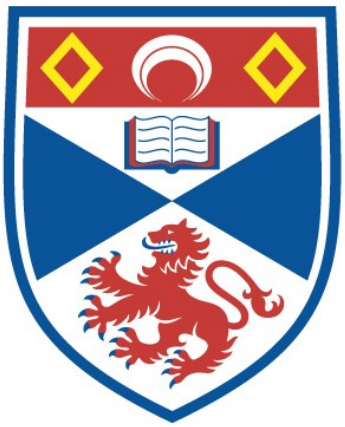
Evaluating *Mycobacterium komossense* JERRO1 as a possible model organism for *Mycobacterium tuberculosis* antibiotic resistance studies

Joanna Esther Rivas

A thesis submitted for the degree of PhD
at the
University of St Andrews



2025

Identifier to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/1047

Table of Contents

List of Figures

List of Tables

List of Abbreviations

°C	Degree Celsius
µg	Microgram
µl	Microliter
ABC	ATP-Binding Cassette
ACT	Artemis Comparison Tool
ADP	Adenosine Diphosphate
AIDS	Acquired immunodeficiency syndrome
AMR	Antimicrobial resistance
ANI	Average Nucleotide Identity
AST	Antimicrobial Susceptibility testing
ATCC	American Type Culture Collection
ATP	Adenosine Trisphosphate
BCG	Bacillus Calmette-Guérin
BDQ	Bedaquiline
BMRC	British Medical Research Council
BSL	Biosafety Level
CD	Cluster of Differentiation
CDS	Protein Coding Sequences
CFU	Colony Forming Units
CLSI	Clinical and Laboratory Standards Institute
COG	Cluster of Orthologous Groups
CSI	Clade of Specific Indels
CSP	Clade of Specific Proteins
DLM	Delamanid
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DR-TB	Drug-Susceptible Tuberculosis
EEA1	Early Endosomal Autoantigen 1
EMB	Ethambutol
EPTB	Extrapulmonary Tuberculosis
ETHO	Ethionamide
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FDA	Food and Drug Administration
HIV	Human Immunodeficiency Virus
INH	Isoniazid
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KH_2PO_4	Monopotassium Phosphate
LEVO	Levofloxacin

LIALC	Lock-In Amplifier, Laser Controller
LJ	Lowenstein-Jensen medium
LMIC	Low-and Middle-Income Countries
LZD	Linezolid
MATE	Multidrug and Toxic Compound Extrusion Family
MBC	Minimum Bactericidal Concentration
MDR-TB	Multi-Drug Resistant Tuberculosis
MFS	Major Facilitator Superfamily
MGIT	Mycobacteria Growth Indicator Tube
MIC	Minimum Inhibitory Concentration
ml	Millilitre
MOXI	Moxifloxacin
MQH_2O	Milli-Q Water
mRNA	Messenger Ribonucleic Acid
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* Complex
NCBI	National Center for Biotechnology Information
NCTC	National Collection of Type Cultures
NIH	National Library of Medicine
NTM	Nontuberculous mycobacteria
OADC	Oleic Albumin Dextrose Catalase
OD	Optical Density
PA	Pretomanid
PAMP	Pathogen-Associated Molecular Patterns
PAS	Para-amino Acid Salt
PBS	Phosphate-buffered saline
PD	Pharmacodynamics
pH	Potential of Hydrogen
PK	Pharmacokinetic
PTB	Pulmonary tuberculosis
PZA	Pyrazinamide
RBS	Ribosomal Binding Site
RD1	Natural Region of Difference 1
RGM	Rapid-Growing Mycobacteria
RIE	Royal Infirmary of Edinburgh
RIF	Rifampicin
RNA	Ribonucleic Acid
RND	Resistance-Nodulation-Cell-division Superfamily
ROI	Reactive Oxygen Intermediates
rpm	Revolutions per Minute
RRDR	Rifampicin Resistance-Determining Region
rRNA	Ribosomal Ribonucleic Acid

SGM	Slow-Growing Mycobacteria
SLIC	Scattered Light Integrated Collection
SM	Streptomycin
SMRL	Scottish Mycobacteria Reference Laboratory
SNP	Single Nucleotide Polymorphisms
TB	Tuberculosis
tRNA	Transfer Ribonucleic Acid
WGS	Whole Genome Sequencing
WHO	World Health Organization
XDR-TB	Extensively Drug Resistant Tuberculosis

Acknowledgments

First and foremost, I would like to praise and thank God for His countless blessings and the opportunity given to submit this thesis. I would like to thank the University of St Andrews for providing me the opportunity to study my PhD on its facilities and to my sponsors, the Panamanian Government through the National Secretary of Science, Technology, and Innovation (SENACYT) for the financial support given to me during these five years, and to IFARHU for processing all the financial matters related to my scholarship.

I would like to thank Professor Matt Holden, Dr Andreas Haag, and Dr Derek Sloan for their help and support during this time, for their encouragement, and their never-ending patience, and to all the members of the labs 249 and 248. Additionally, I would like to thank Dr Robert Hammond for his support in allowing me to use the SLIC device and for his funding so I could sequence the RIF and INH *M.* JERR01 *M.* JERR01 mutants.

I would like to give many thanks to Professor David Harrison, Dr Morven Shearer, Dr Paul Johnston, Dr Evelin Dombay and Krzysztof Mariasnki for their help, and their support. To my dear friends Gina Gnanasampanthan, Hellen Onyango, and Stuart Reid. I would like to also give thanks to Dr Christina Paulus (RIP) for her help, kindness.

To my parents, who have always been there for me and to whom this thesis is fully dedicated to, I would give all my admiration and gratitude, for their never-ending love, support, and for always listening to me during this long journey.

David, thank you for being there for me in the difficult times and being a source of support, for your help, your love, and for being a beautiful part of my life.

Thesis declaration

Candidate's declaration

I, Joanna Esther Rivas Ramos, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 40,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in May 2018.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

01/04/2024

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

Date 1/4/2024 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Joanna Esther Rivas Ramos, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

01/04/2024

Date Signature of candidate

Date 1/4/2024 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Joanna Esther Rivas Ramos, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

01/04/2024

Date Signature of candidate

Abstract

Multi-drug-resistant tuberculosis (MDR-TB) accounted for an estimated 410,000 cases of all TB infections in 2022 and presents a major challenge for global health. Studying TB is complicated due to the fastidiousness and high pathogenicity of *Mycobacterium tuberculosis* (*Mtb*) that requires specialized facilities, expensive equipment, and long experimentation phases. To overcome these challenges, various *Mycobacterium* species have been developed as model organisms that share some of the characteristics of *Mtb* but are less pathogenic and more amenable to experimentation have been proposed in the past. All these models have limitations, therefore new models for *Mtb* that better replicate properties of the pathogen are needed.

This thesis investigates a mycobacterial strain, JERR01, that was originally identified as *Mycobacterium komossense* as a model organism for the development of antimicrobial resistance (AMR) towards key drugs used for the treatment of TB. Through initial phenotypic characterization of the strain that originated from a human urine sample, it was determined that this strain was fast-growing and shared similar antibiotic sensitivity profiles to *Mtb* for six antitubercular drugs, suggesting suitability for studying AMR development. Genomic characterisation of the strain and comparison to a type strain of *M. komossense* revealed that its original taxonomic designation was imprecise. Wider comparison genomic analyses provided evidence that JERR01 belonged to the *Fortuitum-Vaccae* clade of *Mycobacterium* but designated it as an unclassified mycobacterial species. Having identified homologs of genes associated with resistance to antitubercular drugs in *M.* JERR01, *in vitro* studies were conducted to see how

resistance to rifampicin and isoniazid, two key drugs used in tuberculosis treatment, could evolve through *in vitro* experimentation. For rifampicin, similar mutational pathways to *Mtb* were identified in the *rpoB* gene, the primary target of resistance. The mutations associated with decreased sensitivity to isoniazid were almost exclusively associated with inactivating *katG*, which encodes the gene catalase-peroxidase responsible for activating the pro-drug isoniazid. This contrasts to the mutational pathways for isoniazid in *Mtb* which encompass a broader range of loci and mutations. Overall, this work has identified a mycobacterial strain that can safely and rapidly grow in the laboratory to investigate the development of resistance to antituberculosis drugs and has the potential to act as a model organism for studying AMR in *Mtb*.

1 General Introduction

1.1 Tuberculosis

Tuberculosis (TB) is a major global pandemic airborne disease caused by *Mycobacterium tuberculosis* (*Mtb*), a slow-growing bacteria, or other related species within the *Mtb* complex (e.g., *Mycobacterium bovis* and *Mycobacterium africanum*). Up to one quarter of the world's population is infected with *Mtb* (World Health Organization (WHO), 2021). According to the World Health Organization (WHO), a total of 10.6 million people became ill from TB in 2022, and 1.6 million people died of the disease including those who died from HIV co-infection (World Health Organization (WHO), 2023). Disruption in health service provision for TB during the COVID-19 pandemic contributed to the death of almost half a million people (Comella-del-Barrio *et al.*, 2021). TB continues to be a leading killer among individuals with Human Immunodeficiency virus (HIV). If left untreated, all forms of TB disease can be fatal (pulmonary and extrapulmonary TB). With timely diagnosis and effective treatment, TB is curable (Dorbniewski, 2002; Lee, 2015).

TB disproportionately affects people living in low-and middle-income countries (LMICs) and resource-poor settings (Ssengooba *et al.*, 2015). The direct (treatment costs and laboratory tests) and indirect costs (time away from work for recovery, unemployment) of treating TB can reach or exceed 20% of the pre-TB-patient or household annual income (Ghazy *et al.*, 2022), making it a severe socioeconomic problem as well as a physical disease.

1.1.1 The history of tuberculosis in humans

Tuberculosis is one of the oldest diseases known to humans that still causes deaths nowadays. It has been suggested that the genus *Mycobacterium* evolved 150 million years ago (Daniel, 2006). It has also been hypothesized that three million years ago, an early progenitor of *Mtb* might have infected early human ancestors in East Africa, and the common ancestor of modern-day *Mtb* strains emerged 20,000 years ago (Barberis *et al.*, 2017). As early humans migrated out of Africa and began to settle in communities, one can assume that they took their diseases with them, including TB. Despite the discovery and first characterization of the bacilli in the 19th century (Migliori *et al.*, 2007), deformities found in Egyptian mummies dating back to 2,400 BC, and written records describing the disease in China dating back 3,300 years ago suggest that this disease has been affecting humans for much longer. Called schachepheth, which meant a wasting disease in ancient Hebrew and phthisis by the Greeks, TB was well-known for being a fatal disease among young adults (Kestler and Tyler, 2022). There is even evidence of Pott's deformities caused by TB in the bones of Peruvian mummies predating Christopher Columbus, suggesting the presence of the disease before the colonization of the Americas in 1492 (Brites and Gagneux, 2015). Interestingly, in the Middle Ages, it was thought that the disease could be cured by royal touch, a practice that was abolished in 1712 in England but practiced in France until 1825 (Krugman and Terence, 2022). During the 18th century, consumption, as TB was called at the time, became an epidemic, with 900 deaths per 100,000 inhabitants per year. It was not until 1819 that this disease was termed "tuberculosis" by Johann Lukas Schönlein (Armocida and Martini, 2020). The transmission of this disease got worse during the Industrial Revolution due to poor

working conditions, lack of ventilation, malnutrition, primitive sanitation, and overcrowded housing conditions (Silver, 2014; Glaziou, Floyd, and Raviglione, 2018). After centuries of descriptions of the disease, in 1882, the physician Robert Koch identified and isolated the *M. tuberculosis* bacilli and finally identified the causative agent for TB (Migliori *et al.*, 2007). In the first half of the 20th century, with improved living conditions, sanitation, and advances in medicine, TB case reports and deaths decreased considerably. In 1904, the physician Albert Calmette and the veterinarian Camille Guérin, introduced the Bacillus Calmette-Guérin (BCG) vaccine, which was adopted by the Health Committee of the League of Nations (WHO predecessor) in 1928 (Luca and Mihaescu, 2013). In 1943, Selman Waksman, Elizabeth Bugie, and Albert Schatz developed streptomycin, the first antibiotic used for the TB treatment (Quinn *et al.*, 2020).

1.1.2 Epidemiology of tuberculosis

It is estimated that 10.6 million people fell ill with TB In 2022, and 6.3% of these cases were among HIV-positive people. In this year (2022), 1.13 million deaths were also reported, with a further 167,000 deaths amongst individuals living with TB and HIV. Most of the new TB cases (87%) reported in 2022 can be tracked to the following countries: Bangladesh (3.6%), China (7.4%), Democratic Republic of the Congo (2.9%), India (28%), Indonesia (9.2%), Nigeria (4.4%), Pakistan (5.8%) and the Philippines (7.0%) (Figure 1-1). These global data also illustrate a gender difference in the burden of disease with 55% of cases in adult men and 33% in women, with the remaining 12% affecting children (World Health Organization (WHO), 2023).

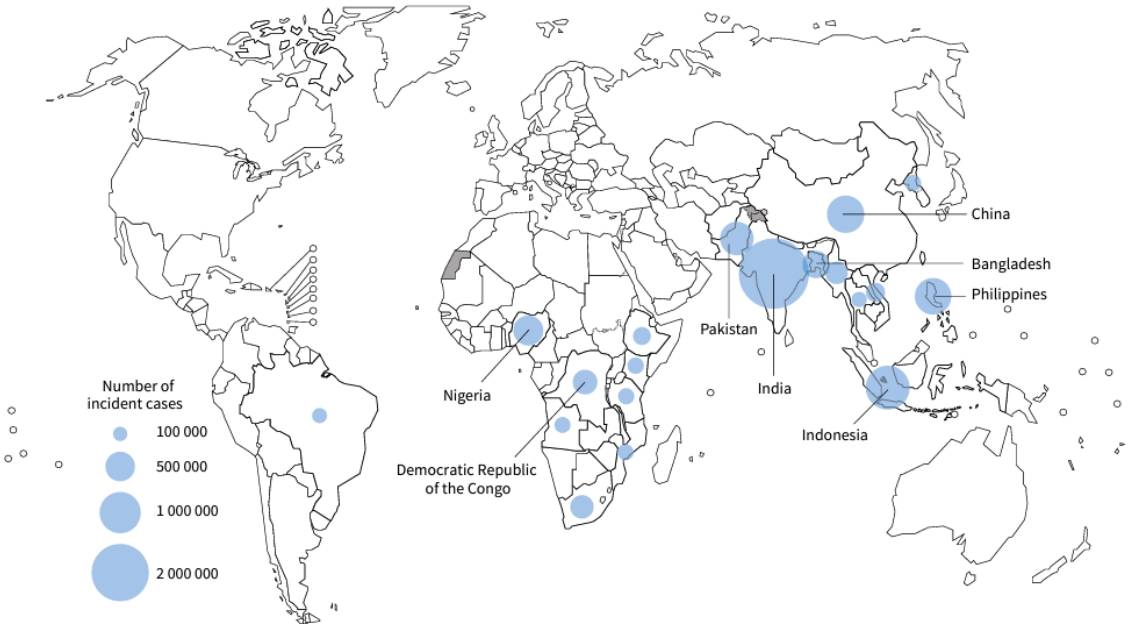


Figure 1-1. Countries with at least 100,000 TB incident cases in 2022. (Source: Global tuberculosis report, WHO, 2023).

Rifampicin-resistant (RIF-R) TB is defined as TB disease caused by a strain of *Mtb* which is resistant to rifampicin (RIF). Multidrug-resistant TB (MDR-TB) is defined as TB disease caused by a strain of *Mtb* which is resistant to RIF but also isoniazid (INH). These are the cornerstone first-line antibiotics for TB treatment. Strains of *Mtb* that are MDR/RIF-R and also resistant to any fluoroquinolone are defined as pre-extensively drug resistant TB (PRE-XDR-TB)(World Health Organization (WHO), 2021). Strains that fulfil the definition of MDR/RIF-R TB which are also resistant to fluoroquinolones and one additional Group A drug (second line anti-TB drug comprising levofloxacin (LEVO), moxifloxacin (MOXI), bedaquiline (BDQ), and linezolid (LZD)) are defined as extensively drug-resistant TB (XDR-TB). In 2022 the approximate number of individuals with MDR/RIF-TB was 410,000 and the number of deaths associated with MDR/RIF-TB was 160,000. The countries bearing the largest share of the current global MDR/RIF-TB burden are India (27%), the

Philippines (7.5%) and the Russian Federation (7.5%)(World Health Organization (WHO), 2023).

1.2 *Mycobacterium tuberculosis*

Mtb is a Gram-positive, rod-shaped, aerobic *bacterium* that produces adenosine triphosphate (ATP) by respiration under the presence of oxygen (although *Mtb* can also survive in the absence of oxygen). It is an acid-fast bacillus with typical dimensions of 0.2-0.7 x 1-10 microns (µm). Its optimum temperature for growth is 37°C, and its colonies are slightly yellowish with a rough and wrinkly appearance when grown on 7H10-11 Middlebrook agar plates (Han *et al.*, 2015).

The first *Mtb* genome was fully sequenced in 1998 (Cole *et al.*, 1998). Whole genome sequencing of the H37Rv strain of *Mtb* revealed a guanine + cytosine rich (G+C content of 65.6%) genome composed of a single circular chromosome of 4,411,529 base pairs (bp) encoding approximately 3,924 proteins coding sequences (CDS). Striking discoveries from the annotation of this first representative genome of *Mtb* were the large proportions of genes that encoded proteins associated with lipogenesis and lipolysis, and two novel families of acidic, glycine-rich proteins (PE and PPE families)(Cole *et al.*, 1998).

A major feature of *Mtb* is its cell wall, which provides a barrier to harmful compounds. The outer portion of the mycobacterial cell wall is waxy, rich in mycolic acids (long chain fatty acids) and plays an important role during virulence (Figure 1-2). The inner portion of the bacterial cell wall contains a complex layer with its core composed of peptidoglycans linked covalently to arabinogalactan. This, in turn, is linked to a mycolic

acid layer (Van Crevel *et al.*, 2011). The cell wall composition makes *Mtb* hydrophobic. It grows in aggregates that clump together and can be seen floating in liquid media. Detergents such as Tween-80 and Tyloxapol are added to mycobacterial broth media to disperse the cells and avoid this aggregation (Percival and Williams, 2013). The lipid-rich cell wall also prevents *Mtb* being stained using the Gram staining technique, however, the Ziehl-Neelsen stain allows its identification under the microscope.

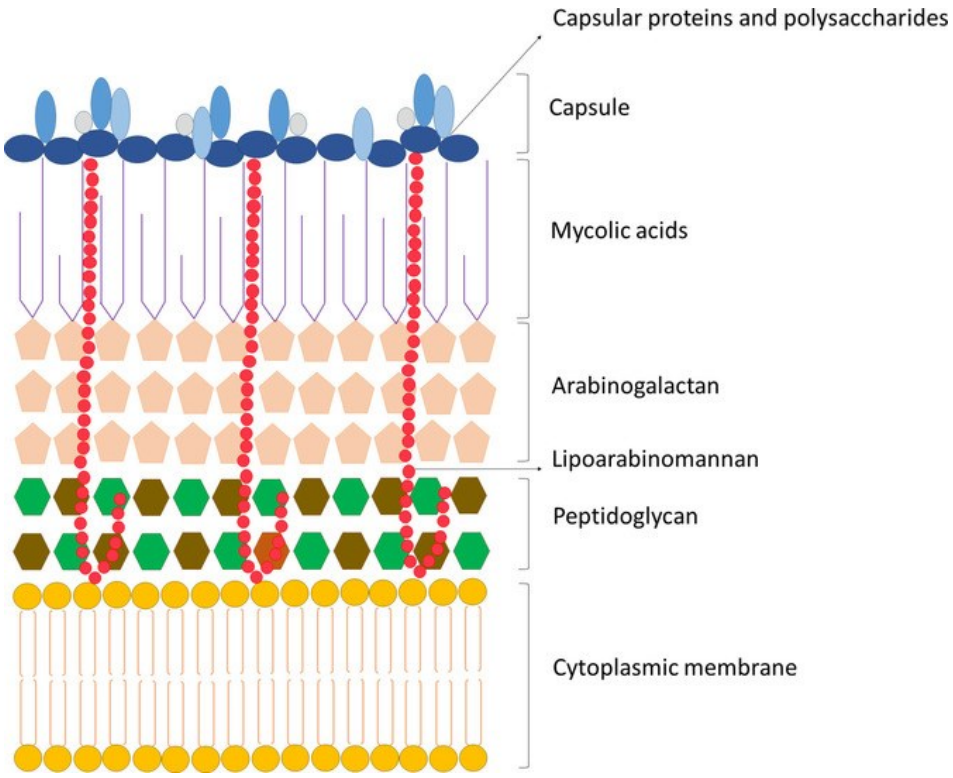


Figure 1-2. Schematic structure of the *Mtb* cell wall. The *Mtb* cell wall is composed of several complex layers.
The inner portion of the cell wall is composed of a peptidoglycan layer linked to arabinogalactan in complex with mycolic acid. Structures like proteins and lipids that interact with the cell wall can be observed in the cell envelope. Image extracted from Squeglia, Ruggiero and Berisio. September (2018). Chemistry A European Journal. Chemistry of Peptidoglycan in *Mycobacterium tuberculosis* Life Cycle: An off-the-wall Balance of Synthesis and Degradation. Volume 24. Issue 11. P2533-2546. DOI: https://doi.org/10.1002/chem.201702973 .

1.2.1 The *Mycobacterium* genus

Mtb is a member of the *Mycobacterium tuberculosis* complex (MTBC), which comprises other species of mycobacteria that can cause TB in humans and animals. The *Mycobacterium* genus, which belongs to the *Actinobacteria* phylum, contains over 190 named species including important human pathogens such as *Mtb* and *Mycobacterium leprae*, the agent that causes leprosy (Gupta, Lo and Son, 2018). The genus also encompasses non-tuberculous mycobacteria (NTMs), which includes members that are human pathogens such as *Mycobacterium abscessus* and *Mycobacterium avium*. However, a considerable number of non-tuberculous mycobacteria are non-pathogenic environmental organisms (Meehan *et al.*, 2021).

Mycobacteria can also be broadly grouped into two groups: rapid-growing mycobacteria (RGM) and slow-growing mycobacteria (SGM) based on their *in vitro* growth characteristics. Species of *Mycobacterium* that take between 3-7 days to form visible cultures on an agar plate are considered rapid growers. RGM are currently subdivided into 6 major taxonomic groups: the *Mycobacterium fortuitum* group, the *Mycobacterium chelonae*/*Mycobacterium abscessus* complex, the *Mycobacterium smegmatis* group, the *Mycobacterium mucogenicum* group, the *Mycobacterium margitense/Mycobacterium wolinsky*i group, and the pigmented RGM (Brown-Elliott and Philley, 2017).

Species of *Mycobacterium* that take longer than 7 days to form visible cultures on agar plates are referred to as slow growers (Bachmann *et al.*, 2020). *Mtb* is a well-known SGM with its 20-hour doubling rate, another example of SGM is *M. leprae*, with a doubling time of 14 days on average (Lambert, Walker and Harnisch, 2013). It has been speculated

that slow growing properties of *Mtb* could be influenced by the nature of host cell environments. When *Mtb* infects the human body, it parasitizes the host macrophage cells. The nutrient-limited environment inside the macrophages might force *Mtb* to adjust its growth rate to make full use of these limited sources of energy and impose minimal metabolic burden to the host (Zhu and Dai, 2018). The intake of nutrients from a limited source is also impaired due to the thick layer of mycolic acid found in the cell wall.

To understand the diversity within the *Mycobacterium* genus and how these organisms are related or divided according to characteristics such as *in vitro* growing characteristics, researchers have undertaken comparative genomic analysis between a growing number of known genome sequences. Pan-genome studies (Pan-genome is a collection of common and unique genes found in a set of genomes (Costa et al., 2020)) allow the comparison of multiple species of mycobacteria and these analyses have proved that although mycobacteria share core proteins that have evolved from a common ancestor, the number of genes present in the genus is expected to keep increasing and evolving (Negrete-Paz *et al.*, 2023). Comparative genomics have helped discover the presence of virulence factors such as the early secretory antigenic target (ESAT6) secretion (ESX) and how some of these protein families are exclusive of certain mycobacteria; for example, ESX-5 has been associated with virulence in pathogenic slow growing mycobacteria.

1.2.2 Division of the genus *Mycobacterium* and four novel genera

Since the genus is composed of 190 species with both clinical and environmental relevance, establishing the taxonomic relationships between species has been an

important task. These relationships have been explored by the comparison of the 16S RNA genes within the group and by using other physical and chemotaxonomic characteristics. Other comparison studies using housekeeping genes such as RNA polymerase beta subunit gene (*rpoB*) that encodes the RNA polymerase enzyme (Goldstein, 2014), the 65-kDa heat shock gene (*hsp65*) which encodes for the 65 kDa heat shock protein (Hsp65) and whose purpose is to prevent misfolding of proteins under stress conditions (Subramani *et al.*, 2017), and the DNA gyrase subunit B gene (*gyrB*), which encodes for the DNA gyrase subunit B (GyrB) protein and plays a role in the process of DNA replication (Marchese and Debbia, 2016), have also been performed to have a deeper understanding of the genetic and evolutionary relationship between members of the genus (Case *et al.*, 2007; Ogier *et al.*, 2019; Pereira *et al.*, 2020).

As previously mentioned, mycobacteria can be grouped and categorized depending on their growth rates, dividing the group broadly into SGM and RGM (Figure 1-3). However, among these groups there are major differences. In 2018, the work by Gupta and colleagues using comparative genomics, led to a proposal to split the genus into four new genera: *Mycolicibacterium, Mycolicibacter, Mycolicibacillus*, and *Mycobacteroides*, and an amended *Mycobacterium* genus (Gupta, Lo and Son, 2018). This division was based on the comparative genomic analysis of 150 genomes of different mycobacterial species and the construction of phylogenetic trees using core protein sequences (Gupta, Lo and Son, 2018).

However, the division of the genus has been met with resistance by microbiologists and clinicians as it is considered unnecessary. The work published by Conor J. Meehan (2021)

considered that there is not enough evidence to support this division as these new genera have overlapping boundaries among each other, making this new division unfounded (Meehan *et al.*, 2021). It has also been noted that although the work Gupta and colleagues (2018), has provided important, and valuable information about the genus, the work only used 150 species for the bioinformatic analysis, and leaving 40 species that kept the name *Mycobacterium* out of the study (Tortoli *et al.*, 2019). This new division might also confuse clinicians as some important pathogens such as *Mycobacterium abscessus* have been renamed *Mycobacteroides abscessus*. The new divisions do not affect treatment regimens so confusion around species naming, and characterisation may be considered disruptive and unjustified (Tortoli *et al.*, 2019). For this thesis, the former names are used when referring to any mycobacteria species to avoid confusion. The name *Mycobacterium* remains valid as it serves as a synonym for the five new genera, and published names of species are never withdrawn (Tindall, 1999).

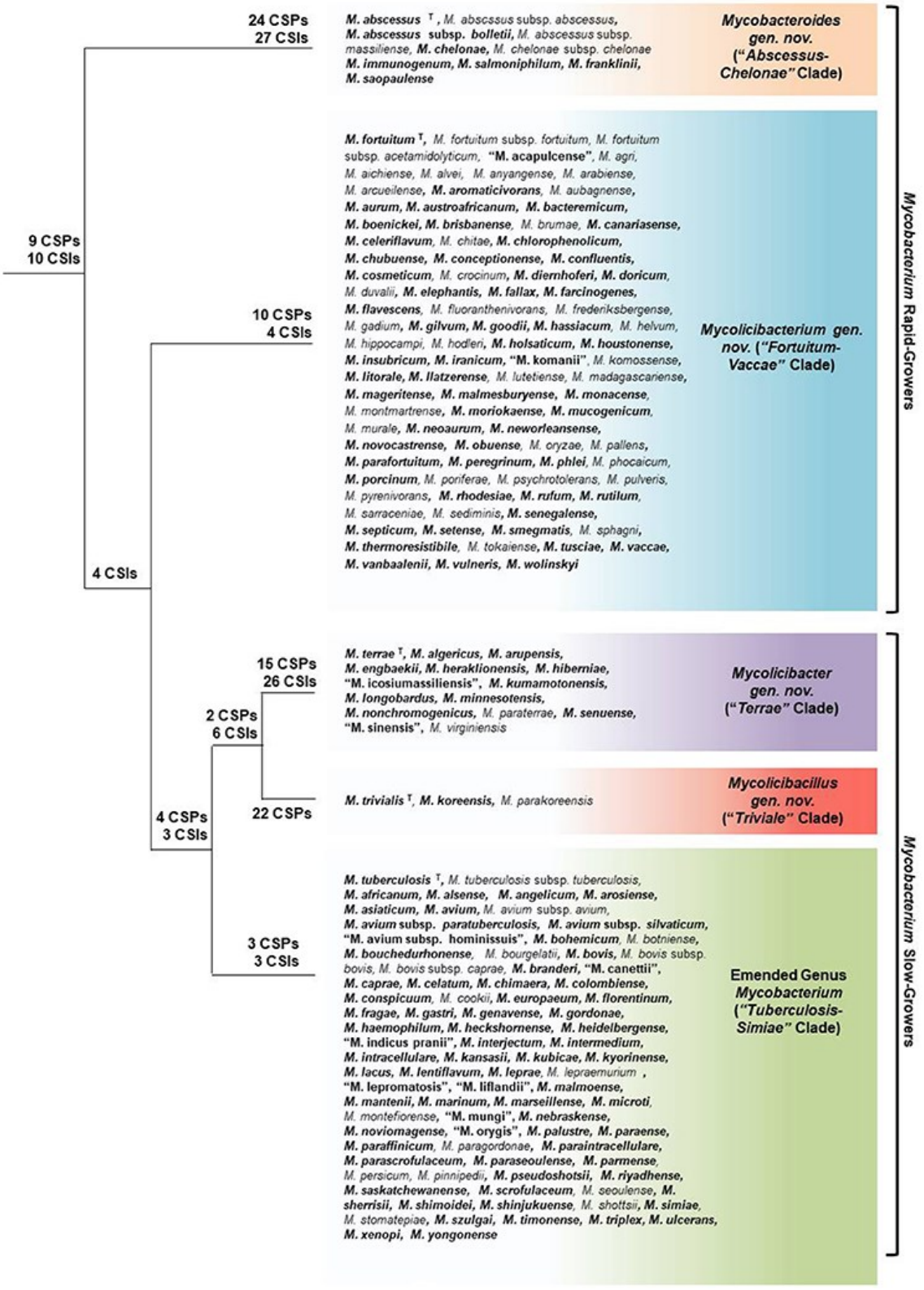

24 CSPs
27 CSIs
M. abscessus T, M. abscssus subsp. abscessus, M. abscessus subsp. bolletii, M. abscessus subsp. massiliense, M. chelonae, M. chelonae subsp. chelonae M. immunogenum, M. salmoniphilum, M. franklinii, M. saopaulense
Mycobacteroides gen. nov. ("Abscessus-Chelonae" Clade)
9 CSPs
10 CSIs
M. fortuitum T, M. fortuitum subsp. fortuitum, M. fortuitum subsp. acetamidolyticum, "M. acapulcense", M. agri, M. aichiense, M. alvei, M. anyangense, M. arabiense, M. arcueilense, M. aromaticivorans, M. aubagnense, M. aurum, M. austroafricanum, M. bacteremicum, M. boenickei, M. brisbanense, M. brumae, M. canariasense, M. celeriflavum, M. chitae, M. chlorophenolicum, M. chubuense, M. conceptionense, M. confluentis, M. cosmeticum, M. crocinum, M. diernhoferi, M. doricum, M. duvalii, M. elephantis, M. fallax, M. farcinogenes, M. flavescens, M. fluoranthenivorans, M. frederiksbergense, M. gadium, M. gilvum, M. goodii, M. hassiacum, M. helvum, M. hippocampi, M. hodleri, M. holsaticum, M. houstonense, M. insubricum, M. iranicum, "M. komanii", M. komossense, M. litorale, M. llatzerense, M. lutetiense, M. madagascariense, M. mageritense, M. malmesburyense, M. monacense, M. montmartrense, M. moriokaense, M. mucogenicum, M. murale, M. neoaurum, M. neworleansense, M. novocastrense, M. obuense, M. oryzae, M. pallens, M. parafortuitum, M. peregrinum, M. phlei, M. phocaicum, M. porcinum, M. poriferae, M. psychrotolerans, M. pulveris, M. pyrenivorans, M. rhodesiae, M. rufum, M. rutilum, M. sarraceniae, M. sediminis, M. senegalense, M. septicum, M. setense, M. smegmatis, M. sphagni, M. thermoresistibile, M. tokaiense, M. tusciae, M. vaccae, M. vanbaalenii, M. vulneris, M. wolinskyi
Mycolicibacterium gen. nov. ("Fortuitum-Vaccae" Clade)
Mycobacterium Rapid-Growers
10 CSPs
4 CSIs
4 CSIs
15 CSPs
26 CSIs
M. terrae T, M. algericus, M. arupensis, M. engbaekii, M. heraklionensis, M. hiberniae, "M. icosiumassiliensis", M. kumamotonensis, M. longobardus, M. minnesotensis, M. nonchromogenicus, M. paraterrae, M. senuense, "M. sinensis", M. virginiensis
Mycolicibacter gen. nov. ("Terrae" Clade)
2 CSPs
6 CSIs
22 CSPs
M. trivialis T, M. koreensis, M. parakoreensis
Mycolicibacillus gen. nov. ("Triviale" Clade)
4 CSPs
3 CSIs
3 CSPs
3 CSIs
M. tuberculosis T, M. tuberculosis subsp. tuberculosis, M. africanum, M. alsense, M. angelicum, M. arosiense, M. asiaticum, M. avium, M. avium subsp. avium, M. avium subsp. paratuberculosis, M. avium subsp. silvaticum, "M. avium subsp. hominissuis", M. bohemicum, M. botniense, M. bouchedurhonense, M. bourgelatii, M. bovis, M. bovis subsp. bovis, M. bovis subsp. caprae, M. branderi, "M. canettii", M. caprae, M. celatum, M. chimaera, M. colombiense, M. conspicuum, M. cookii, M. europaeum, M. florentinum, M. fragae, M. gastri, M. genavense, M. gordonae, M. haemophilum, M. heckshornense, M. heidelbergense, "M. indicus pranii", M. interjectum, M. intermedium, M. intracellulare, M. kansasii, M. kubicae, M. kyorinense, M. lacus, M. lentiflavum, M. leprae, M. lepraemurium, "M. lepromatosis", "M. liflandii", M. malmoense, M. mantenii, M. marinum, M. marseillense, M. microti, M. montefiorense, "M. mungi", M. nebraskense, M. noviomagense, "M. orygis", M. palustre, M. paraense, M. paraffinicum, M. paragordonae, M. paraintracellulare, M. parascrofulaceum, M. paraseoulense, M. parmense, M. persicum, M. pinnipedii, M. pseudoshotsii, M. riyadhense, M. saskatchewanense, M. scrofulaceum, M. seoulense, M. sherrisii, M. shimoidei, M. shinjukuense, M. shottsii, M. simiae, M. stomatepiae, M. szulgai, M. timonense, M. triplex, M. ulcerans, M. xenopi, M. yongonense
Emended Genus Mycobacterium ("Tuberculosis-Simiae" Clade)
Mycobacterium Slow-Growers

Figure 1-3. Phylogenetic tree showing the relationship between the major groups of mycobacteria species, created using core proteins.
The genus *Mycobacterium* has been divided into 5 distinct groups compiling both fast and slow growers. The number of clade specific proteins (CSP) and indels (CSI) are marked on the nodes. The superscript letter T beside a species indicates that it is the type species of the genus, and names not italicized referred to species that have not yet been validated. Image extracted from Gupta, Lo and Son. February (2018). Frontiers in Microbiology. Phylogenomic and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. Volume 9. Issue 67. P1-41. DOI: https://doi.org/10.3389/fmicb.2018.00067.

