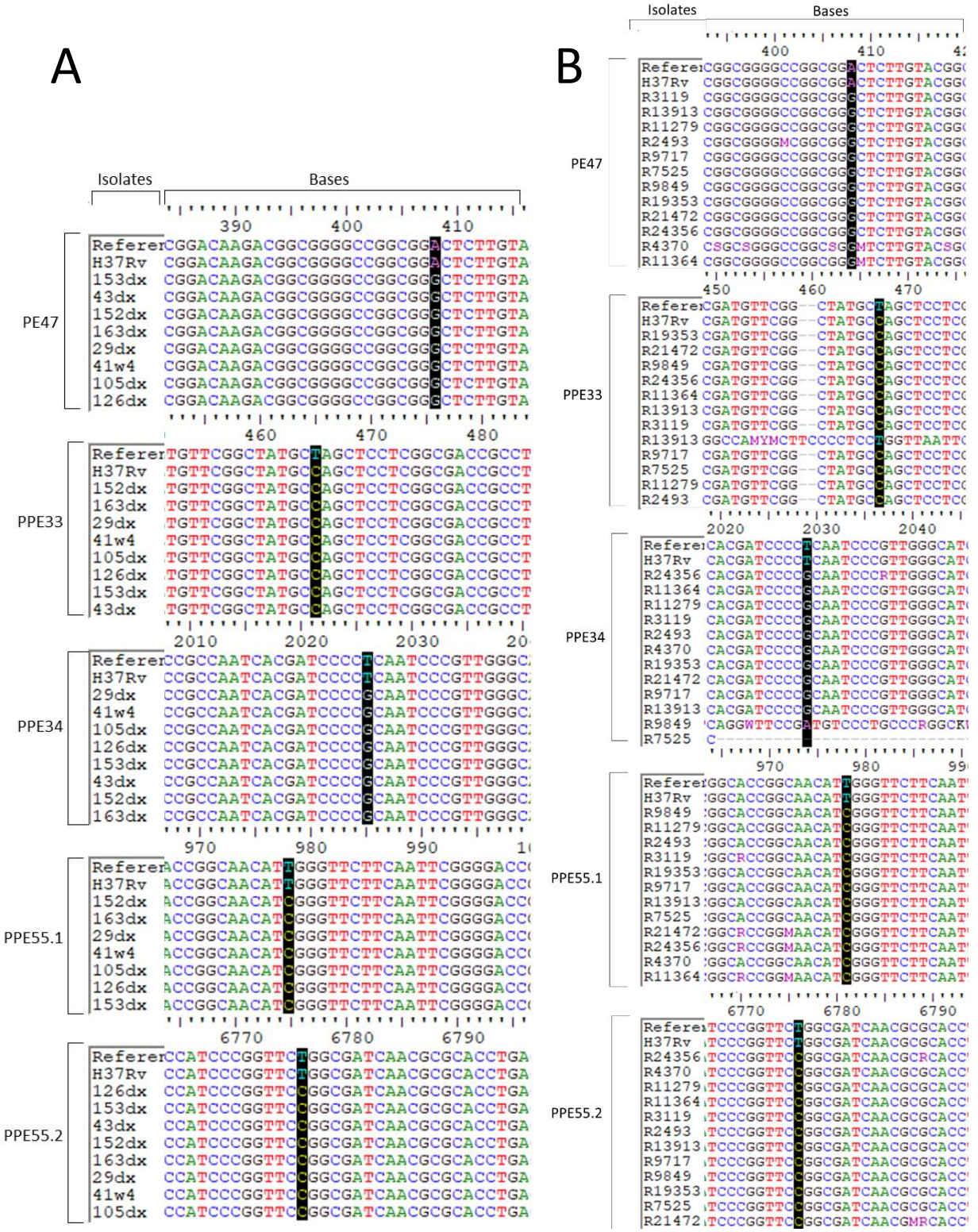


Figure 4.3.5: Sequence alignment of drug susceptible and MR isolates. Images show the cropped sequences obtained from Sanger sequencing; brackets indicate specific gene sequences. **(A)** Sequence alignment of gene fragments amplified from DS isolates. **(B)** Sequence alignment of gene fragments amplified from MR isolates. The *ppe33* R13913, *ppe55.1* R7525 and R9849 sequences were low quality and did not align to the reference sequences, these are noted as Inconclusive. The *pe3* sequencing was unreliable and excluded from these alignments. Each sequence generated from each isolate was aligned to the 'Reference' sequence from the Mycobrowser database (https://mycobrowser.epfl.ch/), as well as a *M. tuberculosis* H37Rv lab strain, from which the same gene fragments were amplified and sequenced.

https://scholar.sun.ac.za

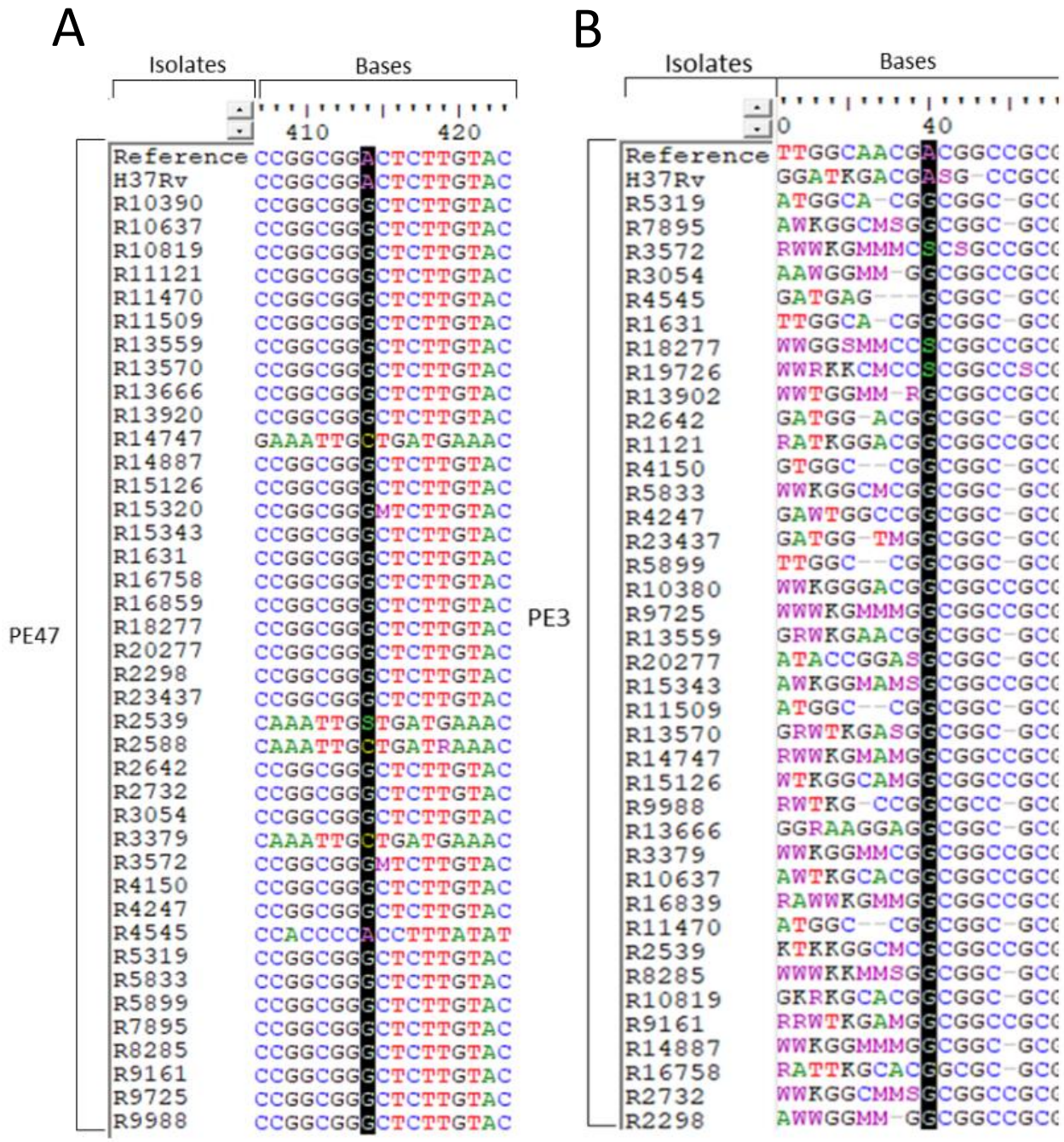


Figure 4.3.6: Sequence alignment of MDR isolates. (A) Image shows the cropped *pe_pgrs47* sequences obtained from Sanger sequencing aligned to the reference genes. Sequences for R14747, R2539, R2588, R3349 and R4545 were low quality and did not align correctly to the reference sequences, these are noted as inconclusive. **(B)** The *pe3* sequencing and subsequent alignment was unreliable, an example of the *pe3* alignment is seen here. Each sequence generated from each isolate was aligned to the 'Reference' sequence from the Mycobrowser database (https://mycobrowser.epfl.ch/), as well as a *M. tuberculosis* H37Rv lab strain, from which the same gene fragments were amplified and sequenced.

https://scholar.sun.ac.za

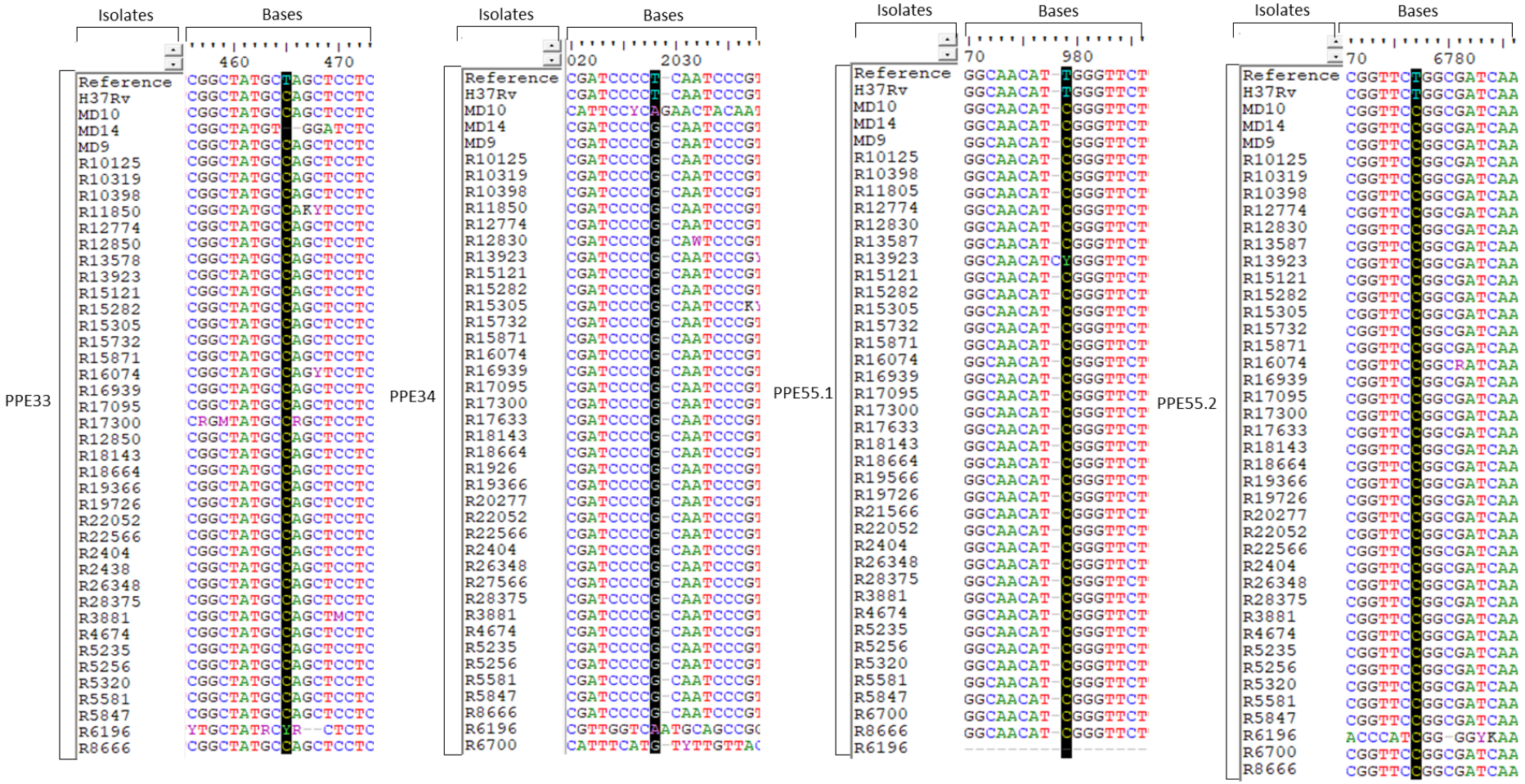


Figure 4.3.7: Sequence alignments of XDR isolates. Image shows the cropped sequences obtained from Sanger sequencing aligned to the reference genes, brackets indicate specific gene sequences. MD14 and R17300 for *ppe33*, MD10 and R6700 for *ppe34*, R19323 and for *ppe55.1*, were all examples of isolated sequences of low quality that did not align correctly, these are noted as inconclusive. The R6196 isolate only delivered low quality sequences that did not align, the whole isolate is noted as inconclusive. Each sequence generated from each isolate was aligned to the 'Reference' sequence from the Mycobrowser database (https://mycobrowser.epfl.ch/), as well as a *M. tuberculosis* H37Rv lab strain, from which the same gene fragments were amplified and sequenced.

https://scholar.sun.ac.za

Table 4.3.2: Variant results for all sequenced isolates.