1.2.3 Pathogenicity of tuberculosis

1.2.3.1 Infection by *M. tuberculosis* and progression to TB disease

Mtb infection is transmitted from person to person through the air. Whenever an infected person talks, coughs, or sneezes, droplets, 1-5 microns in diameter, are released, which contain bacilli. Once these droplets are inhaled by a new host, the bacteria enter the lungs (Pulmonary TB; PTB), where they encounter alveolar macrophages and other immune cells, and set up anatomical lesions called granulomas. The most common route of initial spread is via the lymphatic system to lymph nodes in the lung hila, and the combination of a pulmonary lesion with local lymphadenopathy is traditionally referred to as a Ghon complex (Figure 1-4). Bacilli can also proliferate and spread to other parts of the body (Extrapulmonary TB; EPTB), sometimes through the bloodstream or lymphatic system, causing disseminated or distant foci of disease (Barberis *et al.*, 2017).

1.2.3.2 Entry of *M. tuberculosis*, Alveoli, and Alveolar Macrophages

Once inside the human body, *Mtb* travels to the lower respiratory tract of the host (Delogu, Sali and Fadda, 2013), where *Mtb* will mainly encounter resident macrophages,

but *Mtb* can also be ingested by alveolar epithelial type II pneumocytes, dendritic cells, alveoli epithelial cells and neutrophils. *Mtb* is engulfed by the macrophages through the recognition of pathogen-associated molecular patterns (PAMPs) (Maphasa, Meyer and Dube, 2021). These molecules are located on the bacterial surface and are recognized by cell surface receptors.

Some of the macrophage surface receptors associated with the uptake of *Mtb* are Toll-like receptors, Scavenger receptors, C-type lectin receptors, mannose receptors, surfactant protein A receptors, CD14, cytosolic DNA sensors, immunoglobulin receptors, and complement receptors (Vergne *et al.*, 2004; Songane *et al.*, 2012; Queval, Brosch and Simeone, 2017; Maphasa, Meyer and Dube, 2021).

Once internalized, *Mtb* bacilli reside in the phagosome inside the macrophages, where they encounter a hostile micro-environment with low pH, reactive oxygen intermediates (ROI), lysosomal enzymes, and toxic peptides (Smith, 2003). *Mtb* confronts these hostile conditions by preventing the maturation and acidification of the phagolysosome by blocking the endosomal tethering molecule EEA1, essential for the maturation of the phagolysosome alongside the effector phosphatidylinositol (PI)3 (Levin, Grinstein and Canton, 2016).

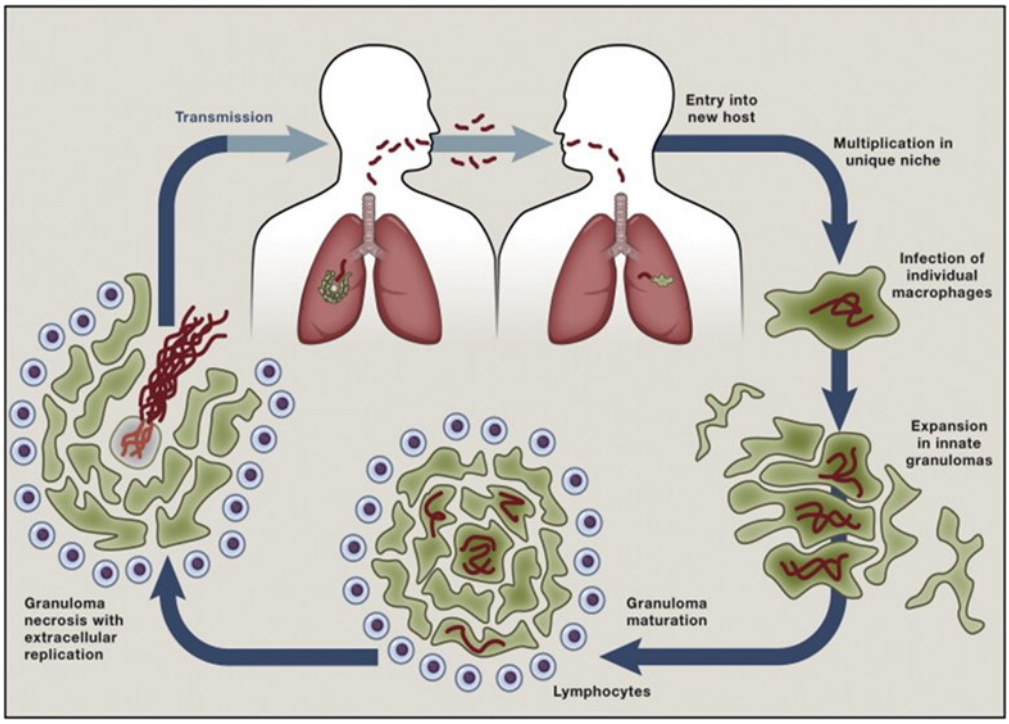


Figure 1-4. The infection by *Mtb* starts when an infected individual releases aerosol particles containing *Mtb*, and these particles are inhaled by a new host.
Mtb travels through the trachea, and then the bacilli are deposited in the lower respiratory tract. Once in the lungs, *Mtb* encounters the resident macrophages and gets engulfed by them, the infected macrophages fail to destroy *Mtb* and instead help *Mtb* by transporting it to various parts of the lung and deeper lung tissue. The infected macrophages will release cytokines to recruit more macrophages and other immune cells to the infected site forming a granuloma. Image extracted from Cambier, Falkow and Ramakrishnan. December 2014. Cell. Pathogenic Life Cycle of *M. tuberculosis*. Volume 159. Issue 7. P1497-1509. DOI: https://doi.org/10.1016/j.cell.2014.11.024.

1.2.3.3 Formation of the caseating granuloma

The infected alveolar macrophages, through the release of cytokines, recruit more cells: granulocytes, dendritic cells, natural killer cells, and T and B lymphocytes (Briken *et al.*, 2018) to the infection site. A slow inflammatory process then starts, during which the infected macrophages begin to form a focal lesion composed of infected and non-infected macrophages, to contain the infection (Smith, 2003), thus creating a granuloma.

The cells that have been recruited (granulocytes, dendritic cells, natural killer cells and T and B lymphocytes) help in the formation of the granuloma.

1.2.3.4 Exit of *M. tuberculosis* from the granuloma

As the disease progresses, the granuloma undergoes a series of changes leading to the formation of the caseum, a cheese-like lipid-rich structure that contains necrotic material at its centre. Progression to active TB disease corresponds with the accumulation of dead macrophages in the caseum of the granuloma centre that leads to a necrotic breakdown (Kim *et al.*, 2010; Ndlovu and Marakalala, 2016; Urbanowski *et al.*, 2020). The destruction of the caseous core results in the liquefaction of the granuloma, which can generate cavities within diseased tissue. Pulmonary cavitation is the result of the complete liquefaction of the necrotic core, which leads to the excavation of an actual cavity within the lung, where the tissue has broken down and exited via an airway.

The presence of pulmonary cavitation is associated with several serious problems: extensive damage to the lung, high concentration of *Mtb* in the cavity walls (10^7-10^9 bacilli), and a direct exit route for aerosolized *Mtb* to spread in the environment to a new potential host (Palaci *et al.*, 2007; Gideon and Flynn, 2011). Individuals with cavitary lesions are the primary transmitters of *Mtb*. The presence of these cavitary lesions is also associated with poor treatment outcomes and relapse after treatment (Urbanowski *et al.*, 2020).

1.3 Treatment for TB

1.3.1 History of the TB treatment

The first antibiotic used for the TB treatment was streptomycin (SM) in 1944 (Zwick and Pepperell, 2020). In the same year, Jorgen Lehman, synthesized para-amino salt or salicylic acid (PAS) and both antibiotics were used against TB (Kerantzas and Jacobs, 2017). In 1952 isonicotinic acid hydrazide (INH) was discovered by Gerhard Domagk, and thus, INH was used in combination with SM and PAS which resulted in improved clinical outcomes (Keshavjee and Farmer, 2012). In 1960 Ethambutol (EMB) was introduced to the TB treatment as a replacement for PAS; this change reduced the treatment time from 24 months to 18 months (Iseman, 2002). The introduction of rifampicin (RIF), extracted from *Streptomyces mediterranei*, in 1970 proved to be a milestone in the history of TB treatment due to its potent sterilizing effect (Iseman, 2002). Early clinical trials by the British Medical Research Council (BMRC) demonstrated that the use of RIF alongside INH, SM, and EMB resulted in a 95% success rate in just 8-9 months (Iseman, 2002). In the 1980s, pyrazinamide (PZA) was introduced, and this new addition was found to accelerate the success rates >95% in 6 months when combined with INH and RIF (Iseman, 2002).

1.3.2 Current treatment

The recommended treatment for new patients with fully drug-susceptible PTB is an intensive phase comprising a multi-drug combination of four first-line antibiotics INH, RIF, EMB, and PZA for 2 months, followed by a continuation phase of INH and RIF for 4 months (World Health Organization, 2020). In cases where the TB diagnosis occurs in a

patient who is also living with HIV, the same 6-month course is recommended, alongside antiretroviral therapy, which may be started immediately (Getahun *et al.*, 2015).

The treatment for patients with MDR-TB is more complicated, including some second line drugs with higher toxicity, often for longer duration but with less favourable outcomes. Research on the best approach to managing DR-TB is ongoing, but Table 1-1 summarizes the current approaches for different types of TB.

Table 1-1. Current treatment regime endorsed why WHO for the treatment of TB disease including drugs susceptible TB, and DR-TB as for 2021 (WHO, 2021).

Type of TB	Brief description	Possible treatment regimens
Drug Susceptible (DS)-TB	Active TB disease, with an infecting *Mtb* isolate susceptible to both anti-TB drugs RIF and INH.	'Standard' regimen: 2 months of RIF-INH-PZA-EMB, then 4 months RIF-INH New 4-month regimen: 2 months of Rifapentine-MOXI-PZA-EMB, then 2 months of Rifapentine-MOXI-EMB
INH resistant TB	Active TB disease with an infecting *Mtb* isolates resistant to INH but susceptible to RIF.	WHO endorsed regimen: 6 months of RIF-LEVO-PZA-EMB
RIF resistant / MDR-TB	Active TB disease with an infecting *Mtb* isolate resistant to RIF +/- resistant to INH. Pre-XDR TB: RIF-R /MDR-TB with added resistance to fluoroquinolones XDR-TB: RIF-R / MDR-TB with added resistance to fluoroquinolones and at least one other 'Group A' second line TB drug	'Standard' 18–20-month regimen: At least 5 effective drugs for 8 months, then at least 4 effective drugs for 12 months (effective drugs and priority order for use defined according to WHO guidelines) Shorter regimens: 6-9 months of BDQ-PA-LZD 6 months of BDQ-PA-MOXI-LZD Several others under ongoing clinical trial investigation

1.3.3 Antimicrobial susceptibility testing in TB treatment

Effective TB treatment and disease control relies on determining the susceptibility of *Mtb* to various anti-tuberculosis drugs. The main objective of antimicrobial susceptibility testing (AST) is to determine if the antibiotic used against the organism will prevent its growth. ASTs also play a crucial role in identifying drug-resistant strains of *Mtb*. This process involves exposing the organism to a gradient of different concentrations of antibiotics followed by an incubation time to allow the antitubercular drug to interact with *Mtb* (AST in *Mtb* usually takes weeks to produce results). After this incubation period, the minimum concentration of the antibiotic tested that achieves at least 90% inhibition of *Mtb* growth (known as the Minimum Inhibitory Concentration, MIC_{90}) is compared to internationally agreed reference standards to determine if the result of the AST can be considered susceptible at normal dosing (S), susceptible at increased exposure (I) or resistant (R) to the agent in question (EUCAST, 2019). This information is essential for patients' treatment design, and therefore better treatment outcomes. International organizations like the World Health Organization (WHO), the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical and Laboratory Standards Institute (CLSI) play a vital role in establishing these protocols. These organizations cover aspects like drug selection, AST methods recommendations (which depend on the organism being tested), media preparation, inoculation procedures, incubation times, and interpretation of results. In the case of *Mtb* and NTMs, the EUCAST recommends the broth microdilution method for MIC determination (Cambau *et al.*, 2019; EUCAST, 2019; Schön *et al.*, 2020). The MIC_{90} values produced by the broth microdilution method have been used by WHO, EUCAST and CLSI to determine

Mtb clinical breakpoints, which are the drug concentrations that differentiate susceptible from resistant organisms and attempt to describe the likelihood that the *Mtb* strain being evaluated will respond to treatment (Wiegand, Hilpert and Hancock, 2008). Clinical breakpoint determination is a meticulous process involving data collection and expert analysis, that requires MIC data collected from large scale studies of various clinical isolates of *Mtb*, and pharmacokinetic/pharmacodynamics (PK/PD) data (dose response, mode of administration, dosing interval, excretion of the drug from the body, etc), and clinical outcomes (data collected regarding the number of successes or failures in the treatment of patients being treated with an antibiotic) (Schön *et al.*, 2017). These considerations are important for more effective treatment and minimization of further emergence of resistance.

The field of AST is constantly evolving. WHO (https://www.who.int/publications/i/item/9789241514842), EUCAST (https://www.eucast.org/mycobacteria/methods_in_mycobacteria), and CLSI (https://clsi.org/standards/products/microbiology/documents/m24/) review *Mtb* relevant information every year. Microbiological protocols and key measurements such as clinical breakpoints and protocols are reviewed accordingly. Researchers are constantly exploring novel approaches to complement established methods such as the gold standard broth microdilution method which is laborious in nature. The WHO provides guidance on standardized AST methods for *Mtb*, including techniques like the resistance ratio, proportion method, and MIC-absolute methods. Implementing these standardized methods ensures consistent and reliable results, allowing for optimal treatment decisions.

1.3.3.1 The Proportion Method

This is the method most used in developing countries due to its low cost and adopted by WHO (Canetti et al., 1963), this method determines the relationship between the number of bacilli growing in a solid media with antibiotics against the number of bacilli growing in a media free of antibiotics, therefore establishing the concentration of *Mtb* who is resistant to antibiotics in a given population.

For this method, *Mtb* cultures are grown in LJ medium, and the results are read on the 28th day; drug containing tubes with more than 1% of colonies will be considered resistant compared to the control-free drug medium. Because of its validity, it is considered one of the most effective methods for AST available nowadays; its low cost and its reproducibility also gives it an advantage; however, the disadvantage of this method is the time for obtaining results that can go anywhere from weeks to months (Dorbniewski, 2002).

1.3.3.2 The Absolute Concentration Method

The absolute concentration method is performed by inoculating *Mtb* into solid-based agar (7H10) or JW medium with different concentrations of drugs as two-fold dilution or is used in a broth microdilution method. The results are read usually after 3-5 weeks. Resistance is determined by the lowest concentration of the antibiotic that inhibits *Mtb* growth. The major source error in this method is variations in inoculum size (Dorbniewski, 2002; World Health Organization, 2018).

1.3.3.3 The Resistance Ratio Method

The resistance ratio method compares a strain of *Mtb* culture that might be drug resistant against a controlled laboratory reference strain (e.g., *Mtb* H37Rv). Both strains are placed in media with antibiotics at different concentrations and their growth is carefully monitored (Acharya & Ghimire, 2008). The *Mtb* strain will be considered susceptible if the ratio is 2 or less, but if the ratio is 8 or more, then the strain will be regarded as highly resistant. One advantage of this method is that it is a simple technique that requires no specialized equipment to be performed (Dorbniewski, 2002; World Health Organization, 2018).

1.4 The challenge of antimicrobial resistance

Antimicrobial resistance (AMR) is when bacteria that were previously susceptible to, and treatable with, specific antibiotics develop genetic mutations or other biological mechanisms that make those antibiotics less effective. Resistance to antitubercular drugs was observed shortly after the introduction of the first antibiotic, SM, in 1944 (Iseman, 2002) (Figure 1-5). Although resistance to specific antibiotics is intrinsic in some bacteria, the increased use of antibiotics in the 20th and 21st century has created an environment which selects for survival of resistant organisms, creating a growing problem acquired resistance. Transmission of infections between people, exacerbated by poor living conditions amplifies the problem further (Ayukekbong, Ntemgwa and Atabe, 2017).

AMR is a problem for many infectious diseases but presents specific issues in TB. The first antitubercular drugs were developed in the 1940s and the current 6-month long

first-line treatment with RIF, INH, EMB, and PZA has been used since the 1980s. As described above, when strains of *Mtb* acquire genetic mutations which reduce the effectiveness of these medications, DR-TB arises necessitating longer treatment with second-line medications that are more expensive, have higher toxicity and are associated with less favourable treatment outcomes (Arbex *et al.*, 2010).

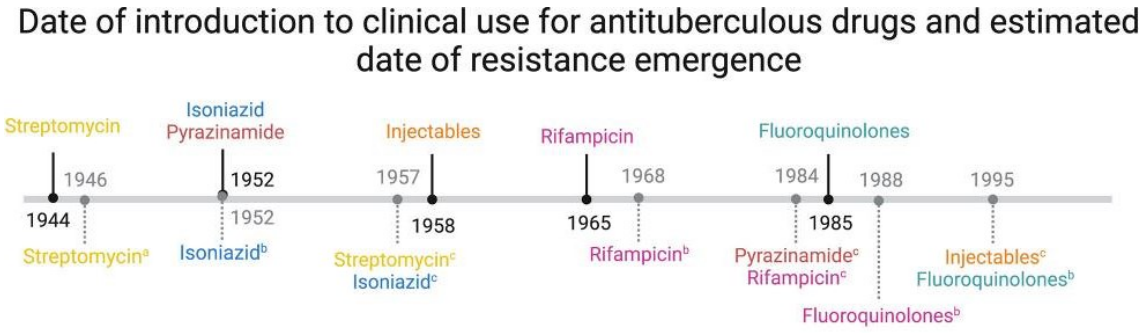


Figure 1-5. The emergence of drug resistance in tuberculosis.
Timeline presenting the date of the introduction of an antitubercular drug (black line) and the year in which resistance to that same drug was first observed (grey dotted line). Image from Nimmo, Millard, Faulkner, Monteserin, Pugh and Johnson. October (2007). Frontiers in Microbiology. Evolution of *Mycobacterium tuberculosis* in the genomic era. Volume 12. 954074. https://doi.org/10.3389/fcimb.2022.954074.

1.4.1 Drug resistance in the genomic era

Traditionally, AST in *Mtb* has been performed using culture-based methods (phenotypic methods). These methods are reliable and are still widely used in clinical microbiology settings, but these methods only provide a visual representation of resistance (growth or no growth in the presence of an antibiotic) and do not provide information on the basis of the resistance (Waddington *et al.*, 2022). The reproducibility and high resolution of whole genome sequencing (WGS) (genotypic method) provides valuable information about an isolate, and in contrast to traditional culture-based methods, this

comprehensive genetic information helps to understand the mechanisms of AMR by identifying the mutations responsible (World Health Organization, 2020). WGS offers rapid results, unlike phenotypic methods that depend on the growth of the organism (taking weeks to produce results). WGS results also produce accurate predictions for resistance while also providing AMR surveillance data (Su, Satola, and Read, 2019).

WGS has also helped researchers to learn more about the evolution of the MTBC and divide it into different lineages (lineages L1-7). These lineages are associated with human migration and the geographic distribution of these lineages can vary, suggesting that specific lineages are adapted to different human populations. Tracking their spread across the population is crucial for TB epidemiology surveillance (O'neill *et al.*, 2019).

Using genomic studies, it was possible to discover that the East Asian lineage (L2) (including the Beijing genotype) and the Euro-American lineage (L4) are presently the most widely spread lineages globally, and that these two lineages (L4) and (L2) are the most virulent lineages of the MTBC (Senghore *et al.*, 2020). Through WGS, it was also discovered that lineage 2 MTBC isolates develop resistance to RIF at a higher rate than lineage 4 (Köser, Ellington and Peacock, 2014).

Whilst WGS has provided great benefits for *Mtb* research, control, and epidemiology surveillance, implementation of these technologies requires complex workflows, can be a costly procedure, personnel with training in laboratory work and expertise in data genome analysis (Mahomed *et al.*, 2017). This is most difficult in low-income countries with a high-TB burden, where it is needed most.

1.4.2 Mechanisms of Antibiotic Resistance in *M. tuberculosis*

The use of antitubercular drugs has provided a great tool to fight infections caused by *Mtb*, but the emergence of highly drug resistant forms of the disease has jeopardized prospects for global TB control (C Reygaert, 2018). Horizontal gene transferring is a rare occurrence in *Mtb,* and most genetic mutations conferring antibiotic resistance involve processes such as deletion, duplication, insertion, and single nucleotide polymorphisms (SNPs) (Woodman, Haeusler and Grandjean, 2019).

1.4.2.1 *M. tuberculosis* cell wall

As already described in Section 1.2, *Mtb,* as a member of the genus mycobacteria, possesses a thick, complex, and lipid-rich cell wall comprising three main layers: peptidoglycan (for rigidity and osmotic stability), arabinogalactan, and mycolic acids (Shaku, Ealand and Kana, 2020).

The *Mtb* cell wall is hydrophobic with a limited permeability, not allowing access to a variety of antitubercular compounds and providing the first barrier of intrinsic resistance to antibiotics, chemotherapeutic agents, and chemical disinfectants (Holzheimer, Buter, and Minnaard, 2021). Lipophilic compounds such as RIF, macrolides, and albamycin can enter the microbial bilayer through a slow diffusion. The peptidoglycan and arabinogalactan cell wall makes the cell wall extremely hydrophobic, preventing the entrance to the cell of hydrophilic compounds, such as INH, EMB, and PZA, but these compounds can access the cell through the very limited porins (Jarher and Nlkaldo, 1994). Porins are water-filled channels that allow the intake of small hydrophilic molecules, and some antibiotics rely on porins to enter bacterial cells. These proteins

that are located across the cellular membrane and serve as routes for the diffusion of molecules and for nutrient intake and play a major role in the influx of hydrophilic drugs such as RIF. However, mutations in these porins can prevent the intake of these antibiotics (Niederweis, 2003). OmpATb was the first porin-like structure identified in *Mtb* and is encoded by the gene Rv0899 (Passiflora Sarathy *et al.*, 2012). *Mtb* has fewer porins than other bacteria, such as *Escherichia coli*, and it is suggested that this prevents entry into the cell of harmful molecules, but a minimum number of porins is required because *Mtb* needs to access nutrients from outside the cell (Raynaud *et al.*, 2002).

1.4.2.2 Drug Efflux pumps in *M. tuberculosis*

Hydrophilic antibiotic molecules can enter the cell through the limited porins located among the cell membrane, and hydrophobic molecules can enter the cell through passive diffusion. Drug efflux pump proteins actively expel antitubercular drug molecules that enter the bacterial cell, thereby preventing the intracellular accumulation necessary to reach therapeutic drug concentrations inside the cell (Szumowski *et al.*, 2012).

Mutations can alter the pump's specificity, allowing the efflux of antibiotics that previously were not transported out. The mutations in Tap (Rv1258c) and MmpL5 genes generate resistance to antitubercular drugs such as INH, PZA, and SM (Liu *et al.*, 2019). Efflux proteins play an important role in the transport of nutrients, waste, bacterial signalling, and bacterial metabolism and are divided into 5 superfamilies based on sequence homology: the ATP-binding cassette (ABC), the major facilitator superfamily (MFS), the multidrug and toxic compound extrusion family (MATE), the resistance-nodulation-cell-division superfamily (RND), and the small multidrug resistance family

(SMR). All superfamily members are associated with resistance to different antibiotics, and different efflux pumps can recognize more than one antibiotic (Laws, Jin, and Rahman, 2022).

1.4.2.3 Drug degradation or inactivation

After penetrating the cell wall, antibiotics can be cleaved by mycobacterial enzymes, which degrade or modify them before they reach the target site and provoke any damage. One of the best examples of enzymatic activity to degrade drugs by *Mtb* is the enzymatic degradation of β-lactam antibiotics by BlaC, the principal β-lactamase of *Mtb*, which hydrolyses the β-lactam ring of the antibiotics (Kurz and Bonomo, 2012; Gygli *et al.*, 2017; Van Alen *et al.*, 2021).

Another way in which *Mtb* can modify antibiotics to inactivate them is by chemically modifying specific sites on the antibiotics through methylation and acetylation (Singh *et al.*, 2020). Antibiotics that have a chemical group added to a specific site are rendered inactive, and this is a process well known to be performed by the enhanced intracellular survival (Eis). This acetyltransferase is the most studied case of drug modification in *Mtb,* where Eis acetylates aminoglycoside/cyclic peptide antibiotics, including the antibiotics gentamicin, amikacin, tobramycin, neomycin, and streptomycin (Nasiri *et al.*, 2017).

1.4.2.4 Enzymatic drug target modification and target mimicry

Many pathogens, including *Mtb,* can avoid the action of antibiotics by structural modifications of the antibiotic's targets (Table 1-2), therefore reducing antibiotic binding affinity (Smith *et al.*, 2012; Palomino and Martin, 2014). An example of this is the

resistance to RIF, one of the most effective anti-TB drugs. *Mtb,* through the substitution of one amino acid in the RIF binding pocket of the beta subunit of the RNA polymerase, can decrease the affinity for the drug, and these changes can lead to different degrees of sensitivity to RIF (Goldstein, 2014). The *rpoB* gene encodes the beta subunit of RNA polymerase (RNAP), an essential enzyme that synthesizes RNA molecules from a template of DNA during the transcriptional process (Adékambi, Drancourt, and Raoult, 2009). Approximately 90-95% of RIF-TB resistant strains have mutations in a small region (81 bp long) inside the *rpoB* gene of *Mtb*, defined as "Rifampicin Resistance-determining region (RRDR)," and the most common mutations in RRDR are in codons 516, 526, and 531 (Zhang *et al.*, 2019). Resistance to RIF can have other collateral effects, such as altered susceptibility towards other antitubercular drugs and reduced fitness and produces durable changes in the lipid content of *Mtb* (Koch, Mizrahi, and Warner, 2014). Compensatory mutations in *rpoA* (which encodes for the RNA polymerase alpha subunit) and *rpoC* (encodes for RNA polymerase beta' subunit) genes alongside with mutations in the *rpoB* gene of *Mtb* have also been observed. These compensatory mutations are associated with higher transmissibility of the infection in some settings (Palomino and Martin, 2014).

Another antimicrobial mechanism observed is target mimicry. This has been observed in *Mtb* to nullify the action of fluoroquinolones, a class of drugs that cause DNA degradation and bacterial cell death by inhibiting DNA gyrase and topoisomerase IV, two enzymes involved in bacterial DNA synthesis (Blondeau, 2004). To nullify the action of the fluoroquinolones, *Mtb* uses MfpA (*Mycobacterium* fluoroquinolone resistance protein A). The MfpA protein mimics the shape, size, and electrostatic surface of DNA,

and competes with bacterial DNA for the binding pocket of the DNA gyrase beta subunit. Since MfpA is already bound to the binding pocket of the GyrB subunit, it prevents fluoroquinolone drugs from reaching their target (Tao *et al.*, 2013; Gautam and Meena, 2020).

Table 1-2. Summary of molecular mechanism of antitubercular drugs and genes associated with resistance in *M. tuberculosis*.

Antibacterial compound	Resistance genes	Gene function	Drug mode of action	Resistance mechanism
Amikacin	*rrs* *eis*	Ribosomal ribonucleic acid (rRNA) is a critical component for the 30S subunit of the ribosome. The enhanced intracellular survival (Eis) protein helps survival of *Mtb* within the macrophages (Pan, Zhao and Ye, 2018)	Disrupt protein synthesis by targeting the 30S subunit of the ribosome.	Modifications in the structure of 16S RNA, a key component of the 30S subunit of the ribosome.
Isoniazid (INH)	*katG* *inhA* *aphC*	Catalase-peroxidase oxidative defence of *Mtb* cells (DeVito and Morris, 2003).	Inhibit mycolic acid synthesis (Khan, Manialawy and Siraki, 2019).	Mutations in *katG* results in failed activation of INH prodrug.
Rifampicin (RIF)	*rpoB*	ß-subunit of RNA polymerase (Taniguchi *et al.*, 1996b)	Inhibition of RNA synthesis (Goldstein, 2014b).	Mutation in codons in RRDR of the *rpoB* prevents interaction with RIF
Ethambutol (EMB)	*embA* *embB* *embC* *embR*	It is essential for the synthesis of arabinogalactan and lipoarabinomannan in the *Mtb* cell wall (Moure *et al.*, 2014).	Inhibition of cell wall synthesis (Zhang *et al.*, 2020).	Overexpression of *embB* allows the continuation of arabinogalactan production.
Pyrazinamide (PZA)	*pncA* *panD*	Production of nicotinamide adenine dinucleotide from nicotinamide which is essential for cellular metabolism and energy	The mode of action is currently unknown, but it is believed that PZA inhibits the synthesis of fatty acids, thus disrupting the *Mtb* cell wall (Zhang *et al.*, 2003).	Loss of *pncA* activity decreases the conversion of PZA into its active form.

		production (Covarrubias *et al.*, 2021).		
Fluoroquinolones (MOXI and LEVO)	*gyrA* *gyrB*	The DNA gyrase enzyme comprises the GyrA and GyrB subunits which are important in DNA replication processes such as initiation, propagation, and termination (Marchese and Debbia, 2016b).	Convert the target genes gyrase and topoisomerase IV into toxic enzymes, thus altering DNA replication (Jang and Chung, 2020).	Mutation of *gyrA* and *gyrB* prevents interaction with fluoroquinolones.
Delamanid (DLM)	*ddn*	Protects *Mtb* against oxidative stress (Gurumurthy *et al.*, 2013).	It inhibits the synthesis of mycolic acid.	Mutation in *ddn* leads to the failure of DLM activation.
Ethionamide (ETHO)	*ethA*	The physiological function of EthA remains known; it may be involved with synthesis of mycolic acid (Ang *et al.*, 2017).	ETHO is a prodrug that is converted into its active form by EthA. Similarly to INH, it inhibits InhA's ability to synthesize mycolic acid.	Mutations in *ethA* leads to a less functional EthA protein, preventing the conversion of ETHO its active form.
Bedaquiline (BDQ)	*atpE*	ATP synthase subunit c (AtpE) catalyses the production of ATP from ADP (Isa *et al.*, 2021).	Targets the adenosine triphosphate synthase enzyme responsible for energy production in *Mtb* (Sarathy, Gruber and Dick, 2019).	Target modification in AtpE prevents the effective binding of BDQ.
Pretomanid (PA)	*ddn* *fbiA* *fbiB* *fbiC* *fbiD* *ddn* *fgd1*	The cellular functions of *ddn* are currently unknown, but it is believed to be a peripheral protein. *Fdg1* and the coenzymes FbiA/B/C/D are involved in coenzyme F_{420}- enzyme biosynthesis, it is suggested that coenzyme F_{420} might protects *Mtb* from macrophage-generated NO_2 under anaerobic	Inhibition of cell wall synthesis (Manjunatha, Boshoff and Barry, 2009).	Target modification. Fail to convert prodrug into its active form.

		conditions(Selengut and Haft, 2010).		
Linezolid (LZD)	*rplC*	Encodes for methyltransferase that modifies the position of adenosine 2503 in the 23S rRNA of *Mtb* for protein synthesis (Toh *et al.*, 2008).	Inhibits bacterial protein synthesis through the binding of 23S ribosomal RNA of the 50S subunit (Long and Vester, 2012).	Target modification of RplC.
Streptomycin (SM)	*rps* *rpsL* *gid*	These genes are important components of the 30S subunit of the ribosome. The 30S subunit plays an important role in decoding the mRNA codons during translation (Carter *et al.*, 2000).	Targets the 30S subunit of the ribosome and prevents protein synthesis.	Mutations prevent interactions with SM.

1.5 Challenges of working with *M. tuberculosis*

Growing *Mtb in vitro* is a challenging task due to its slow generation time, and the risk of fast-growing contaminants in culture media. Different media have been designed to grow *Mtb* and other mycobacterial species. In 1903, the first egg-based media, Lowenstein-Jensen medium (LJ medium), containing egg as a source of nutrients and malachite green to inhibit the growth of undesirable bacteria, was introduced alongside Petragnani medium for the growth of *Mtb* clinical strains (Kim and Ryoo, 2011). Later, in 1947, other media formulations such as Dubos and Middlebrook were designed containing special ingredients such as dextrose, oleic acid, and albumin; these last two components protect *Mtb* from toxic agents (Preece *et al.*, 2016). It has also been noted that Middlebrook with enrichment supplements such as Oleic Albumin Dextrose Catalase (OADC) growth supplement allows faster growth of *Mtb* with a lower likelihood of contamination than egg-based media. Middlebrook media is also used for drug susceptibility testing in *Mtb* samples, since it allows a faster growth of the bacilli. Whilst these modifications have made it easier to do *in vitro* research using *Mtb*, contaminations still represent a risk and OADC supplements are expensive to procure.

In addition to being a slow growing organism, manipulation of *Mtb* in a laboratory setting can be a dangerous task. *Mtb* is regarded as a hazard group 3 pathogen, which can only be handled in dedicated biosafety level three (BSL-3) facilities. These infrastructural requirements generate additional costs, and logistical challenges, but also necessitate extensive training for laboratory staff.

1.5.1 Model organisms used for *M. tuberculosis* studies

In order to study facets of the biology of *Mtb* and overcome some of the challenges of working with it, model organisms have been employed. By using related tractable organisms that are less pathogenetic, but share characteristics of interest with *Mtb*, it is often possible to accelerate research because they are safer to work with, replicate quickly and are easier to maintain.

Desired characteristics for *Mtb* model organisms will depend on the specific characteristics of *Mtb* to observe, i.e., if studying pathogenicity, then the desired model organism must have the capability to cause similar pathological lesions (i.e., granulomas). If studying drug response, then the model organism needs to have the same mechanisms of generating resistance, but ideally with a shorter generation time than *Mtb*.

In recent years, fast-growing mycobacteria have been used in the search to find new antimicrobial molecules in drug resistant TB studies because some fast-growing mycobacteria have similar inherent and acquired patterns of antibiotic resistance as *Mtb* (Gupta and Bhakta, 2012; Phelan *et al.*, 2015). One of the biggest advantages of using carefully selected NTMs as model organisms is that it is possible to work much faster with organisms that are structurally similar to *Mtb* but have less pathogenicity and at a lower cost. Due to the growing threat of MDR-TB strains, mycobacteria models for drug resistance studies are important.

1.5.2 Current organisms used as model organisms for *M. tuberculosis*

While different mycobacteria have been studied in the past to understand the pathogenicity and behaviour of *Mtb*, only a few are commonly selected as model organisms for *Mtb* studies. The most used model organisms for *Mtb* are members of the MTB complex, *Mycobacterium bovis* (BCG) and the *Mtb* strains H37Ra, and the NTMs *Mycobacterium smegmatis* and *Mycobacterium marinum*.