Lineage	Drug Resistance Class	Isolate	SNV PE47 (Y/N)	SNV PPE33 (Y/N)	SNP PPE34 (Y/N)	SNV PPE55 (3746409) (Y/N)	SNV PPE55 (3752207) (Y/N)
Lineage 2	DS	43dx	Y	Y	Y	Inconclusive	Y
Lineage 2	DS	152dx	Y	Y	Y	Y	Y
Lineage 2	DS	163dx	Y	Y	Y	Y	Y
Lineage 2.2	MR	R11279	Y	Y	Y	Y	Y
Lineage 2.2	MR	R2493	Y	Y	Y	Y	Y
Lineage 2.2	MR	R9717	Y	Y	Y	Y	Y
Lineage 2.2	MR	R7525	Y	Y	Inconclusive	Y	Y
Lineage 2.2	MR	R3119	Y	Y	Y	Y	Y
Lineage 2.2	MR	R4370	Y	Y	Y	Y	Y
Lineage 2.2	MR	R9849	Y	Y	Inconclusive	Y	Y
Lineage 2.2	XDR	R10125	-	Y	Y	Y	Y
Lineage 2.2	XDR	R11850	-	Y	Y	Y	Y
Lineage 2.2	XDR	R12774	-	Y	Y	Y	Y
Lineage 2.2	XDR	R3881	-	Y	Y	Y	Y
Lineage 2.2	XDR	R5235	-	Y	Y	Y	Y
Lineage 2.2	XDR	R13587	-	Y	Y	Y	Y
Lineage 2.2	XDR	R15871	-	Y	Y	Y	Y
Lineage 2.2	XDR	R28375	-	Y	Y	Y	Y
Lineage 2.2	XDR	R4674	-	Y	Y	Y	Y
Lineage 2.2	XDR	R5581	-	Y	Y	Y	Y
Lineage 2.2	XDR	R12830	-	Y	Y	Y	Y
Lineage 2.2	XDR	R15282	-	Y	Y	Y	Y
Lineage 2.2	XDR	R16939	-	Y	Y	Y	Y
Lineage 2.2	XDR	R22052	-	Y	Y	Y	Y
Lineage 2.2	XDR	R6196	-	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Lineage 2.2	XDR	R13923	-	Y	Y	Y	Y

https://scholar.sun.ac.za

Lineage	Drug Resistance Class	Isolate	SNV PE47 (Y/N)	SNV PPE33 (Y/N)	SNP PPE34 (Y/N)	SNV PPE55 (3746409) (Y/N)	SNV PPE55 (3752207) (Y/N)
Lineage 2.2	XDR	R15305	-	Y	Y	Y	Y
Lineage 2.2	XDR	R5256	-	Y	Y	Y	Y
Lineage 2.2	XDR	R8666	-	Y	Y	Y	Y
Lineage 2.2	XDR	R18664	-	Y	Y	Y	Y
Lineage 2.2	XDR	R12774	Y	-	-	-	-
Lineage 2.2	XDR	R15871	Y	-	-	-	-
Lineage 2.2	XDR	R16939	Y	-	-	-	-
Lineage 2.2	MDR	R13666	-	Y	Inconclusive	Inconclusive	Y
Lineage 2.2	MDR	R10819	-	Y	Y	Y	Y
Lineage 2.2	MDR	R13920	-	Y	Y	Y	Y
Lineage 2.2	MDR	R11470	Y	-	-	-	-
Lineage 2.2	MDR	R11509	Y	-	-	-	-
Lineage 2.2	MDR	R13666	Y	-	-	-	-
Lineage 2.2	MDR	R2642	Y	-	-	-	-
Lineage 2.2	MDR	R5899	Y	-	-	-	-
Lineage 2.2	MDR	R10380	Y	-	-	-	-
Lineage 2.2	MDR	R10819	Y	-	-	-	-
Lineage 2.2	MDR	R11121	Y	-	-	-	-
Lineage 2.2	MDR	R13570	Y	-	-	-	-
Lineage 2.2	MDR	R9988	Y	-	-	-	-
Lineage 2.2	MDR	R10637	Y	-	-	-	-
Lineage 2.2	MDR	R13920	Y	-	-	-	-
Lineage 2.2	MDR	R2298	Y	-	-	-	-
Lineage 2.2	MDR	R3054	Y	-	-	-	-
Lineage 2.2	MDR	R8285	Y	-	-	-	-
Lineage 2.2	MDR	R14887	Y	-	-	-	-
Lineage 2.2	MDR	R15320	Y	-	-	-	-
Lineage 2.2	MDR	R18277	Y	-	-	-	-

https://scholar.sun.ac.za

Lineage	Drug Resistance Class	Isolate	SNV PE47 (Y/N)	SNV PPE33 (Y/N)	SNP PPE34 (Y/N)	SNV PPE55 (3746409) (Y/N)	SNV PPE55 (3752207) (Y/N)
Lineage 2.2	MDR	R2588	Inconclusive	-	-	-	-
Lineage 2.2	MDR	R9725	Y	-	-	-	-
Lineage 4	DS	H37Rv	N	Y	N	N	N
Lineage 4	DS	29dx	Y	Y	Y	Y	Y
Lineage 4	DS	41w4	Y	Y	Y	Y	Y
Lineage 4	DS	105dx	Y	Y	Y	Y	Y
Lineage 4	DS	126dx	Y	Y	Y	Y	Y
Lineage 4	DS	153dx	Y	Y	Y	Y	Y
Lineage 4.1	MR	R19353	Y	Y	Y	Inconclusive	Y
Lineage 4.1	MR	R21472	Y	Y	Y	Y	Y
Lineage 4.1	MR	R24356	Y	Y	Y	Y	Y
Lineage 4.1	XDR	MD10	-	Y	Inconclusive	Y	Y
Lineage 4.1	XDR	MD14	-	Inconclusive	Y	Y	Y
Lineage 4.1	XDR	MD9	-	Y	Y	Y	Y
Lineage 4.1	XDR	R6700	-	Inconclusive	Y	Y	Y
Lineage 4.1	XDR	R19366	-	Y	Y	Y	Y
Lineage 4.1	XDR	R15732	-	Y	Y	Y	Y
Lineage 4.1	XDR	R5320	-	Y	Y	Y	Y
Lineage 4.1	XDR	R5847	-	Y	Y	Y	Y
Lineage 4.1	XDR	R5320	Y	-	-	-	-
Lineage 4.1	MDR	R15126	-	Y	Y	Y	Y
Lineage 4.1	MDR	R15343	-	Y	Y	Y	Y
Lineage 4.1	MDR	R4545	Inconclusive	-	-	-	-
Lineage 4.1	MDR	R5319	Y	-	-	-	-
Lineage 4.1	MDR	R13559	Y	-	-	-	-
Lineage 4.1	MDR	R16758	Y	-	-	-	-
Lineage 4.1	MDR	R5833	Y	-	-	-	-
Lineage 4.1	MDR	R7895	Y	-	-	-	-

https://scholar.sun.ac.za

Lineage	Drug Resistance Class	Isolate	SNV PE47 (Y/N)	SNV PPE33 (Y/N)	SNP PPE34 (Y/N)	SNV PPE55 (3746409) (Y/N)	SNV PPE55 (3752207) (Y/N)
Lineage 4.1	MDR	R15126	Y	-	-	-	-
Lineage 4.1	MDR	R15343	Y	-	-	-	-
Lineage 4.1	MDR	R16859	Y	-	-	-	-
Lineage 4.1	MDR	R9161	Y	-	-	-	-
Lineage 4.1	MDR	R4247	Y	-	-	-	-
Lineage 4.1	MDR	R14747	Inconclusive	-	-	-	-
Lineage 4.1	MDR	R3379	Inconclusive	-	-	-	-
Lineage 4.3	XDR	R16074	-	Y	Y	Y	Y
Lineage 4.3	XDR	R10398	-	Y	Y	Y	Y
Lineage 4.3	XDR	R19726	-	Y	Y	Y	Y
Lineage 4.3	XDR	R15121	-	Y	Y	Y	Y
Lineage 4.3	XDR	R26348	-	Y	Y	Y	Y
Lineage 4.3	XDR	R10319	-	Y	Y	Inconclusive	Y
Lineage 4.3	XDR	R19726	Y	-	-	-	-
Lineage 4.3	MDR	R20277	-	Y	Y	Y	Y
Lineage 4.3	MDR	R2539	Inconclusive	-	-	-	-
Lineage 4.3	MDR	R2732	Y	-	-	-	-
Lineage 4.3	MDR	R20277	Y	-	-	-	-
Lineage 4.3	MDR	R23437	Y	-	-	-	-
Lineage 4.3	MDR	R4150	Y	-	-	-	-
Lineage 4.9	MR	R11364	Y	Y	Y	Y	Y
Lineage 4.9	MR	R13913	Y	Inconclusive	Y	Y	Y
Lineage 4.9	XDR	R17095	-	Y	Y	Y	Y
Lineage 4.9	XDR	R2404	-	Y	Y	Y	Y
Lineage 4.9	XDR	R17300	-	Y	Y	Y	Y
Lineage 4.9	XDR	R18143	-	Y	Y	Y	Y
Lineage 4.9	XDR	R27566	-	Y	Y	Y	Y
Lineage 4.9	XDR	R17633	-	Y	Y	Y	Y

https://scholar.sun.ac.za

Lineage	Drug Resistance Class	Isolate	SNV PE47 (Y/N)	SNV PPE33 (Y/N)	SNP PPE34 (Y/N)	SNV PPE55 (3746409) (Y/N)	SNV PPE55 (3752207) (Y/N)
Lineage 4.9	XDR	R18143	Y	-	-	-	-
Lineage 4.9	MDR	R1631	Y	-	-	-	-
Lineage 4.9	MDR	R3572	Y	-	-	-	-
Successfully sequenced*			61/67 (91%)	63/67 (94%)	60/67 (90%)	61/67 (91%)	66/67 (99%)

* Where sequencing was attempted
- Sequencing was not attempted
"Inconclusive" Sequencing results were of too low quality to determine the presence of the SNV

https://scholar.sun.ac.za

4.4 Discussion

Genetic variants within the *M. tuberculosis* genome have become commonplace targets for tracking the evolution of the bacteria (Oostvogels *et al.* 2022). WGS data is often analysed to identify the canonical variants associated with specific lineages or drug resistance (Cohen *et al.* 2015; Dixit *et al.* 2019). These variants can then be utilised to genotype strains and become direct targets for the identification of specific infections. As an example, diagnostic kits that exploit known drug resistance mutations for the identification of resistant bacteria have been exploited in clinical settings. However, the sensitivity, specificity, drug coverage range and target gene-loci vary significantly between kits (Forbes *et al.* 2018). As such, WGS of clinical isolates have become an attractive alternative to more time-consuming antimicrobial susceptibility testing for the accurate identification of drug resistant strains of *M. tuberculosis* (Wu *et al.* 2022).

Most WGS pipelines exclude complex, highly repetitive gene regions during analysis (Heupink *et al.* 2021). This stems from wide-scale implementation of short-read sequencing technologies that are limited in their ability to reliably assemble highly repetitive or variable gene regions without significant errors (Ashton *et al.* 2015). Long-read technologies can potentially overcome these limitations by producing longer overlapping strands which negates the difficulty of assembling the genome from short reads in regions of extreme genetic complexity (Heupink *et al.* 2021). Excluded regions in *M. tuberculosis* include the PE/PPE coding regions, which are characterised by their highly repetitive GC-rich sequences (Meehan *et al.* 2019). This has caused a significant gap in our understanding of the potential role of *pe/ppe* genetic variants in the development of new or drug resistant strains of *M. tuberculosis*.

The pipeline whose findings we aimed to validate was developed to address this information gap. It was designed to analyse available *M. tuberculosis* WGS data, specifically screening PE/PPE coding regions, to identify unique genetic variants within and between different drug resistant classes. Previous studies have carried out extensive WGS on a single XDR isolate, which included the PE/PPE coding regions (Bainomugisa *et al.* 2018). This allowed for the identification of variants in *pe/ppe* genes, that otherwise would be missed in more conventional pipelines that would exclude these regions (Bainomugisa *et al.* 2018). In contrast, the pipeline that informed our study utilised a more inclusive approach. It was developed to screen and filter archived WGS data, analysing and comparing high-quality *pe/ppe* sequence variants to identify true and unique variants in drug resistant *M. tuberculosis*.

https://scholar.sun.ac.za

All SNVs identified by this pipeline, which were selected for investigation, were found to be true variants. However, a limitation of the pipeline was revealed when predicting unique variants within each drug susceptibility class. Variants in *pe3* and *pe_pgrs47* were predicted as unique to MDR isolates, while variants in *ppe33*, *ppe34* and -*55* were predicted as unique to XDR *M. tuberculosis*. None of these hypotheses were true following targeted Sanger sequencing of in-house clinical isolates. All investigated variants were found in DS clinical isolates, as well as in a selected subset of MDR and XDR isolates, all of which were represented by bacteria from both Lineage 2 and 4. Previous studies have noted that a significant portion of drug-resistance associated mutations are shared across MDR and XDR isolates, likely stemming from shared drug-resistance profiles (Zhang *et al.* 2013). However, in the case of the pipeline validated within this study, the false identification of unique variants stemmed from over-stringent filtering of the sample set, aiming to pull out the SNVs present in all 30 sequences of each drug category.

Notably, the variant identified within *ppe33* was not unique to clinical isolates, but also present in the laboratory strain of *M. tuberculosis* H37Rv, while being absent in the online reference genome. Recent studies have noted that the original WGS carried out on H37Rv, might not have captured the complexity within GC-rich repetitive regions, such as those encoding the PE/PPE (Chitale *et al.* 2022). This lack of sufficient coverage is suggested to also be due to the original reliance on short-read sequencing for generating reads and genome assembly (Chitale *et al.* 2022). When considering our result, it suggests an error in the online reference of *M. tuberculosis* H37Rv, which could potentially stem from limitations surrounding the original sequencing of the strain.

As previously mentioned, emergence and accumulation of genetic variants are the targets of WGS when determining the spread and evolution of new *M. tuberculosis* strains across the globe (Xu *et al.* 2018; Oostvogels *et al.* 2022). Here we showed that four of our target variants were present in clinical *M. tuberculosis* isolates from the Western Cape, while absent in the laboratory strain and online reference of *M. tuberculosis* H37Rv. These variants can thus be seen as indicative of the genetic drift of *M. tuberculosis* and how the genome has changed over time, reflecting more modern strains of the bacteria. Generally, *pe/ppe* genes have been excluded from analyses of WGS, due to their repetitive, GC-rich regions that are prone to erroneously map during genome assembly. This includes studies that aim to identify novel mutations that can be used to barcode or be associated with drug-resistant strains. The rapid

https://scholar.sun.ac.za

rate of genetic substitutions within *pe/ppe* genes can cause neutral mutations to become fixed during population bottlenecks in the process of antibiotic treatments (Farhat *et al.* 2013). Exclusion of these gene regions as a whole on this basis does run the risk of missing mutations that may have a legitimate contribution to particular phenotypes, including drug resistance profiles.

Classical drug resistance mutations that occur in genes encoding protein targets of drugs or drug metabolising enzymes are the focus of many studies tracking resistance-associated genetic variants (Zhang and Yew 2015). Mutations in other classes of genes that effect cell wall permeability, efflux pump activity, act as compensatory mutations that increase survivability or promote mutational phenotypes can also promote drug resistance (Nikaido 1994; Schrag *et al.* 1997; Denamur and Matic 2006). Several PE/PPE proteins have previously been implicated in modulating cell-wall permeability and regulating nutrient uptake during mycobacterial growth and infection, which suggests some *pe/ppe* gene variants may be able to fill the role of non-classical mutations that could promote resistance (Farhat *et al.* 2013; Wang *et al.* 2020c; Ehtram *et al.* 2021).

From the proteins investigated in this study, PPE33, in combination with PE20, has been suggested to contribute to the transport of a yet to be defined nutrient source across the mycobacterial membrane (Ehtram *et al.* 2021). PPE34 and PPE55 are known to be highly polymorphic and immunogenic, PPE55 specifically appears significantly expressed in sub-clinical tuberculosis (TB) infections (Singh *et al.* 2005b; Sharma *et al.* 2022). PE47 has been shown to inhibit autophagy (Saini *et al.* 2016), while PE3 has been associated with bacterial persisters in murine models and was significantly upregulated under hypoxic stress or during chronic infection states (Singh *et al.* 2013). Whether the variants identified within these genes may have any functional consequence for the encoded proteins, is beyond the scope of this project. Future studies could exploit *in silico* models to predict the changes in protein folding and structure that these variants may introduce, however this may first require a more detailed understanding of the structure of individual PE/PPE proteins (Portelli *et al.* 2018; Williamson *et al.* 2020). The *pe/ppe* variants could also be introduced into *M. tuberculosis* H37Rv to observe any effects on the bacteria's survival under different growth conditions, antibiotic regimens or any changes to the bacterial morphology and membrane permeability (Wong *et al.* 2018; Kumar *et al.* 2022).

https://scholar.sun.ac.za

For the proteins of interest (PE3, PE47, PPE33, PPE34 and PPE55), more targeted studies would be needed to understand whether they might be able to play a role in promoting drug-resistant phenotypes, either cumulatively or in isolation. This does however highlight how new procedures, such as combination long-short read sequencing, can be used to target PE/PPE coding regions (Bainomugisa *et al.* 2018). Long-read technologies can sequence whole genetic regions, negating the limitations of short read genome assembly, but can then be screened with short-reads to adjust for sequencing errors (Bainomugisa *et al.* 2018). These technologies could reveal more genes/proteins for targeted study, exploring their role in the virulence and resistance of clinical *M. tuberculosis* strains.

4.5 Limitations and Future Approaches

The DNA databank within the Division of Molecular Biology and Human Genetics presented a unique opportunity to utilise a curated sample set of drug-resistant clinical *M. tuberculosis* for the validation of the genetic variants identified by the computational pipeline (Oostvogels *et al.* 2022). Overall, the study was limited in the size of the sample set that was used both for computational identification of target variants, as well as the isolates collected for validation (100 isolates). Isolates included representatives from both Lineage 2 and 4, respectively representing Beijing and Euro-American lineages, clinically relevant lineages in the South African setting. Forty isolates were selected to represent MDR and XDR *M. tuberculosis*, respectively. However, only eight isolates were readily available to represent true DS *M. tuberculosis*, which were isolated from a previous project within the HPM group, outside of the established databank (Julian Coetzee, MSc). An additional twelve isolates, which had previously been computationally classified as DS, but phenotypically presented as mono-resistant were also included. These isolates represent an opportunistic and localised sample population, focusing solely on samples collected from the Western Cape province of South Africa, in contrast to the computational pipeline which utilised publicly available WGS data. Future work on refining and validating the pipeline could benefit from including a more robust sample set, both in the analysis and in the final validation to increase the diversity of the represented genomes (Dixit *et al.* 2019). Further targeted validation can also expand the pool of identified variants that are addressed, as our study primarily focussed on variants called as unique to MDR and XDR *M. tuberculosis*, while we excluded those called as unique to DS *M. tuberculosis* or were suggested to overlap between drug resistant classes. Different filtering parameters could be implemented to ensure the integrity of variants called as unique from WGS data (Heupink *et al.* 2021).

https://scholar.sun.ac.za

Chapter 5

Conclusions

Many unknowns remain regarding the fundamental components of the PE/PPE proteins, especially regarding their secretion and localisation. Cell surface localisation may position PE/PPE proteins to directly interact at the host-pathogen interface during infection.

We firstly aimed to express specific PPE_MPTR proteins in *Mycobacterium smegmatis* to determine their localisation by fluorescent microscopy. The PPE_MPTR subfamily was of specific interest as it is unique to pathogenic strains of mycobacteria, but still remains relatively understudied in comparison to other PE/PPE proteins. We selected six PPE_MPTR proteins to study based on their known association with the cell membrane or previous identification in culture filtrates, suggesting a potential to interact at the host-pathogen interface. We successfully amplified all six *ppe_mptr* genes from *M. tuberculosis* H37Rv DNA and cloned *ppe_mptr10* into the pCG expression plasmid, as confirmed by Sanger sequencing. However, we were unable to obtain PPE_MPTR10-GFP expression from this construct within *M. smegmatis* and were unable to determine the protein's localisation. Although expression could not be obtained within the timeframe of this project, future work could optimise cloning and expression to produce target proteins within *M. smegmatis* or *M. tuberculosis*. Once achieved, the localisation of the proteins could be determined using fluorescent microscopy, but it would also open other opportunities for study. The constructs could allow for the expression of these PPE_MPTR proteins in mutant *M. tuberculosis* strains, allowing a direct comparison between bacterial phenotypes. Constructs would also allow a direct way to introduce specific mutations into the *ppe_mptr* gene, to determine their effect on the bacterial phenotype when expressed in strains lacking the native expression of that gene.

The second half of this project focussed more broadly on the association of *pe/ppe* genetic variations with drug resistance. Genetic variants that accumulate within the genome of clinical *M. tuberculosis* act as valuable markers to track and diagnose different strains. Variants that are associated with drug resistance profiles carry unique importance with the ongoing spread of MDR and XDR TB. Several studies have noted the biological significance of PE/PPE proteins, yet in the same instance exclude *pe/ppe* coding regions from analytical pipelines designed to screen for variants associated with mycobacterial virulence and resistance.

https://scholar.sun.ac.za

Although the hurdles of sequencing high-GC, repetitive regions are significant, the development of newer long read and dual sequencing technologies have begun to make these regions more accessible. Our results showed that the variants identified as unique to MDR and XDR sequences were not restricted to these strains. However, these variants were present across all classes of clinical *M. tuberculosis* isolates and, with the exception of variants in *ppe33*, absent from wildtype *M. tuberculosis* H37Rv. This highlights the potential for future studies to screen biologically significant PE/PPE coding sequences for variants associated with new phenotypic strains, where classical mutations may not be able to account for differences. Ongoing efforts to further elucidate the biological components that allow *M. tuberculosis* to function and continue to develop as an efficient pathogen will need to explore the PE/PPE proteins in further detail to expand our understanding of their role in mycobacterial pathogenesis.

https://scholar.sun.ac.za

Chapter 6

References

Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJP, Vandenbroucke-Grauls CMJE, Appelmelk BJ, Luirink J, Bitter W (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol Microbiol 62:667–679. https://doi.org/10.1111/j.1365-2958.2006.05409.x

Abdallah AM, Verboom T, Weerdenburg EM, Gey Van Pittius NC, Mahasha PW, Jiménez C, Parra M, Cadieux N, Brennan MJ, Appelmelk BJ, Bitter W (2009) PPE and PE-PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. Mol Microbiol 73:329–340. https://doi.org/10.1111/j.1365-2958.2009.06783.x

Ahmed A, Das A, Mukhopadhyay S (2015) Immunoregulatory functions and expression patterns of PE/PPE family members: Roles in pathogenicity and impact on anti-tuberculosis vaccine and drug design. IUBMB Life 67:414–427. https://doi.org/10.1002/iub.1387

Ahmed A, Dolasia K, Mukhopadhyay S (2018a) Mycobacterium tuberculosis PPE18 Protein Reduces Inflammation and Increases Survival in Animal Model of Sepsis . The Journal of Immunology 200:3587–3598. https://doi.org/10.4049/jimmunol.1602065

Ahmed S, Kabir M, Arif M, Ali Z, Ali F, Swati ZNK (2018b) Improving secretory proteins prediction in Mycobacterium tuberculosis using the unbiased dipeptide composition with support vector machine. Int J Data Min Bioinform 21:212–229. https://doi.org/10.1504/IJDMB.2018.097682

Akopian D, Shen K, Zhang X, Shan S (2013) Signal Recognition Particle: An essential protein targeting machine. Annu Rev Biochem 82:693–721. https://doi.org/10.1038/jid.2014.371

Ansari MY, Batra SD, Ojha H, Dhiman K, Ganguly A, Tyagi JS, Mande SC (2020) A novel function of Mycobacterium tuberculosis chaperonin paralog GroEL1 in copper homeostasis. FEBS Lett 594:3305–3323. https://doi.org/10.1002/1873-3468.13906

Asaad M, Kaisar Ali M, Abo-kadoum MA, Lambert N, Gong Z, Wang H, Uae M, Nazou SAE, Kuang Z, Xie J (2021) Mycobacterium tuberculosis PPE10 (Rv0442c) alters host cell apoptosis and cytokine profile via linear ubiquitin chain assembly complex HOIP-NF-κB signaling axis. Int Immunopharmacol 94:107363. https://doi.org/10.1016/j.intimp.2020.107363

Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J, O'Grady J (2015) MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol 33:296–302

Ates LS (2019) New insights into the mycobacterial PE and PPE proteins provide a framework for future research. Mol Microbiol 113:4–21. https://doi.org/10.1111/mmi.14409

Ates LS, Brosch R (2017) Discovery of the type VII ESX-1 secretion needle? Mol Microbiol 103:7–12. https://doi.org/10.1111/mmi.13579

Ates LS, Dippenaar A, Sayes F, Pawlik A, Bouchier C, Ma L, Warren RM, Sougakoff W, Majlessi L, van Heijst JWJ, Brossier F, Brosch R (2018a) Unexpected Genomic and Phenotypic Diversity of Mycobacterium africanum Lineage 5 Affects Drug Resistance, Protein Secretion, and Immunogenicity. Genome Biol Evol 10:1858–1874. https://doi.org/10.1093/gbe/evy145

Ates LS, Dippenaar A, Ummels R, Piersma SR, Van Der Woude AD, Van Der Kuij K, Le Chevalier F, Mata-Espinosa D, Barrios-Payán J, Marquina-Castillo B, Guapillo C, Jiménez CR, Pain A, Houben ENG, Warren RM, Brosch R, Hernández-Pando R, Bitter W (2018b) Mutations in ppe38 block PE-PGRS secretion and increase virulence of Mycobacterium tuberculosis. Nat Microbiol 3:181–188. https://doi.org/10.1038/s41564-017-0090-6

Ates LS, Sayes F, Frigui W, Ummels R, Damen MPM, Bottai D, Behr MA, van Heijst JWJ, Bitter W, Majlessi L, Brosch R (2018c) RD5-mediated lack of PE_PGRS and PPE-MPTR export in BCG vaccine strains results in strong reduction of antigenic repertoire but little impact on protection. PLoS Pathog 14:1–29. https://doi.org/10.1371/journal.ppat.1007139

Ates LS, van der Woude AD, Bestebroer J, van Stempvoort G, Musters RJP, Garcia-Vallejo JJ, Picavet DI, Weerd R van de, Maletta M, Kuijl CP, van der Wel NN, Bitter W (2016) The ESX-5 System of Pathogenic Mycobacteria

https://scholar.sun.ac.za

Is Involved In Capsule Integrity and Virulence through Its Substrate PPE10. PLoS Pathog 12:1–26. https://doi.org/10.1371/journal.ppat.1005696