1.5.2.1 *Mycobacterium bovis* – bacillus Calmette-Guérin (BCG strain)

Currently, of the various mycobacteria that are in use as model organisms for *Mtb* studies, the most used has been *M. bovis* (a member of the TB complex), which also can cause TB in humans but is a more important health problem in the agricultural sector since *M. bovis* can cause TB in cattle and in other mammals. An attenuated strain of *M. bovis,* BCG, was created for the development of the BCG vaccine, which is the only vaccine used against TB and is regarded as the most used vaccine in history (Horwitz *et al.*, 2010; Asadian *et al.*, 2022).

M. bovis BCG has been widely used as a model organism for *Mtb* studies due to its close genetic relationship with *Mtb*; conserved genes are 99% identical to *Mtb*. There are, however, significant differences between BCG and *Mtb* (Bottaro, Larsen, 2010). Among the differences between these two organisms, is that BCG contains a natural region of difference 1 (RD1) deletion; this locus encodes for at least 9 genes in *Mtb*, some of which are related to virulence, induction of cell death, and cytokine secretion in infected macrophages (Daugelat *et al.*, 2003; Yang *et al.*, 2014). Another major difference is that all *M. bovis* isolates, including BCG, are intrinsically resistant to PZA, and additionally, *M.*

bovis isolates have reduced cord formation compared to *Mtb*. BCG is also a slow-growing organism, with a 20-hour doubling time, and requires enrichment supplements for its growth, such as OADC, making it as slow to work with as *Mtb*.

1.5.2.2 *M. tuberculosis* strain H37Ra

M. tuberculosis strain H37Ra is an attenuated strain of *Mtb*. Both *Mtb* H37Ra and its virulent counterpart H37Rv were both isolated from the parent strain H37; this parent strain was obtained from a male patient with chronic pulmonary TB in 1905 (He, Du and Du, 2013).

Mtb H37Ra ("a" for avirulent) has been used extensively as a reference for protein and pathogenicity studies due to its similarities to its counterpart H37Rv ("v" for virulent) as well as for resistance studies as both strains have equivalent MICs when exposed to antitubercular drugs (Heinrichs *et al.*, 2017). There are also phenotypic differences between the two sister strains; H37Ra has a different colony morphology (rough colonies), compared to H37Rv (smooth colonies), does not exhibit cord formation, and has a lower survival rate inside macrophages and under anaerobic conditions (Zheng *et al.*, 2008).

In different countries such as China, the strain H37Ra has been considered a safe model to work with, and it has been classified as a biosafety level 2 (BSL-2) organism due to studies in guinea pigs and mice (Xu, Wang and Xu, 2020) where the loss of virulence was observed.

However, in the United Kingdom, the strain H37Ra is still classified as a BSL-3 organism by the UK Health Security Agency. The study by D'Arcy Hart & Armstrong (1974) (D'Arcy Hart and Armstrong, 1974) has observed the survival of H37Ra strain within guinea pig macrophages and in *in vitro* experiments (Heinrichs *et al.*, 2017), these studies also suggested that H37Ra should be considered a low-virulence strain rather than a truly avirulent strain.

1.5.2.3 *Mycobacterium smegmatis*

In 1889, scientists Karl Bernhard Lehmann and Rudolf Otto Neuman isolated a new species of *Mycobacterium* and named it *Mycobacterium smegmatis* (T *et al.*, 2020). This *Mycobacterium* is mainly found in soil, water, and plants. When grown on 7H10 agar, it presents as a shiny white colony with wrinkled edges, and after 48-hours of incubation, its colour changes to a velvety yellow followed with a smoother appearance.

M. smegmatis is another *Mycobacterium* widely used for *Mtb* studies. *M. smegmatis* has an advantage over *M. bovis* BCG because its doubling time is only 4 hours, much faster than the 20 hours doubling time of *M. bovis* BCG. *M. smegmatis* is also a BSL-2 organism, making it safer for manipulation, and it is not known to cause any severe disease in humans. *M smegmatis* is also considered a transformable *Mycobacterium* and has been used to study the expression of vectors and the expression of luciferase-reporter phages. It was used for the initial compound screening for the compound that later became bedaquiline, the first new anti-TB drug to be licensed for use in humans in over 50 years (Andries *et al.*, 2005).

However, the safety that makes *M. smegmatis* easy to work with, is also its most significant disadvantage as a model organism for *Mtb. M. smegmatis* does not possess essential virulence genes found in *Mtb*, therefore it cannot be used to study pathogenesis; nonetheless, it provides a model to study mycobacteria in general (Reyrat and Kahn, 2001).

Another disadvantage of *M. smegmatis* as a model organism is that most of its strains are either intrinsically resistant to TB drugs such as PZA or have a high resistance to other TB drugs such as RIF and INH (Lelovic *et al.*, 2020). Nevertheless, it has played a role in discovering the mechanism of resistance for drugs such as INH and EMB (Sparks *et al.*, 2023).

1.5.2.4 *Mycobacterium marinum*

Mycobacterium marinum is a *Mycobacteriu*m first described in 1929 by Joseph D. Aronson and is a close relative of *Mtb* (Aronson, 1926) with ~3000 shared orthologs proteins with 85% amino acid identity with *Mtb* (Stinear *et al.*, 2008). This *Mycobacterium* was extracted from the spleen and liver of fish found dead in the Philadelphia Aquarium in 1926. *M. marinum* is found in plants and, soil but is known to infect fish, reptiles, and amphibians (convenient laboratory animals), producing granuloma-like structures in these animals similar to the granuloma-like structures found in human lungs produced by *Mtb* (Ramakrishnan, 2004).

M. marinum is also a rapid grower when compared with *Mtb* with a doubling time of 7 hours. It can cause cutaneous lesions in both immunocompromised and immunocompetent humans, with lesions most observed in fishermen and others at

occupational risk. The lesions produced by *M. marinum* in humans are usually localized in hands, feet, or arms (exposed skin), and it presents as smooth or warty nodules observed between 2 to 8 weeks after inoculation (Canetti *et al.*, 2022).

M. marinum is not known to spread to internal organs in humans, making its pathogenicity lower than that of *Mtb*. This may be partly explained by the observation, in some studies, that infection in mammals by *M. marinum* is strictly temperature dependant, and its growth is inhibited at 37°C unlike *Mtb* (*M. marinum* has an optimal temperature for growth of 30°C) (Ramakrishnan, 2004). *M. marinum*, has not been used for AMR resistance studies since this organism has shown *in vitro* resistance to antitubercular drugs such as INH, RIF, EMB, and PZA (Haberecht, Cull and Wieland, 2024). Like *Mtb*, *M. marinum* can survive inside the host cells of fish, making *M. marinum* a proposed model organism for pathogenicity studies (Petrini, 2006). Important advances in *Mtb* research have been possible thanks to *M. marinum*, a goldfish model has been developed for the study of mycobacterial pathogenicity (Talaat *et al.*, 1998) and zebrafish embryos infected with *M. marinum* have allowed the study of granuloma formation (Swaim *et al.*, 2006).

1.5.2.5 *Mycobacterium komossense*

Mycobacterium komossense is an environmental *Mycobacterium*, first identified in the Komosse sphagnum bog, in south Sweden (Kazda and Muller, 1979). NTMs, such as *M. komossense* can grow in demanding habits, such as low pH soils, and in low temperatures, and it has shown to be less fastidious compared to other mycobacteria.

J. Kazda and K. Muller were the first to describe *M. komossense* in 1979. They performed a series of taxonomic studies and reported their findings: among these experiments were lipid production, and enzymatic activity experiments. A pathogenicity test was done in which 5 rabbits, and 10 guinea pigs were infected intravenously with 10 mg of *M. komossense* bacterial mass, and 5 mice were infected intravenously with 1 mg of bacterial mass of *M. komossense*. After 8 weeks the animals were sacrificed and examined, and the results concluded that the different strains of *M. komossense* did not produce any lesions in any of the infected animals (Kazda and Muller, 1979). The Kazda and Muller study was performed using 18 strains of *M. komossense*, and one of which, Ko2, was deposited in the American Type Culture Collection (ATCC) under the number 33013.

Kazda and Muller also performed different experiments to see at what temperature *M. komossense* could grow (22°C, 31°C, and 37°C) and how many days it took for the observed growth in cultures (7 days), unfortunately some of the laboratory methods used by Kazda and Muller were omitted from the paper.

The strain of *M. komossense* used for this thesis, JERR01, was obtained from the Scottish Mycobacteria Reference Laboratory (SMRL) and was isolated from a urine sample obtained from an elderly male patient in 2010. The strain was categorized as *M. komossense* based on sequencing of its 65 kDa heat-shock protein (*hsp65*) gene. Since this organism was isolated from a human, and the initial characterization of this organism from the Royal Infirmary revealed sensitivity of this organism towards antitubercular drugs (e.g., RIF and EMB), this raised the possibility that this strain would

have properties that would make it a suitable model organism to study antibiotic response, and mechanisms of resistance in *Mtb*.

1.6 Aims and Objectives of the research in this thesis

The emergence and propagation of MDR-TB strains pose a threat to global health. The overarching aim of this thesis is to examine whether the clinical strain of *M. komossense* obtained from SMRL (strain JERR01) exhibits properties that would make it a suitable model organism to study *Mtb*. Given the global burden of AMR and the importance of antibiotics for the effective treatment and management of *Mtb* infections, an additional aim of this thesis will be to demonstrate the suitability of *M. komossense* JERR01 to study resistance to clinically relevant antibiotics used in the treatment of TB. This thesis proposes to achieve this goal by pursuing the following objectives:

1. Perform phenotypic comparisons between *M. komossense* JERR01 and the reference strain of *M. komossense,* ATCC 33013, by comparing physical traits, growth patterns, and their MIC_{90} profiles against antitubercular drugs.
2. Genetically characterize *M. komossense* JERR01 through the sequencing of its genome and investigation of its annotation; and undertake comparative genomic analysis to explore its genetic and functional relationships with *M. komossense* ATCC 33013 and other mycobacteria, including *Mtb*.
3. Undertake selection experiments using INH and RIF to demonstrate the evolution of antibiotic resistance in *M. komossense* JERR01 and characterize the genetic basis of resistance; comparing this to known mechanisms resistance in

Mtb will help to assess the suitability of *M. komossense* JERR01 as a model to study AMR.

2 Materials and Methods

2.1 Bacterial strains

M. komossense strain JERR01 was obtained from the SMRL, Royal Infirmary of Edinburgh (RIE) where it was isolated from a clinical urine sample in 2010. This strain had been identified as *M. komossense* by sequencing of the *hsp65* gene. *M. komossense* strain Ko2 was obtained from the American Type Culture Collection (ATCC; *M. komossense* strain Ko2 is subsequently referred to as *M. komossense* ATCC 33013) (Kazda and Muller, 1979), *M. smegmatis* strain MC2 155 (ATCC 700084) from Dr Hua Wang, University of Glasgow, *M. abscessus* subsp. *massiliense* from Dr Daniela Lima, University of Dundee, and a representative strain of *Bacillus subtilis* from Dr Robert Hammond, University of St Andrews.

2.2 Media and supplements

The culture media used during this work were Middlebrook 7H9 Broth Base (Sigma-Aldrich, UK, M0178) and Middlebrook 7H10 Agar Base (Sigma-Aldrich, UK, M0303). Both broth and agar media were prepared according to manufacturer instructions. The media was sterilized by autoclaving at 15 lb pressure (121°C) for 10 minutes. In addition, the broth was supplemented with 4 ml of glycerol 50% (Sigma-Aldrich, UK, G6279) per 450 ml of broth media, and 0.05% (v/v) Tween-80 (Fisher Scientific, Fisher BioReagents, UK, 11473601) to avoid clumping. Solid media was supplemented with 10 ml of glycerol 50% (Sigma-Aldrich, UK) per 450 ml of agar media. Additionally, 0.75 M of monopotassium phosphate (KH_2PO_4) (Sigma-Aldrich, UK, P0662) was used to acidified 7H9 Middlebrook broth media to a pH of 6.0 for PZA AST.

2.3 Mycobacterial culture and preparation of glycerol stocks

To create glycerol stocks of mycobacterial strains, 10 µl of a pre-grown culture at a OD_{600} of 1.0 were spread onto Middlebrook 7H10 agar (Sigma-Aldrich, UK, M0303), supplemented with 10 ml glycerol 50%, and incubated until colonies were visible to the naked eye. Once colonies were visible, single colonies were inoculated back into in 10 ml of 7H9 Middlebrook broth, 120 rpm and incubated at 30°C for *M. komossense* JERR01 and *M. komossense* ATCC 33013, and 37°C for *M. smegmatis* MC2 155 for 5 days or until reaching an OD_{600} of 1.0. Once the bacterial cultures reached an OD_{600} of 1.0, 800 µl of bacterial culture was mixed with 200 µl of 75% (v/v) glycerol solution (diluted with milli-Q water and solution autoclaved). The bacterial suspension was left at room temperature for 30 minutes, to allow the glycerol to penetrate the cells. Then the stock tubes with mycobacteria were stored at -80°C until required. To start a new culture, 5 µl of the bacterial suspension was plated on Middlebrook 7H10 agar. One colony was picked up, using a disposable plastic loop, to start the new liquid culture. After two passages, a new culture was started from the glycerol stock.

2.4 Gram staining

From an agar plate incubated for 5 days, a single colony was picked up and smeared onto a slide creating a thin layer. The bacteria were heat-fixed to the slide by passing the slide twice over a Bunsen burner. Heat-fixed slides were left to cool to room temperature and then two drops of crystal violet were added to the smear and left for 1 minute. The slides were then washed with distilled water until the water running from the slides was clear. Using a clean Pasteur pipette, two drops of Gram's Iodine were added to the slides and left for 1 minute before rinsing off with distilled water. The slides were de-stained by

adding 10 drops of acetone-alcohol solution, with the glass-slide tilted at an angle. The excess acetone-alcohol was washed off with distilled water, two drops of Safranin were added to the slides for 1 minute and then washed off with distilled water, again enough water was used to wash-off the excess safranin until the water from the slides ran clean. The slides were carefully dried with filter paper and viewed under the microscope. The microscope used throughout this thesis to examine these stained slides was the Olympus Bx51 Transmitted Light Microscope and the software used to process of the images was Leica Application Suite (LAS) version 4.6.0.

2.5 Acid fast stain (Ziehl-Neelsen technique)

A single colony was picked from the agar plate and smeared onto the glass slide creating a thin layer. Once the slide had dried, bacteria were heat-fixed with a Bunsen burner and set aside to cool down. Using a clean Pasteur pipette, two drops of carbol fuchsin were added to the cells for 20 minutes before washing off with distilled water. The slide was rinsed with acid alcohol for 1 minute and excess acid-alcohol was washed off from the slide with distilled water. Two drops of methylene blue were added for 1 minute to counterstain non-acid-fast cells. The slides were rinsed with distilled water to remove excess stain and dried gently using filter paper before inspection under the microscope.

2.6 Capsule stain

A single colony was picked from the agar plate and smeared unto a glass-slide creating a thin layer and allowed to air dry. On this occasion, bacteria were not heat-fixed as heat can destroy the capsules. The smear was covered with 1% crystal violet for 2 minutes and rinsed with 20% solution copper sulphate (Sigma-Aldrich, UK, MKC6873). The glass-

slides were left until fully air-dried. No blotting was used to speed up the drying process as this action could remove non-heat-fixed bacteria from the slide and disrupt the capsule. Bacteria and stained capsules were observed under the microscope.

2.7 Miles-Misra technique for Colony Forming Unit (CFU) counting

Serial 10-fold dilutions of the original culture were prepared in Phosphate-buffered saline (PBS) as shown in

Figure **2-1**. Three 10 µl drops from each dilution were then spotted onto a 7H10 agar plate and left to dry completely in a microbiological safety cabinet. The plates were sealed with parafilm to prevent dessication and incubated at 30°C until colonies were visible (5 days for *M. komossense* JERR01) (Miles, Misra and Irwin, 1938; Datta, 2021). Colony forming units (CFU) per ml were calculated using the Equation 2.1:

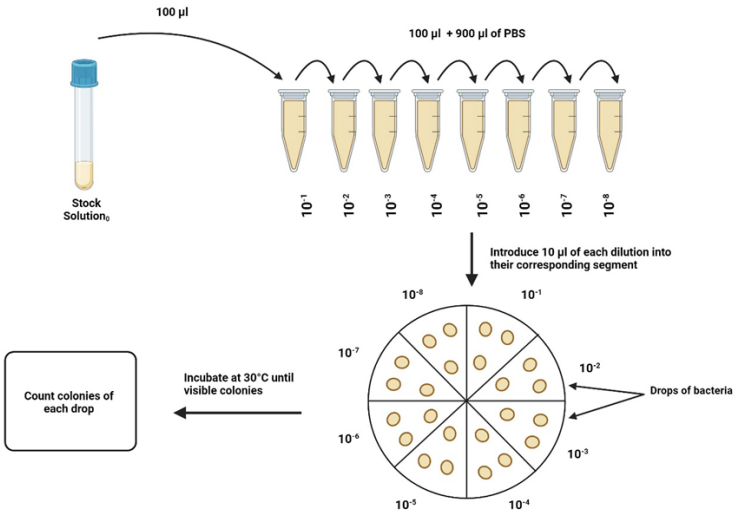


Figure 2-1. Schematic representation of the Miles and Misra method.

From the original culture, a series ten-fold dilutions were prepared in 900 µl of PBS. Three 10 µl drops of each dilution were spotted onto the appropriate segment of a 7H10 agar plate and left to dry completely. The agar was incubated at 30°C and the colonies were counted when colonies became visible.

Equation 2.1

$$CFU/mL = \ average\ number\ of\ colonies \times \ dilution\ factor \times 100$$

2.8 Determination of bacterial growth kinetics

To monitor bacterial growth in broth, a single colony of *M. komossense* JERR01 or *M. komossense* ATCC 33013 from a 7H10 Middlebrook agar plate (after 5-day incubation at 30°C) was transferred into 5 ml of 7H9 Middlebrook broth supplemented with 4 ml of glycerol and 0.05 g of Tween-80 and grown at 30°C, with constant shaking at 120 rpm until reaching an OD_{600} of 2.0. Once this culture had reached the desired OD_{600}, a sufficient volume was transferred to 100 ml of Middlebrook broth in a 250 ml baffled flask to reach a starting OD_{600} of 0.05 and grown at 30°C, with constant shaking at 120 rpm until an OD_{600} of 2.0 (approximately 5 days) was reached. Three new 100 ml cultures in 250 ml baffled flasks were prepared from this culture with a starting OD_{600} of 0.05 and grown at the following temperatures: 30°C, 37°C, and 42°C, at 120 rpm. Two 1 ml samples were taken from each bacterial culture twice a day (6 hours apart) at the same time for 7 days to monitor bacterial growth by measuring the OD_{600} and viable bacterial CFUs as shown in Figure 2-2.

Bacterial growth kinetics experiments were conducted using three independent biological repeats and one technical repeat per sample. Growth curves were plotted of the OD_{600} and CFU/ml measurements against time of sampling. The results provided by this experiment were used to create growth curves using both measurements CFUs and OD_{600}.

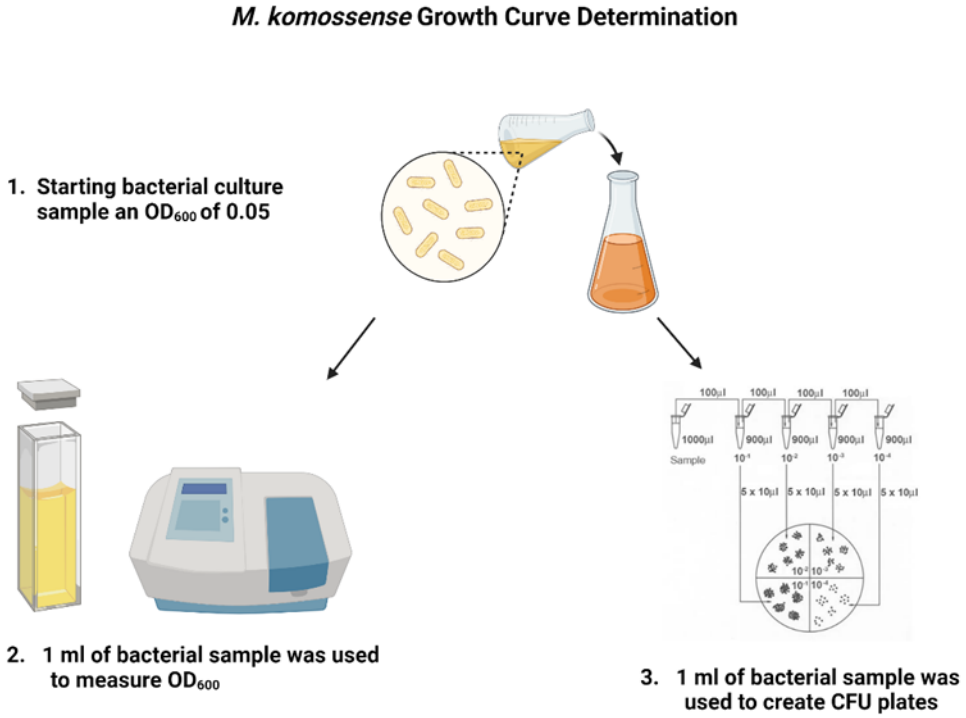


Figure 2-2. Workflow for growth rate determination experiment.
From a pre-grown culture, enough bacterial suspension is added to 100 ml of Middlebrook to reach an initial OD_{600} of 0.05. For 7 seven days, two different type or measurements were taken, 1 ml for optical density (measuring turbidity) and 1 ml for CFU counting (counting of viable cells).

2.9 Generation time

To calculate the mean generation time of these mycobacteria, data from the exponential phase of the CFU growth curves at 30°C and 37°C were used. The mean generation time (G) was calculated for each temperature using Equation 2.2 (Labster Generation time, 2024, https://theory.labster.com/generation_time_bacteria/. Accessed 2 of March 2024), in which "G" is the generation time, "t" is the time interval, and "n" is the number of generations within the exponential phase.

Equation 2.2

$$G = t/n$$

The time interval for this equation was determined by selecting the two points with the steepest slope in the exponential phase, corresponding to the initial number of bacteria (N_0) and the final number of bacteria (N_t). The value for n (number of generations) can be calculated using Equation 2.3:

Equation 2.3

$$n = 3.3\ x \log\left(\frac{N_t}{N_0}\right)$$

2.10 Minimum Inhibitory Concentration (MIC_{90}) determination method

2.10.1 Antibiotic preparation

Table 2-1. Preparation of antitubercular agents that were evaluated against *M. komossense* JERR01.

Antimicrobial agent	Sigma catalogue number of drug	Solvent	Range of final concentrations (µg/ml)
Isoniazid (INH)	I3377	MQH2O	2-512
Rifampicin (RIF)	SBR00065	DMSO	0.0007-512
Ethambutol (EMB)	E4630	MQH2O	0.0156-512
Pyrazinamide (PZA)	P4050000	DMSO	0.003-512
Moxifloxacin (MOXI)	SML1581	MQH2O	0.0156-512
Levofloxacin (LEVO)	28266	DMSO	0.0156-512
Bedaquiline (BDQ)	ATE517247923	DMSO	0.003-512
Pretomanid (PA)	SML1290	DMSO	0.003-512
Linezolid (LZD)	PZ0014	DMSO	0.0156-512

All nine antitubercular agents studied (summarised in Table 2-1) were obtained from Sigma-Aldrich, the same licensed drug manufacturer, together with the appropriate

documentation for quality assurance. The drugs were dissolved according to the manufacturer's instructions. Some antitubercular drugs used for MIC_{90} estimation were not soluble in milli-Q water (MQH2O). Where solvents other than water were used to dissolve the antitubercular drug, I tested whether these solvents had any antibacterial or inhibitory effect by adding the solvent (DMSO) at identical concentrations to those found in the drug-containing assay and assessing any impact on bacterial growth.

Stock solutions of each antibiotic were prepared using a high concentration of the drug (~50 mg/ml) which was then diluted tenfold, transferred to 10 ml vials, then stored at –20°C. From these original stock solutions, working solutions were created. To avoid using a stock solution that might have lost some of its anti-bacterial properties, dates were written on vials containing each stock solution and stocks were not retained for longer than one week.

Standard stock solutions of INH, EMB, and MOXI were prepared by dissolving in Milli-Q water and filtering with a 0.22 µm pore sized cellulose membrane, while RIF, PZA, LEVO, BDQ, PA and LZD stock solutions were prepared in DMSO (Sigma-Aldrich, UK). Working solutions were prepared on a final volume of 10 ml in Middlebrook 7H9 broth. For every drug tested against any of the mycobacteria used in this thesis, the highest concentration used was 512 µg/ml.

2.10.2 EUCAST Reference Method

The EUCAST reference method used to test antimicrobial was the broth microdilution AST method and was used as a comparative standard AST method in the study. Bacteria were cultured from 10 µl of glycerol stock in 10 ml of Middlebrook 7H9. The bacterial concentration culture was adjusted to an OD_{600} reading of 0.05 (~10^7 CFU/ml), (Biochrom™ Ultrospec™ 10, Fisher Scientific, UK) and standardised to a final bacterial concentration of 10^5 CFU/ml in a microwell plate (Nunc™ MicroWell™ 96-Well, Nunclon Delta-Treated, Flat-Bottom Microplate, Thermo Scientific, UK). Using a multichannel pipette, two-fold serial dilutions were made by adding 200 µl of the antibiotic solution present in the highest concentration row, and moving 100 µl from each row to wells in the next row which already contained 100 µl of Middlebrook 7H9 media (thus attaining the two-fold dilution). The antibiotic concentrations used were based on EUCAST/CLSI/WHO/European Centre for Disease Control clinical breakpoints and critical concentrations (Niward et al., 2016; European Centre for Disease Control, 2018; Woods et al., 2018; World Health Organization, 2018; EUCAST, 2024). The plates were then covered with Breathe-Easier sealing membranes for multi-well plates (Scientific Laboratory Supplies, UK) to prevent dehydration and contamination. The plates were set to incubate at 30°C for 5 days and the turbidity observed in the plates was measured in a spectrophotometer (CLARIOstar PLUS microplate reader, BMG Labtech) at an absorbance of 600 nm, with the results recorded (Figure 2-3). The percentage inhibition for each treated sample was calculated using Formula 2 (Rekha, Kulandhaivel and Hridhya, 2018). The MIC_{90} for each drug was reported as the lowest antibiotic

concentration which achieved more than 90% inhibition of bacterial growth. These MIC_{90} determination experiments were run in triplicates.

Formula 2.4

$$Percentage\ Inhibition\ (\%) = \left(\frac{Control\ OD - treated\ sample\ OD}{Control\ OD}\right) x\ 100$$

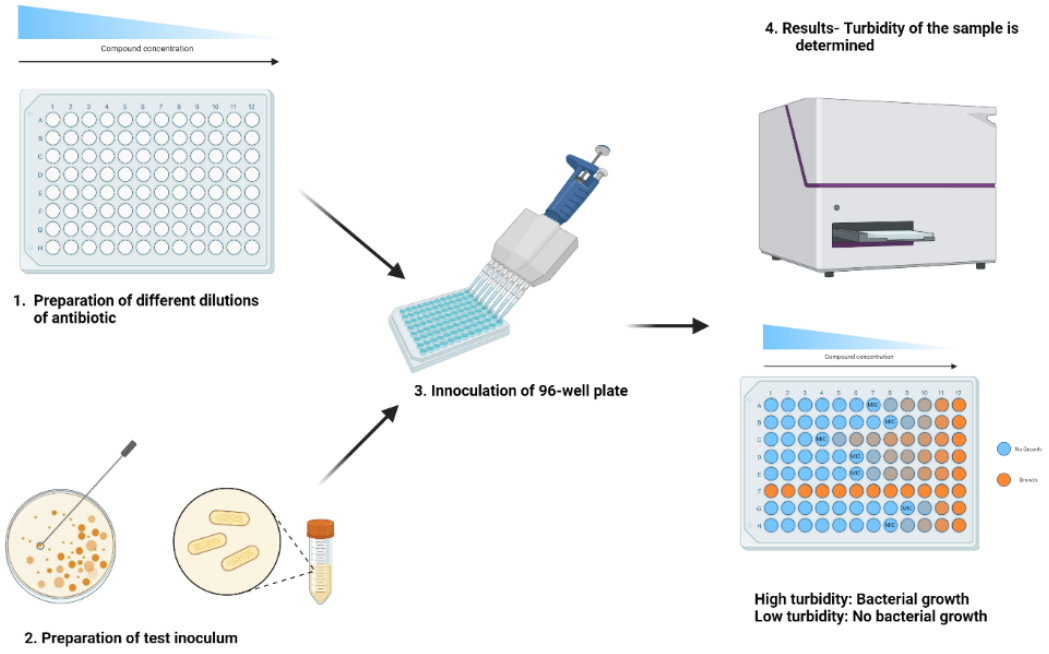


Figure 2-3. Workflow diagram for the MIC_{90} determination of *M. komossense* JERR01 for 9 antitubercular drugs.
Starting with the preparation of the antitubercular agents, followed by the preparation of the inoculum, both preparations were added to the 96-microdilution well plates and sealed to avoid dehydration. Plates were set to incubate and after a period of 5 days, the results were then recorded in a plate reader at a wavelength of 600 nm.

2.11 Bactericidal activity of RIF and INH against wild type *M. komossense* JERR01

To determine the minimum bactericidal concentration (MBC), which is the minimum concentration of a drug required to kill 99.9% of the initially inoculated organism, a sample of 10 µl was taken from each well containing bacteria exposed at all concentrations of INH and RIF equal to or above the MIC_{90} and added to 90 µl of PBS buffer (Sigma-Aldrich, UK). Serial dilutions were made from 10^{-1} to 10^{-8} and CFU/ml were determined according to the Miles and Misra method plates on Middlebrook 7H10 agar plates after 5 days of incubation at 30°C.

2.12 Mutation frequency of *M. komossense* JERR01 against RIF and INH

To determine the mutation frequency of *M. komossense* JERR01 to RIF and INH, 100 ml of 7H9 Middlebrook broth were inoculated with bacteria to a starting OD_{600} of 0.05 and incubated at 30°C in a 250-ml baffled flask at 120 rpm until reaching an OD_{600} of 3.0. These cultures were then centrifuged at 4000g for 30 minutes. The supernatant was discarded, and the concentrated pellet was re-suspended in 1 ml of PBS buffer (Sigma-Aldrich, UK). CFU/ml were determined as described above (Chapter 2.7). The resuspended pellet was diluted 10-fold in PBS and 200 µl aliquots of this dilution were spread equally onto ten 7H10 Middlebrook agar plates for each antibiotic (for RIF=0.1 µg/ml, 0.2 µg/ml, and 0.4 µg/ml and for INH=128 µg/ml, 256 µg/ml, 512 µg/ml). Plates were sealed with parafilm to avoid dehydration and incubated at 30°C until colonies were visible to the naked eye (Figure 2-4). The mutation frequency (µ) was determined using Equation 2.5, where N is the average of number of cells present in the culture and m is the number of mutant colonies grown on the selective plates (Lang & Murray, 2008).

Equation 2.5

$$\mu = m/N$$

For each experiment one colony was selected and plated onto both non-selective 7H10 plates (no antibiotic added) and selective 7H10 agar plates (containing the antibiotic used for the evolution experiment). Each confirmed mutant isolate was then stored at -80°C until required for later genome sequencing.

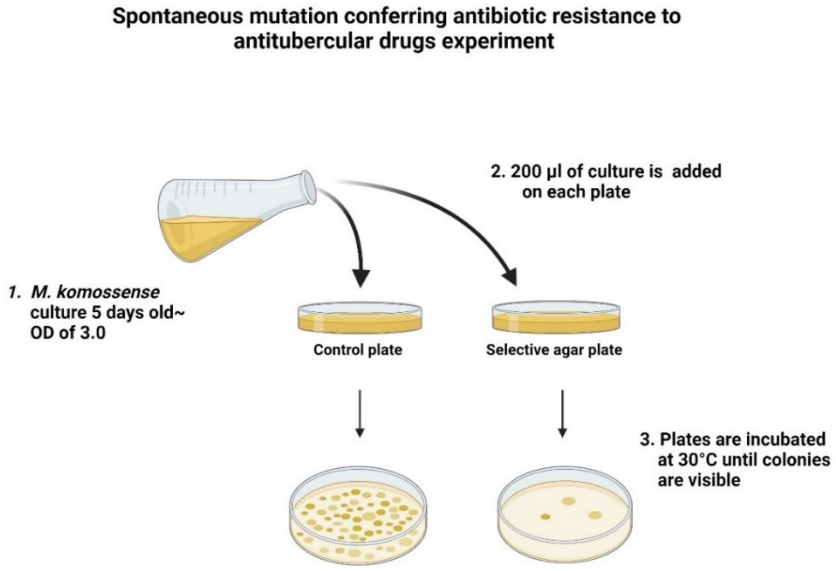


Figure 2-4. Simplified diagram followed for the evolution experiment for the determination of the mutation frequency of *M. komossense* JERR01.
Selective agar plates containing RIF (0.1, 0.2, 0.4 µg/ml) and INH (128, 256, 512 µg/ml) were created, and to these plates, *M. komossense* JERR01 at a concentration of 10^{10} was added. After the incubation period, colonies in selective plates were counted and a number of isolates picked and stored for later genome sequencing.

2.13 Scattered Light Integrating Collector (SLIC)

Growth rate determination using SLIC.

SLIC is a novel diagnostic tool being developed at the University of St Andrews to measure bacterial growth and antibiotic response (Hammond *et al.*, 2017, 2022; Falconer *et al.*, 2024). Exploratory use with *M. komossense* JERR01 will be described in Chapter 3. The total light scatter (mV) of bacterial cells was measured using a laser at a wavelength of 635 nm. The laser was split across 6 sample wells and the light scatter was converted to an electrical signal (decibels (dB)) by 6 individual photodetectors. The signal to noise ratio was minimised by a Lock-In Amplifier, Laser Controller (LIALC). The LIALC also controls the laser power for the entirety of the run (Figure 2-5).

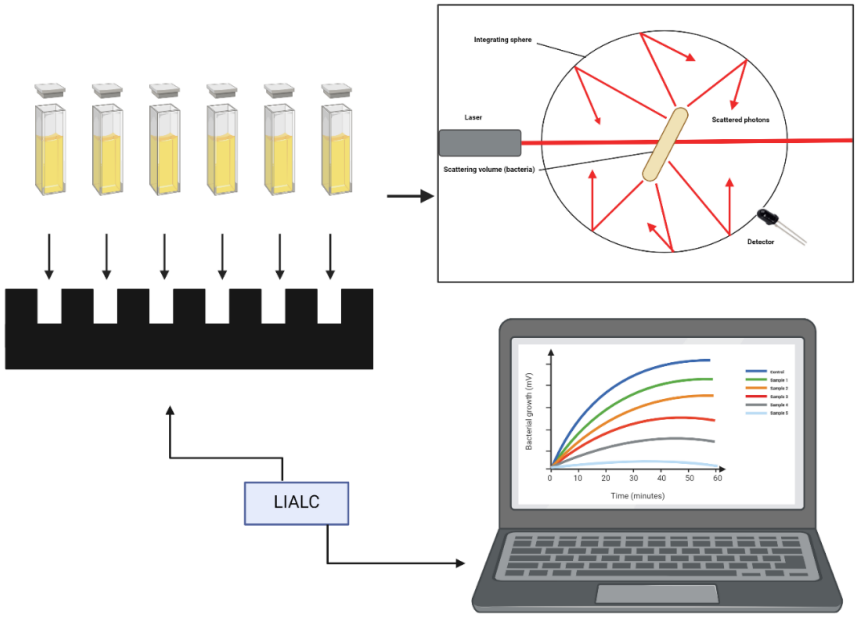


Figure 2-5. Schematic laboratory workflow for the use of SLIC.
SLIC is controlled via a laptop using the HyperTerminal software. The LIALC acts as a feedback loop to the software based on the output of the photodetectors. The sample carriage contains 6 sample wells which can be simultaneously measured by a laser at a wavelength of 635nm. The heating element allows the sample carriage to incubate at 30°C for the duration of the experiment (24 hours).