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

Baldwin SL, Reese VA, Larsen SE, Beebe E, Guderian J, Orr MT, Fox CB, Reed SG, Coler RN (2021) Prophylactic efficacy against Mycobacterium tuberculosis using ID93 and lipid-based adjuvant formulations in the mouse model. PLoS One 16:1–20. https://doi.org/10.1371/journal.pone.0247990

Bansal K, Sinha AY, Ghorpade DS, Togarsimalemath SK, Patil SA, Kaveri S v., Balaji KN, Bayry J (2010) Src homology 3-interacting domain of Rv1917c of Mycobacterium tuberculosis induces selective maturation of human dendritic cells by regulating PI3K-MAPK-NF-κB signaling and drives Th2 immune responses. Journal of Biological Chemistry 285:36511–36522. https://doi.org/10.1074/jbc.M110.158055

Bashiri G, Baker EN (2015) Production of recombinant proteins in Mycobacterium smegmatis for structural and functional studies. Protein Science 24:1–10

Basu D, Khare G, Singh S, Tyagi A, Khosla S, Mande SC (2009) A novel nucleoid-associated protein of Mycobacterium tuberculosis is a sequence homolog of GroEL. Nucleic Acids Res 37:4944–4954. https://doi.org/10.1093/nar/gkp502

Beckham KSH, Ciccarelli L, Bunduc CM, Mertens HDT, Ummels R, Lugmayr W, Mayr J, Rettel M, Savitski MM, Svergun DI, Bitter W, Wilmanns M, Marlovits TC, Parret AHA, Houben ENG (2017) Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. Nat Microbiol 2:1–7. https://doi.org/10.1038/nmicrobiol.2017.47

Berks BC, Sargent F, Palmer T (2000) The Tat protein export pathway. Mol Microbiol 35:260–274. https://doi.org/10.1046/j.1365-2958.2000.01719.x

Berthet FX, Lagranderie M, Gounon P, Laurent-Winter C, Ensergueix D, Chavarot P, Thouron F, Maranghi E, Pelicic V, Portnoï D, Marchal G, Gicquel B (1998) Attenuation of virulence by disruption of the Mycobacterium tuberculosis erp gene. Science (1979) 282:759–762. https://doi.org/10.1126/science.282.5389.759

Bertholet S, Ireton GC, Ordway DJ, Windish HP, Pine SO, Kahn M, Phan T, Orme IM, Vedvick TS, Baldwin SL, Coler RN, Reed SG (2010) A defined tuberculosis vaccine candidate boosts BCG and protects against multidrug-resistant mycobacterium tuberculosis. Sci Transl Med 2. https://doi.org/10.1126/scitranslmed.3001094

Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K (2002) Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol 43:717–731. https://doi.org/10.1046/j.1365-2958.2002.02779.x

Bitter W, Houben ENG, Bottai D, Brodin P, Brown EJ, Cox JS, Derbyshire K, Fortune SM, Gao LY, Liu J, van Pittius NCG, Pym AS, Rubin EJ, Sherman DR, Cole ST, Brosch R (2009) Systematic genetic nomenclature for type VII secretion systems. PLoS Pathog 5:8–13. https://doi.org/10.1371/journal.ppat.1000507

Bordes P, Cirinesi AM, Ummels R, Sala A, Sakr S, Bitter W, Genevaux P (2011) SecB-like chaperone controls a toxin-antitoxin stress-responsive system in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 108:8438–8443. https://doi.org/10.1073/pnas.1101189108

Bottai D, Brosch R (2009) Mycobacterial PE, PPE and ESX clusters: Novel insights into the secretion of these most unusual protein families. Mol Microbiol 73:325–328. https://doi.org/10.1111/j.1365-2958.2009.06784.x

Bottai D, di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, Bitter W, Brennan MJ, Leclerc C, Batoni G, Campa M, Brosch R, Esin S (2012) Disruption of the ESX-5 system of Mycobacterium tuberculosis causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol 83:1195–1209. https://doi.org/10.1111/j.1365-2958.2012.08001.x

Braunstein M, Espinosa BJ, Chan J, Belisle JT, Jacobs WR (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium tuberculosis. Mol Microbiol 48:453–464. https://doi.org/10.1046/j.1365-2958.2003.03438.x

Brennan MJ (2017) The Enigmatic PE / PPE Multigene Family Vaccination. Infect Immun 85:1–8

https://scholar.sun.ac.za

Brennan PJ (1995) The Envelope of Mycobacteria. Annu Rev Biochem 64:29–63. https://doi.org/10.1146/annurev.biochem.64.1.29

Brodin P, Poquet Y, Levillain F, Peguillet I, Larrouy-Maumus G, Gilleron M, Ewann F, Christophe T, Fenistein D, Jang J, Jang MS, Park SJ, Rauzier J, Carralot JP, Shrimpton R, Genovesio A, Gonzalo-Asensio JA, Puzo G, Martin C, Brosch R, Stewart GR, Gicquel B, Neyrolles O (2010) High content phenotypic cell-based visual screen identifies Mycobacterium tuberculosis acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathog 6. https://doi.org/10.1371/journal.ppat.1001100

Burg-Golani T, Pozniak Y, Rabinovich L, Sigal N, Paz RN, Herskovits AA (2013) Membrane chaperone SecDF plays a role in the secretion of Listeria monocytogenes major virulence factors. J Bacteriol 195:5262–5272. https://doi.org/10.1128/JB.00697-13

Cai YD, Zhou GP, Chou KC (2003) Support vector machines for predicting membrane protein types by using functional domain composition. Biophys J 84:3257–3263. https://doi.org/10.1016/S0006-3495(03)70050-2

Chakhaiyar P, Nagalakshmi Y, Aruna B, Murthy KJR, Katoch VM, Hasnain SE (2004) Regions of High Antigenicity within the Hypothetical PPE Major Polymorphic Tandem Repeat Open-Reading Frame, Rv2608, Show a Differential Humoral Response and a Low T Cell Response in Various Categories of Patients with Tuberculosis. J Infect Dis 190:1237–1244. https://doi.org/10.1086/423938

Champion PAD, Stanley SA, Champion MM, Brown EJ, Cox JS (2006) C-Terminal Signal Sequence Promotes Virulence Factor Secretion in Mycobacterium tuberculosis. Science (1979) 313:1632–1637

Chen JM, Boy-Röttger S, Dhar N, Sweeney N, Buxton RS, Pojer F, Rosenkrands I, Cole ST (2012) EspD is critical for the virulence-mediating ESX-1 secretion system in Mycobacterium tuberculosis. J Bacteriol 194:884–893. https://doi.org/10.1128/JB.06417-11

Chen X, Cheng HF, Zhou J, Chan CY, Lau KF, Tsui SKW, Au SW ngor (2017) Structural basis of the PE–PPE protein interaction in Mycobacterium tuberculosis. Journal of Biological Chemistry 292:16880–16890. https://doi.org/10.1074/jbc.M117.802645

Chilukoti N, M. Santosh Kumar C, Mande SC (2016) GroEL2 of Mycobacterium tuberculosis reveals the importance of structural pliability in chaperonin function. J Bacteriol 198:486–497. https://doi.org/10.1128/JB.00844-15

Chitale P, Lemenze AD, Fogarty EC, Shah A, Grady C, Odom-Mabey AR, Evan Johnson W, Yang JH, Murat Eren A, Brosch R, Kumar P, Alland D (2022) A comprehensive update to the Mycobacterium tuberculosis H37Rv reference genome. bioRxiv. https://doi.org/10.1101/2022.07.15.500236

Cohen KA, Abeel T, Mcguire AM, Desjardins A, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida D V, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J (2015) Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades : Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from. 1–22. https://doi.org/10.1371/journal.pmed.1001880

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S V., Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream M-A, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 396:190–190. https://doi.org/10.1038/24206

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S (2013) Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. Nat Genet 45:1176–1182. https://doi.org/10.1038/ng.2744.Out-of-Africa

Copin R, Coscollá M, Seiffert SN (2014) Sequence Diversity in the pe_pgrs Genes of Mycobacterium tuberculosis Is Independent of Human T Cell Recognition. mBio 5:1–11. https://doi.org/10.1128/mBio.00960-13.Editor

https://scholar.sun.ac.za

Coros A, Callahan B, Battaglioli E, Derbyshire KM (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in Mycobacterium smegmatis. Mol Microbiol 69:794–808. https://doi.org/10.1111/j.1365-2958.2008.06299.x

Cosma CL, Sherman DR, Ramakrishnan L (2003) The Secret Lives of the Pathogenic Mycobacteria. Annu Rev Microbiol 57:641–676. https://doi.org/10.1146/annurev.micro.57.030502.091033

Cristóbal S, de Gier JW, Nielsen H, von Heijne G (1999) Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. EMBO Journal 18:2982–2990. https://doi.org/10.1093/emboj/18.11.2982

Cui ZJ, Yang QY, Zhang HY, Zhu Q, Zhang QY (2016) Bioinformatics identification of drug resistance-associated gene pairs in mycobacterium tuberculosis. Int J Mol Sci 17. https://doi.org/10.3390/ijms17091417

Daim S, Kawamura I, Tsuchiya K, Hara H, Kurenuma T, Shen Y, Dewamitta SR, Sakai S, Nomura T, Qu H, Mitsuyama M (2011) Expression of the Mycobacterium tuberculosis PPE37 protein in Mycobacterium smegmatis induces low tumour necrosis factor alpha and interleukin 6 production in murine macrophages. J Med Microbiol 60:582–591. https://doi.org/10.1099/jmm.0.026047-0

Daleke MH, Cascioferro A, de Punder K, Ummels R, Abdallah AM, van der Wel N, Peters PJ, Luirink J, Manganelli R, Bitter W (2011) Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. Journal of Biological Chemistry 286:19024–19034. https://doi.org/10.1074/jbc.M110.204966

Daleke MH, Ummels R, Bawono P, Heringa J, Vandenbroucke-Grauls CMJE, Luirink J, Bitter W (2012a) General secretion signal for the mycobacterial type VII secretion pathway. Proc Natl Acad Sci U S A 109:11342–11347. https://doi.org/10.1073/pnas.1119453109

Daleke MH, van der Woude AD, Parret AHA, Ummels R, de Groot AM, Watson D, Piersma SR, Jiménez CR, Luirink J, Bitter W, Houben ENG (2012b) Specific chaperones for the type VII protein secretion pathway. Journal of Biological Chemistry 287:31939–31947. https://doi.org/10.1074/jbc.M112.397596

Damen XMPM, Phan TH, Ummels R, Rubio-Canalejas A, Bitter W, Houben XENG (2020) Modification of a pe/ppe substrate pair reroutes an esx substrate pair from the mycobacterial esx-1 type VII secretion system to the esx-5 system. Journal of Biological Chemistry 295:5960–5969. https://doi.org/10.1074/jbc.RA119.011682

Dang G, Cao J, Cui Y, Song N, Chen L, Pang H, Liu S (2016) Characterization of Rv0888, a Novel Extracellular Nuclease from Mycobacterium tuberculosis. Sci Rep 6:1–11. https://doi.org/10.1038/srep19033

Dejesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, Rubin EJ, Schnappinger D, Ehrt S, Fortune SM, Sassetti CM, Ioerger TR (2017) Comprehensive essentiality analysis of the Mycobacterium tuberculosis genome via saturating transposon mutagenesis. mBio 8:1–17. https://doi.org/10.1128/mBio.02133-16

Delogu G, Brennan MJ (2001) Comparative immune response to PE and PE_PGRS antigens of Mycobacterium tuberculosis. Infect Immun 69:5606–5611. https://doi.org/10.1128/IAI.69.9.5606-5611.2001

Delogu G, Brennan MJ, Manganelli R (2017a) PE and PPE Genes: A Tale of Conservation and Diversity. Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control, Advances in Experimental Medicine and Biology 135–153. https://doi.org/10.1007/978-3-319-64371-7

Delogu G, Brennan MJ, Manganelli R (2017b) PE and PPE Genes: A Tale of Conservation and Diversity. Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control, Advances in Experimental Medicine and Biology 135–153. https://doi.org/10.1007/978-3-319-64371-7

Denamur E, Matic I (2006) Evolution of mutation rates in bacteria. Mol Microbiol 60:820–827. https://doi.org/10.1111/j.1365-2958.2006.05150.x

Deng A, Boxer SG (2018) Structural Insight into the Photochemistry of Split Green Fluorescent Proteins: A Unique Role for a His-Tag. J Am Chem Soc 140:375–381. https://doi.org/10.1021/jacs.7b10680

Deng W, Xie J (2012) Ins and outs of Mycobacterium tuberculosis PPE family in pathogenesis and implications for novel measures against tuberculosis. J Cell Biochem 113:1087–1095. https://doi.org/10.1002/jcb.23449

Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A, Bukau B (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature 400:693–696. https://doi.org/10.1038/23301

https://scholar.sun.ac.za

Dheda K, Barry CE, Maartens G (2016) Tuberculosis. The Lancet 387:1211–1226. https://doi.org/10.1016/S0140-6736(15)00151-8

Digiuseppe Champion PA, Cox JS (2007a) Protein secretion systems in Mycobacteria. Cell Microbiol 9:1376–1384. https://doi.org/10.1111/j.1462-5822.2007.00943.x

Digiuseppe Champion PA, Cox JS (2007b) Protein secretion systems in Mycobacteria. Cell Microbiol 9:1376–1384. https://doi.org/10.1111/j.1462-5822.2007.00943.x

DiMaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D (2013) Improved low-resolution crystallographic refinement with Phenix and Rosetta. Nat Methods 10:1102–1104. https://doi.org/10.1038/nmeth.2648.Improved

Ding C, Yuan LF, Guo SH, Lin H, Chen W (2012) Identification of mycobacterial membrane proteins and their types using over-represented tripeptide compositions. J Proteomics 77:321–328. https://doi.org/10.1016/j.jprot.2012.09.006

Dixit A, Freschi L, Vargas R, Calderon R, Sacchettini J, Drobniewski F, Galea JT, Contreras C, Yataco R, Zhang Z, Lecca L, Kolokotronis SO, Mathema B, Farhat MR (2019) Whole genome sequencing identifies bacterial factors affecting transmission of multidrug-resistant tuberculosis in a high-prevalence setting. Sci Rep 9:1–10. https://doi.org/10.1038/s41598-019-41967-8

Djinovic K, Carugo O (2015) Missing strings of residues in protein crystal structures. Intrinsically Disord Proteins 3:1–7. https://doi.org/10.1080/21690707.2015.1095697

Dutta NK, Mehra S, Didier PJ, Roy CJ, Doyle LA, Alvarez X, Ratterree M, Be NA, Lamichhane G, Jain SK, Lacey MR, Lackner AA, Kaushal D (2010) Genetic requirements for the survival of tubercle bacilli in primates. Journal of Infectious Diseases 201:1743–1752. https://doi.org/10.1086/652497

Egea PF, Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. Proc Natl Acad Sci U S A 107:17182–17187. https://doi.org/10.1073/pnas.1012556107

Eguchi Y, Makanae K, Hasunuma T, Ishibashi Y, Kito K, Moriya H (2018) Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins. Elife 7. https://doi.org/10.7554/eLife.34595.001

Ehtram A, Shariq M, Ali S, Quadir N, Sheikh JA, Ahmad F, Sharma T, Ehtesham NZ, Hasnain SE (2021) Teleological cooption of Mycobacterium tuberculosis PE/PPE proteins as porins: Role in molecular immigration and emigration. International Journal of Medical Microbiology 311:151495. https://doi.org/10.1016/j.ijmm.2021.151495

Ekiert DC, Cox JS, Hultgren SJ (2014a) Structure of a PE-PPE-EspG complex from mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. Proc Natl Acad Sci U S A 111:14758–14763. https://doi.org/10.1073/pnas.1409345111

Ekiert DC, Cox JS, Hultgren SJ (2014b) Structure of a PE-PPE-EspG complex from mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. Proc Natl Acad Sci U S A 111:14758–14763. https://doi.org/10.1073/pnas.1409345111

Fan M, Rao T, Zacco E, Ahmed MT, Shukla A, Ojha A, Freeke J, Robinson C v., Benesch JL, Lund PA (2012) The unusual mycobacterial chaperonins: Evidence for in vivo oligomerization and specialization of function. Mol Microbiol 85:934–944. https://doi.org/10.1111/j.1365-2958.2012.08150.x

Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PKC, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M (2013) Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet 45:1183–1189. https://doi.org/10.1038/ng.2747

Fay A, Glickman MS (2014) An Essential Nonredundant Role for Mycobacterial DnaK in Native Protein Folding. PLoS Genet 10. https://doi.org/10.1371/journal.pgen.1004516

Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature 431:590–596. https://doi.org/10.1038/nature02944.1.

https://scholar.sun.ac.za

Fishbein S, van Wyk N, Warren RM, Sampson SL (2015) Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. Mol Microbiol 96:901–916. https://doi.org/10.1111/mmi.12981

Forbes BA, Hall GS, Miller MB, Novak SM, Rowlinson M-C, Salfinger M, Somosközi A, Warshauer DM, Wilson ML (2018) Practice Guidelines for Clinical Microbiology Laboratories: Mycobacteria. https://doi.org/10.1128/CMR

Franco-Paredes C, Marcos LA, Henao-Martínez AF, Rodríguez-Morales AJ, Villamil-Gómez WE, Gotuzzo E, Bonifaz A (2019) Cutaneous mycobacterial infections. Clin Microbiol Rev 32:1–25. https://doi.org/10.1128/CMR.00069-18

Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO Journal 29:1552–1564. https://doi.org/10.1038/emboj.2010.52

Gallant J, Heunis T, Beltran C, Schildermans K, Bruijns S, Mertens I, Bitter W, Sampson SL (2021) PPE38-Secretion-Dependent Proteins of M. tuberculosis Alter NF-kB Signalling and Inflammatory Responses in Macrophages. Front Immunol 12:1–22. https://doi.org/10.3389/fimmu.2021.702359

Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, Zeller K, Andrews J, Friedland G (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368:1575–1580. https://doi.org/10.1016/S0140-6736(06)69573-1

Garman EF, Owen RL (2006) Cryocooling and radiation damage in macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 62:32–47. https://doi.org/10.1107/S0907444905034207

Gaur RL, Ren K, Blumenthal A, Bhamidi S, Gibbs S, Jackson M, Zare RN, Ehrt S, Ernst JD, Banaei N (2014) LprG-Mediated Surface Expression of Lipoarabinomannan Is Essential for Virulence of Mycobacterium tuberculosis. PLoS Pathog 10. https://doi.org/10.1371/journal.ppat.1004376

Gey Van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions

Goldstone RM, Moreland NJ, Bashiri G, Baker EN, Shaun Lott J (2008) A new Gateway® vector and expression protocol for fast and efficient recombinant protein expression in Mycobacterium smegmatis. Protein Expr Purif 57:81–87. https://doi.org/10.1016/j.pep.2007.08.015

Goosens VJ, Monteferrante CG, van Dijl JM (2014) The Tat system of Gram-positive bacteria. Biochim Biophys Acta Mol Cell Res 1843:1698–1706. https://doi.org/10.1016/j.bbamcr.2013.10.008

Grāve K, Bennett MD, Högbom M (2022) High-throughput strategy for identification of Mycobacterium tuberculosis membrane protein expression conditions using folding reporter GFP. Protein Expr Purif 106132. https://doi.org/10.1016/j.pep.2022.106132

Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7:1–9. https://doi.org/10.1371/journal.ppat.1002251

Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R (2016) ESX secretion systems: Mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14:677–691. https://doi.org/10.1038/nrmicro.2016.131

Gupta S, Kumar A, Singh K, Kumari R, Sharma A, Singh RK, Pandey SK, Anupurba S (2020) Rv1273c, an ABC transporter of Mycobacterium tuberculosis promotes mycobacterial intracellular survival within macrophages via modulating the host cell immune response. Int J Biol Macromol 142:320–331. https://doi.org/10.1016/j.ijbiomac.2019.09.103

Halle B (2004) Biomolecular cryocrystallography: Structural changes during flash-cooling. Proc Natl Acad Sci U S A 101:4793–4798. https://doi.org/10.1073/pnas.0308315101

Hang NT le, Hijikata M, Maeda S, Thuong PH, Ohashi J, van Huan H, Hoang NP, Miyabayashi A, Cuong VC, Seto S, van Hung N, Keicho N (2019) Whole genome sequencing, analyses of drug resistance-conferring mutations, and correlation with transmission of Mycobacterium tuberculosis carrying katG-S315T in Hanoi, Vietnam. Sci Rep 9. https://doi.org/10.1038/s41598-019-51812-7

https://scholar.sun.ac.za

Harnagel A, Lopez Quezada L, Park SW, Baranowski C, Kieser K, Jiang X, Roberts J, Vaubourgeix J, Yang A, Nelson B, Fay A, Rubin E, Ehrt S, Nathan C, Lupoli TJ (2020) Non-redundant functions of Mycobacterium tuberculosis chaperones promote survival under stress

Henao-Tamayo M, Junqueira-Kipnis AP, Ordway D, Gonzales-Juarrero M, Stewart GR, Young DB, Wilkinson RJ, Basaraba RJ, Orme IM (2007) A mutant of Mycobacterium tuberculosis lacking the 19-kDa lipoprotein Rv3763 is highly attenuated in vivo but retains potent vaccinogenic properties. Vaccine 25:7153–7159. https://doi.org/10.1016/j.vaccine.2007.07.042

Hermans PWM, van Soolingen D, van Embden JDA (1992) Characterization of a major polymorphic tandem repeat in Mycobacterium tuberculosis and its potential use in the epidemiology of Mycobacterium kansasii and Mycobacterium gordonae. J Bacteriol 174:4157–4165. https://doi.org/10.1128/jb.174.12.4157-4165.1992

Hett EC, Chao MC, Steyn AJ, Fortune SM, Deng LL, Rubin EJ (2007) A partner for the resuscitation-promoting factors of Mycobacterium tuberculosis. Mol Microbiol 66:658–668. https://doi.org/10.1111/j.1365-2958.2007.05945.x

Heupink TH, Verboven L, Warren RM, van Rie A (2021) Comprehensive and accurate genetic variant identification from contaminated and low-coverage Mycobacterium tuberculosis whole genome sequencing data. Microb Genom 7. https://doi.org/10.1099/mgen.0.000689

Hickey TBM, Thorson LM, Speert DP, Daffé M, Stokes RW (2009) Mycobacterium tuberculosis Cpn60.2 and DnaK are located on the bacterial surface, where Cpn60.2 facilitates efficient bacterial association with macrophages. Infect Immun 77:3389–3401. https://doi.org/10.1128/IAI.00143-09

Hinchey J, Lee S, Jeon BY, Basaraba RJ, Venkataswamy MM, Chen B, Chan J, Braunstein M, Orme IM, Derrick SC, Morris SL, Jacobs WR, Porcelli SA (2007) Enhanced priming of adaptive immunity by a proapoptotic mutant of Mycobacterium tuberculosis. Journal of Clinical Investigation 117:2279–2288. https://doi.org/10.1172/JCI31947

Hinsley AP, Stanley NR, Palmer T, Berks BC (2001) A naturally occurring bacterial Tat signal peptide lacking one of the "invariant" arginine residues of the consensus targeting motif. FEBS Lett 497:45–49. https://doi.org/10.1016/S0014-5793(01)02428-0

Houben ENG, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jiménez CR, Ottenhoff THM, Luirink J, Bitter W (2012) Composition of the type VII secretion system membrane complex. Mol Microbiol 86:472–484. https://doi.org/10.1111/j.1365-2958.2012.08206.x

Houben ENG, Korotkov K v., Bitter W (2014) Take five - Type VII secretion systems of Mycobacteria. Biochim Biophys Acta Mol Cell Res 1843:1707–1716. https://doi.org/10.1016/j.bbamcr.2013.11.003

Houry WA (2001) Mechanism of substrate recognition by the chaperonin GroEL. Biochemistry and Cell Biology 79:569–577. https://doi.org/10.1139/bcb-79-5-569

Ilghari D, Lightbody KL, Veverka V, Waters LC, Muskett FW, Renshaw PS, Carr MD (2011) Solution structure of the Mycobacterium tuberculosis Esxg•EsxH complex: Functional implications and comparisons with other M. tuberculosis Esx family complexes. Journal of Biological Chemistry 286:29993–30002. https://doi.org/10.1074/jbc.M111.248732

Jajou R, van der Laan T, de Zwaan R, Kamst M, Mulder A, de Neeling A, Anthony R, van Soolingen D (2019) WGS more accurately predicts susceptibility of Mycobacterium tuberculosis to first-line drugs than phenotypic testing. Journal of Antimicrobial Chemotherapy 74:2605–2616. https://doi.org/10.1093/jac/dkz215

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Kanji A, Hasan Z, Ali A, McNerney R, Mallard K, Coll F, Hill-Cawthorne G, Nair M, Clark TG, Zaver A, Jafri S, Hasan R (2015) Characterization of genomic variations in SNPs of PE_PGRS genes reveals deletions and insertions in extensively drug resistant (XDR) M. tuberculosis strains from Pakistan. Int J Mycobacteriol 4:73–79. https://doi.org/10.1016/j.ijmyco.2014.11.049

Khan M, Hayat M, Khan SA, Ahmad S, Iqbal N (2017a) Bi-PSSM: Position specific scoring matrix based intelligent computational model for identification of mycobacterial membrane proteins. J Theor Biol 435:116–124. https://doi.org/10.1016/j.jtbi.2017.09.013

https://scholar.sun.ac.za

Khan M, Hayat M, Khan SA, Iqbal N (2017b) Unb-DPC: Identify mycobacterial membrane protein types by incorporating un-biased dipeptide composition into Chou's general PseAAC. J Theor Biol 415:13–19. https://doi.org/10.1016/j.jtbi.2016.12.004

Kim DN, Sanbonmatsu KY (2017) Tools for the cryo-EM gold rush: Going from the cryo-EM map to the atomistic model. Biosci Rep 37:1–10. https://doi.org/10.1042/BSR20170072

Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, Wilmanns M, Bitter W, Parret AHA, Houben ENG, Korotkov K v. (2014) Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. Mol Microbiol 94:367–382. https://doi.org/10.1111/mmi.12770

Kramarska E, Squeglia F, de Maio F, Delogu G, Berisio R (2021) Pe_pgrs33, an important virulence factor of mycobacterium tuberculosis and potential target of host humoral immune response. Cells 10:1–20

Kruh NA, Troudt J, Izzo A, Prenni J, Dobos KM (2010) Portrait of a pathogen: The Mycobacterium tuberculosis proteome In vivo. PLoS One 5. https://doi.org/10.1371/journal.pone.0013938

Kudryashev M, Wang RYR, Brackmann M, Scherer S, Maier T, Baker D, Dimaio F, Stahlberg H, Egelman EH, Basler M (2015) Structure of the Type VI secretion system contractile sheath. Cell 160:952–962. https://doi.org/10.1016/j.cell.2015.01.037