Bacterial cultures were grown in 10 ml of 7H9 Middlebrook broth for 5 days, 120 rpm at 30°C. Bacteria were then diluted in fresh, pre-warmed Middlebrook 7H9 broth (30°C) to an OD_{600} of 0.05, corresponding to 10^7 CFU/ml. A single bacterial preparation at a concentration of 10^5 CFU/ml in 50 ml of 7H9 Middlebrook broth was used for the devices. For each cuvette, 2 ml of bacterial preparation was added to warm media broth (10^5 CFU/ml per cuvette).

As cell aggregation and sedimentation at the bottom of culture tubes can severely impact growth measurements using SLIC, and since this frequently is the case with mycobacteria, the Middlebrook 7H9 broth used for growth was supplemented with 4 ml of glycerol, 0.5 g Tween-80 per 450 ml of broth and 1 mg/ml of D-arabinose (Anton, Rougé and Daffé, 1996). The cuvettes were capped and loaded into the pre-warmed (30°C) SLIC device and light scattering recorded every second over a period of 24 hours. The data acquisition was set to 1 second. Real time data was saved as a text comma-separated file (text.csv).

2.13.1 Dose response of *M. komossense* JERR01 to RIF and INH using SLIC

Preparation of cuvettes

For each cuvette, 200 µl of antibiotic solution was added at 10x the desired concentration alongside a 1780 µl aliquot of sterile 7H9 Middlebrook broth and 20 µl of diluted culture stock (final concentration ~ 10^5 CFU/ml) to make a final volume of 2 ml. These cuvettes were capped, then loaded into SLIC and scattering measured for up to 24 hours (1,440 minutes). As before, all samples were warmed to 30°C, and transferred directly to the SLIC device for measurement.

2.13.1.1 Data Analysis and presentation

At the end of the run, the text.csv file was converted to an excel worksheet. Data were manually normalised (transformed from dB to mV) by subtracting the initial signal output from each time point, making the first time point equal to zero. The following formula was used to calculate the growth percentage for each treated sample:

Formula 3

$$Growth\ percentage\ \ (\%) = \left(\frac{Control\ mV - treated\ sample\ mV}{Control\ mV}\right) x\ 100\%$$

The software GraphPad Prism 10 was used for the SLIC graphs. Data shown comprise the mean of the reading taken from three independent experiments with 1 standard deviation (SD) in error bars.

2.14 Bioinformatics analysis

2.14.1 Preparation of samples for Whole Genome Sequencing of *M. komossense* JERR01

Bacterial liquid cultures were grown in liquid broth, until the cultures reached an OD_{600} of 0.5-1.0. Thereafter, the cultures were centrifuged for 30 minutes at 4000 g and the supernatants were discarded. Pellets were placed into bead tubes supplied by MicrobesNG containing 500 µl of cryopreservant liquid. The tubes were mixed by inverting them 10 times. The tubes were sealed and sent to MicrobesNG (https://microbesng.com), University of Birmingham. DNA extraction and Library preparations were performed by MicrobesNG.

2.14.2 Genome sequencing for *M. komossense* JERR01 - Short read sequencing

Sequencing was performed by MicrobesNG using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA is increased 2-fold, and PCR elongation time is increased to 45 seconds. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Libraries are sequenced on an lllumina NovaSeq 6000 (Illumina, San Diego, USA) using a 250 bp paired end protocol. Reads were adapter trimmed using Trimmomatic version 0.30 with a sliding window quality cutoff of Q15. This protocol was obtained from the microbesNG website (https://microbesng.com/microbesng-faq/).

2.14.3 Genome sequencing for *M. komossense* JERR01 - Long read sequencing

For the long-read data, MicrobesNG generated the sequencing by using the Rapid Barcoding kit 96. Sequencing was performed on a GridION (Oxford Nanopore Technologies) using an R10.4.1 flowcell, with basecalling model r1041_e82_400bps_hac_v4.2.0. Reads were randomly sub-sampled to 50X coverage using Rasusa (V 0.7.1) (protocol extracted from MicrobesNG website (https://microbesng.com/microbesng-faq/).

2.14.4 Genome Assembly

From MicrobesNG we obtained Illumina short short-read and nanopore long-read data. The Nanopore long reads were assembled using Unicycler v0.5.4, and using the optimizer SPAdes. The nanopore reads were also assembled using Unicycler, which runs miniasm and Racon pipelines for the best possible assembly (https://github.com/rrwick/Unicycler). The following script was used to run the assembly in unicycler:

```
unicycler -l long_reads.fastq.gz -o output_dir
```

To determine the statistics on the assembly the following script was used:

```
quast -o output_dir –threads 4 assembly_file
```

The completeness of the assembly was assessed by using BUSCO v5.6.1 (https://busco.ezlab.org/busco_userguide.html). This program was carried out by using the recently performed assembly against BUSCO created datasets (which can be downloaded from the BUSCO website). The script used for BUCSCO was:

```
busco -m MODE -i INPUT -o OUTPUT --auto-lineage-prok
```

2.14.5 Annotation and genomic analysis

Long read assemblies were annotated with Prokka v1.14 to identify protein coding sequences (CDSs), tRNAs, Insertion sequences (ISs), Cluster of Orthologous Genes (COGs) and rRNA:

```
prokka sequence.fasta
```

Prokka produced the following files: *gbk, *err, *faa, *ffn, *fsa, *gff, *sqn, *tbl, *tsv, *log and a text file with annotation statistics. Circular images of the *M. komossense* JERR01 chromosome and possible plasmids were generated using Proksee and annotations generated by Prokka (*gbk files) were added to the circular maps. Phylogenetic relatedness to other member of the genus *Mycobacterium* was determined through average nucleotide identities (ANIs) of the full genomes using OrthoANI (http://www.ezbiocloud.net/sw/oat) (Lee et al., 2016).

Genomic comparisons between *M. komossense* JERR01 and other mycobacteria were created using Artemis Comparison Tool (ACT) (Carver *et al.*, 2005) (https://sanger-pathogens.github.io/Artemis/ACT/). To use ACT, a pairwise alignment of the two genomes in comparison was performed by firstly ordering a query assembly against the reference chromosomes, using abacas through the following script:

```
abacas.1.3.1.pl -r ref.dna -q quary_sequence.fasta -p nucmer -b
-d -a -c -o quary_sequence.ordered
```

Then the script:

```
formatdb -p F -i ref.dna
```

was used to format the reference as a blast database and then the following script:

```
blastall -p blastn -m 8 ref.dna -i quary_sequence.ordered.fasta
-o ref.dna_vs_query_sequence.ordered.fasta
```

to run a blast comparison between the two genomes. The outputs of these scripts were visualized in ACT using the following script:

```
act ref.dna ref.dna_vs_quary_sequence.ordered.fasta
quary_sequence.ordered.fasta &
```

Subsystem analysis and comparison between mycobacterial species was also performed using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Sato, 2000). Protein sequences from *M. komossense* JERR01, *M. komossense* ATCC 33013, *M. vanbaalenii* PYR-1, *M. aurum* NCTC 10437, *M. vaccae* ATCC 95051, *M. gilvum* Spyr1, and *Mtb* H37Rv were annotated using GhostKoala v2.0 (https://www.kegg.jp/ghostkoala/) (Kanehisa, Sato and Morishima, 2016) against the KEGG gene database, to obtain a subsystem distribution and the number of proteins per genome present on each subsystem. The subsystem distribution histograms were created using Microsoft Excel from percentage values generated by GhostKoala.

Comparison between AMR associated genes present in *M. komossense* JERR01 and *Mtb* H37Rv was performed by using Needleman-Wunsch Global Protein alignment algorithm (Needleman and Wunsch, 1970) available in the National Library of Medicine (NIH) from the Center of Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/. Global alignment was used, as this method compares the whole content of a query sequence against the reference.

2.14.6 Plasmid detection

Three criteria were used to determine the presence of plasmids. The first was based on detecting the homology of our contigs against known plasmid sequences already available in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). If contigs suspected to be plasmids from *M. komossense* JERR01 contained a query coverage higher than 75% with

reported and confirmed plasmids from the NCBI plasmids database (https://blast.ncbi.nlm.nih.gov), these contigs were considered potential plasmids. Secondly, contigs were analysed for circularity using Unicycler. During assembly, Unicycler evaluates contig ends for significant overlap, which would indicate that the contig was circular. And the third criterion used to identify was through the use of Ori-Finder 2022 (https://tubic.org/Ori-Finder2022/public/index.php/index), to predict origin of replication site *(oriC)* regions in *M. komossense* JERR01 with the default settings (Dong, Luo and Gao, 2022). Both the complete chromosome and the four contigs (two linear and two closed), were analysed using their *gbk files. Since Ori-Finder 2022 can predict more than one *ori*, the origin of replication with the highest score (based on the Ori-Finder 2022 databases calculation) was chosen as the origin of replication.

2.14.7 Pan-genome analysis

A Pan-genome analysis was constructed using Panaroo v1.4.0 (Tonkin-Hill *et al.*, 2020) (https://github.com/gtonkinhill/panaroo). Panaroo was used to identify the core genes shared by 99–100% of the genomes studied and was run using the following script:

```
panaroo -i gff/*.gff -o results_panaroo_l50 --clean-mode strict
--remove-invalid-genes -t 12 -a core --aligner clustal --
core_threshold 0.80 --threshold 0.5 --len_dif_percent 0.5
```

This script identified proteins based on a minimum of 50% that were present in at least 80% of the selected 20 genomes (Table 2-2). The maximum likelihood phylogeny tree was created by IQ-TREE using 1000 ultrafast bootstraps, using the following script:

```
iqtree -s core_gene_aligment_filtered.aln -m GTR+F+I -T 2 -mem
2G -B 1000 -0 M66
```

The data for this tree was obtained from the core genome alignment output file generated by Panaroo and visualized.

This tree was mid-rooted and edited using the interactive tree of life (iTOL) server (Letunic and Bork, 2021) (https://itol.embl.de/#). A presence/absence matrix of genes was created using roary_plots.py using the gene_presence_absence.csv file output produced by Panaroo containing all 20,384 genes detected in this study. The script to produce this matrix is not included by default with the download for Roary, but as an additional tool that can be found in the repository files of Roary (Fleres *et al.*, 2018) (https://sanger-pathogens.github.io/Roary/). The following script was used to create the matrix:

```
roary_plots.py input_newick_tree_file.tre
gene_presence_absence.csv
```

Table 2-2. Overview of selected mycobacterial species and strains for comparative bioinformatics analysis.

Clade	Species name	Accession number
Tuberculosis-Simiae	*M. marinum* E11	NZ_HG917972.2
	M. avium subsp. *hominissuis* OCU889s_P11-4s	NZ_CP018019.1
	M. canettii CIPT 140010059	NC_015848.1
	M. leprae MRHRU-235-G	NZ_CP029543.1
	M. kansasii ATCC 12478	NC_022663.1
	M. tuberculosis H37Rv	AL123456
Fortuitum-Vaccae	*M. smegmatis* MC2 155	NZ_CP009494
	M. aurum NCTC10437	NZ_LR134356
	M. fortuitum CT6	CP011269.1
	M. vaccae 95051	NZ_CP011491
	M. gilvum Spyr1	NC_014814
	M. vanbaalenii PYR-1	NC_008726
	M. komossense ATCC 33013	JACKTY010000001
	Mycobacterium komossense strain JERR01	
Terrae	*M. algericus* strain JCM 30723	NZ_BLKY01000001.1
	M. hiberniae JCM 13571	NZ_AP022609
	M. kumamotenensis strain DSM 45093	NZ_MVHU01000100.1
	M. terrae NCTC 10856	NZ_LT906469.1
	M. arupensis strain DSM 44942	NZ_MVHH01000001.1
Abscessus-Chelonae	*M. salmoniphilum* strain DSM 43276	NZ_CP024633.1
	M. abscessus strain GZ002	NZ_CP034181.1

2.14.8 Rapid haploid variant calling

Variant calling on data derived from nanopore sequencing was performed using Medaka v1.11.3 (Sanderson *et al.*, 2020) (https://github.com/nanoporetech/medaka) using the following script:

```
medaka_haploid_variant -i <reads.fastq> -r <ref.fasta>
```

Variant calling analysis was carried out in short read generated from Illumina sequencing was performed using Snippy v4.6.0 (Bush *et al.*, 2020) (https://github.com/tseemann/snippy) using the following script:

```
snippy --cpus 16 –outdir --ref *gbk --R1 *fastq.gz --R2
*fastq.gz
```

Both tools perform this task by comparing a sample sequence against a draft assembly and finding substitutions (SNPs) and insertions/deletions (indels). SNPs and indels in genes known to confer antimicrobial resistance were identified by manual inspection of the BAM files on the Integrative Genome Browser (IGV) v2.16.0 (Robinson *et al.*, 2011).

3 Phenotypic characteristics of *M. komossense* JERR01 and comparison with reference strain *Mycobacterium komossense* ATCC 33013

3.1 Introduction

Mtb requires a biosafety level 3 (BSL-3) laboratory for research. It is a fastidious slow-growing organism which requires special growth supplements and has a doubling time of 18-24 hours (Flynn, Chan and Lin, 2011). Appearance of visible *Mtb* colonies on agar plates can take 3-5 weeks, so experimental work takes a long time. In such a setting there is always a higher risk of bacterial or fungal contamination of *in vitro* cultures. AST in clinical settings can also take weeks to produce results (Rancoita et al., 2018; Von Groll et al., 2010). The laborious nature of working with *Mtb,* including the transmissibility and pathogenicity characteristics which necessitate high-grade containment facilities, has resulted in the use of environmental non-pathogenic members of the genus *Mycobacterium* as model systems to create safer research conditions (Bottaro, Larsen, 2010).

However, current mycobacteriology research only focusses on a limited number of mycobacterial species for this purpose (e.g., *M. smegmatis*, *M. bovis*, *M. marinum*) (Smeulders *et al.*, 2017). The current mycobacteria model organisms used for *Mtb* drug resistance studies each have their strengths and weaknesses. For example, *M. smegmatis* has a different antibiotic profile to INH and RIF, *M. marinum* is temperature dependant (30^{o}C) (Petrini, 2006), and *M. bovis* takes as much time as *Mtb* to grow (16-20h generation time)(Sawyer *et al.*, 2023). Whilst current models have provided some insight into different aspects of the *Mtb* infection such as pathogenicity (*M. marinum*

produces granuloma like structures when infecting) (Tobin and Ramakrishnan, 2008) and drug discovery (*M. smegmatis* was used for the initial screening of the compound that later became the drug BDQ)(Sparks *et al.*, 2023), and these model are fast growing mycobacteria allowing rapid acquisition of data, these models do not share the same resistance profile as *Mtb*. For these reasons, the introduction of new mycobacterial models can help address the limitations of the currently existing ones.

Here, *M. komossense* JERR01, a strain of *M. komossense* is presented as a potential model organism for *Mtb* resistance studies. This new strain was first isolated in 2010 by the Scottish Mycobacterial Reference Laboratory (SMRL). The strain was isolated from a human urine sample, and beyond an initial molecular characterisation based on the sequence of the *hsp65* gene, little information is known about the properties of the strain and its relationship to other members of the genus.

Describing the properties of a new organism involves description of its growth kinetics. Conventionally, this involves serial OD_{600} measurements and CFU counts as described in Chapter 2. Whilst these techniques are robust, the slow growing nature of mycobacteria means that it takes 7 days to generate results even for fast growing species in this genus.

Currently, at the University of St Andrews, a novel system named SLIC (The Scattered Light Integrated Collector) has been developed to study bacterial growth, including the effect of antibiotics on this. This system has provided promising results for fast growing non-mycobacterial organisms such as *Escherichia coli* and *Staphylococcus aureus* and has been able to provide rapid AST results for these organisms in hours (Hammond *et al.*,

2017; Falconer *et al.*, 2024). Whilst working *with M. komossense* JERR01, this thesis took the opportunity to assess the performance of SLIC on a *Mycobacterium* for the first time.

3.2 Chapter aims and objectives.

The aim of this chapter was to describe the phenotypic characteristics of *M. komossense* JERR01 and compare this organism with other relevant mycobacteria to help determine its suitability as a new model organism for *Mtb*, particularly in relation to antibiotic resistance studies. Specific objectives were to:

1. Describe selected phenotypic characteristics of *M. komossense* JERR01, including macroscopic and microscopic morphology, growth rate, generation time, and susceptibility to anti-tubercular antibiotics.
2. Where feasible and appropriate, compare these characteristics of *M. komossense* JERR01 with (a) the reference *M. komossense strain* ATCC 33013, (b) *M. smegmatis* MC2 155, an alternative *Mycobacterium* which is used as an experimental model of *Mtb,* and (c) relevant information on the AST profile of *Mtb*.
3. Evaluate whether the novel SLIC tool could be used to generate information on growth kinetics and antibiotic response in mycobacteria more quickly than standard methodologies.

3.3 Results

3.3.1 Phenotypic traits of *M. komossense* JERR01 and comparison with reference strain *M. komossense* strain ATCC 33013.

The initial steps taken to describe *M. komossense* JERR01 were to examine its phenotypic characteristics and compare them to the reference strain of *M. komossense* (ATCC 33013). Single colonies of both organisms were grown in 7H9 Middlebrook broth with an initial OD_{600} of 0.05. Ten microlitres were then plated on 7H10 Middlebrook plates enriched with glycerol, using the Miles and Misra method, and incubated at 30°C and 37°C. Colonies were observed and photographed every day for 7 days to observe *M. komossense* JERR01 and *M. komossense* ATCC 33013 growth over time at both temperatures (Figure 3-1 and Figure 3-2).

In Figure 3-1, *M. komossense* JERR01 and *M. komossense* ATCC 33013 were incubated at 30°C. Colonies were lightly visible from day 2, and fully visible from day 3. *M. komossense* JERR01 produced a bright yellow pigment that became more prominent over time. *M. komossense* ATTCC 33013 had visible colonies from day 3 and reached full coverage of the agar plate on day 7. Unlike *M. komossense* JERR01, it did not produce any pigment. In Figure 3-2*, M. komossense* JERR01 was grown at 37°C. Like Figure 3-1, colonies were lightly visible from day 2 and fully visible from day 3, reaching full coverage of the plate by day 7. *M. komossense* ATCCC 33013 did not grow at 37°C as shown in Figure 3-2.

Both *M. komossense* JERR01 and *M. komossense* ATCC 33013 were also inoculated onto Middlebrook 7H10 agar plates at 42°C, but no growth was seen after 7 days of

incubation. These plates were then further incubated at 30°C for an additional 7 days, but no cells were recovered.

The *M. komossense* JERR01 colonies presented an irregular shape with rough edges, a dry appearance, and a light yellowish colour (Figure 3-3 B). The morphology of *M. komossense* ATCC 33013 colonies was different. The reference strain colonies were more rounded with less prominent edges (Figure 3-3 C). The morphology of both strains was determined to be smooth as they lacked the bundle-forming fimbriae, which are usually observed in species such as *M. abscessus* (Figure 3-3 A) which is known to produce both smooth and rough colonies (Clary *et al.*, 2018).

Another difference between *M. komossense* JERR01 and the reference strain was pigment production. *M. komossense* JERR01 started producing a bright yellow pigment regardless of the presence of light after 3 days of incubation whilst *M. komossense* ATCC 33013 did not (Figure 3-1 and Figure 3-2). This pigment produced by *M. komossense* JERR01 was brighter at 37°C than at 30°C when the organism was grown in liquid culture, as shown in Figure 3-4. No pigment was seen when *M. komossense* JERR01 was incubated at 42°C, consistent with no growth of the organism at this temperature.

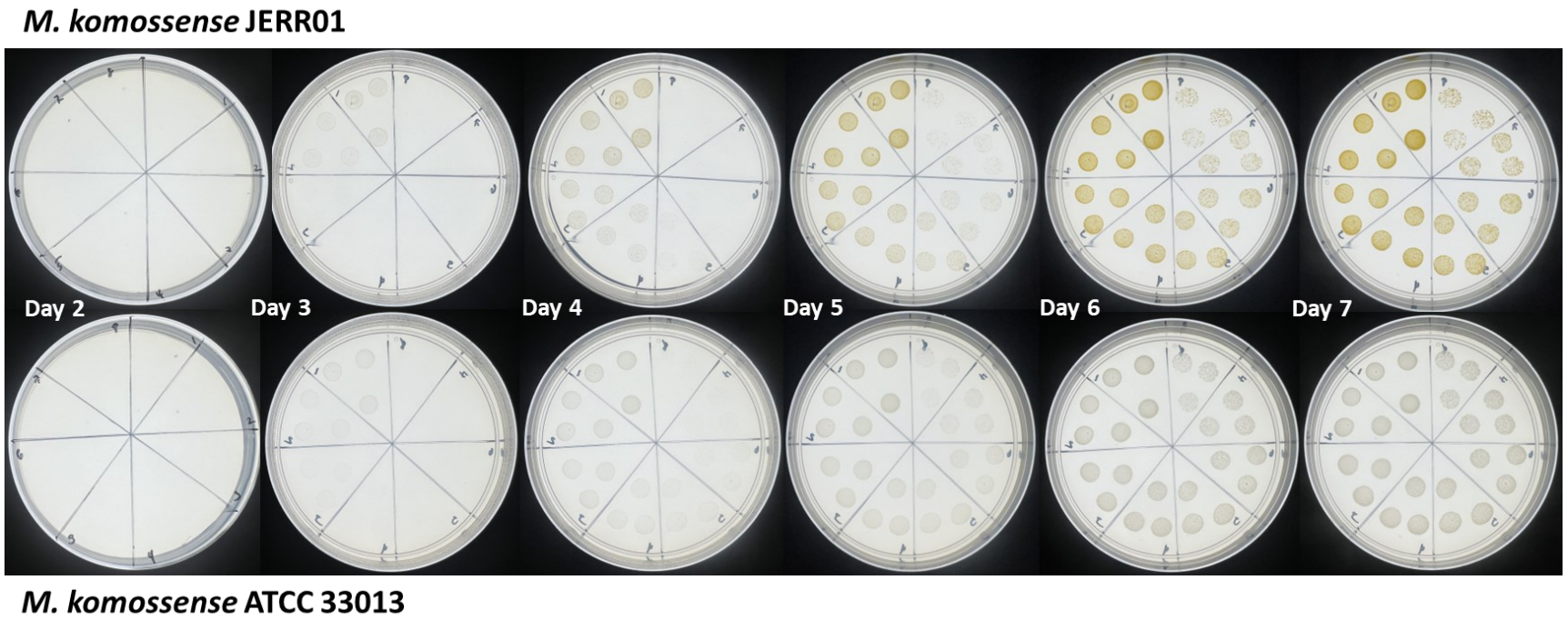


Figure 3-1. Non-selective 7H10 Middlebrook agar plates of *M. komossense* JERR01 and *M. komossense* ATCC 33013 incubated at 30°C for 7 days.

Both strains were diluted to an initial OD_{600} of 0.05 and plated using the Miles and Misra method. Photographs of bacterial colonies on agar were taken. On day 1, images are not presented as no colonies were observed. Photographs from the subsequent 6 days illustrate the progressive growth of bacterial colonies.

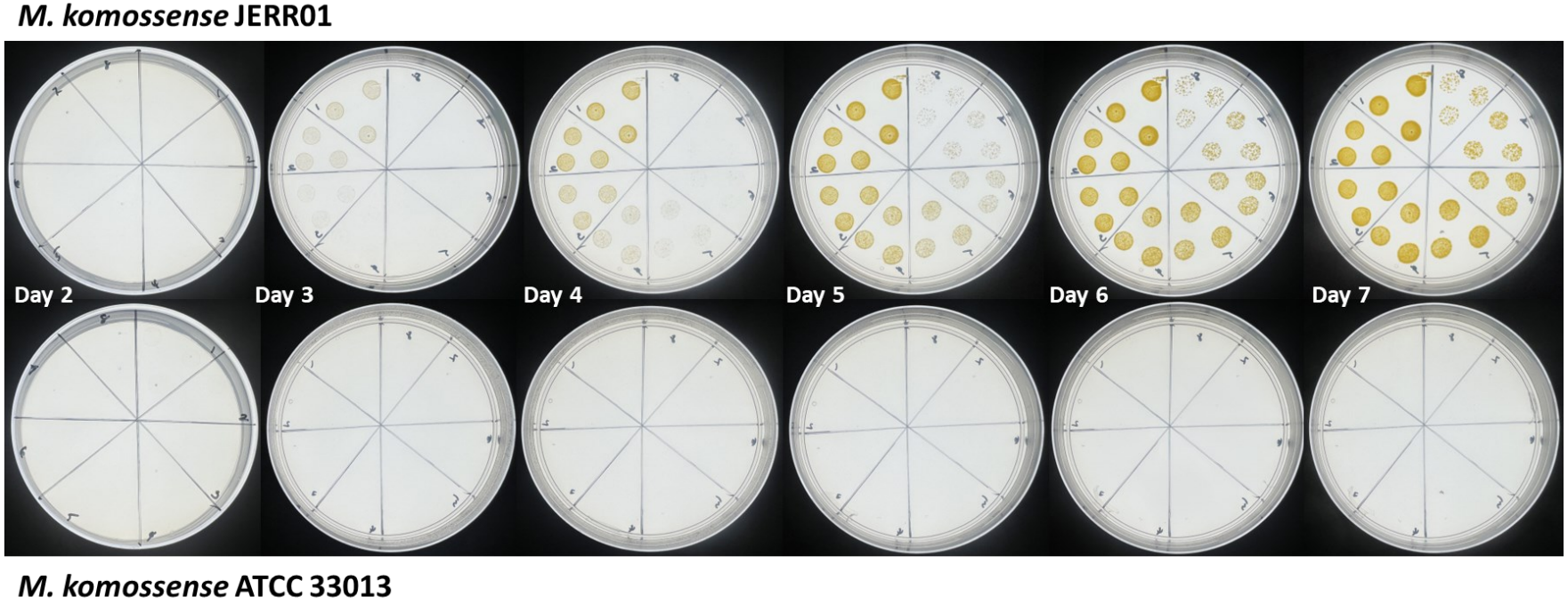


Figure 3-2. Non-selective 7H10 Middlebrook agar plates of *M. komossense* JERR01 and *M. komossense* ATCC 33013 incubated at 37°C for 7 days.

Both strains were diluted to an initial OD_{600} of 0.05 and plated using the Miles and Misra method. Photographs of bacterial colonies on agar were taken. On day 1, images are not presented as no colonies were observed. Photographs from the subsequent 6 days illustrate the progressive growth of bacterial colonies.

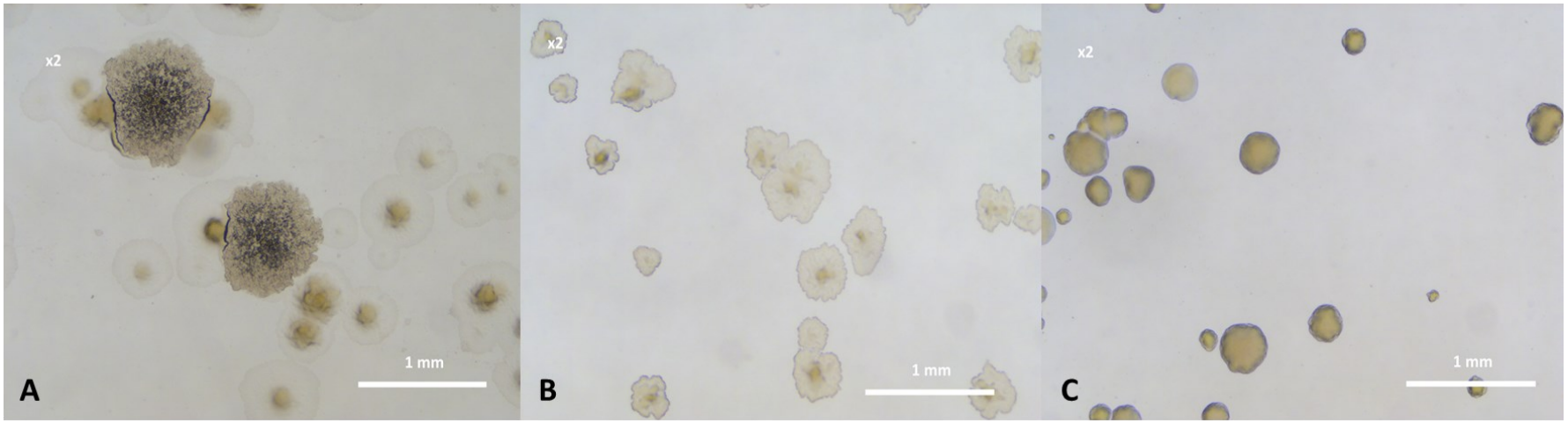


Figure 3-3. The colony appearance of *M. komossense* JERR01 and *M. komossense* ATCC 33013 with *M. abscessus* subsp. *massiliense* as a reference.

(A) *M. abscessus* subsp. *massiliense* presenting both types of morphologies: rough and smooth as a visual reference. (B) *M. komossense* JERR01 presenting a smooth morphology with prominent and rough edges and (C) *M. komossense* ATCC 33013 showing a smooth morphology with round appearance. Pictures A-C were taken at a magnification of 2x using the Olympus Bx51 Transmitted Light Microscope.

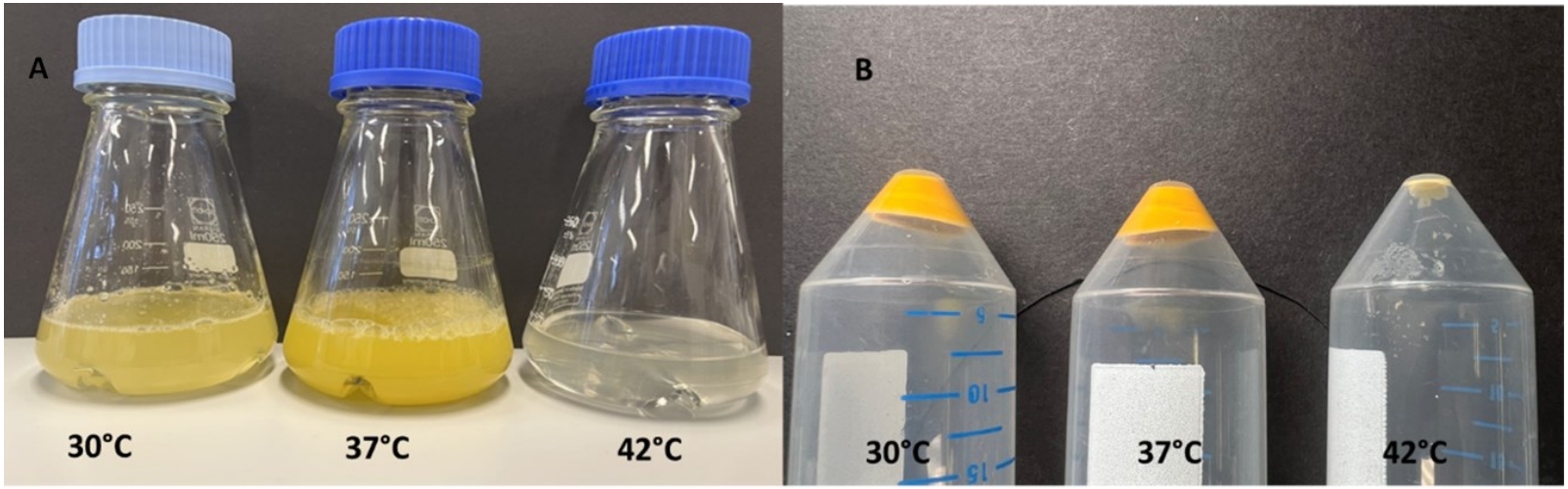


Figure 3-4. Pigments produced by *M. komossense* JERR01 at different temperatures.
Picture taken in regular daylight of *M. komossense* JERR01 cultured in Middlebrook 7H9 media broth at, (A) from left to right, 30°C, 37°C and 42°C, in a regular atmosphere and displaying pigmentation. In (B), cell pellets from cultures of *M. komossense* JERR01 grown at different temperatures (30°C, 37°C and 42°C) are shown after centrifugation.

3.3.2 Microscopy of *M. komossense* JERR01 and *M. komossense* ATCC 33013

Microscopic examination of *M. komossense* JERR01 involved the staining with the Ziehl-Neelsen acid fast procedure (see chapter 2.5 for details). As a procedural control, *M. smegmatis* MC2 155 was stained and compared to *M. komossense* JERR01 and *M. komossense* ATCC 33013. *M. smegmatis* cells were red and rod-shaped cells when viewed under the microscope as expected of a *Mycobacterium*, Figure 3-5 A. Similarly, *M. komossense* JERR01 had a higher retention of the primary stain (Carbol fuchsin), leading to more bacterial cells being observed under the microscope and a clear observation of rod-shaped cells as shown in Figure 3-5 B. In contrast, *M. komossense* ATCC 33013 poorly retained the primary stain, which was washed off by the acid alcohol, leading to fewer red bacterial cells being observed and more cells being stained by the counterstain (Methylene blue) under the microscope, as showed in Figure 3-5 C.

M. komossense ATCC 33013 presented a white/grey appearance resembling a capsule. Some bacteria produce a protective layer surrounding the cell as a measure against desiccation and harmful compounds, and this protective layer can extend a distance many times that of the cell wall. This protective layer is composed of mostly polysaccharides and is termed a capsule (Bian *et al.*, 2021; Santos-López, Rodríguez-Beltrán and Millán, 2021). To determine whether this layer was present in *M. komossense* JERR01 or *M. komossense* ATCC 33013 these organisms were stained alongside a capsule producing bacteria, *Bacillus subtilis*. No bacterial capsule was observed in *M. komossense* JERR01 nor *M. komossense* ATCC 33013.

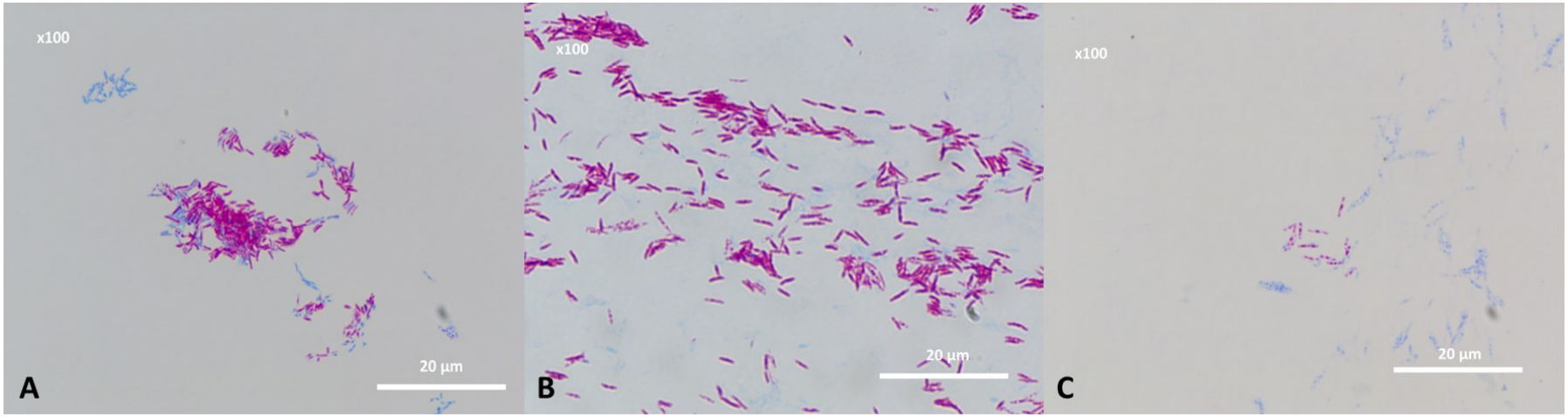


Figure 3-5. Acid-fast staining examination for mycobacteria.
For reference, we used (A) *M. smegmatis* MC2 155, (B) *M. komossense* JERR01 and (C) *M. komossense* ATCC 33013. Microscopic examination of acid fastness in mycobacteria was performed on a bright field microscope using an oil immersion lens at a magnification of 100x using the Olympus Bx51 Transmitted Light Microscope.

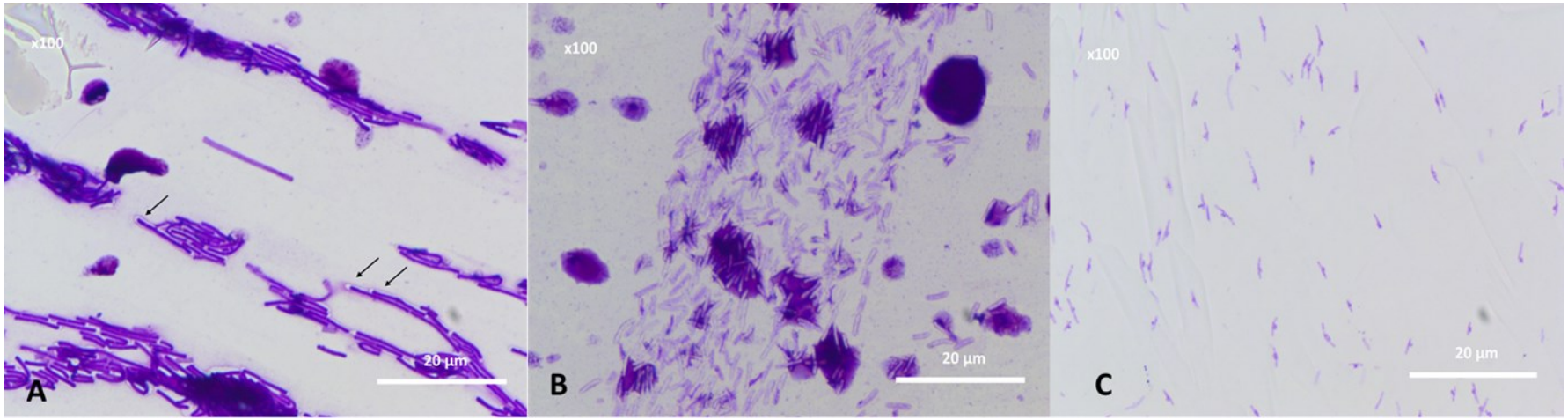


Figure 3-6. Bacterial capsule staining to detect the presence of polysaccharide outer layer.
Picture of *Bacillus subtilis* (A) as a procedural control of how bacterial capsule looks followed by (B) *M. komossense* JERR01 where no capsule was observed and (C) *M. komossense* ATCC 33013. Microscopic examination of bacterial capsule was performed using an oil immersion lens at a magnification of 100x using the Olympus Bx51 Transmitted Light Microscope.

3.3.3 Growth curve of *M. komossense* JERR01 and *M. komossense* ATCC 33013

After describing the macroscopic and microscopic appearance of *M. komossense* JERR01 and the reference strain *M. komossense* ATCC 33013, phenotypic growth characteristics were assessed. Growth kinetics at a range of environmentally and physiologically relevant temperatures were investigated in liquid and solid media. Determination of the rate of growth is important because fast-growing mycobacteria have some advantages as model organisms for *Mtb* studies; they allow for quicker experimentation and faster data acquisition, with less time for culture contamination.

In liquid media, *M. komossense* JERR01 grows well at 30°C (blue line) and 37°C (red line) when monitored by changes in OD measurement until 78 hours as shown in Figure 3-7 A. This mainly represents the period of lag and exponential phase growth. However, in the stationary phase (after 80 hours), the *M. komossense* JERR01 strain incubated at 37°C continues to reach progressively higher OD_{600} compared to the same organism incubated at 30°C. The flask containing *M. komossense* JERR01 incubated at 42°C shows no bacterial growth at all.

When grown on solid agar to measure growth by calculating changes in the number of viable CFU/ml over time, *M. komossense* JERR01 strains incubated at 30°C and 37°C showed no difference in the number of CFU/ml during the first 72 hours (Figure 3-7 B). However, in the stationary phase (after 78 hours) CFU/ml counts of the *M. komossense* JERR01 strain incubated at 37°C slightly declined.

At 78-80 hours, Figure 3-7A shows that OD_{600} at 37°C measured more turbidity in the culture compared to 30°C but (Figure 3-7 A and B) shows a higher concentration of viable

CFU/ml during stationary phase at 30°C compared to 37°C. The reason for this is not known, but a possible explanation could be that while OD_{600} reflects total cell mass, dead or degraded cells can still cause turbidity in the broth as this method does not completely differentiate between viable and non-viable cells. At 37°C, *M. komossense* JERR01 cultures may contain more cellular material but fewer viable organisms. This could explain the higher OD_{600} but lower CFU count at 37°C.

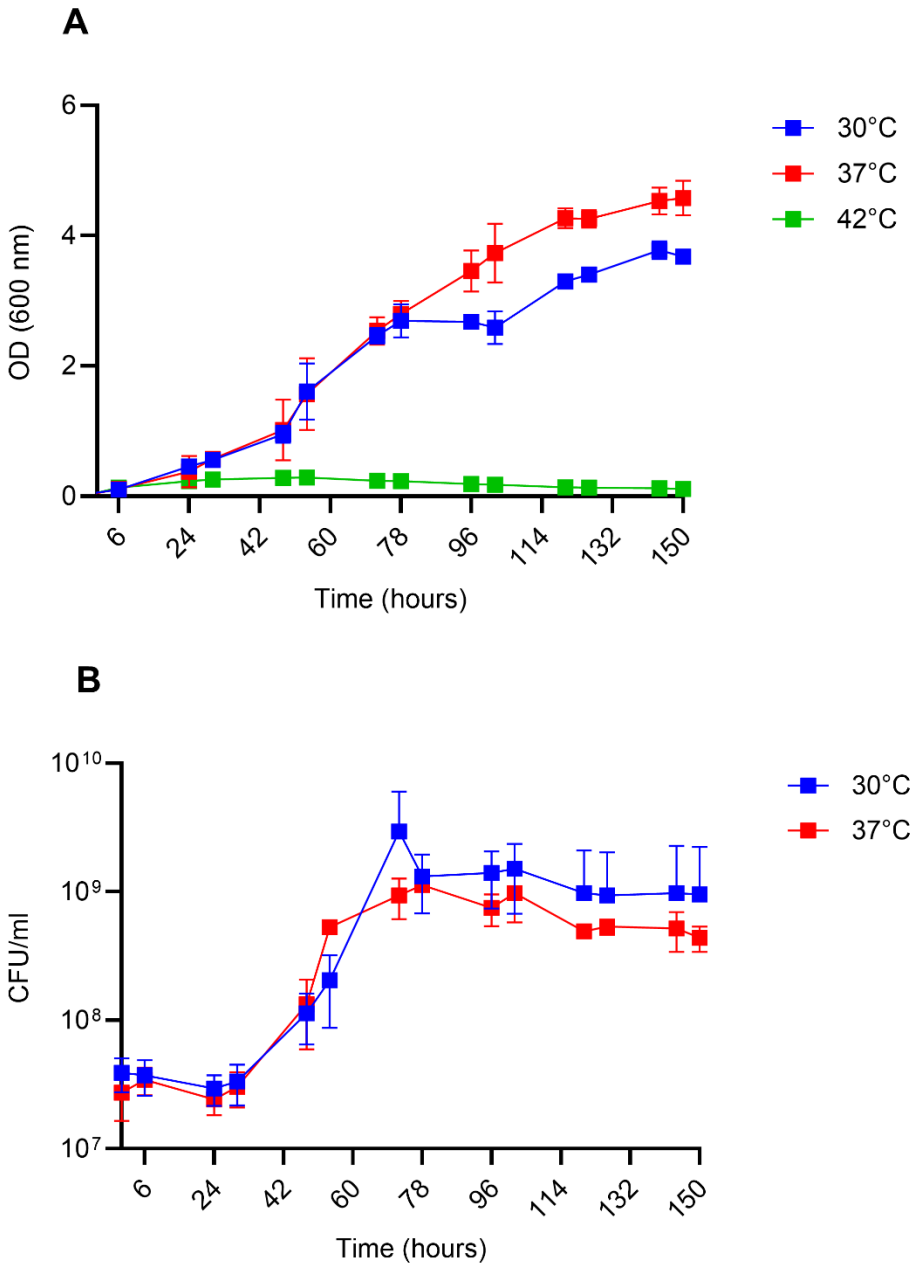


Figure 3-7. Growth curves of *M. komossense* JERR01 over the course of 7 days (150h) at two different temperatures (30°C and 37°C) using two different measuring techniques.

The different measurements used to display the growth of *M. komossense* JERR01 were: (A) Growth curve using OD_{600} measurements (spectrophotometer) and (B) CFU/ml measurements (Miles & Misra method). Each point corresponds to the average of three biological repeats and error bars represent the standard deviation from the mean.

An attempt to perform the growth kinetics experiments conducted in *M. komossense* JERR01 was undertaken with *M. komossense* ATCC 33013, so that growth rates could be compared. However, it was not possible to determine a growth curve for *M. komossense* ATCC 33013 in liquid media as no bacterial growth was observed for this strain after 100 hours of incubation at 30°C (Figure 3-8).

M. komossense ATCC 33013 could grow in solid agar media as shown in Figure 3-1. Different experimental approaches were taken to try to grow *M. komossense* ATCC 33013 in liquid broth. Among these was the addition of 2 ml of glycerol per 450 ml of 7H9 Middlebrook broth as a source of energy and the removal of Tween-80 from the media. It is known that Tween-80, as a surfactant, can inhibit the growth of certain mycobacteria, such as *Mtb,* by altering the composition of the cell wall and disrupting the cell membrane integrity by solubilizing the lipids present in the cell membrane (Pietersen *et al.*, 2020). It was possible that Tween-80 could have led to the cell membrane damage of *M. komossense* ATCC 33013, so it was removed from the media. *M. komossense* ATCC 33013 was also grown at different temperatures and at different shaking conditions (40, 80, and 120 rpm). Shaking bacterial culture during incubation helps with aeration, providing more oxygen to the bacteria and promoting bacterial growth. However, after all these changes, no growth for *M. komossense* ATCC 33013 was observed in 7H9 Middlebrook broth as presented in Figure 3-8. Viewed together, this suggests that this organism is preferentially culturable in the solid agar format.

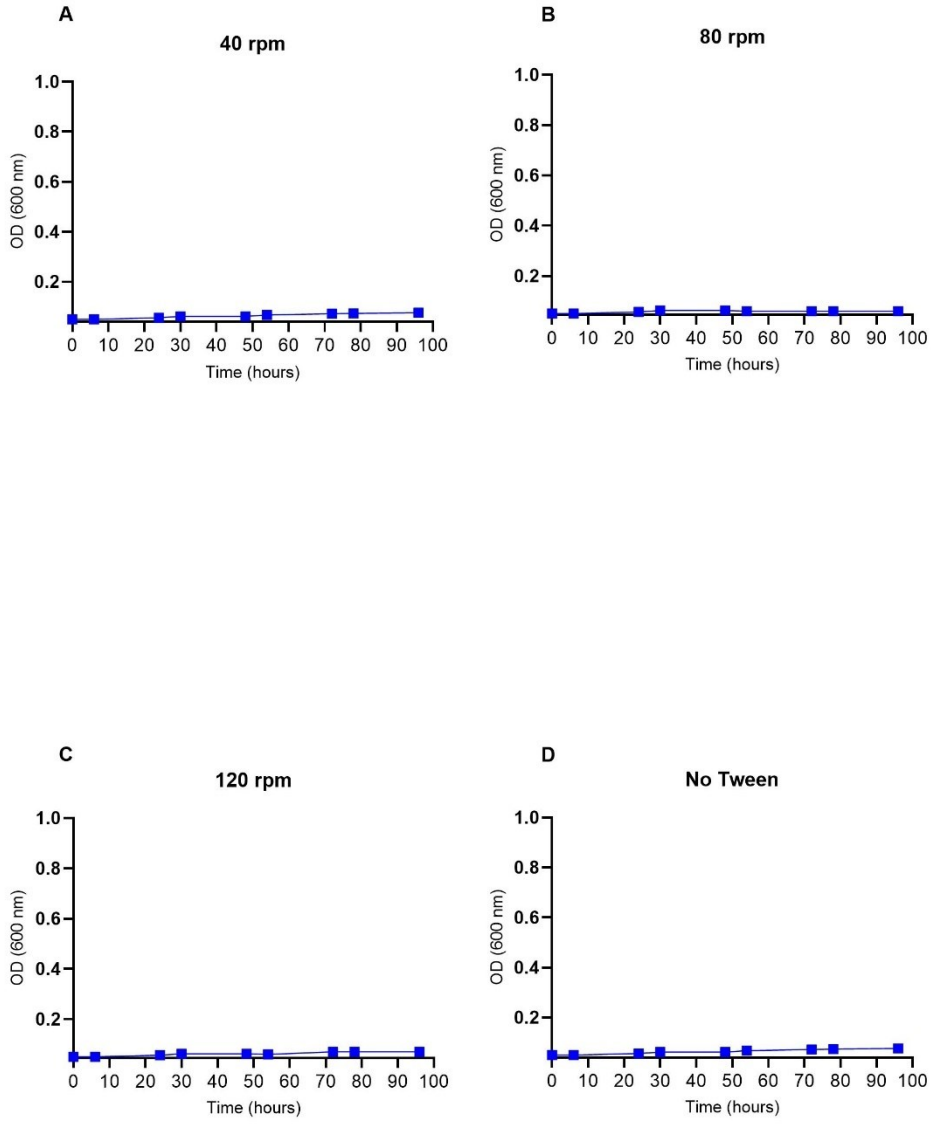


Figure 3-8. Growth curves of *Mycobacterium komossense* ATCC 33013 for 4 days (96h). *M. komossense* ATCC 33013 was grown in 100 ml of 7H9 Middlebrook media supplemented with glycerol in a 250 ml baffled flask at 40, 80, and 120 rpm at 30°C with and without the presence of Tween-80. OD_{600} measurements of *M. komossense* ATCC 33013 were taken twice daily for 96 hours; however, no turbidity was observed in flasks containing bacterial cultures throughout the course of the experiment.

3.3.4 Generation time for *M. komossense* JERR01

After investigating the growth kinetics of *M. komossense* JERR01, experiments were conducted to calculate the bacterial generation time and provide a numerical value for the rate at which *M. komossense* JERR01 divides when incubated at 30°C and 37°C. Building upon the growth kinetics data, the determination of a generation time for *M. komossense* JERR01 offers a direct comparison with other model organisms, including *Mtb* itself (18-20 hours doubling time) as one of the advantages of having a rapid grower as a model organism is the reduction of time for experimentation and the production of faster results.

Generation times for *M. komossense* JERR01 at both 30^{o}C and 37^{o}C were calculated using data from Figure 3-7 B, which reported CFU/ml measurements of cultures of *M. komossense* JERR01. Solid agar growth data were used for these calculations in preference to liquid broth because only viable CFUs from viable bacteria are observed and counted using this method. To calculate the generation time for *M. komossense* JERR01 at both temperatures, two CFU/ml readings at an exponential rate were chosen. Equations 2.2 and 2.3 (see Section 2.9) were used as shown below:

$$G = \frac{t_1 - t_2}{\frac{3.3\ log\ Nt}{N0}}$$

$$G = \frac{72 - 48}{\frac{3.3\ log\ (2.95 \times 10^9)}{(1.13 \times 10^8)}}$$

$$G = \frac{24}{4.68}$$

$$G = 5.13\ hours\ (SD \pm 0.326\ hours)$$

The generation time of *M. komossense* JERR01 was estimated to be 5.13 hours at 30°C in liquid. The same mathematical equation as used for two data points at 37°C.

$$G = \frac{t_1 - t_2}{\frac{3.3\ log\ Nt}{N0}}$$

$$G = \frac{72 - 48}{\frac{3.3\ log\ (5.30 \times 10^8)}{(3.02 \times 10^7)}}$$

$$G = \frac{24}{4.10}$$

$$G = 5.84\ hours\ (SD \pm 0.862\ hours)$$

The generation time of *M. komossense* JERR01 was estimated to be 5.84 hours at 37°C in liquid. These results allow the direct comparison with the generation time of other mycobacteria: *Mtb* in liquid (18-20h), *M. smegmatis* in liquid (3-4h), *M. marinum* in liquid (10-12h), and *M. bovis* in liquid (16-20h) (Smeulders *et al.*, 2017).

3.3.5 Determining the Minimum Inhibitory Concentration (MIC_{90}) of *M. komossense* JERR01 for 9 antitubercular drugs.

Having described the physical appearance and growth characteristics of *M. komossense* JERR01, evaluating whether this *Mycobacterium* could be used as a model of *Mtb* in antibiotic resistance studies turned to the assessment of anti-mycobacterial drug susceptibility. From this perspective, Minimum Inhibitory Concentration (MIC_{90}) determination, defined as the minimum concentration of an antibiotic that inhibits the growth of 90% of bacteria, is an important phenotypic measurement.

A standard method for determining MIC_{90} was selected, using the broth microdilution method in 96-broth microdilution well plates. The protocol used for MIC_{90} determination followed the EUCAST standardized guidelines used for antibiotic susceptibility testing (AST) for *Mtb* and NTMs (Cambau *et al.*, 2019; Schön *et al.*, 2020). Comparing the MICs of *M. komossense* JERR01 with the clinical breakpoints of *Mtb* for the same antimicrobials could help assess how well *M. komossense* JERR01 reflects the drug susceptibility of the target pathogen (*Mtb*). For the following drugs: PZA and LZD, clinical breakpoints are still under investigation therefore, it was decided to use the critical concentration of these antibiotics as a reference; the critical concentration (CC) is the lowest concentration of a drug that will inhibit 99% of wild-type strains of *Mtb* (European Centre for Disease Control, 2018). If the MIC_{90} were similar, *M. komossense* JERR01 may be advocated as a reliable model for studying *Mtb* drug resistance.

The drugs selected for this series of experiments were as outlined in section 2.10, Table 2.1. These drugs were chosen for the following reasons:

- RIF, INH, PZA, EMB: standard first-line therapy for DS-TB (Nahid *et al.*, 2016; World Health Organization, 2020)
- MOXI, LEVO: two 8-methoxyfluoroquinolones, conventionally used for DR-TB patients with recent evidence suggesting they may also be used to shorten DS-TB treatment (Pranger *et al.*, 2019).
- BDQ, LZD, and PA: new or repurposed drugs recently classified as being important components of effective therapy for DR TB (Krishnan Hyderabad, 2019; Oelofse *et al.*, 2021).

Throughout these experiments, *M. smegmatis* MC2 155 was used as procedural control. The clinical breakpoints and critical concentrations for *Mtb* used as a reference were obtained from the Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis by the World Health Organization (WHO), and European Committee of Antimicrobial Susceptibility Testing (EUCAST) Clinical Breakpoint Tables v. 14.0, valid from 2024-01-01. The clinical breakpoint is defined as the concentration or concentrations of an antibiotics agent that separates strains that will likely respond to treatment from those strains which likely will not respond to treatment (Niward *et al.*, 2016; European Centre for Disease Control, 2018; Woods *et al.*, 2018; World Health Organization 2018, 2018; EUCAST, 2024).

The MIC_{90} of *M. komossense* JERR01 for 6 antitubercular drugs (RIF, EMB, MOXI, LEVO, BDQ, and LZD) were lower than the clinical breakpoints for *Mtb* against these drugs

(Table 3-1). The MIC_{90} of *M. komossense* JERR01 for INH (32 µg/ml) was higher than the clinical breakpoint of *Mtb* for INH (0.1 µg/ml). The MIC_{90} of PZA for *M. komossense* JERR01 was found to be above 512 µg/ml, exceeding the critical concentration established for PZA in *Mtb* for these drugs, 100 µg/ml. The MIC_{90} for PA was >512 µg/ml, exceeding the clinical breakpoint recommended for this antibiotic by EUCAST (2 µg/ml), suggesting an intrinsic resistance from *M. komossense* JERR01 towards these antibiotics.

Table 3-1. Table providing the MIC_{90} of *M. komossense* JERR01 and *M. smegmatis* MC2 155 determined using a gold standard method, and the clinical breakpoints of *Mtb* against 9 antitubercular drugs.

Antibiotic	*M. smegmatis* MC2 155	*M. komossense* JERR01		Clinical breakpoints[1]/Critical concentrations[1]
	Microdilution broth MIC_{90} (µg/ml)	Microdilution broth MIC_{90} (µg/ml)	Interpretation	*Mtb* (µg/ml)
INH	64	32	R	0.12[2]
RIF	4	0.002	S	1.0[2]
EMB	0.25	0.5	S	2.0[2]
PZA	>512	>512	R	100[3]
MOXI	0.0625	0.0625	S	1.0[4]
LEVO	0.125	0.125	S	1.0[4]
BDQ	0.0625	0.125	S	0.25[5]
LZD	1.0	0.5	S	1.0[3]
PA	>512	>512	R	2.0[5]

MIC_{90} of *M. komossense* JERR01 at 30°C and *M. smegmatis* MC2 155 at 37°C against 9 antitubercular drugs, followed by the interpretation of the MIC_{90} of *M. komossense* JERR01 when compared to the clinical breakpoints for *Mtb* for the treatment of TB.

[1] The clinical breakpoints and critical concentrations used as references were determined using the Mycobacterial Growth Indicator Tube (MGIT).
[2] Clinical breakpoints extracted according to CLSI for the treatment of DR-TB.
[3] CC Extracted from Handbook on tuberculosis laboratory diagnostic methods.
[4] Clinical Breakpoints recommended from WHO for drug susceptibility testing of medicines used in the treatment of tuberculosis.
[5] Clinical breakpoints according to EUCAST.

3.3.6 Bactericidal activity of RIF and INH in *M. komossense* JERR01

With the MIC_{90} of *M. komossense* JERR01 determined through an established standard method, the focus has now shifted to the Minimum Bactericidal Concentration (MBC). As the antibiotics INH and RIF have potent bactericidal activity on *Mtb* with MBC of 25 µg/ml for INH and <1.0 for RIF (Nair *et al.*, 2015), experiments were undertaken to determine whether the same bactericidal activity was observed in *M. komossense* JERR01.

After determining the MIC_{90} for *M. komossense* JERR01, at 30°C, 10 µl of bacterial suspension were taken from each 96-microdilution well plate from concentrations 0.06-0.005 for RIF and 512-4 for INH from the MIC_{90} plates and inoculated into nonselective 7H10 Middlebrook agar before measuring viable CFU/ml growth using the Miles and Misra method. After 5 days of incubation, it was possible to recover *M. komossense* JERR01 cells exposed to high doses of INH at the following concentrations: 512 µg/ml ($2.58x10^4$ CFU/ml $\pm 1.34x10^4$), 256 µg/ml ($6.79x10^4$ CFU/ml $\pm 7.09x10^3$), and 128 µg/ml ($2.68x10^4$ CFU/ml $\pm 1.81x10^4$). Therefore, it was not possible to determine an MBC for *M. komossense* JERR01 when exposed to INH as a small number of CFUs were still growing at a concentration as high as 512 µg/ml, as presented in Figure 3-9. However, when the cells exposed to RIF were plated, the MBC was observed to be 0.03 µg/ml, the lowest antibiotic concentration at which no *M. komossense* JERR01 cells were observed to grow, as shown in Figure 3-10.

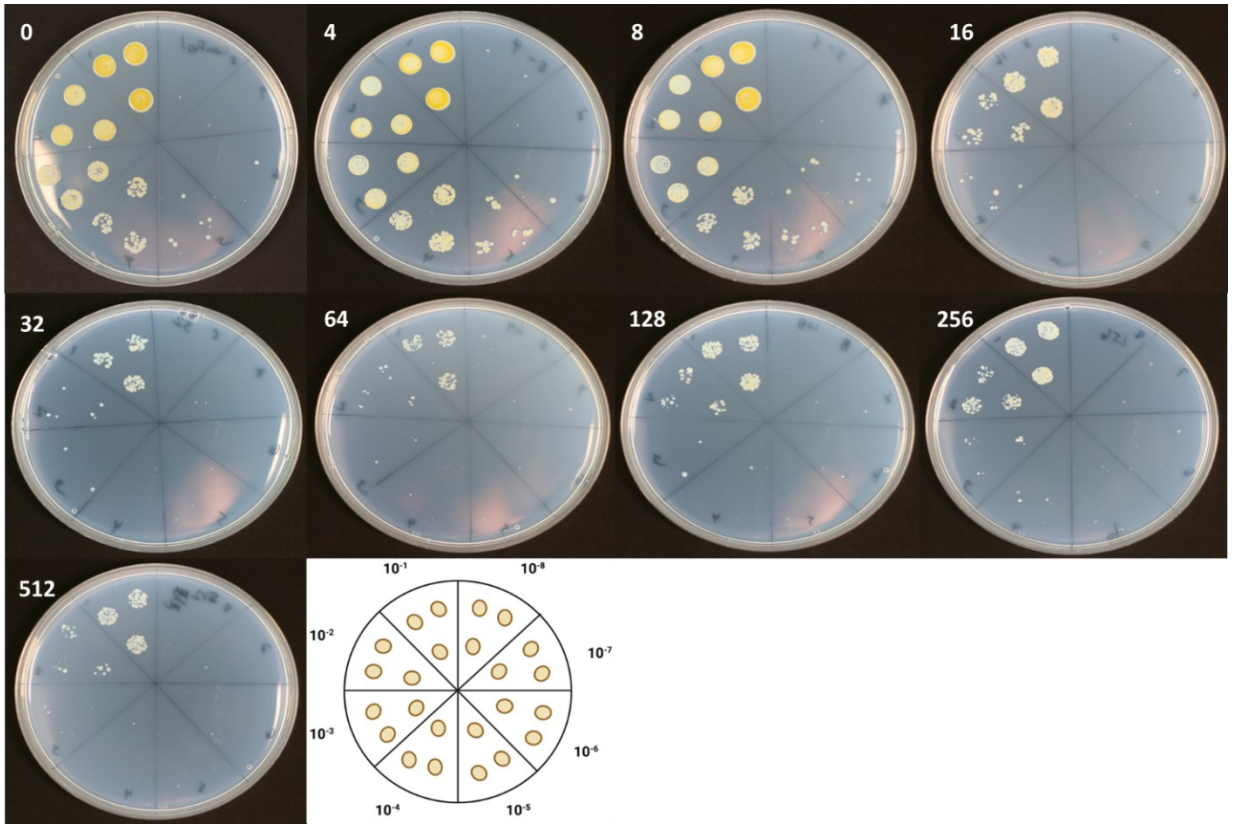


Figure 3-9. Representative nonselective agar plates showing the regrowth of *M. komossense* JERR01 after exposure to different concentrations of INH.

Each plate corresponds to a specific microdilution well from which 10 µl was inoculated on non-selective agar after the MIC_{90} incubation period. The plates were incubated at 30°C for 5 days, allowing the assessment of bactericidal activity. Figure contains cells recovered from INH microdilution plates, starting with a drug free control (0), followed by the concentration 4-512 µg/ml *M. komossense* JERR01 in ascending order.

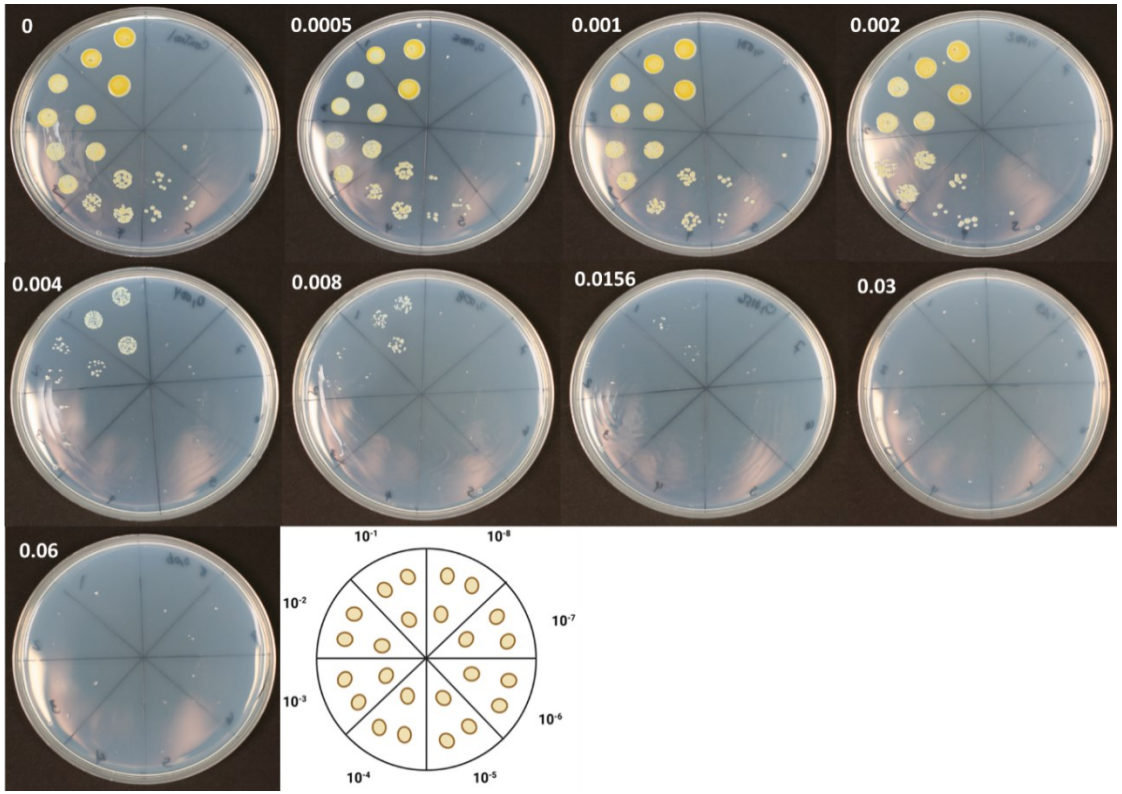


Figure 3-10. Representative nonselective agar plates showing the regrowth of *M. komossense* JERR01 after exposure to different concentrations of RIF.

Each plate corresponds to a specific microdilution well from which 10 µl was inoculated on non-selective agar after the MIC_{90} incubation period. The plates were incubated at 30°C for 5 days, allowing the assessment of bactericidal activity. The figure contains the cells recovered from RIF microdilution plates starting with a control plate, followed by the concentrations 0.0005-0.06 µg/ml of *M. komossense* JERR01 in ascending order.

3.3.7 Using SLIC for *M. komossense* JERR01 for bacterial growth determination.

Whilst SLIC offers the potential to study bacterial growth characteristics, including the effect of antibiotics, it is a recent technology which has not previously been used to study mycobacteria. Three SLIC devices (v6.1.3, v7.0.2, and v7.1.3) were available for use but experiments to assess consistency of results from them had never been performed. Therefore, the variability in the results of the three available SLIC devices was investigated by comparing the bacterial growth curves that they generate.

Growth curves from SLIC were based on the amount of light scatter (mV) produced by the interaction between the laser and bacterial cells present in the sample. A single *M. komossense* JERR01 preparation at a concentration of 10^5 CFU/ml in 50 ml of 7H9 Middlebrook broth was used for all three devices. Two milliliters of this preparation were loaded into wells 2-6 of each device, with well 1 containing 2 ml of broth only, as a bacteria-free control. Both the SLIC devices and the media were preheated to 30°C, and *M. komossense* JERR01 bacterial growth was measured by total light scatter (mV). Runs were conducted for a total of 24 hours (see chapter 2.13). Results from all three machines are shown in Figure 3-11.

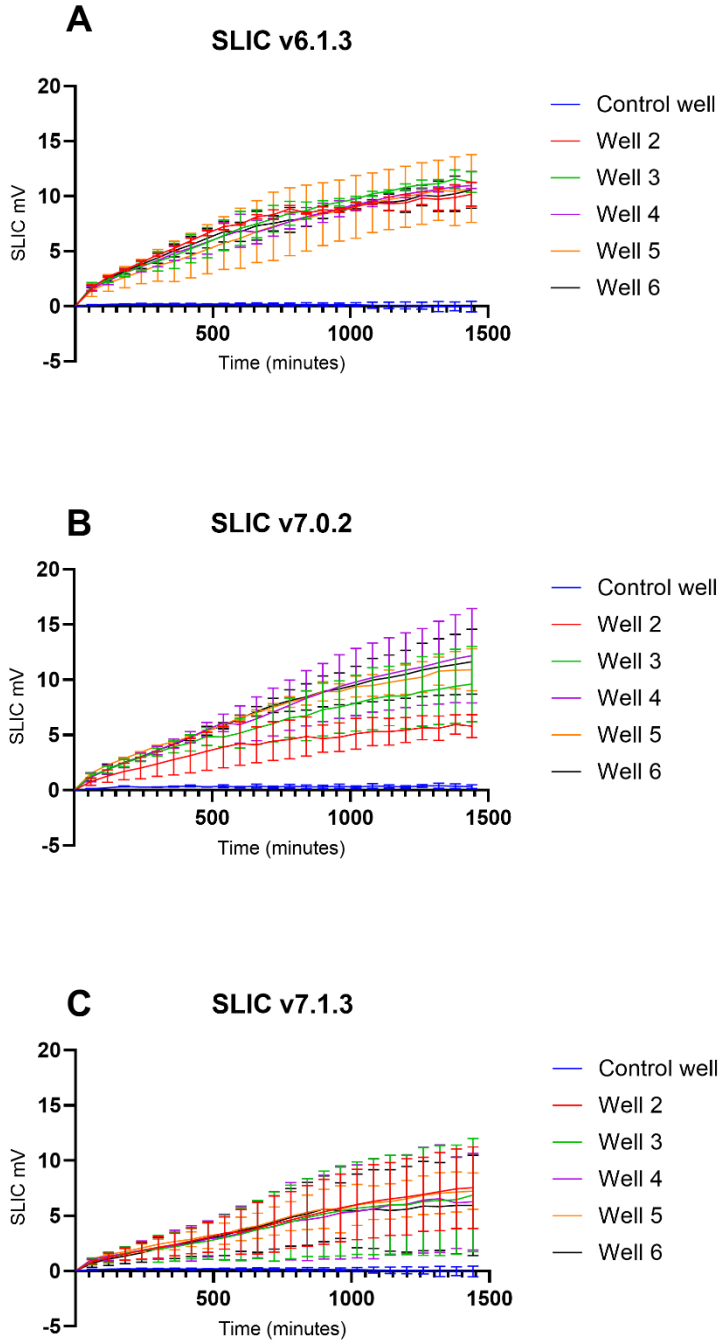


Figure 3-11. Growth detection of *M. komossense* JERR01 at 30°C using SLIC: (A) v6.1.3, (B) v7.0.2, and (C) v7.1.3.
Well 1 contained broth, Wells 2-6 contained *M. komossense* JERR01 at 10^5 CFU/ml. Light scatter signal was plotted in mV. Error bars represent the standard deviation from the mean of 3 biological repeats.