Kumar CMS, Khare G, Srikanth C v., Tyagi AK, Sardesai AA, Mande SC (2009) Facilitated oligomerization of mycobacterial GroEL: Evidence for phosphorylation-mediated oligomerization. J Bacteriol 191:6525–6538. https://doi.org/10.1128/JB.00652-09

Kumar GS, Sobhia ME, Ghosh K (2022) Binding affinity analysis of quinolone and dione inhibitors with Mtb-DNA gyrase emphasising the crystal water molecular transfer energy to the protein–ligand association. Mol Simul 48:631–646. https://doi.org/10.1080/08927022.2022.2042530

Kurtz S, McKinnon KP, Runge MS, Ting JPY, Braunstein M (2006) The SecA2 secretion factor of Mycobacterium tuberculosis promotes growth in macrophages and inhibits the host immune response. Infect Immun 74:6855–6864. https://doi.org/10.1128/IAI.01022-06

Lamichhane G, Zignol M, Blades NJ, Geiman DE, Dougherty A, Grosset J, Broman KW, Bishai WR (2003) A postgenomic method for predicting essential genes at subsaturation levels of mutagenesis: Application to Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 100:7213–7218

Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301

Leversen NA, de Souza GA, Målen H, Prasad S, Jonassen I, Wiker HG (2009) Evaluation of signal peptide prediction algorithms for identification of mycobacterial signal peptides using sequence data from proteomic methods. Microbiology (N Y) 155:2375–2383. https://doi.org/10.1099/mic.0.025270-0

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li W, Deng W, Xie J (2019a) Expression and regulatory networks of Mycobacterium tuberculosis PE/PPE family antigens. J Cell Physiol 234:7742–7751. https://doi.org/10.1002/jcp.27608

Li Z, Wu D, Zhan B, Hu X, Gan J, Ji C, Li J (2019c) Structural insights into the complex of trigger factor chaperone and ribosomal protein S7 from Mycobacterium tuberculosis. Biochem Biophys Res Commun 512:838–844. https://doi.org/10.1016/j.bbrc.2019.03.166

Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expr Purif 41:98–105. https://doi.org/10.1016/j.pep.2005.01.019

Ligon LS, Hayden JD, Braunstein M (2012) The ins and outs of Mycobacterium tuberculosis protein export. Tuberculosis 92:121–132. https://doi.org/10.1016/j.tube.2011.11.005

Lou Y, Rybniker J, Sala C, Cole ST (2017) EspC forms a filamentous structure in the cell envelope of Mycobacterium tuberculosis and impacts ESX-1 secretion. Mol Microbiol 103:26–38. https://doi.org/10.1111/mmi.13575

https://scholar.sun.ac.za

Lü L, Cao HD, Zeng HQ, Wang PL, Wang LJ, Liu SN, Xiang TX (2009) Recombinant Mycobacterium smegmatis mc2 155 vaccine expressing outer membrane protein 26 kDa antigen affords therapeutic protection against Helicobacter pylori infection. Vaccine 27:972–978. https://doi.org/10.1016/j.vaccine.2008.12.003

Lu Z, Wang H, Yu T (2016) The SecB-like chaperone Rv1957 from Mycobacterium tuberculosis: Crystallization and X-ray crystallographic analysis. Acta Crystallographica Section:F Structural Biology Communications 72:457–461. https://doi.org/10.1107/S2053230X16007287

Luirink J, Sinning I (2004) SRP-mediated protein targeting: Structure and function revisited. Biochim Biophys Acta Mol Cell Res 1694:17–35. https://doi.org/10.1016/j.bbamcr.2004.03.013

Machowski EE, Senzani S, Ealand C, Kana BD (2014) Comparative genomics for mycobacterial peptidoglycan remodelling enzymes reveals extensive genetic multiplicity. BMC Microbiol 14. https://doi.org/10.1186/1471-2180-14-75

Maciuca S, Elias CDO, McVean G, Iqbal Z (2016) A natural encoding of genetic variation in a burrows-wheeler transform to enable mapping and genome inference. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9838 LNCS:222–233. https://doi.org/10.1007/978-3-319-43681-4_18

Målen H, de Souza GA, Pathak S, Søfteland T, Wiker HG (2011) Comparison of membrane proteins of Mycobacterium tuberculosis H37Rv and H37Ra strains. BMC Microbiol 11. https://doi.org/10.1186/1471-2180-11-18

Martin-Garcia JM, Conrad CE, Coe J, Roy-Chowdhury S, Fromme P (2016) Review: Serial Femtosecond Crystallography: A Revolution in Structural Biology. Arch Biochem Biophys 32–47. https://doi.org/10.1016/j.abb.2016.03.036.Review

Mazandu GK, Mulder NJ (2012) Function prediction and analysis of mycobacterium tuberculosis hypothetical proteins. Int J Mol Sci 13:7283–7302. https://doi.org/10.3390/ijms13067283

McDonough JA, Hacker KE, Flores AR, Pavelka MS, Braunstein M (2005) The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial β-lactamases. J Bacteriol 187:7667–7679. https://doi.org/10.1128/JB.187.22.7667-7679.2005

McDonough JA, McCann JR, Tekippe EME, Silverman JS, Rigel NW, Braunstein M (2008) Identification of functional Tat signal sequences in Mycobacterium tuberculosis proteins. J Bacteriol 190:6428–6438. https://doi.org/10.1128/JB.00749-08

McEvoy CRE, Cloete R, Müller B, Schürch AC, van Helden PD, Gagneux S, Warren RM, Gey van Pittius NC (2012) Comparative analysis of mycobacterium tuberculosis pe and ppe genes reveals high sequence variation and an apparent absence of selective constraints. PLoS One 7. https://doi.org/10.1371/journal.pone.0030593

Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, Farhat MR, Guthrie JL, Laukens K, Miotto P, Ofori-Anyinam B, Dreyer V, Supply P, Suresh A, Utpatel C, van Soolingen D, Zhou Y, Ashton PM, Brites D, Cabibbe AM, de Jong BC, de Vos M, Menardo F, Gagneux S, Gao Q, Heupink TH, Liu Q, Loiseau C, Rigouts L, Rodwell TC, Tagliani E, Walker TM, Warren RM, Zhao Y, Zignol M, Schito M, Gardy J, Cirillo DM, Niemann S, Comas I, van Rie A (2019) Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. Nat Rev Microbiol 17:533–545. https://doi.org/10.1038/s41579-019-0214-5

Mi Y, Bao L, Gu D, Luo T, Sun C, Yang G (2017) Mycobacterium tuberculosis PPE25 and PPE26 proteins expressed in Mycobacterium smegmatis modulate cytokine secretion in mouse macrophages and enhance mycobacterial survival. Res Microbiol 168:234–243. https://doi.org/10.1016/j.resmic.2016.06.004

Miao J, Ishikawa T, Robinson IK, Murnane MM (2015) Beyond crystallography: Diffractive imaging using coherent X-ray light sources. Science (1979) 348:530–535. https://doi.org/10.1126/science.aaa1394

Miller BK, Zulauf KE, Braunstein M (2017) The Sec Pathways and Exportomes of Mycobacterium tuberculosis. Microbiol Spectr 5. https://doi.org/10.1128/microbiolspec.tbtb2-0013-2016

Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinger CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNerney R, Ezewudo M, Cirillo DM, Schito M, Köser CU, Rodwell TC (2017) A

https://scholar.sun.ac.za

standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. European Respiratory Journal 50. https://doi.org/10.1183/13993003.01354-2017

Mitra A, Speer A, Lin K, Ehrt S, Niederweisa M (2017) PPE Surface Proteins Are Required for Heme Utilization by Mycobacterium tuberculosis. mBio 8:1–14

Miyashita O, Joti Y (2017) X-ray free electron laser single-particle analysis for biological systems. Curr Opin Struct Biol 43:163–169. https://doi.org/10.1016/j.sbi.2017.03.014

Modlin SJ, Elghraoui A, Gunasekaran D, Zlotnicki AM, Dillon NA, Dhillon N, Kuo N, Robinhold C, Chan CK, Baughn AD, Valafar F (2021) Structure-Aware Mycobacterium tuberculosis Functional Annotation Uncloaks Resistance, Metabolic, and Virulence Genes

Mouton JM, Heunis T, Dippenaar A, Gallant JL, Kleynhans L, Sampson SL (2019) Comprehensive characterization of the attenuated double auxotroph mycobacterium tuberculosisΔleudΔpanCD as an alternative to h37Rv. Front Microbiol 10:1–13. https://doi.org/10.3389/fmicb.2019.01922

Nair S, Pandey AD, Mukhopadhyay S (2011) The PPE18 Protein of Mycobacterium tuberculosis Inhibits NF-κB/rel–Mediated Proinflammatory Cytokine Production by Upregulating and Phosphorylating Suppressor of Cytokine Signaling 3 Protein . The Journal of Immunology 186:5413–5424. https://doi.org/10.4049/jimmunol.1000773

Nair S, Ramaswamy PA, Ghosh S, Joshi DC, Pathak N, Siddiqui I, Sharma P, Hasnain SE, Mande SC, Mukhopadhyay S (2009) The PPE18 of Mycobacterium tuberculosis Interacts with TLR2 and Activates IL-10 Induction in Macrophage . The Journal of Immunology 183:6269–6281. https://doi.org/10.4049/jimmunol.0901367

Nakayama H, Kurokawa K, Lee BL (2012) Lipoproteins in bacteria: Structures and biosynthetic pathways. FEBS Journal 279:4247–4268. https://doi.org/10.1111/febs.12041

Newton GL, Koledin T, Gorovitz B, Rawat M, Fahey RC, Av-Gay Y (2003) The glycosyltransferase gene encoding the enzyme catalyzing the first step of mycothiol biosynthesis (mshA). J Bacteriol 185:3476–3479. https://doi.org/10.1128/JB.185.11.3476-3479.2003

Newton-Foot M, Warren RM, Sampson SL, van Helden PD, Gey Van Pittius NC (2016) The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems. BMC Evol Biol 16. https://doi.org/10.1186/s12862-016-0631-2

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Nikaido H (1994) Prevention of Drug Access to Bacterial Targets: Permeability Barriers and Active Efflux. Science (1979) 264:382–388

Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS (2010) Mycobacterium tuberculosis MycP1 Protease Plays a Dual Role in Regulation of ESX-1 Secretion and Virulence. Cell Host Microbe 7:210–220. https://doi.org/10.1016/j.chom.2010.02.006

Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF (2005) GroEL1: A dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. Cell 123:861–873. https://doi.org/10.1016/j.cell.2005.09.012

Oldfield CJ, Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem 83:553–584. https://doi.org/10.1146/annurev-biochem-072711-164947

Ong E, He Y, Yang Z (2020) Epitope promiscuity and population coverage of Mycobacterium tuberculosis protein antigens in current subunit vaccines under development. Infection, Genetics and Evolution 80:104186. https://doi.org/10.1016/j.meegid.2020.104186

Oostvogels S, Ley SD, Heupink TH, Dippenaar A, Streicher EM, de Vos E, Meehan CJ, Dheda K, Warren R, van Rie A (2022) Transmission, distribution and drug resistance-conferring mutations of extensively drug-resistant tuberculosis in the Western Cape Province, South Africa. Microb Genom 8. https://doi.org/10.1099/mgen.0.000815

https://scholar.sun.ac.za

Pajuelo D, Tak U, Zhang L, Danilchanka O, Tischler AD, Niederweis M (2021) Toxin secretion and trafficking by Mycobacterium tuberculosis. Nat Commun 12:1–13. https://doi.org/10.1038/s41467-021-26925-1

Palmer T, Berks BC (2012) The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol 10:483–496. https://doi.org/10.1038/nrmicro2814

Park HD, Guinn KM, Harrell MI, Liao R, Voskuil MI, Tompa M, Schoolnik GK, Sherman DR (2003) Rv3133c/dosR is a transcription factor that mediates the hypoxic response of Mycobacterium tuberculosis. Mol Microbiol 48:833–843. https://doi.org/10.1046/j.1365-2958.2003.03474.x

Patzelt H, Rüdiger S, Brehmer D, Kramer G, Vorderwülbecke S, Schaffitzel E, Waitz A, Hesterkamp T, Dong L, Schneider-Mergener J, Bukau B, Deuerling E (2001) Binding specificity of Escherichia coli trigger factor. Proc Natl Acad Sci U S A 98:14244–14249. https://doi.org/10.1073/pnas.261432298

Phan TH, Houben ENG (2018) Bacterial secretion chaperones: The mycobacterial type VII case. FEMS Microbiol Lett 365:1–8. https://doi.org/10.1093/femsle/fny197

Phan TH, Ummels R, Bitter W, Houben ENG (2017) Identification of a substrate domain that determines system specificity in mycobacterial type VII secretion systems. Sci Rep 7:1–9. https://doi.org/10.1038/srep42704

Phelan JE, Coll F, Bergval I, Anthony RM, Warren R, Sampson SL, Gey van Pittius NC, Glynn JR, Crampin AC, Alves A, Bessa TB, Campino S, Dheda K, Grandjean L, Hasan R, Hasan Z, Miranda A, Moore D, Panaiotov S, Perdigao J, Portugal I, Sheen P, de Oliveira Sousa E, Streicher EM, van Helden PD, Viveiros M, Hibberd ML, Pain A, McNerney R, Clark TG (2016) Recombination in pe/ppe genes contributes to genetic variation in Mycobacterium tuberculosis lineages. BMC Genomics 17:1–12. https://doi.org/10.1186/s12864-016-2467-y