The results obtained from the SLIC runs show differences in reproducibility of results between the devices. In SLIC v6.1.3, (Figure 3-11), there is a slight variation among the different biological repeats, however, the device produced graphs that could be considered replicable as the growth curves obtained from wells 2-6 overlap very closely with narrow, overlapping error bars. SLIC devices v7.0.2 and v7.1.3, presented a lack of reproducibility in their results (revealed through divergent growth curves and wide error bars in wells 2-6). Therefore, it was decided to use the device SLIC v6.1.3 to test the effects of antitubercular drugs in *M. komossense* JERR01.

3.3.8 Utilizing SLIC to assess the impact of antibiotics on *M. komossense* JERR01.

After selecting one SLIC machine to describe the bacterial growth curve of *M. komossense* JERR01, the effect of two antibiotics (INH and RIF) on the organism was assessed. These drugs were selected for their key roles in the treatment of DS-TB due to their potent bactericidal activity and also, because the identification of MDR-TB (INH resistance and RIF resistance) through early AST improves the assignment of TB, facilitating better outcomes and less transmission (Jhun & Koh, 2020; Kim et al., 2023; Vogensen et al., 2022).

In each SLIC run, a panel of 5 different concentrations of either INH or RIF were used with 1 bacteria-free control well. The antibiotics concentrations used were doubling dilutions in the following ranges: INH (8-128 µg/ml) and RIF (0.001-0.0156 µg/ml). Data from Section 3.3.5 reported MIC_{90}s for INH and RIF of 32 µg/ml and 0.002 µg/ml, respectively. When determining antibiotic ranges for SLIC, the MIC_{90} concentration was used as a middle point, and test concentrations below and above this were used. All SLIC

runs continued for 24 hours (1,440 minutes). The response of *M. komossense* JERR01 to the antibiotics was based on the level of scattering (mV) of the antibiotic treated sample and compared to the antibiotic-free control. The growth percentage for each sample was calculated using equation 2.6 (see Chapter 2, section 2.13.2.2).

For INH (Figure 3-12 A) growth of bacteria at the highest antibiotic concentration (128 µg/ml) started to deviate from the antibiotic-free control after 300 minutes. By 24 hours, the wells containing the concentrations of 64 µg/ml and 128 µg/ml had inhibited approximately 90% of the growth of *M. komossense* JERR01. Although this pattern of inhibition at higher INH concentrations is observed, wide error bars indicate a high degree of variability between biological repeats and suggest the need for caution in data interpretation.

For RIF (Figure 3-12 B), similar to the INH dose response graph, there is an inhibition at the two highest concentrations of RIF, with an 80% reduction in the growth of *M. komossense* JERR01 compared to the antibiotic-free control by 24 hours. Again, this data must be approached with caution as the wider error bars indicate variability between biological repeats.

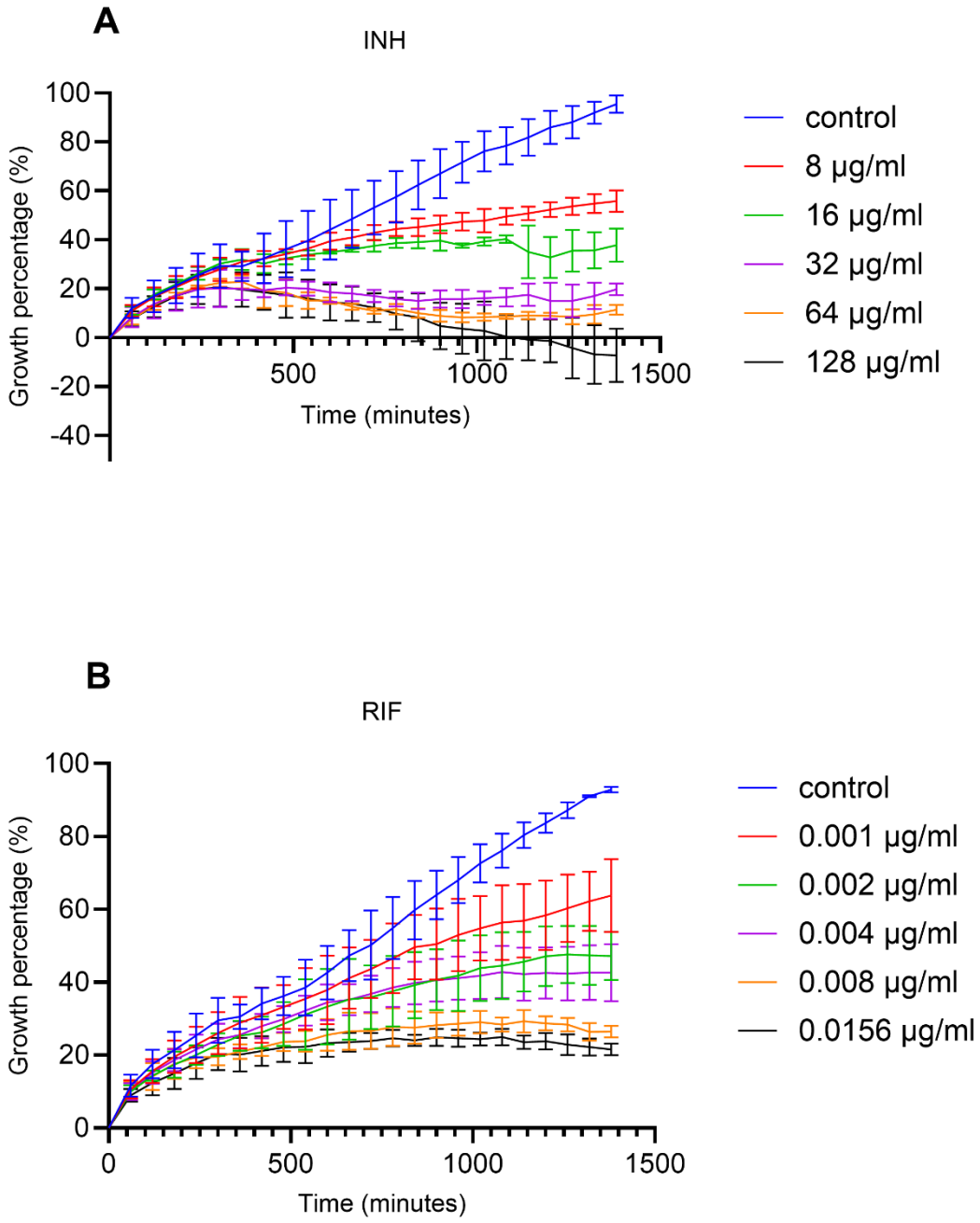


Figure 3-12. Response of *M. komossense* JERR01 grown at 30°C in SLIC in the presence of A-INH (8 – 128 µg/ml), and B-RIF (0.001 – 0.0156 µg /ml).
Light scattering was detected and plotted in bacterial growth rate (mV) against time. mV values were added to observe the percentage growth change between the control and the treated samples. Error bars represent the standard deviation from the mean of three biological replicates.

3.4 Discussion

In this chapter, a targeted exploration of the growth rate and sensitivity towards antibiotics of *M. komossense* JERR01 strain was performed, with the final aim of evaluating its potential as a model organism. By comparing its phenotypic traits with those of a reference strain of *M. komossense* (ATCC 33013), differences were observed that highlight distinct properties that *M. komossense* JERR01 has in comparison to the reference strain. The investigation into the phenotypic characteristics of *M. komossense* JERR01 produced information regarding its growth kinetics, colony morphology, and responses to various antibiotics with different modes of action.

M. komossense JERR01 grows at 30°C and 37°C, and this corresponds to the environment on which it was isolated from (human body), while *M. komossense* ATCC 33013 is able to grow from 21°C to 30°C, which corresponds to its usual environment (moss vegetation)(Kazda and Muller, 1979). *M. komossense* JERR01 is a fast-growing *Mycobacterium* that is accessible, easy to manipulate, and has a status as a BSL-2 organism compared to *Mtb,* and these are advantageous factors for practical use in research settings. The doubling time of *M. komossense* JERR01 is comparable to what is has been observed with other fast-growing mycobacteria used as model organisms for drug resistance studies such as *M. smegmatis* MC2 155 (4 hours). But similarly to *M. smegmatis*, *M. komossense* JERR01 has its limitations. The slow growing rate of *Mtb* has been associated with its pathogenicity (Bachmann *et al.*, 2020), thus, *M. komossense* JERR01 would be an unsuitable model for host-pathogen interactions and pathogenicity, as is *M. smegmatis.*

It was also noted that *M. komossense* JERR01 produces a pigment which is not observed in the type strain. The yellow pigment produced by *M. komossense* JERR01 is present in cells grown at both 30°C and 37°C, notably, at the latter temperature, the type strain of *M. komossense* does not grow. Most of the pigments observed in mycobacteria are carotenoids, and their main function is to serve as free radical scavengers to protect bacteria against photooxidative damage (Tapiero, D.M. Townsend and Tew, 2018). It is unknown if those carotenoid pigments present in mycobacteria possess any role in virulence (Robledo, Murillo and Rouzaud, 2011). It is also unknown if pigment production plays a role in antimicrobial resistance. Despite the scarcity of literature on mycobacterial pigments and antimicrobial resistance, some historical studies suggest a potential correlation. In the work published by (Mayer, 1944), it was noted a particular strain of *Mtb var. hominis* that produced a bright yellow pigment when grown in the presence of p-aminobenzoic acid, which is a compound found in folic acid and in certain foods. Since *M. komossense* JERR01 is also a pigment producing *Mycobacterium*, this organism could perhaps be used to determine if there is a correlation between carotenoid production and AMR.

It was not possible to determine a growth curve of *M. komossense* ATCC 33013 strain in liquid media, despite growth on agar. The lack of growth on liquid media by *M. komossense* ATCC 33013 could be due to several factors such as oxygen availability. *M. komossense* ATCC 33013 might require a higher level of oxygenation to grow than is available in broth media, therefore allowing the growth of *M. komossense* only on the surface of Middlebrook agar media (Bonnet *et al.*, 2020). However, when growing *M.*

komossense ATCC 33013 at different rpm to increase aeration in the culture, no growth was observed.

Another explanation for the lack of growth of *M. komossense* in liquid broth could be due to surface adherence. *M. komossense* ATCC 33013 might prefer to adhere to solid surfaces like agar, the attachment to solid surfaces might be essential for the growth and reproduction of this strain (Zheng *et al.*, 2021). Although some studies have used this organism on agar-based media, there is a paucity of information regarding its cultivation in broth culture (Kazda and Muller, 1979; Boddinghaus *et al.*, 1990; Tay *et al.*, 1998; Vuorio *et al.*, 1999). Since *M. komossense* ATCC 33013 was not able to grow in liquid media, regardless of the different experimental approaches taken, other experiments such as growth rate estimation and MIC_{90} determination using antitubercular drugs were not possible as these are reliant on being able to grow the strain in liquid culture.

Since *M. komossense* JERR01 and *M. komossense* ATCC 33013, exhibited different growth characteristics and morphological features, it raises the possibility that the classification of *M. komossense* JERR01 as *M. komossense* ATCC 33013 may be imprecise; it may belong to a different *Mycobacterium* species, indicating the need for further taxonomic investigation and genotypic characterization to elucidate its distinct classification within the genus. Further work on this will be described in Chapter 4.

M. komossense JERR01 presented an MIC_{90} for the drugs RIF, EMB, MOXI, LEVO, BDQ, and LZD below the clinical breakpoints for *Mtb*, suggesting its potential as a model organism for *Mtb* resistance studies for these drugs, however, further experimentation is required to confirm this. The observed MIC_{90} for INH in *M. komossense* JERR01 was 32

µg/ml, higher than the clinical breakpoint determined for *Mtb* (0.1 µg/ml). INH is a prodrug that needs to be activated by a catalase enzyme coded by the *katG* gene to work. When activated, INH inhibits mycobacterial replication by inhibiting the function of *inhA* (responsible for the synthesis of mycolic acid) and *ahpC* (involved in oxidative stress response) (Master *et al.*, 2002). Other environmental mycobacteria used as model organisms also have a high MIC_{90} when exposed to INH. For example, *M. smegmatis* MC2 155 has an MIC_{90} to INH of 128 µg/ml. These results are different from what is observed in *Mtb*, where INH has a high early bactericidal potency at low concentrations by binding at various positions in KatG. Among the different binding positions in *Mtb* KatG, the most prominent and clinically relevant is the position S315T and C-15T region of *fabG/inhA* (Unissa *et al.*, 2016a; Maladan *et al.*, 2021). Interestingly, when performing a Needleman-Wunch global alignment of the *M. smegmatis* MC2 155 KatG protein with *Mtb*, there is a lack of conservation at the position S315T of the KatG proteins, which may explain the high MIC_{90} observed in *M. smegmatis* MC2 155. Similar genomic features could explain the high MIC_{90} of *M. komossense* JERR01 to INH. However, to corroborate if there are differences in conservation between both KatG proteins (*Mtb* and *M. komossense* JERR01), the protein sequence of the *M. komossense* KatG would be required for evaluation. In chapter 4 of this thesis, the whole genome sequencing and genotypic characterization of *M. komossense* AMR associated genes will be described.

The observed MIC_{90} for pyrazinamide (>512 µg/ml) shows an intrinsic resistance from *M. komossense* JERR01 towards this antibiotic. PZA is an antibiotic that can only act in the acidic intracellular environment (Fontes *et al.*, 2020). Replicating the acid environment where PZA works is difficult to accomplish *in vitro*, and false positives for

resistance can commonly be reported in error for this drug, however, a protocol was followed in which acidified media was used, suggesting that the finding of intrinsic resistance was genuine (Biadglegne et al., 2014).

PA is another antibiotic that it is thought to inhibit cell wall formation by blocking the synthesis of mycolic acid, but its molecular mechanism of action is still not fully understood (Manjunatha, Boshoff and Barry, 2009). As a prodrug, PA requires activation by deazaflavin (F_{420}) dependent nitroreductase Ddn through the action of key markers: *ddn, fgd1, fbiA, fbiB, fbiC, and fbiD* (Gómez-González *et al.*, 2021). Differences in these key markers between *M. komossense* JERR01 and *Mtb* could lead to resistance to PA.

M. komossense JERR01 RIF MBC suggests a similar mechanism of action for RIF resistance compared to *Mtb*, however, further investigations are required to determine how RIF inhibits *M. komossense* JERR01. No MBC could be measured for *M. komossense* JERR01 to INH because bacteria could be revived from cultures even at the highest antibiotic concentration tested. This contrasts with the previously reported MBC activity of *Mtb*, where early bactericidal activity of INH is observed during the lag and log phase growth (Yamori *et al.*, 1992; Nair *et al.*, 2015). In chapter 5 a more in-depth look into RIF and INH induced mutations will be undertaken. Spontaneous mutations that confer resistance to RIF and INH at a range of concentrations in *M. komossense* JERR01 will be induced to determine whether it develops mutations in the same manner as *Mtb.*

SLIC as a novel technology, was evaluated to determine whether it can provide information on growth characteristics and antibiotic responses of mycobacteria more quickly than conventional methods (which take up to 7 days, even for fast-growing

organisms in the genus). In previous studies, it has shown potential to detect bacterial growth of several Gram-positive and Gram-negative bacteria and has been used to test antimicrobial susceptibility (Hammond *et al.*, 2017, 2022; Falconer *et al.*, 2024).

When tested using *M. komossense* JERR01, limitations of SLIC were observed, including variability in growth curves of the same organism between devices or, in some cases, when run at the same starting inoculation in different wells of the same device. To try and minimise this variability, the single device which provided the most consistent drug-free growth curves was used to assess drug response *in M. komossense* JERR01. Although a dose-response relationship between higher INH and RIF exposures and greater inhibition of bacterial growth could be seen, further work to standardise the methodology is necessary before it could be used to report MIC_{90} measurements or describe antibiotic susceptibility in mycobacteria, especially when members of this genus possess different doubling times and their tendency to form aggregates and clump at the bottom of culture tubes.

3.5 Summary

Although the initial speciation of the organism under study in this thesis was *M. komossense*, phenotypic differences from the *M. komossense* ATCC 33013 type strain (different growth kinetics on solid and liquid media, different degrees of acid-fastness on microscopy, and different pigment production) indicate a need for further study of the identity of the organism, including genotypic characterisation which will follow in Chapter 4.

The delineation of the generation time of *M. komossense* JERR01 as 5.13 hours at 30°C and 5.84 hours at 37°C, confirms this to be a rapidly growing *Mycobacterium*, like *M. smegmatis*. This may confer some advantages for the read-out time of experiments using *in vitro* model organisms.

Considering its potential use as a model of *Mtb* in antibiotic resistance studies, *M. komossense* JERR01 had an MIC_{90} profile which would confer 'susceptibility' based on clinical breakpoints or critical concentrations for *Mtb* to all drugs tested except INH, PA, and PZA. Further genomic studies of the organism's interaction with drugs INH and RIF may improve understanding of resistance mechanisms for those first line drugs. This will also be advanced in Chapters 5.

Finally, the characterisation of the growth rate of *M. komossense* JERR01 by SLIC was challenging, suggesting that further optimisation of this technology is required before it can be used for mycobacteria.

4 Genomic characterization of *M. komossense* JERR01

4.1 Introduction.

In the previous chapter, a phenotypic comparison between *M. komossense* JERR01 and *M. komossense* ATCC 33013 was performed. The results provided interesting phenotypic differences between these two organisms. The differences observed, encompass colony morphology, microscopic features, and differences in growth rates. Due to the inability of *M. komossense* ATCC 33013 to grow in liquid media, it was not possible to perform an AST using the EUCAST/CLSI recommended method for NTMs, broth microdilution; therefore it was not possible to extend the phenotypic comparison to antibiotic sensitivity as planned.

When performing an AST on *M. komossense* JERR01, however, while 6 MIC_{90} values fell below the *Mtb* clinical breakpoint range (RIF, EMB, MOXI, LEVO, BDQ, and LZD), suggesting sensitivity according to the *Mtb* clinical breakpoints, 3 MIC_{90} values (INH, PZA, and PA-824) exceeded these thresholds, suggesting an intrinsic resistance towards these antibiotics.

To address these observations, a genotypic characterization of *M. komossense* JERR01 was performed, and genomic comparisons between *M. komossense* JERR01, *M. komossense* ATCC 33013, and *Mtb* H37Rv were also performed.

4.2 Chapter aims and objectives

Given the differences presented in previous chapter (Chapter 3) between the *M. komossense* strain JERR01 and *M. komossense* ATCC 33013, it provided an opportunity to genetically characterise *M. komossense* JERR01 through the sequencing of its genome. The aim of this chapter was to provide an overview of the genome architecture of *M. komossense* JERR01 and explore its annotation, and undertake comparative genomic analysis to explore its genetic and functional relationships with *M. komossense* ATCC 33013, and other mycobacteria, including *Mtb*. Specific objectives were to:

1. Determine the genomic structure of the *M. komossense* JERR01 genome through whole genome sequencing and provide an overview of functional annotation.
2. Compare the genome content and genome structure of *M. komossense* JERR01 with the reference strain *M. komossense* ATCC 33013 to describe their relatedness.
3. Describe the taxonomic relationship of *M. komossense* JERR01 among the *Mycobacteriaceae* using a subset of representative members of the genus and using comparative analysis of their whole genome data (e.g. ANI, shared genes, genome structure) and subsequently, create a phylogenetic tree to visualize the evolutionary relationship of *M. komossense* JERR01 within the genus.
4. To identify and compare AMR related genes present in the genome of *M. komossense* JERR01 compared to *Mtb*.

4.3 Results

4.3.1 *M. komossense* JERR01 sequencing and assembly.

In order to characterize the genomic structure of *M. komossense* JERR01 its whole genome was sequenced. Bacteria colonies of *M. komossense* JERR01 were sent to MicrobesNG (Birmingham, UK) for this purpose. MicrobesNG, extracted the DNA from the *M. komossense* JERR01 samples, constructed the DNA library and performed an Oxford nanopore sequencing on the extracted DNA. The result of this sequencing produced long read data and sequencing files in FASTA and GenBank format. With the long-read data, an assembly using Unicycler was performed (Wick *et al.*, 2017), and the visualization of the assembly graphs was performed using Bandage (Wick *et al.*, 2015). The purpose of the assembly was to align and merge the long DNA fragments in order to reconstruct the original genome, including chromosomes and any mobile genetic elements. The results of this assembly produced one circularized chromosome and 4 fragments of DNA (Figure 4-1). A quality assessment was also performed in the assembly using QUAST 5.2.0 (Quality Assessment Tool)(Table 4-1), to evaluate the quality of the newly assembled genome (Gurevich *et al.*, 2013).

Table 4-1. Genome assembly quality assessment: Genome statistics.

Assembly details	
Contigs	5
Total length (bp)	6,778,417
GC (%)	67.14
Plasmids	4
N50	5,995,690
L50	1
Chromosomes	1

The assembly analysis performed by Quast concluded that the assembly of *M. komossense* JERR01 consisted of 5 contigs with a total length of 6,778,417 bp and a GC content of 67.14%.

To assess the accuracy and completeness of the assembly, a BUSCO v5.4.7 analysis was performed, the results provided by BUSCO were of 97% completeness, meaning that the vast majority (97%) of the expected single-copy orthologs were found in the *M. komossense* JERR01 data (Simão *et al.*, 2015), thus indicating a likely high-quality dataset. The size of the chromosome of *M. komossense* JERR01 was reported to be 5,995,690 bp, and additionally to this chromosome, four additional Contigs were identified, two of these Contigs were linear with sizes of 261,146 bp and 220,722 bp, and two of the contigs circularized with sizes 191,510 bp, and 109,349 bp, respectively.

The whole genome was annotated using Prokka to identify and annotate various features like protein-coding sequences, RNA genes (tRNA and rRNA) and predict protein functions through similarity searches in databases (Table 4-2)(Triebel *et al.*, 2023). This provides a comprehensive description of the genome content and potential functions encoded within it.

Table 4-2. *M. komossense* JERR01 genome annotation.

Annotated genome features	
Contigs	5
Bases (bp)	6,778,417
CDS	6,400
rRNA	6
tRNA	51
tmRNA	2

This table summarizes the annotation of the *M. komossense* JERR01 genome using Prokka. It shows the number of contigs, total number of bases, protein coding sequences (CDS), ribosomal RNAs (rRNA), and transfer RNAs (tmRNA).

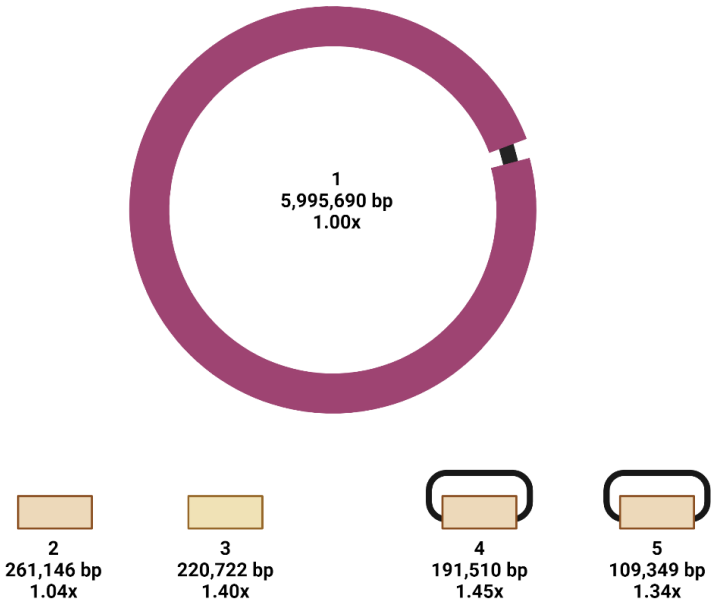


Figure 4-1. Structure of the *M. komossense* JERR01 genome. Complete genome assembly of *M. komossense* strain JERR01.
Unicycler nanopore assembly generates an assembly graph that is visualized by Bandage and adapted using Biorender. The assembly graph provides a simplistic schematic of the genome, indicating a closed chromosome or plasmids by a linker connecting two ends of a single contig or incomplete chromosomes and plasmids which are not connected by a linker.

4.3.2 Chromosomal characterization

Having described the general characteristics of the *M. komossense* JERR01 draft genome, the next step was to characterize its different components: chromosomes and potential plasmids individually. The chromosome of *M. komossense* JERR01 was circularized during the assembly and, therefore, predicted to be completed. Based on the Prokka annotation, the chromosome was predicted to have the following features: 5,620 coding sequence proteins (CDSs), 50 tRNAs, 6 rRNAs, and 2 tmRNA. Among the CDSs annotated, hypothetical proteins make most of the proteins annotated (Figure 4-2). This provides a resource to investigate the genome content and functions encoded within.

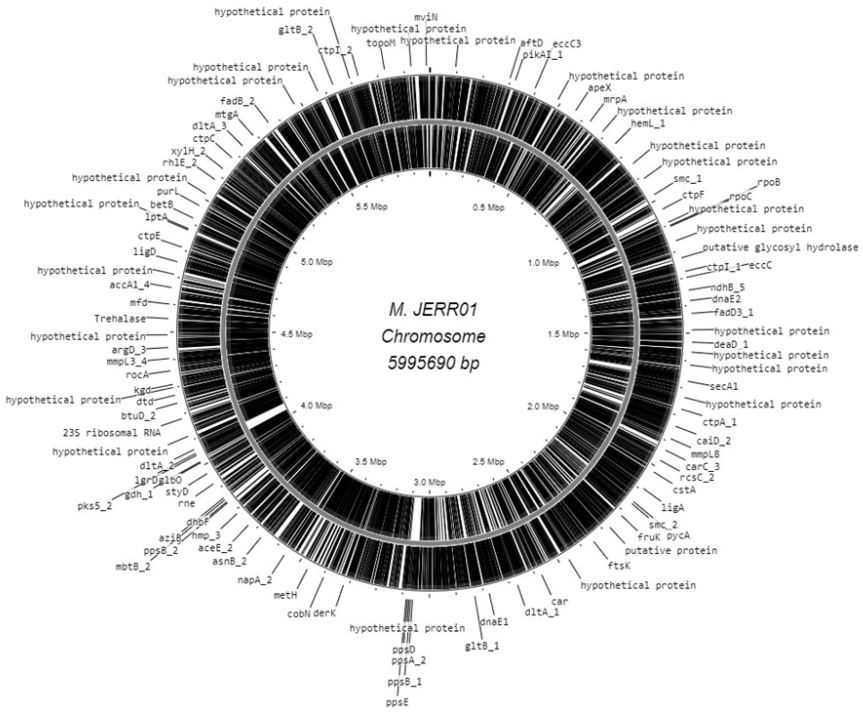


Figure 4-2. Map of *M. komossense* JERR01 chromosome.
The chromosome of *M. komossense* JERR01 was circularized during the assembly and was reported to have 5,995,690 bp. The circles in the map represent, from outside to inside, the encoded CDSs on the forward strand and the reverse strand; tRNAs and rRNAs. Figure produced using Proskee (Grant et al., 2023).

Since the chromosome of *M. komossense* JERR01 has 5,620 CDS, a subsystem distribution was performed on the chromosome proteins to categorize them based on their predicted functions using GhostKoala (Kanehisa, Sato and Morishima, 2016). The CDSs were compared to the GhostKoala database, which matches each protein to a subsystem. Upon conducting a subsystem analysis of the chromosomal composition, the results indicate a prevalence of genes within the "Carbohydrate metabolism" subsystem category (284 proteins), showing the high reliance of *M. komossense* JERR01 to carbohydrates as a primary source of energy as shown in Figure 4-3 and the distribution of proteins in different subsystem categories. This subsystem analysis provides valuable functional context that complements my aim of analysing the *M. komossense* JERR01 annotations.

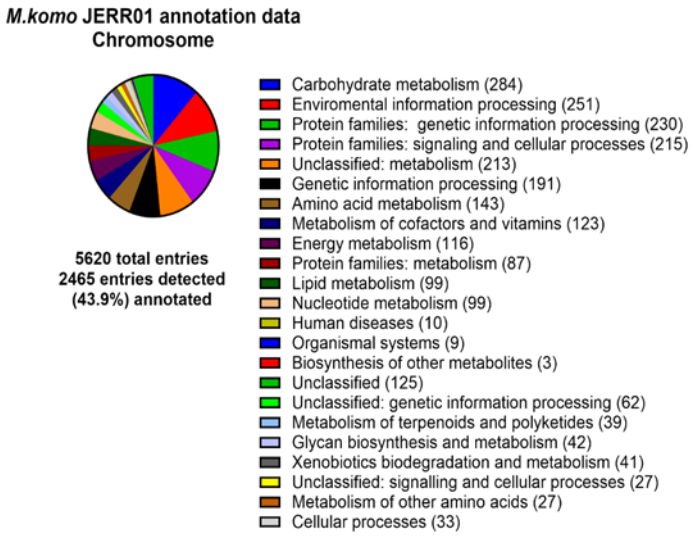


Figure 4-3. Subsystem distribution of the annotated functions encoded in Contig 1. GhostKoala results of subsystem and functional traits identified from *M. komossense* JERR01 chromosome that were connected with KEGG pathway database. The subsystem analysis detected 5,620 entries, but only 2,465 (43.9%) entries were annotated by KEGG and distributed into different KEGG pathways.

The GenBank file sequence of the *M. komossense* JERR01 chromosome generated by the initial Prokka annotation was analysed using Ori-Finder 2022 (Dong, Luo and Gao, 2022) to identify the origin of replication (*oriC*) related elements and terminus (*terC*). The characterization of *oriC* elements was done to further confirm the quality of the genome sequencing and assembly as a high-quality assembled genome should contain a complete and identifiable *ori* sequence. The results provided by Ori-Finder 2022 revealed the presence of different *oriC* indicator genes, such as chromosomal replication initiator protein DnaA (*dnaA*) in the coordinates 1-1,488 bp, Beta sliding clamp (*dnaN*) in the coordinates 2,199-3,395 bp and Integration host factor subunit alpha (*ihfA*) at the coordinates 2,119,440-2,120,102 bp. Figure 4-4 A illustrates the disparity of distribution of adenine/thymine (A/T), guanine/cytosine (G/C), purine/pyrimidine (R/Y), and amino/keto (M/K) bases, DnaA boxes clusters, *oriC* related indicator genes, potential *oriCs*, from which we chose the origin of replication with the higher score as *oriC*. In Figure 4-4 B, the *oriC* sequence is displayed on the x-axis, with different elements related to the origin of replication depicted. A notable feature is the presence of a red-colored peak in the x-axis; this peak represents an AT-rich region that might serve as AT-rich unwinding elements; also in the x-axis, it is possible to observe one Dam methylation site (GATC) region, 4 DnaA boxes and one DnaA-trio (repeated 3-mer motif).

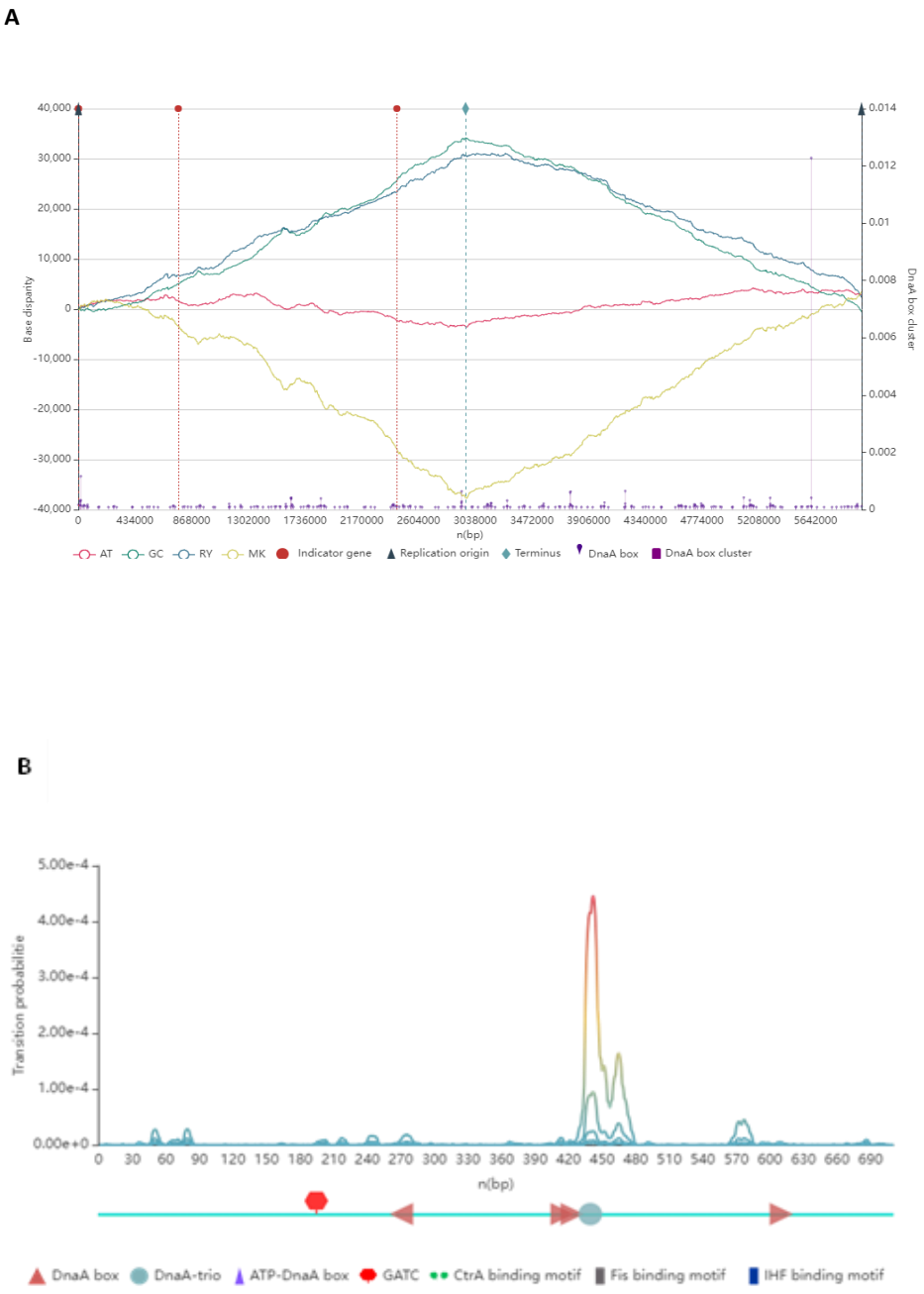


Figure 4-4. Identification of replichores in *M. komossense* JERR01.
(A) The interactive Z-curve figure of the *M. komossense* JERR01 chromosome indicating origin of replication (*oriC*) and terminus (*terC*) for the chromosome and indicator genes depicted as a red-dotted line. The red, green, blue and yellow line graphs indicate AT, GC, RY and MK disparity. (B) characteristic visualization of *oriC* sequence with different *oriC* elements alongside the x-axis, the read peak represents the AT-rich region.

4.3.3 Plasmid characterization

The four additional contigs detected in the assembly (Contigs 2-5; Figure 4-1) were investigated as potential plasmids. Contigs 2 and 3, which were assembled as linear contigs, and contigs 4 and 5 were circularized during the assembly by Unicycler. Contigs 2 and 3 were not circularized during assembly because neither contig had significant overlap of the ends. Thus, additional evidence such as replication initiation sites (*ori*) and plasmid partitioning systems were searched to uncover if these contigs 2 and 3 were indeed plasmids.

4.3.3.1 Contig annotations

Contig 2 is the largest among the linear pieces of DNA and the second largest after the chromosome. The annotation of this contig resulted in 264 predicted CDSs and a single tRNA. Among the CDSs annotated in this contig, hypothetical proteins that constitute a considerable portion of the contig. Functional annotation of these hypothetical proteins is necessary to find the possible roles in these proteins in the *bacterium*. Contig 3 is the second largest piece of linear DNA identified during the assembly, this contig was predicted to contain 214 CDSs. Contig 4 was circularized during the assembly, and therefore, it was considered a circular plasmid. The annotation by Prokka identified 191 predicted CDSs. Contig 5 is the smallest of the contigs assembled by Unicycler. In this contig were identified 111 predicted CDSs, among these CDSs, hypothetical proteins constitute the highest percentage of proteins in the contig (Figure 4-5).

4.3.3.2 Contig subsystem analysis

Functional annotation of these contigs was necessary to find the possible roles in these contigs in the bacterium. For Contig 2, out of the 264 CDSs, 42 (15.9%) were distributed among 12 different subsystem categories by GhostKoala. The presence of 13 proteins categorized under "genetic information processing" suggests this subsystem is well-represented in the plasmid compared to other subsystems. This might indicate that the plasmid's primary function is related to processes like replication, transcription, translation, and DNA repair, which are core aspects of genetic information processing. In Contig 3, 49 CDSs (22.9%) were assigned to 13 different subsystem categories using GhostKoala. Similar to Contig 2, most of the CDSs mapped in Contig 3 (16 proteins in total) were related to genetic information and processing. Contig 4 was circularized during the assembly, and from the predicted 191 CDSs, 35 (18.3%) were distributed among 11 different subsystem categories using GhostKoala. Contig 5 subsystem distribution analysis using GhostKoala, mapped 18 proteins (16.2%) into 8 different subsystem distributions, as shown in Figure 4-15, with most of these proteins related to important signalling and cellular processes such as transport and catabolism, cell growth, and death, cellular community (biofilm formation)(Figure 4-6).

4.3.3.3 Origins of Replication

For Contig 2, BLAST comparisons were conducted to identify overall matches to plasmids in the public sequence databases. No significant extended matches or hits were identified. Since BLAST did not produce any hits, the next option was to identify on the contig if there were elements encoded that were associated with plasmid machinery. The software tool Ori-Finder 2022 was used for the determination of the origin of replication (*ori*) and origin of replication elements. Ori-finder 2022 was able to identify the presence of *parA* in the coordinates 212,021-213,890 bp. ParA, although not involved in DNA replication *per se*, binds to specific DNA sequences near the origin of replication. The *ori* sequence identified by Ori-Finder 2022, in this sequence, it is possible to observe a red peak that represents AT-rich regions that might serve as AT-rich unwinding elements. The evidence provided by Ori-Finder 2022, strongly suggests that this contig is an independently replicating genetic element and, therefore likely to belong to a plasmid (Figure 4-7).

Since Contig 3 was also reported as linear during the assembly, it was analysed using BLAST to find similarities with other published plasmids, however, not extended matches/hits were reported. As for the other contigs Ori-finder 2022 was used to identify potential origin of replication site region and *ori*-related elements. Figure 4-8 A presents the distribution of nucleotides alongside the genome sequence and position of predicted *ori* regions. The specific *ori* sequence with the presence of several peaks suggesting AT-rich regions. Ori-Finder 2022 identified two Dam methylation sites (GATC) regions alongside the x-axis. Similarly to Contig 2, the results produced by Ori-Finder 2022

strongly suggest that Contig 3 is also an independent replicating element and, therefore, likely to be part of the plasmid.

When using BLAST to determine the closeness of Contig 4 to plasmids already published, but only low percentage matches (<40%) were reported. The circularization of this plasmid during the genome assembly, the identification of a match by BLAST, and the identification of *ori* related elements by Ori-Finder 2022 strongly support the identification of this Contig 4 as a plasmid (Figure 4-9). When using BLAST with Contig 5, the highest match obtained was with *Mycobacterium* spp. SMC-8 plasmid pSMC-8_2 (CP079867.1) with a query coverage of 85% and percentage identity of 96%. A closer look into the genes that matched between Contig 5 and plasmid pSMC-8_2 revealed that most of the gene matches were between hypothetical proteins with high percentage similarities 95-100% (Figure 4-11). Contig 5 was the only contig from the *M. komossense* JERR01 assembly that obtained a high match/hit when compared to the NCBI database of microbial genomes. Ori-Finder 2022 was used to identify the *ori* regions in Contig 5, with the distribution of nucleotides alongside the genome sequence and the black dotted lines representing the possible origins of replication. Thus, confirming the identity of this Contig as a plasmid (Figure 4-10).

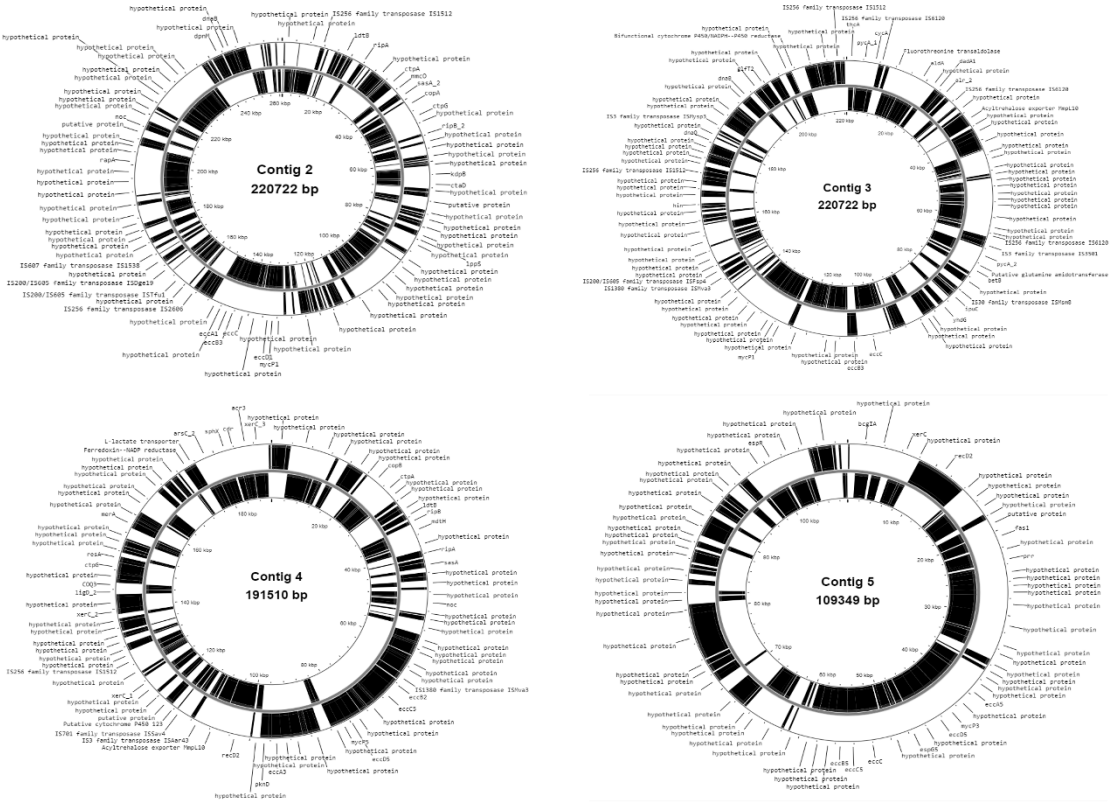


Figure 4-5. Map of *M. komossense* JERR01 identified contigs.
Circular map of *M. komossense* JERR01 Contigs 2 to 5. Rings from the inner to outer layer represent the location of annotated CDSs reverse and forward. Most of the annotated genes in these contigs are hypothetical proteins.

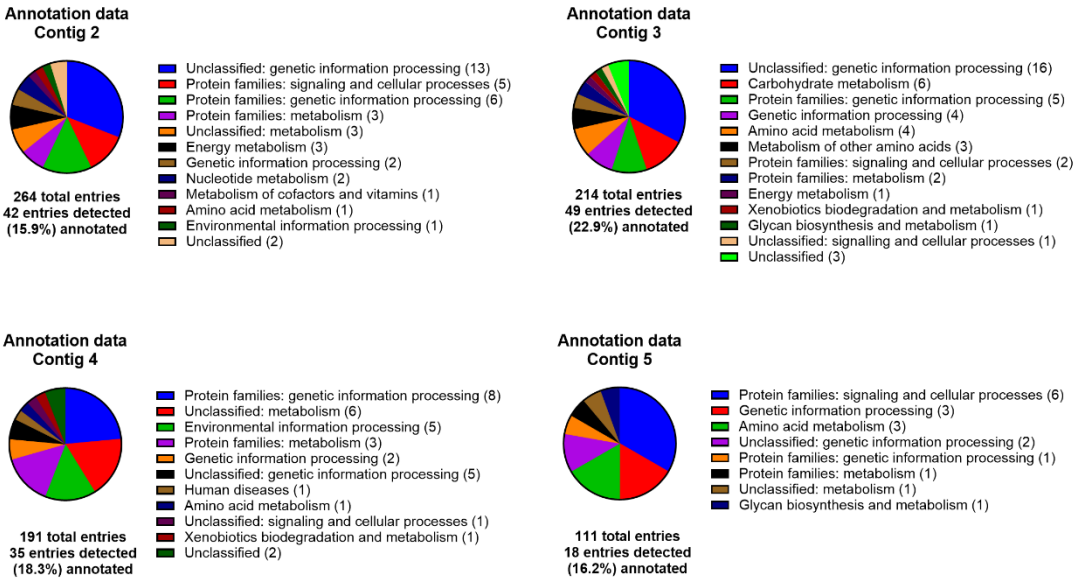


Figure 4-6. Subsystem distribution of the annotated functions encoded of contigs found in *M. komossense* JERR01 genome.

GhostKoala results of subsystem and functional distribution of the different proteins annotated in the *M. komossense* JERR01 Contigs 2 to 5.

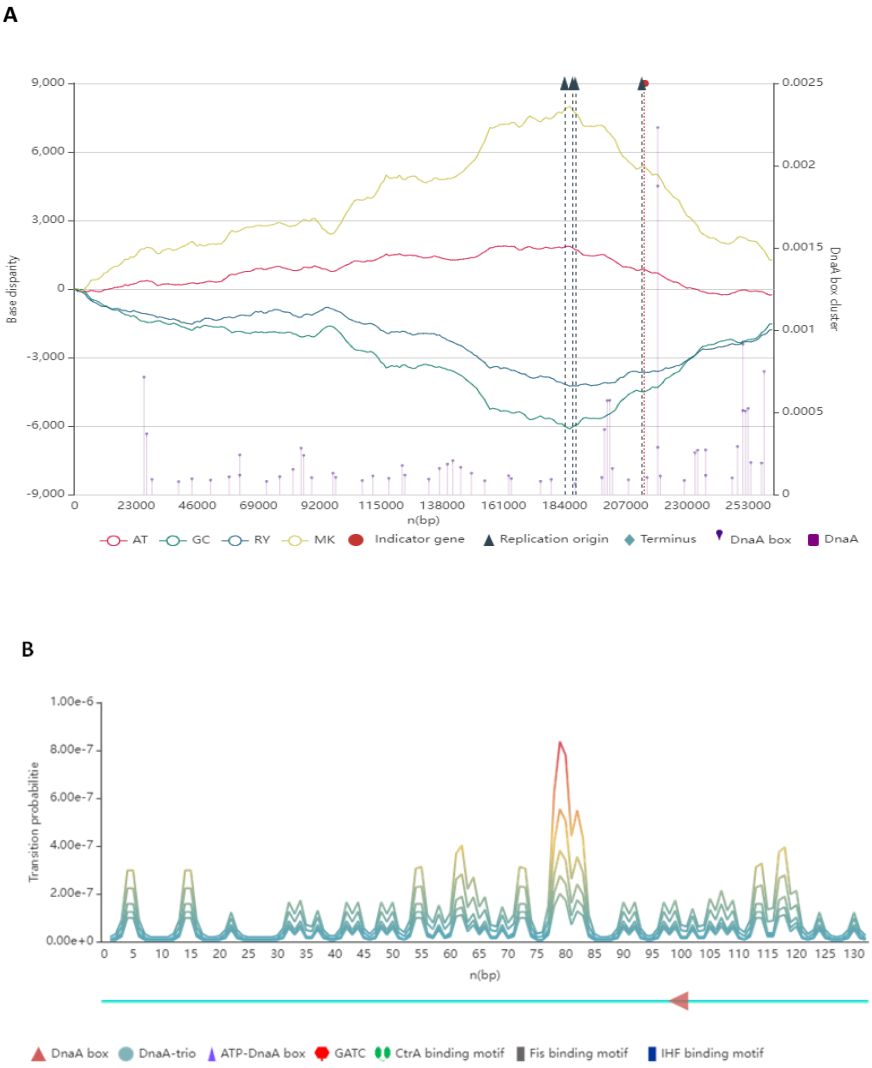


Figure 4-7. Identification of replichores in Contig 2, linear plasmid.
(A) The interactive Z-curve figure of *M. komossense* JERR01 Contig 2 indicating the origin of replication (*ori*) of the linear contig and indicator genes (DnaA box). The red, green, blue, and yellow line graphs indicate AT, GC, RY, and MK disparity. (B) characteristic visualization of *ori* sequence, alongside the x-axis, the presence of a DnaA box and one red peak representing an AT-rich region.

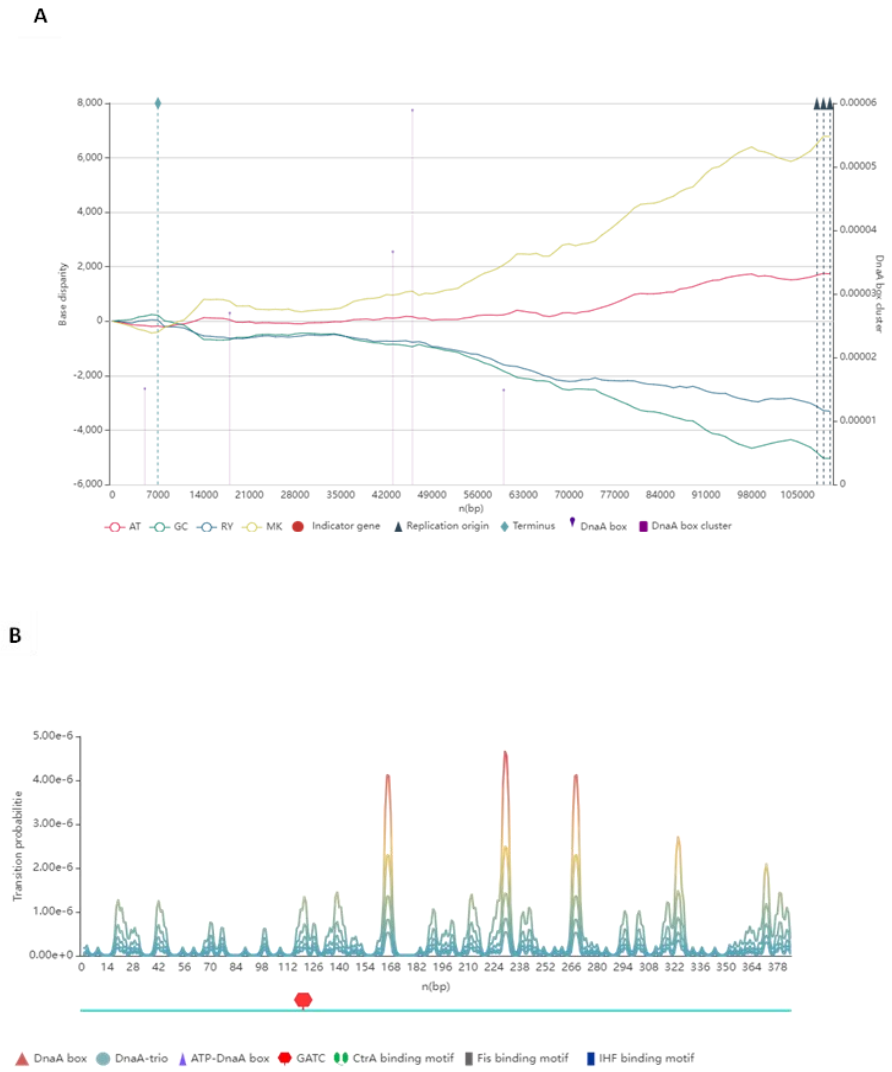


Figure 4-8. Identification of replichores in Contig 3 linear plasmid.
(A) The interactive Z-curve figure of *M. komossense* JERR01 Contig 3 indicating origin of replication (*ori*) the linear contig. The red, green, blue and yellow line graphs indicate AT, GC, RY and MK disparity. (B) characteristic visualization of *ori* sequence and alongside the x-axis the presence of two GATC regions and two red peaks representing two AT-rich regions.

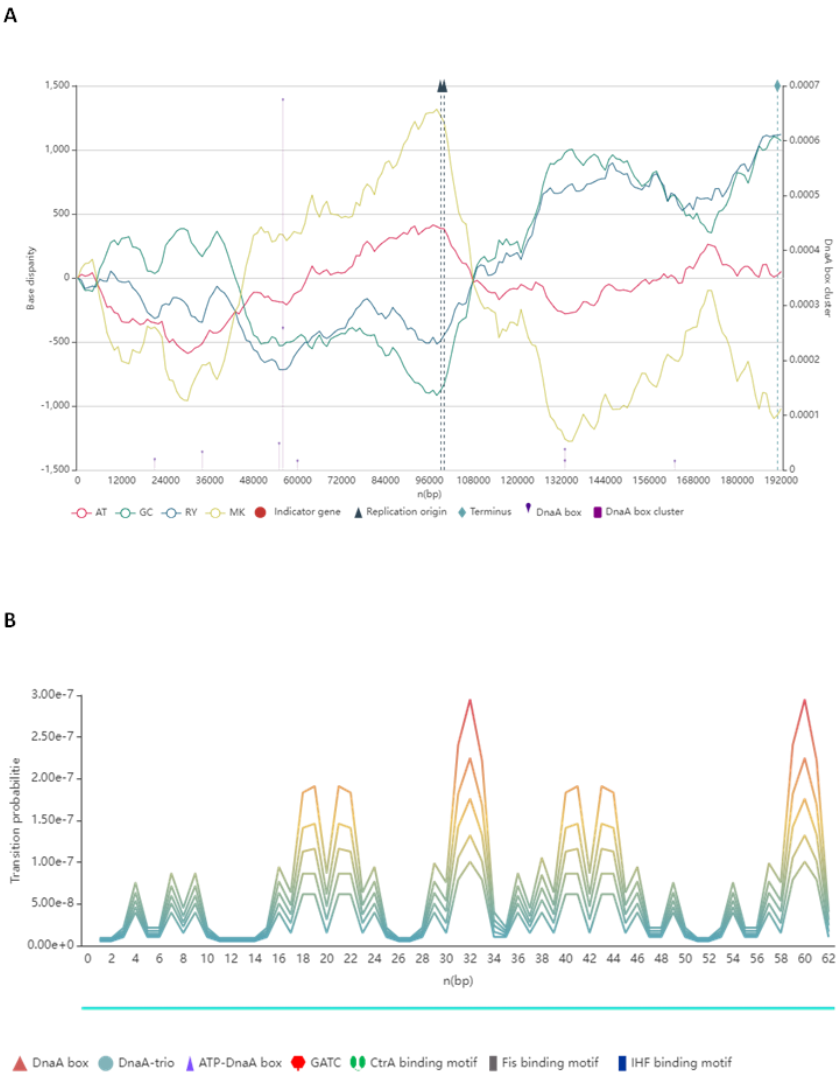


Figure 4-9. Identification of replichores in Contig 4 using Ori-Finder 2022.
(A) The interactive Z-curve figure of *M. komossense* JERR01 and Contig 4 indicating origin of replication (*ori*) and terminus (*ter*) in the circular contig. The red, green, blue and yellow line graphs indicate AT, GC, RY and MK disparity. (B) characteristic visualization of *ori* sequence, on the x-axis, no *ori* elements were displayed, however, there are two AT-rich regions represented as two red peaks.

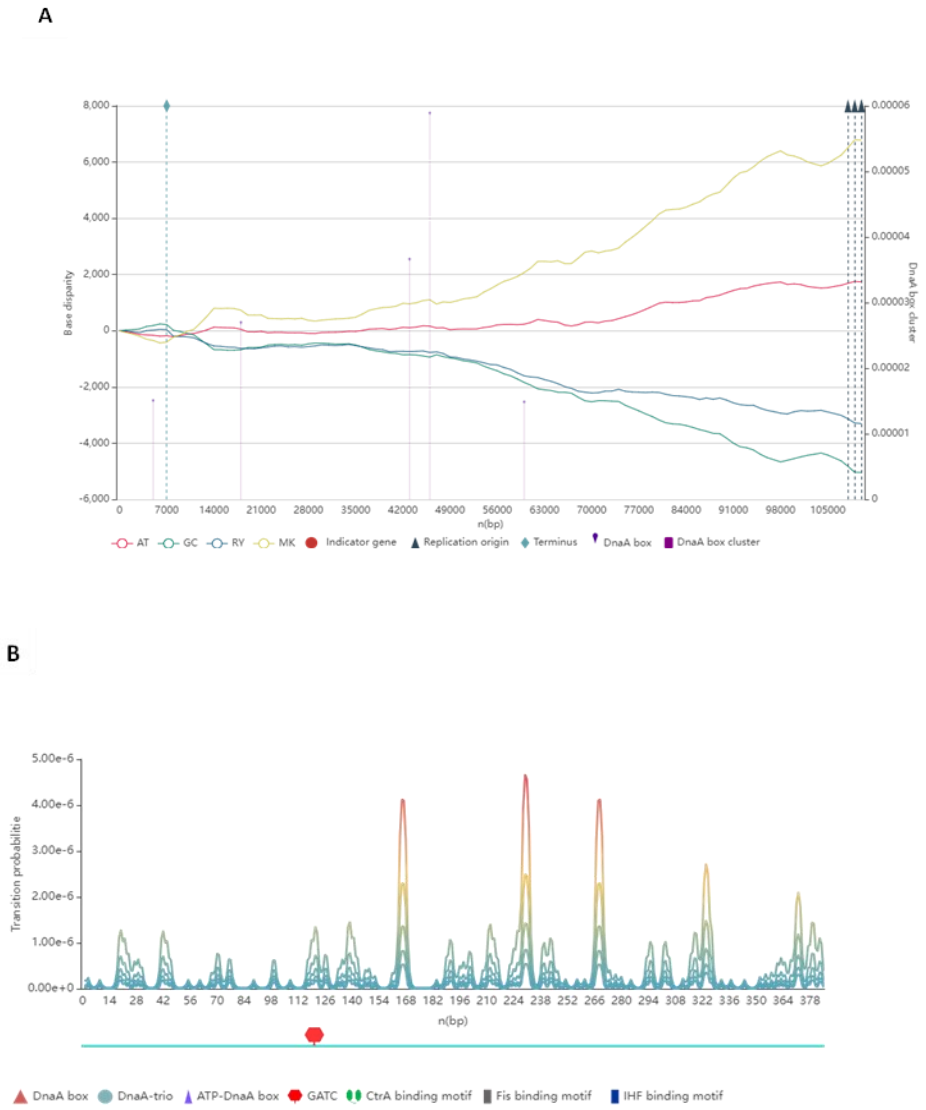


Figure 4-10. Identification of replichores in Contig 5 using Ori-Finder 2022.
(A) The interactive Z-curve figure of *M. komossense* JERR01chromosome Contig 5 indicating origin of replication (*ori*) and terminus (*ter*) in the circular contig. The red, green, blue and yellow line graphs indicate AT, GC, RY and MK disparity. (E) characteristic visualization of *oriC* sequence with a GATC element in the x-axis and three red peaks representing AT-rich regions.

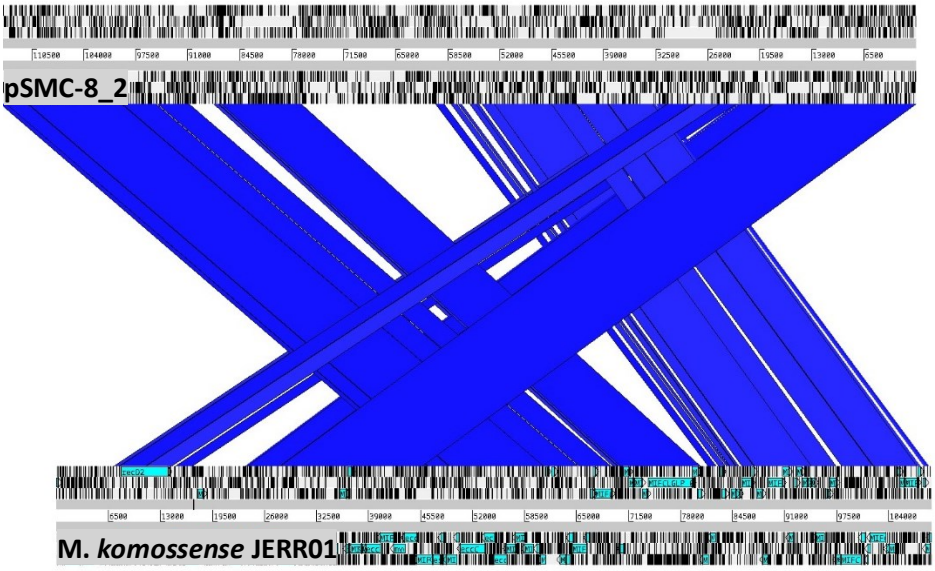


Figure 4-11. Pairwise comparison between Contig 5 and *Mycobacterium* spp. plasmid pSMC-8_2.
BLAST search comparison was performed between Contig 5 and the NCBI nucleotide database identifying a significant hit with *Mycobacterium* spp. SMC-8 plasmid pSMC-8_2 showing a query cover of 85% and a percentage identity of 96.02%. and visualised using ACT comparison tool.

4.3.3.4 Subsystem distribution of *M. komossense* JERR01 using KEGG.

With the annotated main chromosome and plasmids, a metabolic network was constructed. By performing this analysis, it is possible to predict the potential functional capabilities of the different components that make up the entire genome and determine if the plasmids identify in *M. komossense* JERR01 carry unique sets of genes such as antibiotic resistance factors compared to the chromosome.

Protein sequences in FASTA format were obtained from the chromosome, and the 4 different contigs and GhostKoala was used to assign ID numbers to each protein. The

percentage of proteins assigned to that subsystem relative to the total number of protein-coding sequences (CDSs) detected by KEGG was calculated. This approach using percentages allows for a standardized comparison of functional distribution across different genome components (chromosome and contigs). Since the total number of identified CDSs can vary between the chromosome and contigs, using percentage controls for these variations and provides a more accurate representation of the relative abundance of proteins associated with each functional category within each compartment (Dias *et al.*, 2018).

The results of this analysis revealed the presence of chromosomal encoded proteins across all KEGG generated subsystems as shown in Figure 4-12. Most of the KEGG annotated proteins of *M. komossense* JERR01 chromosome were assigned to the subsystem "Carbohydrate metabolism" (11.5%), followed by "Environmental information processing" (10.18%) subsystem, and the smallest number of KEGG annotated proteins from the chromosome (0.12 %) were assigned to "Biosynthesis of other metabolites". CDSs from Contig 2 are not as well represented across the different subsystems as the chromosome (Contig 2 present in 12 subsystems out of 23), with most of its annotated CDSs (21.95%) assigned to "Unclassified: genetic information processing" followed by "Protein families: signalling and cellular processes" (19.15%) and "Protein families: genetic information processing" and "Protein families: metabolism" with 12.19%. Contig 2 also has CDSs assigned to "Human diseases" (2.43%), suggesting that this Contig of *M. komossense* JERR01 possesses machinery that can be traced to pathogenic mycobacteria. Contig 3 (present in 7 subsystems out of 23) has its highest percentage of CDSs (57.14%) assigned to "Carbohydrates metabolism" and is the

only Contig apart from the *M. komossense* JERR01 chromosome that has CDSs assigned to this subsystem. Contig 3 also has 7.14% of its CDSs assigned to the following subsystems: “Unclassified: metabolism”, “Amino acid metabolism”, “Protein families: metabolism”, “Lipid metabolism”, “Metabolism of other amino acids”, and “Biosynthesis of other metabolites”. Contig 4 (present in 12 subsystems out of 23), has the highest distribution of its CDSs (19.51%) assigned to “Protein families: genetic information processing”, and “Protein families: signalling and cellular processes”. Contig 4 also has KEGG annotated CDSs assigned to “Human diseases” (2.43%) apart from the *M. komossense* JERR01 chromosome and Contig 2. Contig 5 (present in 7 subsystems out of 23) has its highest percentage of KEGG annotated CDSs (35.29%) to “Protein families: signalling and cellular processes”, followed by “Protein families: Genetic information processing” (17.64%) and “Unclassified: metabolism”, “Genetic information processing” and, “Unclassified: genetic information processing”, all three with 11.76% of CDSs assigned.

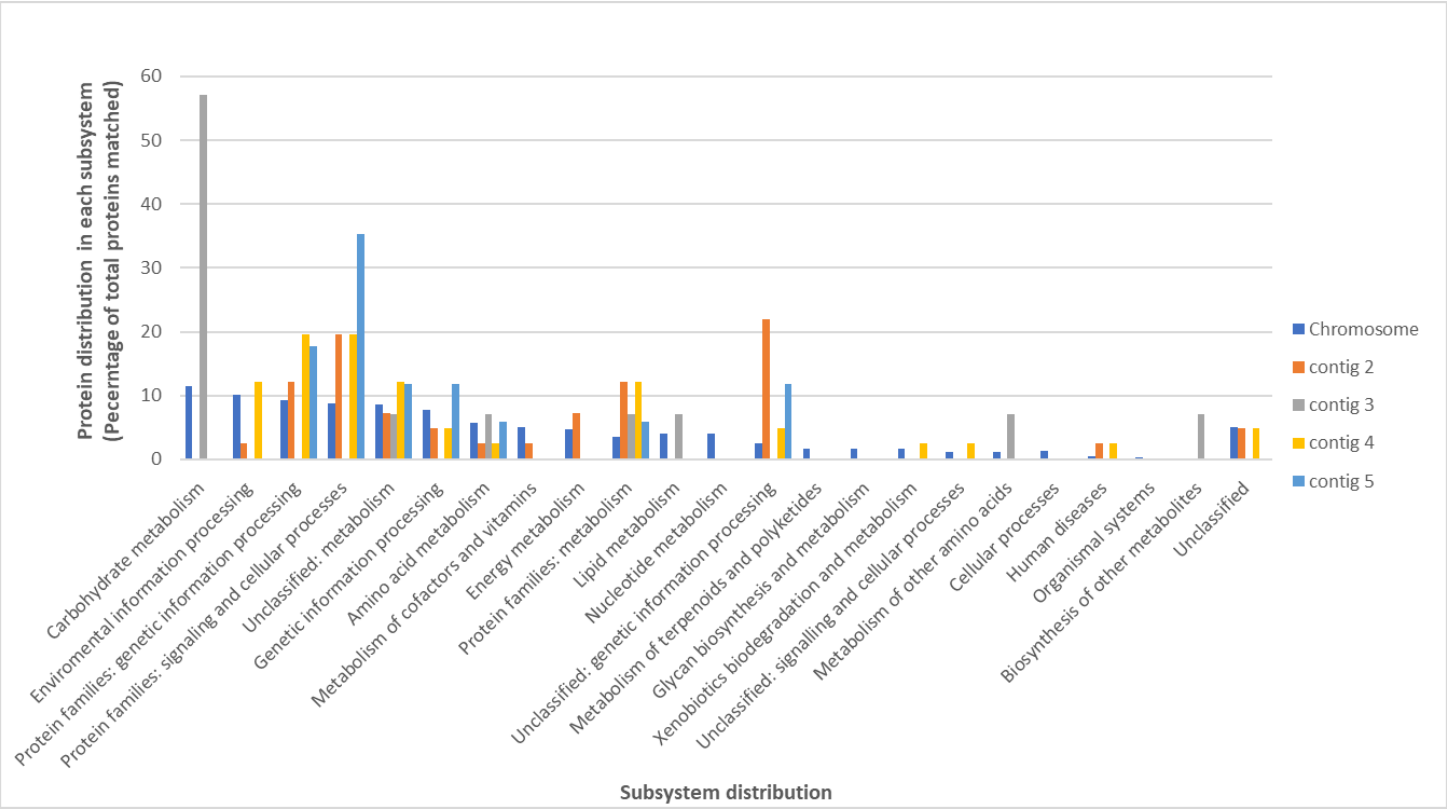


Figure 4-12. Distribution of proteins from the chromosome and predicted plasmids mapped to different subsystems using KEGG.
The Y-axis represents the percentages of mapped proteins in relation to the number of proteins annotated by KEGG. The X-axis represents the different KEGG subsystems. The different colours of the histogram bars represent the chromosome and the different contigs.

4.3.4 Comparative Genomics

4.3.4.1 *M. komossense* JERR01 and *M. komossense* ATCC 33013

Following the characterization of *M. komossense* JERR01s genetic components (chromosome and plasmids) and metabolic distribution of the annotated proteins in the previous section of this chapter, the analysis moved to comparative genomics with the reference strain *M. komossense* ATCC 33013. The genome assembly of *M. komossense* ATCC 33013 was also sequenced, and similarly to *M. komossense* JERR01, it was assembled, and its quality was assessed using QUAST and BUSCO. The reference strains had 6,844,114 bp in length and were predicted to contain 6,519 CDSs, 53 tRNAs, 1 tmRNAs, and 6 rRNAs. No plasmids were detected after the assembly of the *M. komossense* ATCC 33013 genome.