Portelli S, Phelan JE, Ascher DB, Clark TG, Furnham N (2018) Understanding molecular consequences of putative drug resistant mutations in Mycobacterium tuberculosis. Sci Rep 8. https://doi.org/10.1038/s41598-018-33370-6

Posey JE, Shinnick TM, Quinn FD (2006) Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis. J Bacteriol 188:1332–1340. https://doi.org/10.1128/JB.188.4.1332-1340.2006

Poulet S, Cole ST (1995) Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in Mycobacterium tuberculosis. Arch Microbiol 163:87–95. https://doi.org/10.1007/BF00381781

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Res 40:290–301. https://doi.org/10.1093/nar/gkr1065

Qamra R, Mande SC (2004) Crystal structure of the 65-kilodalton heat shock protein, chaperonin 60.2, of Mycobacterium tuberculosis. J Bacteriol 186:8105–8113. https://doi.org/10.1128/JB.186.23.8105-8113.2004

Qamra R, Srinivas V, Mande SC (2004) Mycobacterium tuberculosis GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. J Mol Biol 342:605–617. https://doi.org/10.1016/j.jmb.2004.07.066

Qian J, Chen R, Wang H, Zhang X (2020) Role of the PE/PPE Family in Host–Pathogen Interactions and Prospects for Anti-Tuberculosis Vaccine and Diagnostic Tool Design. Front Cell Infect Microbiol 10

Quiblier C, Zinkernagel AS, Schuepbach RA, Berger-Bächi B, Senn MM (2011) Contribution of SecDF to Staphylococcus aureus resistance and expression of virulence factors. BMC Microbiol 11. https://doi.org/10.1186/1471-2180-11-72

Rachman H, Strong M, Timo U, Leander G, Johannes S, Mollenkopf H, Kosmiadi G, Eisenberg D, Kaufman S (2006) Unique Transcriptome Signature of Mycobacterium tuberculosis in Pulmonary Tuberculosis. American Society for Microbiology 74:1233–1242. https://doi.org/10.1128/IAI.74.2.1233

Ramasamy S, Abrol R, Suloway CJM, Clemons WM (2013) The glove-like structure of the conserved membrane protein tatc provides insight into signal sequence recognition in twin-arginine translocation. Structure 21:777–788. https://doi.org/10.1016/j.str.2013.03.004

Reddy PV, Puri RV, Khera A, Tyagi AK (2012) Iron storage proteins are essential for the survival and pathogenesis of Mycobacterium tuberculosis in THP-1 macrophages and the guinea pig model of infection. J Bacteriol 194:567–575. https://doi.org/10.1128/JB.05553-11

https://scholar.sun.ac.za

Reed SG, Coler RN, Dalemans W, Tan E v., dela Cruz EC, Basaraba RJ, Orme IM, Skeiky YAW, Alderson MR, Cowgill KD, Prieels JP, Abalos RM, Dubois MC, Cohen J, Mettens P, Lobet Y (2009) Erratum: Defined tuberculosis vaccine, Mtb72F/AS02A, evidence of protection in cynomolgus monkeys (Proceedings of the National Academy of Sciences of the United States of America (2009) 106 (2301-2306) DOI:10.1073/pnas.0712077106). Proc Natl Acad Sci U S A 106:7678. https://doi.org/10.1073/pnas.0901917106

Renshaw PS, Lightbody KL, Veverka V, Muskett FW, Kelly G, Frenkiel TA, Gordon S v., Hewinson RG, Burke B, Norman J, Williamson RA, Carr MD (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO Journal 24:2491–2498. https://doi.org/10.1038/sj.emboj.7600732

Renshaw PS, Panagiotidou P, Whelan A, Gordon S v., Glyn Hewinson R, Williamson RA, Carr MD (2002) Conclusive evidence that the major T-cell antigens of the Mycobacterium tuberculosis complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6·CFP-10 complex. Implications for p. Journal of Biological Chemistry 277:21598–21603. https://doi.org/10.1074/jbc.M201625200

Rezaei MA, Abdolmaleki P, Karami Z, Asadabadi EB, Sherafat MA, Abrishami-Moghaddam H, Fadaie M, Forouzanfar M (2008) Prediction of membrane protein types by means of wavelet analysis and cascaded neural networks. J Theor Biol 254:817–820. https://doi.org/10.1016/j.jtbi.2008.07.012

Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, Jäger F, Johnson S, Krehenbrink M, Liu SM, Lukey MJ, Marcoux J, McDowell MA, Rodriguez F, Roversi P, Stansfeld PJ, Robinson C v., Sansom MSP, Palmer T, Högbom M, Berks BC, Lea SM (2012) Structure of the TatC core of the twin-arginine protein transport system. Nature 492:210–214. https://doi.org/10.1038/nature11683

Roy S, Ghatak D, Das P, BoseDasgupta S (2020) ESX secretion system: The gatekeepers of mycobacterial survivability and pathogenesis. Eur J Microbiol Immunol (Bp) 10:202–209. https://doi.org/10.1556/1886.2020.00028

Rüdiger S, Germeroth L, Schneider-Mergener J, Bukau B (1997) Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. EMBO J 16:1501–1507

Saini NK, Baena A, Ng TW, Venkataswamy MM, Kennedy SC, Kunnath-Velayudhan S, Carreño LJ, Xu J, Chan J, Larsen MH, Jacobs WR, Porcelli SA (2016) Suppression of autophagy and antigen presentation by Mycobacterium tuberculosis PE_PGRS47. Nat Microbiol 1:139–148. https://doi.org/10.1038/nmicrobiol.2016.133.Suppression

Sampson SL (2011) Mycobacterial PE/PPE proteins at the host-pathogen interface. Clin Dev Immunol 2011. https://doi.org/10.1155/2011/497203

Sampson SL, Lukey P, Warren RM, van Helden PD, Richardson M, Everett MJ (2001) Expression, characterization and subcellular localization of the Mycobacterium tuberculosis PPE gene Rv1917c. Tuberculosis 81:305–317. https://doi.org/10.1054/tube.2001.0304

Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. Mol Microbiol 48:77–84. https://doi.org/10.1046/j.1365-2958.2003.03425.x

Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100:12989–12994. https://doi.org/10.1073/pnas.2134250100

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: Insights into the phagosomal environment. Journal of Experimental Medicine 198:693–704. https://doi.org/10.1084/jem.20030846

Schrag SJ, Perrot V, Levin BR (1997) Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. Proceedings of the Royal Society B: Biological Sciences 264:1287–1291. https://doi.org/10.1098/rspb.1997.0178

Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, Gengenbacher M, Maiolica A, Kaufmann SHE, Moritz RL, Aebersold R (2013) The Mtb proteome library: A resource of assays to quantify the complete proteome of mycobacterium tuberculosis. Cell Host Microbe 13:602–612. https://doi.org/10.1016/j.chom.2013.04.008

Scott JR, Barnett TC (2006) Surface Proteins of Gram-Positive Bacteria and How They Get There. Annu Rev Microbiol 60:397–423. https://doi.org/10.1146/annurev.micro.60.080805.142256

https://scholar.sun.ac.za

Shamu S, Kuwanda L, Farirai T, Guloba G, Slabbert J, Nkhwashu N (2019) Study on knowledge about associated factors of Tuberculosis (TB) and TB/HIV co-infection among young adults in two districts of South Africa. PLoS One 14:1–13. https://doi.org/10.1371/journal.pone.0217836

Sharma T, Alam A, Ehtram A, Rani A, Grover S, Ehtesham NZ, Hasnain SE (2022) The Mycobacterium tuberculosis PE_PGRS Protein Family Acts as an Immunological Decoy to Subvert Host Immune Response. Int J Mol Sci 23. https://doi.org/10.3390/ijms23010525

Sharma V, Arockiasamy A, Ronning DR, Savva CG, Holzenburg A, Braunstein M, Jacobs WR, Sacchettini JC (2003) Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase. Proc Natl Acad Sci U S A 100:2243–2248. https://doi.org/10.1073/pnas.0538077100

Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc Natl Acad Sci U S A 106:18792–18797. https://doi.org/10.1073/pnas.0900589106

Sielaff B, Lee KS, Tsai FTF (2011) Structural and functional conservation of mycobacterium tuberculosis GroEL paralogs suggests that GroEL1 Is a chaperonin. J Mol Biol 405:831–839. https://doi.org/10.1016/j.jmb.2010.11.021

Sierra RG, Gati C, Laksmono H, Dao EH, Gul S, Fuller F, Kern J, Chatterjee R, Ibrahim M, Brewster AS, Young ID, Michels-clark T, Aquila A, Liang M, Hunter MS, Koglin JE, Boutet S, Junco EA, Hayes B, Bogan MJ, Hampton CY, Puglisi E v, Sauter NK, Stan CA, Yano J, Yachandra VK, Soltis SM, Puglisi JD, Demirci H (2016) Concentric-Flow Electrokinetic Injector Enables Serial Crystallography of Ribosome and Photosystem-II. Nat Methods 13:59–62. https://doi.org/10.1038/nmeth.3667.Concentric-Flow

Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, Enninga J (2012) Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. PLoS Pathog 8. https://doi.org/10.1371/journal.ppat.1002507

Singh KK, Dong Y, Patibandla SA, Arora VK, Laal S (2005a) Immunogenicity of the. 73:5004–5014. https://doi.org/10.1128/IAI.73.8.5004

Singh KK, Dong Y, Patibandla SA, McMurray DN, Arora VK, Laal S (2005b) Immunogenicity of the Mycobacterium tuberculosis PPE55 (Rv3347c) protein during incipient and clinical tuberculosis. Infect Immun 73:5004–5014. https://doi.org/10.1128/IAI.73.8.5004-5014.2005

Singh SK, Kumari R, Singh DK, Tiwari S, Singh PK, Sharma S, Srivastava KK (2013) Putative roles of a proline-glutamic acid-rich protein (PE3) in intracellular survival and as a candidate for subunit vaccine against Mycobacterium tuberculosis. Med Microbiol Immunol 202:365–377. https://doi.org/10.1007/s00430-013-0299-9

Skeiky YAW, Alderson MR, Ovendale PJ, Guderian JA, Brandt L, Dillon DC, Campos-Neto A, Lobet Y, Dalemans W, Orme IM, Reed SG (2004) Differential Immune Responses and Protective Efficacy Induced by Components of a Tuberculosis Polyprotein Vaccine, Mtb72F, Delivered as Naked DNA or Recombinant Protein. The Journal of Immunology 172:7618–7628. https://doi.org/10.4049/jimmunol.172.12.7618

Smith J, Manoranjan J, Pan M, Bohsali A, Xu J, Liu J, McDonald KL, Szyk A, LaRonde-LeBlanc N, Gao LY (2008) Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in Mycobacterium marinum escape from the vacuole. Infect Immun 76:5478–5487. https://doi.org/10.1128/IAI.00614-08

Snapper SB, Melton RE, Mustafa S, Kieser T, Jr WRJ (1990) Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Mol Microbiol 4:1911–1919. https://doi.org/10.1111/j.1365-2958.1990.tb02040.x

Soldini S, Palucci I, Zumbo A, Sali M, Ria F, Manganelli R, Fadda G, Delogu G (2011) PPE-MPTR genes are differentially expressed by Mycobacterium tuberculosis in vivo. Tuberculosis 91:563–568. https://doi.org/10.1016/j.tube.2011.08.002

Solomonson M, Setiaputra D, Makepeace KAT, Lameignere E, Petrotchenko E v., Conrady DG, Bergeron JR, Vuckovic M, Dimaio F, Borchers CH, Yip CK, Strynadka NCJ (2015) Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. Structure 23:571–583. https://doi.org/10.1016/j.str.2015.01.002

https://scholar.sun.ac.za

Sorensen AL, Nagai S, Houen G, Andersen P, Andersen AB (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. Infect Immun 63:1710–1717. https://doi.org/10.1128/iai.63.5.1710-1717.1995

Spertini F, Audran R, Lurati F, Ofori-Anyinam O, Zysset F, Vandepapelière P, Moris P, Demoitié MA, Mettens P, Vinals C, Vastiau I, Jongert E, Cohen J, Ballou WR (2013) The candidate tuberculosis vaccine Mtb72F/AS02 in PPD positive adults: A randomized controlled phase I/II study. Tuberculosis 93:179–188. https://doi.org/10.1016/j.tube.2012.10.011

Spriestersbach A, Kubicek J, Schäfer F, Block H, Maertens B (2015) Purification of His-Tagged Proteins. In: Methods in Enzymology. Academic Press Inc., pp 1–15

Srivastava A, Nagai T, Srivastava A, Miyashita O, Tama F (2018) Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics. Int J Mol Sci 19. https://doi.org/10.3390/ijms19113401

Srivastava R, Kumar D, Waskar MN, Sharma M, Katoch VM, Srivastava BS (2006) Identification of a repetitive sequence belonging to a PPE gene of Mycobacterium tuberculosis and its use in diagnosis of tuberculosis. J Med Microbiol 55:1071–1077. https://doi.org/10.1099/jmm.0.46379-0

Stanley NR, Palmer T, Berks BC (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli. Journal of Biological Chemistry 275:11591–11596. https://doi.org/10.1074/jbc.275.16.11591

Stanley SA, Raghavan S, Hwang WW, Cox JS (2003) Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proceedings of the National Academy of Sciences 100:13001–13006. https://doi.org/10.1016/0020-7519(79)90047-X

Stewart GR, Patel J, Robertson BD, Rae A, Young DB (2005) Mycobacterial mutants with defective control of phagosomal acidification. PLoS Pathog 1:0269–0278. https://doi.org/10.1371/journal.ppat.0010033

Su X-D, Zhang H, Terwilliger TC, Liljas A, Xiao J, Dong Y (2015) Protein Crystallography from the Perspective of Technology Developments. Crystallogr Rev 21:122–153. https://doi.org/10.1080/0889311X.2014.973868.Protein

Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M, Sacchettini JC (2016) Structural similarities and differences between two functionally distinct SecA proteins, Mycobacterium tuberculosis SecA1 and SecA2. J Bacteriol 198:720–730. https://doi.org/10.1128/JB.00696-15

Talaat AM, Lyons R, Howard ST, Johnston SA (2004) The temporal expression profile of Mycobacterium tuberculosis infection in mice. Proc Natl Acad Sci U S A 101:4602–4607. https://doi.org/10.1073/pnas.0306023101

Thi Hoang D, Chernomor O, von Haeseler A, Quang Minh B, Sy Vinh L, Rosenberg MS (2017) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522. https://doi.org/10.5281/zenodo.854445

Tilton RF, Dewan JC, Petsko GA (1992) Effects of Temperature on Protein Structure and Dynamics: X-ray Crystallographic Studies of the Protein Ribonuclease-A at Nine Different Temperatures from 98 to 320 K. Biochemistry 31:2469–2481. https://doi.org/10.1021/bi00124a006

Tiwari B, Ramakrishnan UM, Raghunand TR (2015) The Mycobacterium tuberculosis protein pair PE9 (Rv1088)-PE10 (Rv1089) forms heterodimers and induces macrophage apoptosis through Toll-like receptor 4. Cell Microbiol 17:1653–1669. https://doi.org/10.1111/cmi.12462

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet De La Salmoniere YOL, Aman K, Kato-Maeda M, Small PM (2004) Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101:4865–4870. https://doi.org/10.1073/pnas.0305634101

Tufariello JAM, Mi K, Xu J, Manabe YC, Kesavan AK, Drumm J, Tanaka K, Jacobs WR, Chan J (2006) Deletion of the Mycobacterium tuberculosis resuscitation-promoting factor Rv1009 gene results in delayed reactivation from chronic tuberculosis. Infect Immun 74:2985–2995. https://doi.org/10.1128/IAI.74.5.2985-2995.2006

https://scholar.sun.ac.za

Valent QA, de Gier JL, Heijne G von, Kendall DA, ten Hagen-Jongman CM, Oudega B, Luirink J (1997) Nascent membrane and presecretory proteins synthesized in Escherichia coli associate with signal recognition particle and trigger factor . Mol Microbiol 25:53–64. https://doi.org/10.1046/j.1365-2958.1997.4431808.x

Voskuil MI, Visconti KC, Schoolnik GK (2004) Mycobacterium tuberculosis gene expression during adaptation to stationary phase and low-oxygen dormancy. Tuberculosis 84:218–227. https://doi.org/10.1016/j.tube.2004.02.003

Wagner S, Baarst L, Ytterberg AJ, Klussmerer A, Wagner CS, Nord O, Nygren PÅ, van Wijks KJ, de Gier JW (2007) Consequences of membrane protein overexpression in Escherichia coli. Molecular and Cellular Proteomics 6:1527–1550. https://doi.org/10.1074/mcp.M600431-MCP200

Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I (2006) The Mycobacterium tuberculosis PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol 60:312–330. https://doi.org/10.1111/j.1365-2958.2006.05102.x

Wandersman C (1993) The general secretory pathway in bacteria. Trends Microbiol 1:249–250. https://doi.org/10.1016/0966-842X(93)90043-Q

Wang L, Wu J, Li J, Yang H, Tang T, Liang H, Zuo M, Wang J, Liu H, Liu F, Chen J, Liu Z, Wang Y, Peng C, Wu X, Zheng R, Huang X, Ran Y, Rao Z, Ge B (2020a) Host-mediated ubiquitination of a mycobacterial protein suppresses immunity. Nature 577:682–688. https://doi.org/10.1038/s41586-019-1915-7

Wang Q, Boshoff HIM, Harrison JR, Ray PC, Green SR, Wyatt PG, Barry CE (2020b) PE/PPE proteins mediate nutrient transport across the outer membrane of Mycobacterium tuberculosis. Science (1979) 367:1147–1151. https://doi.org/10.1126/science.aax3072

Wards BJ, de Lisle GW, Collins DM (2000) An esat6 knockout mutant of Mycobacterium bovis produced by homologous recombination will contribute to the development of a live tuberculosis vaccine. Tubercle and Lung Disease 80:185–189. https://doi.org/10.1054/tuld.2000.0244

WHO (2021) Global Tuberculosis Report 2021

Wiker HG, Wilson MA, Schoolnik GK (2000) Extracytoplasmic proteins of Mycobacterium tuberculosis - Mature secreted proteins often start with aspartic acid and proline. Microbiology (N Y) 146:1525–1533. https://doi.org/10.1099/00221287-146-7-1525

Williamson ZA, Chaton CT, Ciocca WA, Korotkova N, Korotkov K v. (2020) PE5-PPE4-EspG3 trimer structure from mycobacterial ESX-3 secretion system gives insight into cognate substrate recognition by ESX systems

Wolfe LM, Mahaffey SB, Kruh NA, Dobos KM (2010) Proteomic definition of the cell wall of mycobacterium tuberculosis. J Proteome Res 9:5816–5826. https://doi.org/10.1021/pr1005873

Wong D, Li W, Chao JD, Zhou P, Narula G, Tsui C, Ko M, Xie J, Martinez-Frailes C, Av-Gay Y (2018) Protein tyrosine kinase, PtkA, is required for Mycobacterium tuberculosis growth in macrophages. Sci Rep 8. https://doi.org/10.1038/s41598-017-18547-9

Wu X, Tan G, Sha W, Liu H, Yang J, Guo Y, Shen X, Wu Z, Shen H, Yu F (2022) Use of Whole-Genome Sequencing to Predict Mycobacterium tuberculosis Complex Drug Resistance from Early Positive Liquid Cultures. Microbiol Spectr 48–53

Xu Y, Liu F, Chen S, Wu J, Hu Y, Zhu B, Sun Z (2018) In vivo evolution of drug-resistant Mycobacterium tuberculosis in patients during long-term treatment. BMC Genomics 19. https://doi.org/10.1186/s12864-018-5010-5

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)7 chaperonin complex. Nature 388:741–750

Yang D, Klebl DP, Zeng S, Sobott F, Prévost M, Soumillion P, Vandenbussche G, Fontaine V (2020a) Interplays between copper and: Mycobacterium tuberculosis GroEL1. Metallomics 12:1267–1277. https://doi.org/10.1039/d0mt00101e

Yang H, Tang H, Chen XX, Zhang CJ, Zhu PP, Ding H, Chen W, Lin H (2016) Identification of Secretory Proteins in Mycobacterium tuberculosis Using Pseudo Amino Acid Composition. Biomed Res Int 2016. https://doi.org/10.1155/2016/5413903

https://scholar.sun.ac.za

Yang XG, Luo RY, Feng ZP (2007) Using amino acid and peptide composition to predict membrane protein types. Biochem Biophys Res Commun 353:164–169. https://doi.org/10.1016/j.bbrc.2006.12.004

Yang Y, Xu P, He P, Shi F, Tang Y, Guan C, Zeng H, Zhou Y, Song Q, Zhou B, Jiang S, Shao C, Sun J, Yang Y, Wang X, Song H (2020b) Mycobacterial PPE13 activates inflammasome by interacting with the NATCH and LRR domains of NLRP3. FASEB Journal 34:12820–12833. https://doi.org/10.1096/fj.202000200RR

Yu H, Lupoli TJ, Kovach A, Meng X, Zhao G, Nathan CF, Li H (2018) ATP hydrolysis-coupled peptide translocation mechanism of Mycobacterium tuberculosis ClpB. Proc Natl Acad Sci U S A 115:E9560–E9569. https://doi.org/10.1073/pnas.1810648115

Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L (2013) Genome sequencing of 161 Mycobacterium tuberculosis isolates from China identifies genes and intergenic regions associated with drug resistance. Nat Genet 45:1255–1260. https://doi.org/10.1038/ng.2735

Zhang XL, Li DF, Fleming J, Wang LW, Zhou Y, Wang DC, Zhang XE, Bi LJ (2015) Core component EccB1 of the Mycobacterium tuberculosis type VII secretion system is a periplasmic ATPase. FASEB Journal 29:4808–4814. https://doi.org/10.1096/fj.15-270843

Zhang Y, Wang L, Hu Y, Jin C (2014) Solution structure of the TatB component of the twin-arginine translocation system. Biochim Biophys Acta Biomembr 1838:1881–1888. https://doi.org/10.1016/j.bbamem.2014.03.015

Zhang Y, Yew WW (2015) Mechanisms of drug resistance in Mycobacterium tuberculosis: Update 2015 [Mecanismos de farmacorresistencia en Mycobacterium tuberculosis: Actualización 2015]. International Journal of Tuberculosis and Lung Disease 19:1276–1289

Zheng H, Handing KB, Zimmerman MD, Shabalin IG, Almo SC, Minor W (2015) X-ray crystallography over the past decade for novel drug discovery -where are we heading next? Expert Opin Drug Discov 10:975–989. https://doi.org/10.1517/17460441.2015.1061991

Zhu PP, Li WC, Zhong ZJ, Deng EZ, Ding H, Chen W, Lin H (2015) Predicting the subcellular localization of mycobacterial proteins by incorporating the optimal tripeptides into the general form of pseudo amino acid composition. Mol Biosyst 11:558–563. https://doi.org/10.1039/c4mb00645c

Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, Hendrickson WA (1996) Structural Analysis of Substrate Binding by the Molecular Chaperone DnaK. Science (1979) 272:1606–1614

Zvi A, Ariel N, Fulkerson J, Sadoff JC, Shafferman A (2008) Whole genome identification of Mycobacterium tuberculosisvaccine candidates by comprehensive data mining and bioinformatic analyses. BMC Med Genomics 1:1–25. https://doi.org/10.1186/1755-8794-1-18

https://scholar.sun.ac.za

Supplementary Materials

Supplementary Table 6.2: Selected clinical DNA isolates.

Lineage	Drug Resistant Class	Isolate
Lineage 2	DS	43diag, S152, 163dx
Lineage 2.2	MR	R11279, R2493, R9717, R7525, R3119, R4370, R9849
	MDR	R2298, R3054, R2642, R8285, R11121, R10637, R4863, R13920, R15320, R14887, R18277, R5899, R10380, R11509, R13570, R13666, R11470, R9988, R10819, R9725
	XDR	R3881, R4731, R5235, R10125, R4674, R15871, R16939, R5581, R6196, R12774, R13587, R15282, R22052, R28375, R11850, R12830, R13923, R15305
Lineage 4	DS	29dx, 41w4, 105dx, 126dx, 153dx
Lineage 4.1	MR	R19353, R21472, R24356
	MDR	R4545, R5319, R5833, R13559, R15126, R15343, R16859, R16758, R7895, R4247, R9161, R14747
	XDR	R5847, R6700, R10512, R19366, MD10, MD14, MD9
Lineage 4.3	MDR	R23437, R4150, R20277
	XDR	R26348, R32646, R16074, R19713, R19726, R10398
Lineage 4.9	MR	R11364, R13913
	MDR	R1152, R1631, R2311, R3572
	XDR	R2404, R4971, R17095, R17300, R18143

https://scholar.sun.ac.za

Appendix

Appendix Table 6. 2: Stock Solutions used during the project.

Stock Solution	Preparation
EDTA Stock [0.5 M] (500 ml)	Dissolve 93.05 g of EDTA disodium salt (MW=372.24 g/mol) in 400 ml distilled H_2O. Adjust pH to 8.0, then adjust to final volume of 500 ml with distilled H_2O and autoclave at 121 °C for 20 min. Store at 4°C in dark.
TAE [50x] (2 L)	Dissolve 484 g of Tris-base (MW = 121.14 g/mol) in 1.5 L of distilled H_2O. Add 114.2 ml of glacial ascetic acid [100%] and 200 ml of EDTA [0.5 M]. Adjust to final volume of 2 L using distilled H_2O. Store at room temperature. Dilute 1:50 using distilled H_2O when 1x TAE is required.
Glycerol (10%) (1 L)	Dilute 100 ml of molecular-biology-grade glycerol [100%] in 900 ml of distilled H_2O. Autoclave at 121 °C for 20 min. Store at 4°C.
$MgCl_2$ [100 mM] (100ml)	Dissolve 0.95 g of $MgCl_2$ (MW = 95.211 g/mol) in 90 ml distilled H_2O. Once fully incorporated add distilled H_2O to 100ml. Autoclave at 121 °C for 20 min. Store at 4°C.
$CaCl_2$ [100 mM] – Glycerol [15%] (100ml)	Dissolve 1.11 g of $CaCl_2$ in 80 ml of distilled H_2O. Add 15 ml Glycerol [100%]. Once incorporated increase volume to 100 ml with distilled H_2O. Autoclave at 121 °C for 20 min. Store at 4°C.
TBST [10x] (1 L)	Dissolve 24.23 g Tris-base (MW = 121.14 g/mol) and 87.66 g of NaCl (MW = 58.44 g/mol) in 800 ml of distilled H_2O. Add 50ml of Tween-20 [20%]. Adjust pH to 7.5 using HCl. Adjust volume to 1 L using distilled H_2O. Store at 4°C. Dilute 1:10 using distilled H_2O when 1x TBST is required.
Protein Gel Loading Buffer	Mix NuPAGE, LDS Sample buffer [4x] with β-Mercaptoethanol (10%) in a 1:1 ratio. Store at room temperature.