The first comparison analysis performed was by using the nucleotide FASTA files of both organisms to perform a relatedness analysis through the Average nucleotide identities (ANIs) computational analysis using OrthoAni (Lee *et al.*, 2016). This analysis specifically considers orthologous pairs that share a common evolutionary ancestor in order to measure the relatedness between genomes. The expected result from this analysis is a percentage numerical score between 0-100% representing the percentage of average nucleotide identity between the orthologous genes in the two genomes being compared. The percentage obtained as a result is then compared with the threshold for species demarcation which is 95-96% OrthoANI score (Lee *et al.*, 2016). Scores above this threshold suggest the genomes belong to the same species, while lower scores indicate increasing levels of divergence.

M. komossense JERR01 and *M. komossense* ATCC 33013 exhibit a moderate degree of nucleotide identity based on an ANI analysis, yielding a score of 75.13%. This value falls below the commonly accepted threshold (95-96%) for species demarcation using OrthoAni. Therefore, despite sharing some evolutionary relationship as evidenced by the obtained score, these results suggest that *M. komossense* JERR01 and *M. komossense* ATCC 33013 likely represent distinct species.

Building upon the initial ANI analysis results suggesting that *M. komossense* JERR01 and *M. komossense* ATCC 33013 are members of distinct species, a subsequent Pan-genome analysis using Panaroo provided a more in-depth comparison analysis (Tonkin-Hill *et al.*, 2020). While an ANI analysis looks over the overall content of the orthologous proteins and compares them providing a percentage similarity, a Pan-genome analysis reports the exact number of shared and unique proteins between the two organisms. To do so, Pan-genome software such as Panaroo requires the annotated protein files (*gff files) from each genome. These files are automatically generated when a Prokka annotation is performed in an organism.

The Pan-genome analysis revealed that while both strains share a substantial core genome (60.4% or 3612 proteins), a substantial number of unique genes were identified (2,373 in *M. komossense* JERR01 and 2,553 in *M. komossense* ATCC 33013; Figure 4-13). These results aligned with the initial assessment from the ANI analysis indicating diversity beyond the species level, confirming the presence of extensive sets of unique proteins beyond the core set of genes. The presence of a substantial number of unique proteins suggesting functional divergence between the strains, possibly due to

environments from which each of these strains was isolated from (*M. komossense* JERR01 from a human and *M. komossense* ATCC from Nordic moss). Therefore, combining the results from both analyses strengthens the conclusion that *M. komossense* JERR01 and *M. komossense* ATCC 33013 represent distinct species with potential functional differences driven by unique genes in their respective genomes.

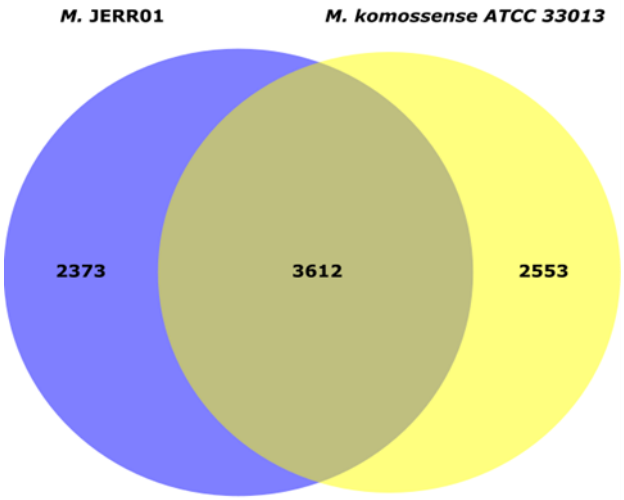


Figure 4-13. Shared and unique proteins comparison a between of *M. komossense* JERR01 and *M. komossense* ATCC 33013.
This is a proportional Venn diagram depicting the shared and unique number of proteins of *M. komossense* JERR01 and *M. komossense* ATCC 33013. Analysis performed using Panaroo, numbers represent the count of identified proteins in each region.

Followed these two analyses, a comparison of genome synteny was performed using the Artemis Comparison Tool (ACT). While the previous two analysis provided a quantitative analysis based on DNA and protein similarity of the genomes, ACT allows a visual representation of BLAST alignment of the genome sequences, allowing the identification of structural variations such as insertions, deletions, and rearrangements. The ACT comparison between *M. komossense* JERR01 and *M. komossense* ATCC 33013 (Figure

4-14) shows that whilst there is diversity in the BLAST match between the two sequences as illustrated by the gaps in the matches, there is evidence of conserved sequence identity and synteny across the whole sequence/chromosome.

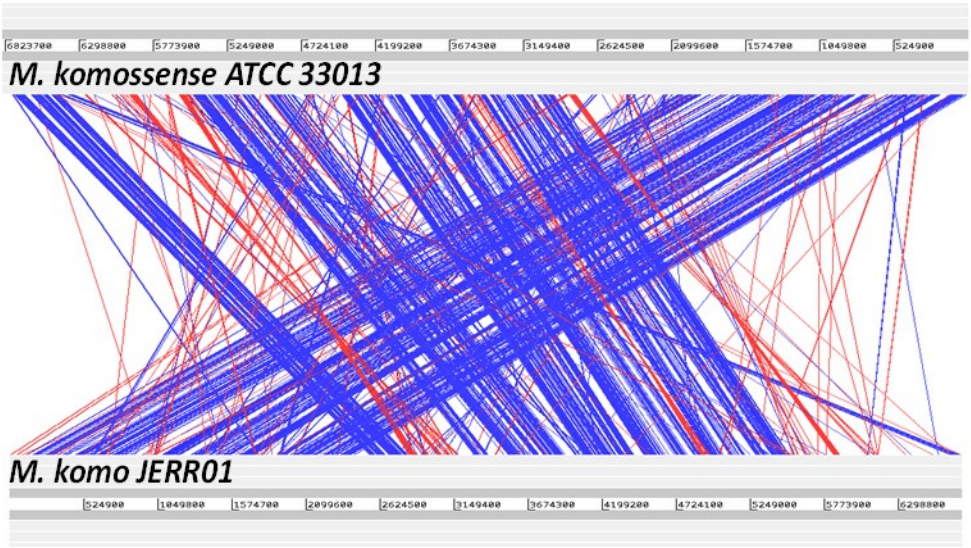


Figure 4-14. Pairwise comparison of DNA sequences of between of *M. komossense* JERR01 and *M. komossense* ATCC 33013.

Pairwise comparison of Chromosome sequences of *M. komossense* JERR01 and *M. komossense* ATCC 33013 genomic structures displayed in ACT. Red and blue bars indicate matches in the same orientation (red) and reverse orientation (blue), and regions of uniqueness (white).

Analysis of the protein distribution across KEGG subsystems between the two bacterial genomes, provides an insight into the functional specializations of *M. komossense* JERR01 and *M. komossense* ATCC 33013. This analysis was performed by using the protein FASTA files of both organisms and using GhostKoala to investigate potential functional divergence between *M. komossense* JERR01 and *M. komossense* ATCC 33013. To determine the proportional representation, the number of proteins in each category

was divided by the total number of proteins per strain and multiplied by 100%. This data was then visualized as a histogram, revealing the distribution of proteins across different KEGG subsystems for each strain.

KEGG subsystem analysis revealed subtle differences in the functional distribution of proteins between *M. komossense* JERR01 and *M. komossense* ATCC 33013 (Table 3 and Figure 4-15). Notably, the *M. komossense* ATCC 33013 genome possesses a slightly higher proportion of proteins associated with carbohydrate metabolism (13.4% vs. 11.1%), Amino acid metabolism (6.9% vs 5.8%), Metabolism of cofactors and vitamins (5.2% vs 4.7%). Conversely, *M. komossense* JERR01 possessed a higher proportion of proteins assigned to Environmental information processing (9.9% vs 8.9%), Protein families: genetic information processing (9.6% vs 8.9%), Unclassified: metabolism (8.7% vs 8.0%), Genetic information processing (8.5% vs 7.8%), Protein families: signalling and cellular processes (7.7% vs 7.2%) and Unclassified: genetic information processing (3.7% vs 2.3%). These findings illustrate the functional divergence between these strains and adding more evidence to the fact that these two mycobacteria do not belong to the same species.

Table 4-3. Percentages of proteins present per KEGG subsystem distribution in *M. komossense* JERR01 and *M. komossense* ATCC 33013.

Subsystem distribution	Protein (%) *M. komossense* JERR01	Protein (%) *M. komossense* ATCC 33013
Carbohydrate metabolism	11.1	13.4
Environmental information processing	9.9	8.9
Protein families: genetic information processing	9.6	8.9
Unclassified: metabolism	8.7	8.0
Genetic information processing	8.5	7.8
Protein families: signalling and cellular processes	7.7	7.2
Amino acid metabolism	5.8	6.9
Metabolism of cofactors and vitamins	4.7	5.2
Energy metabolism	4.6	5.1
Lipid metabolism	3.8	3.8
Protein families: metabolism	3.8	4.2
Nucleotide metabolism	3.7	3.4
Unclassified: genetic information processing	3.7	2.3
Xenobiotics biodegradation and metabolism	1.5	2.9
Glycan biosynthesis and metabolism	1.7	1.8
Cellular processes	1.6	1.7
Metabolism of terpenoids and polyketides	1.1	1.0
Unclassified: signalling and cellular processes	1.2	1.1
Metabolism of other amino acids	1.3	0.4
Human diseases	0.4	0.8
Organismal systems	0.3	0.3
Biosynthesis of other secondary metabolites	0.1	0.2
Unclassified	5.1	5.0

Percentage of proteins present per subsystem distribution present in each genome.

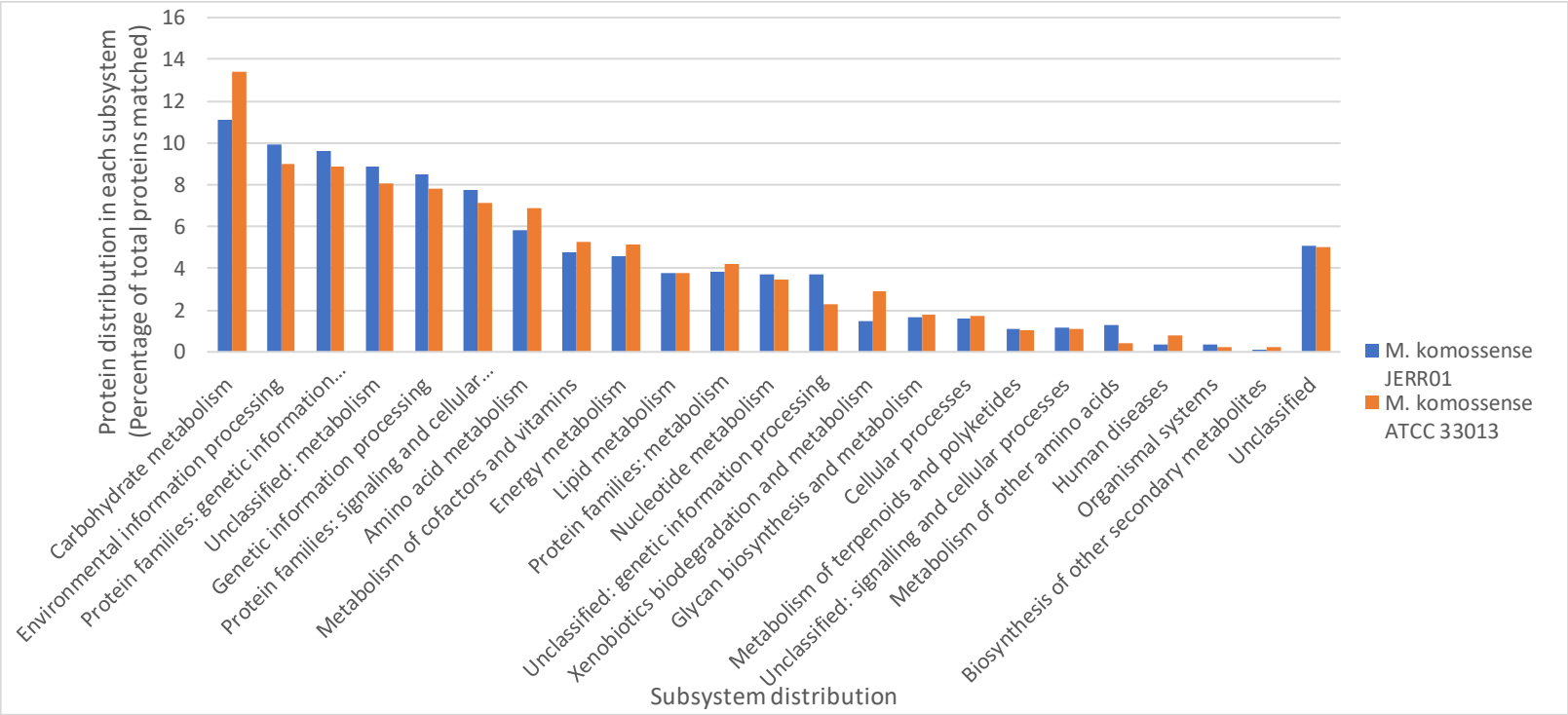


Figure 4-15. KEGG subsystem distribution of *M. komossense* JERR01 and *M. komossense* ATCC 33013 genomes.
This histogram depicts the distribution of functional proteins in relation to KEGG pathways between the whole genome of *M. komossense* JERR01 and the reference strain *M. komossense* ATCC 33013. The analysis, based on protein FASTA files processed by GhostKoala and the KEGG database, employed percentages instead of raw gene encoding protein counts. Each bar represents the difference in the proportion of proteins associated with each subsystem distribution between the two strains.

4.3.5 Comparison of *M. komossense* JERR01 to other *Mycobacterium* species.

Comparative genomic analysis of *M. komossense* JERR01 and *M. komossense* ATCC 33013 highlighted differences indicating that these two organisms are more distantly related than would be typical of members of the same species. To discover where *M. komossense* JERR01 was taxonomically positioned and to determine its closest mycobacterial relatives, a comparison with species from each of the major *Mycobacterium* clades was performed. This ensures a comprehensive analysis by incorporating representatives from the various evolutionary clades within the *Mycobacterium* genus (Gupta, Lo and Son, 2018).

These 20 species complete genomes were obtained from the National Center for Biotechnology Information (NCBI) database. Selection criteria focused on complete mycobacterial genomes with high quality genome assemblies. Preference was given to well-characterized strains that have been already published in other studies, including human pathogens and environmental isolates to capture broader diversity within the *Mycobacterium* genus (See chapter 2.14.7)**.**

An average nucleotide identity (ANI) analysis was performed between *M. komossense* JERR01 and all 20 mycobacterial species to identify closely related species. The analysis revealed the following closest matches was *M. vanbaalenii* PYR-1 (83.32%) followed by *M. vaccae* ATCC 95051 (82.45%), *M. gilvum* Spyr1 (80.9%), and *M. aurum* NCTC 10437 (80.32%) (Table 4.4). These results indicate that on a DNA:DNA identity level *M. komossense* JERR01 is taxonomically close to members of the *Fortuitum-Vaccae* clade, with *M. vanbaalenii* PYR-1 being its closest relative among the 20 genomes examined.

A Pan-genome analysis was used to identify the core genes shared among all 20 species, hypothesising that closely related species will have a greater number of shared genes. As with the ANI analysis, *M. komossense* JERR01 shared the highest percentage of its genes (70.7%, 4,233 genes) with *M. vanbaalenii* PYR-1, followed by *M. vaccae* ATCC 95051 (67.1%, 4,014 genes) and *M. gilvum* Spyr1 (66.03%, 3,952 genes), all members of the *Fortuitum-Vaccae* clade. By contrast, strains belonging to the other clades exhibited lower ANI values and shared fewer shared genes with *M. komossense* JERR01 compared to members of the *Fortuitum-Vaccae* clade (Table 4.4). The results provided by the pan-genome analysis strengthen the results provided by the ANI analysis highlighting the closer genetic relationship with *M. komossense* JERR01 and members of the *Fortuitum-Vaccae* clade.

Since the following species: *M. vanbaalenii* PYR-1, *M. vaccae* ATCC 95051, *M. gilvum* Spyr1, and *M. aurum* NCTC 10437, were considered close relatives to *M. komossense* JERR01 (percentage similarity above 80%) based on the results produced by the ANI and Pan-genome analysis, a pairwise blastn comparisons of the chromosomes of these mycobacteria to *M. komossense* JERR01 were undertaken. ACT comparisons of the blastn results revealed a high degree of synteny between the *M. komossense* JERR01 DNA and the environmental mycobacteria sequences (Figure 4-16), with 74-100% identity across their entire DNA sequence. The comparison between *M. komossense* JERR01 with *M. vaccae* ATCC 95051 demonstrated continuous synteny across the whole sequences, albeit with discreate regions of difference peppering the chromosomes. The alignment *M. gilvum and M. komossense* JERR01 identified an inversion (blue hourglass section) that straddled the terminus of replication. Smaller inversions were also

observed in the comparison between *M. komossense* JERR01 and *M. vanbaalenii* PYR-1 and *M. aurum* NCTC 10437. The high levels of nucleotide conservation and synteny observed across the chromosomes supported the close relationship of *M. komossense* JERR01 with the species that are representatives of the *Fortuitum-Vaccae* clade.

While the results provided by OrthoAni, Panaroo, and ACT indicate a close evolutionary relationship, this does not necessarily indicate close genomic functional relationships. A subsystem analysis of the functional distribution analysis among highlighted that carbohydrate, amino acid, and energy metabolism were prominent in the species; although a trend is followed, these presented different levels of enrichment in the different subsystems, suggesting adaptations to specific niches. Perhaps the trend represents core functions, while variations point to adaptations for specific environments. The presence of a large number of proteins in “Unclassified” category highlights a portion of genes with unknown functions. Novel species like *M. komossense* JERR01 and the other members of the *Fortuitum-Vaccae* clade have higher percentages in this category, indicating potentially unique genes or those requiring further characterization as shown in Figure 4-17. Subsystems like "Xenobiotics biodegradation and metabolism" and "Metabolism of terpenoids and polyketides" have a small proportion of proteins and these subsystems might be indicative of adaptations to specific environmental niches.

Genome	Average nucleotide identity (%)	Number of shared genes	Number of unique genes	Mycobacteria clade	Growth	Pathogen	Intrinsically host restricted?
***M. vanbaalenii* PYR-1**	83.3	4233	1752	*Fortuitum-Vaccae*	Rapid	No	No
***M. aurum* NCTC 10437**	80.3	4069	1916	*Fortuitum-Vaccae*	Rapid	No	No
***M. gilvum* Spyr1**	80.9	3952	2033	*Fortuitum-Vaccae*	Rapid	No	No
***M. smegmatis* MC2 155**	76.6	3897	2088	*Fortuitum-Vaccae*	Rapid	No	No
***M. vaccae* ATCC 95051**	82.5	3743	2242	*Fortuitum-Vaccae*	Rapid	No	No
***M. komossense* ATCC 33013**	75.1	3613	2372	*Fortuitum-Vaccae*	Rapid	No	No
***M. fortuitum* CT6**	76.1	3588	2397	*Fortuitum-Vaccae*	Rapid	Yes	No
***M. avium* subsp. *hominissuis* OCU889s_P11-4s**	75.7	3181	2804	*Tuberculosis-Simiae*	Slow	Yes	No
***M. kansasii* ATCC 12478**	74.2	3146	2839	*Tuberculosis-Simiae*	Slow	Yes	No
M. marinum E11	73.8	3109	2876	*Tuberculosis-Simiae*	Slow	Yes	No
***M. kumamotonensis* DMS 45093**	75.9	3056	2929	*Terrae*	Slow	Yes	No
***M. terrae* NCTC 10856**	75.2	3004	2981	*Terrae*	Slow	Yes	No

***M. algericus* JCM 30723**	75.2	3001	2984	*Terrae*	Slow	Yes	No
***M. hiberniae* JCM 13571**	75.9	2990	2995	*Terrae*	Slow	Yes	No
***M. arupensis* DSM 44942**	73.3	2951	3034	*Terrae*	Slow	Yes	No
***Mtb* H37Rv**	74.1	2802	3183	*Tuberculosis-Simiae*	Slow	Yes	Yes
***M. abscessus* GZ002**	72.7	2709	3276	*Abscessus-Chelonae*	Rapid	Yes	No
***M. canettii* CIPT 140010059**	74.2	2704	3281	*Tuberculosis-Simiae*	Slow	Yes	Yes
***M. salmoniphilum* DSM 43276**	72.4	2628	3357	*Abscessus-Chelonae*	Rapid	Yes	Yes
***M. leprae* MRHRU-235-G**	71.3	1432	4553	*Tuberculosis-Simiae*	Slow	Yes	Yes

Table 4-4. Comparative genomic content of *M. komossense* JERR01 and representative *Mycobacterium* species:
Shared and unique genes. This table summarizes the number of genes shared between *M. komossense* JERR01 and 20 other mycobacterial species identified by the Pan-genome analysis, alongside the number of genes unique to each of these species. Additionally, information on the species clade affiliation, growth rate, pathogenicity and host restriction were also described.

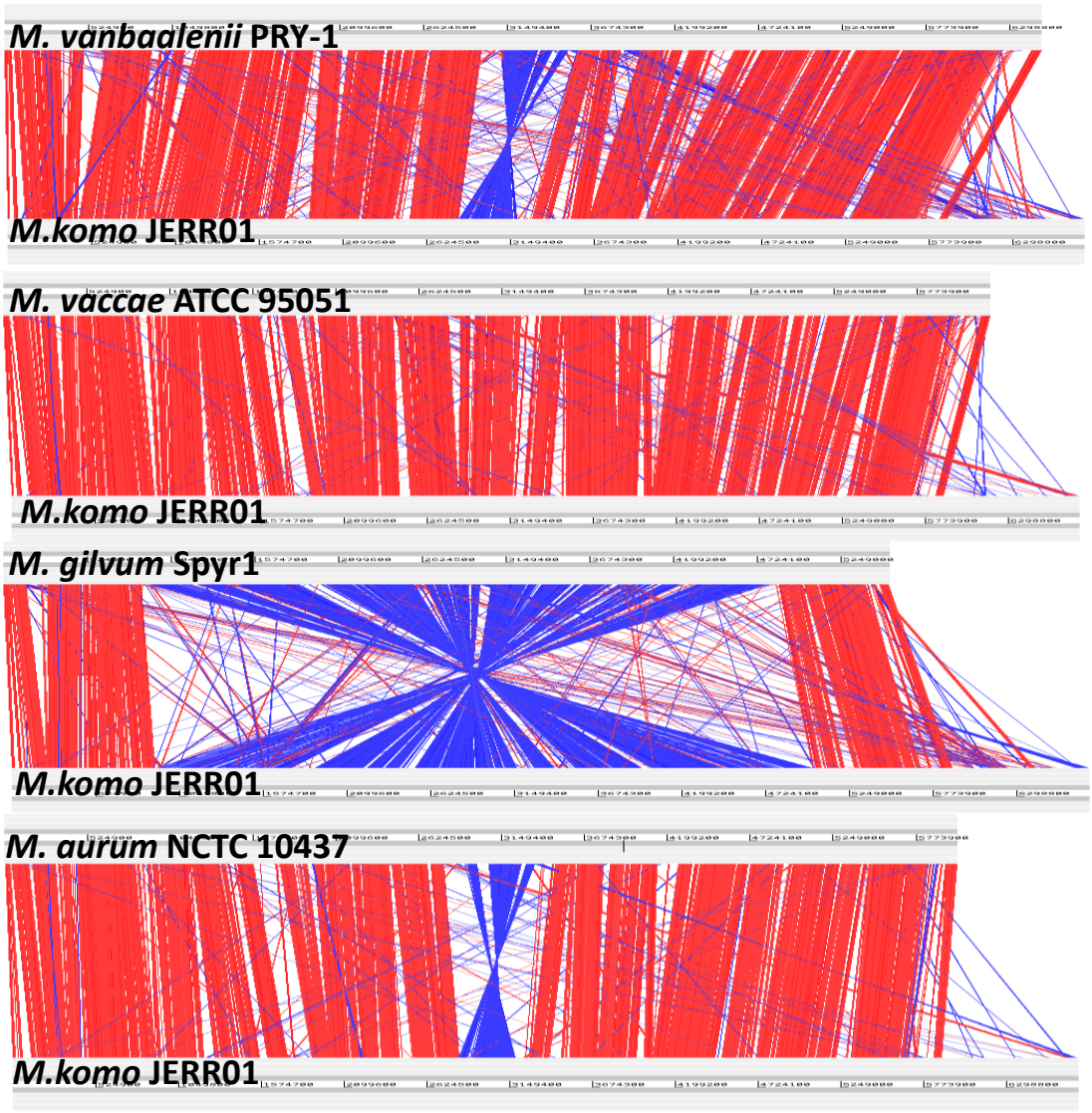


Figure 4-16. Pairwise genome comparison between *M. komossense* JERR01, *M. vanbaalenii* PRY-1, *M. vaccae* ATCC 95051, *M. gilvum* Spyr1 and *M. aurum* NCTC 10437. The Artemis Comparison Tool (ACT) was used to display pairwise blastn comparisons of the genomes of *M. komossense* JERR01 against four environmental mycobacteria. Genome sequences were aligned from the predicted *oriC* and visualised in ACT with a cut-off set to blast scores >600. Red and blue bars indicate matches in the same orientation (red) and reverse orientation (blue), and regions of uniqueness (white).

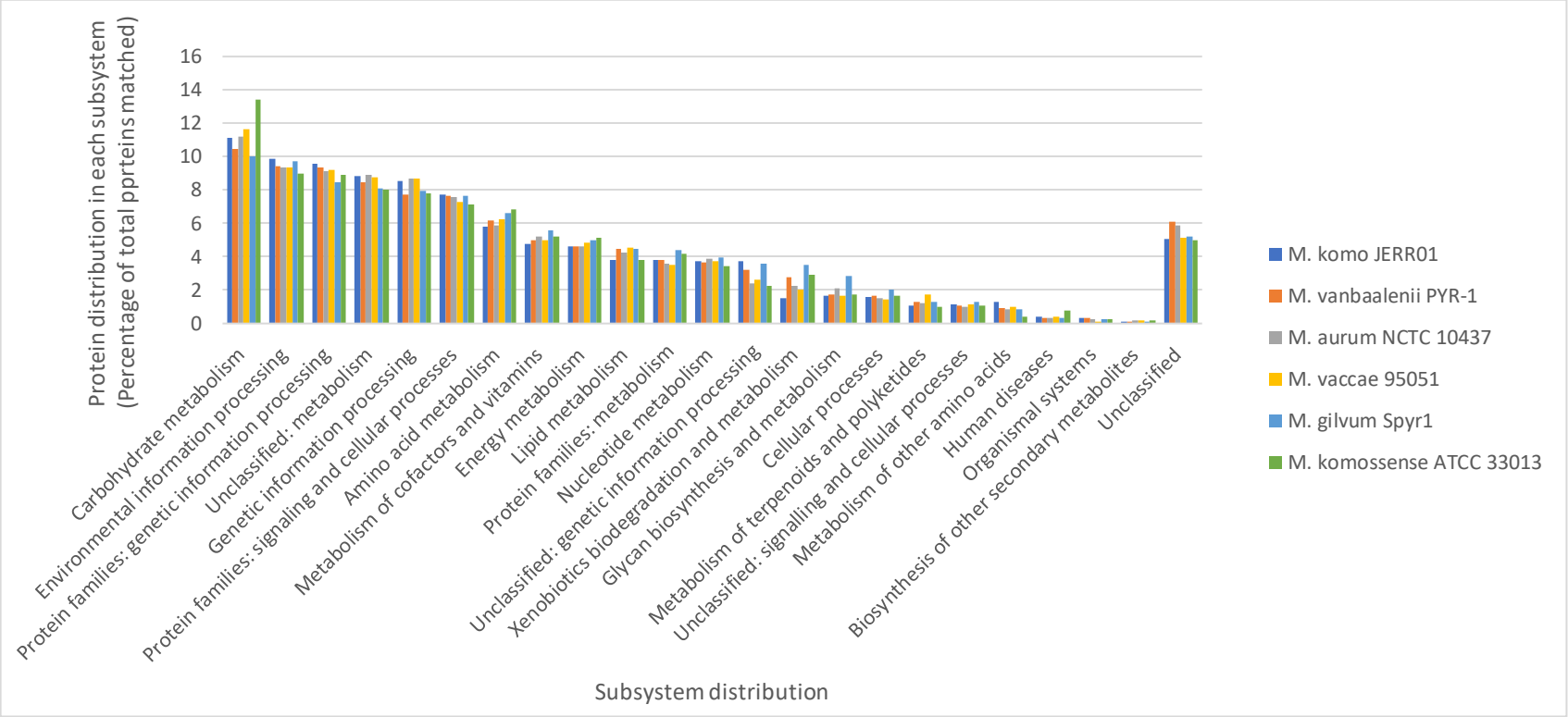


Figure 4-17. Difference in KEGG subsystem distribution between *M. komossense* JERR01, *M. komossense* ATCC 33013 and the four closest 4 environmental mycobacteria.

Distribution of subsystems in *M. komossense* JERR01, *M. vanbaalenii* PYR-1, *M. aurum* NCTC 10437, *M. vaccae* ATCC 95051, *M. gilvum* Spyr1 and *M. komossense* ATCC 33013. Functional subsystems were identified using KEGG. The y-axis represents the percentage of proteins assigned to each subsystem relative to the total number of entries detected by KEGG and the x-axis has the different KEGG subsystem distribution. Protein FASTA files were used for analysis, and KEGG IDs were added using GhostKoala.

Using the genomes of the 20 representative species *M. komossense* JERR01 Pan-genome analysis was undertaken to identify core genes and accessory genes. Figure 4.18 Illustrates a presence/absence matrix of core and accessory genes (20,384 genes clusters) ordered according to a core gene phylogeny. The phylogenetic tree (Figure 4-19) positions *M. komossense* JERR01 with members of the *Fortuitum-Vaccae* clade of fast-growing mycobacteria, close to *M. vanbaalenii* PYR-1 and *M. vaccae* ATCC 95051 and more distant from *M. komossense* ATCC 33013, also a member of the *Fortuitum-Vaccae* clade (Figure 4-19). Slow-growing mycobacteria such as *Mtb* H37Rv and *M. leprae* MRHRU-235-G were clustered in a distant clade which also encompassed other members of the *Tuberculosis-Simiae* clade, as well as members of the *Abcessus-Chelonae* clade and the *Terrae* clade.

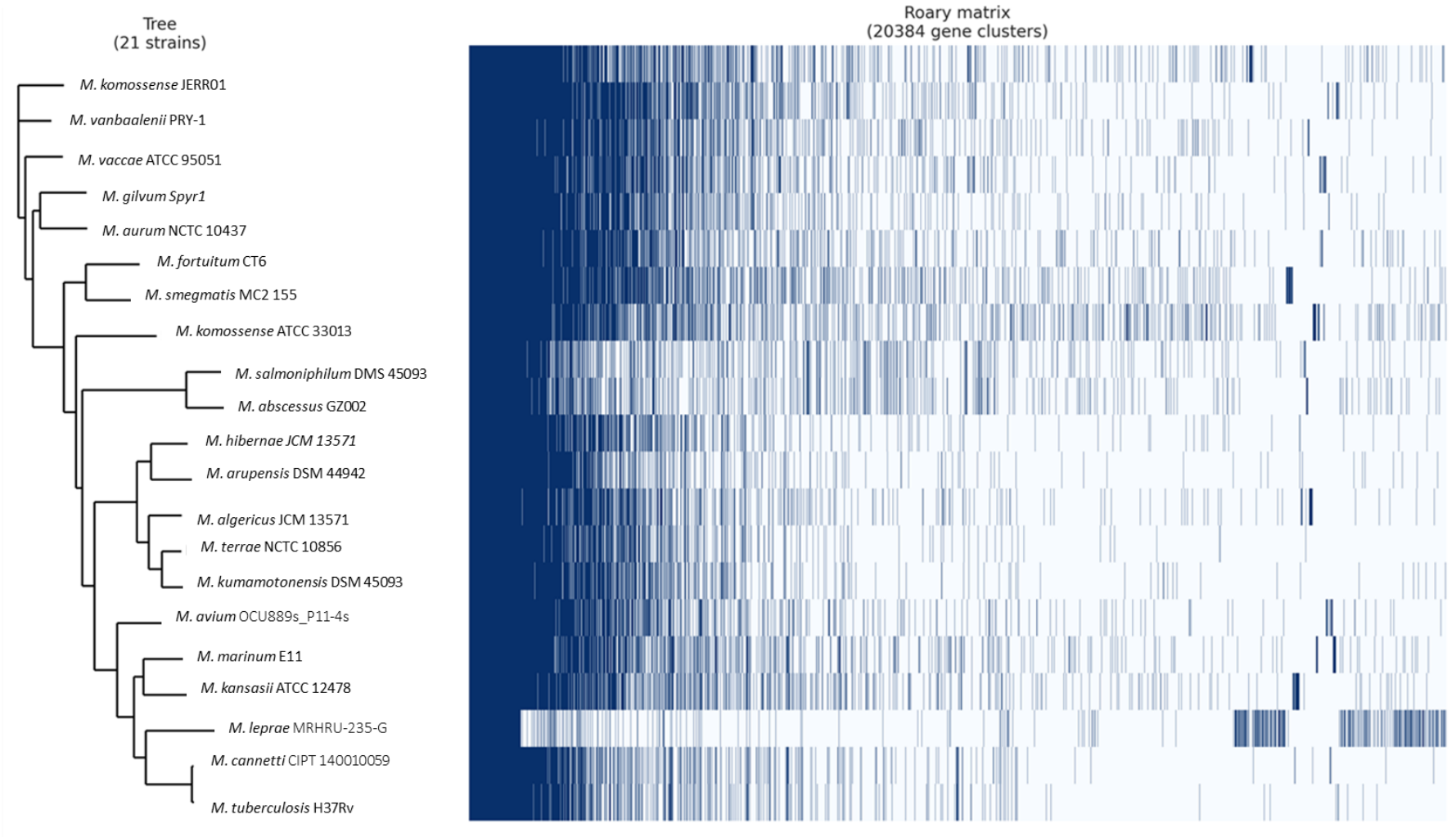

Tree
(21 strains)
M. komossense JERR01
M. vanbaalenii PRY-1
M. vaccae ATCC 95051
M. gilvum Spyr1
M. aurum NCTC 10437
M. fortuitum CT6
M. smegmatis MC2 155
M. komossense ATCC 33013
M. salmoniphilum DMS 45093
M. abscessus GZ002
M. hibernae JCM 13571
M. arupensis DSM 44942
M. algericus JCM 13571
M. terrae NCTC 10856
M. kumamotonensis DSM 45093
M. avium OCU889s_P11-4s
M. marinum E11
M. kansasii ATCC 12478
M. leprae MRHRU-235-G
M. cannetti CIPT 140010059
M. tuberculosis H37Rv
Roary matrix
(20384 gene clusters)

Figure 4-18. Pan-genome analysis of *M. komossense* JERR01 and representative *Mycobacterium* species.
Heatmap displaying the presence/absence of genes in the Pan-genome of *M. komossense* JERR01 and 20 other selected *Mycobacterium* species. Blue lines indicate gene presence, while white lines indicate absence. Darker colours in the heatmap represent on the left of the graph counts for all the genes present in their corresponding genomes, while lighter colours indicate their absence. The phylogenetic tree presented was built by using Panaroo output of all the genes used for the Pan-genome (gene_presence_gene_absence file) analysis (20,384 genes total detected by Panaroo) and utilizing the roary_plots_py script available within the Roary package.

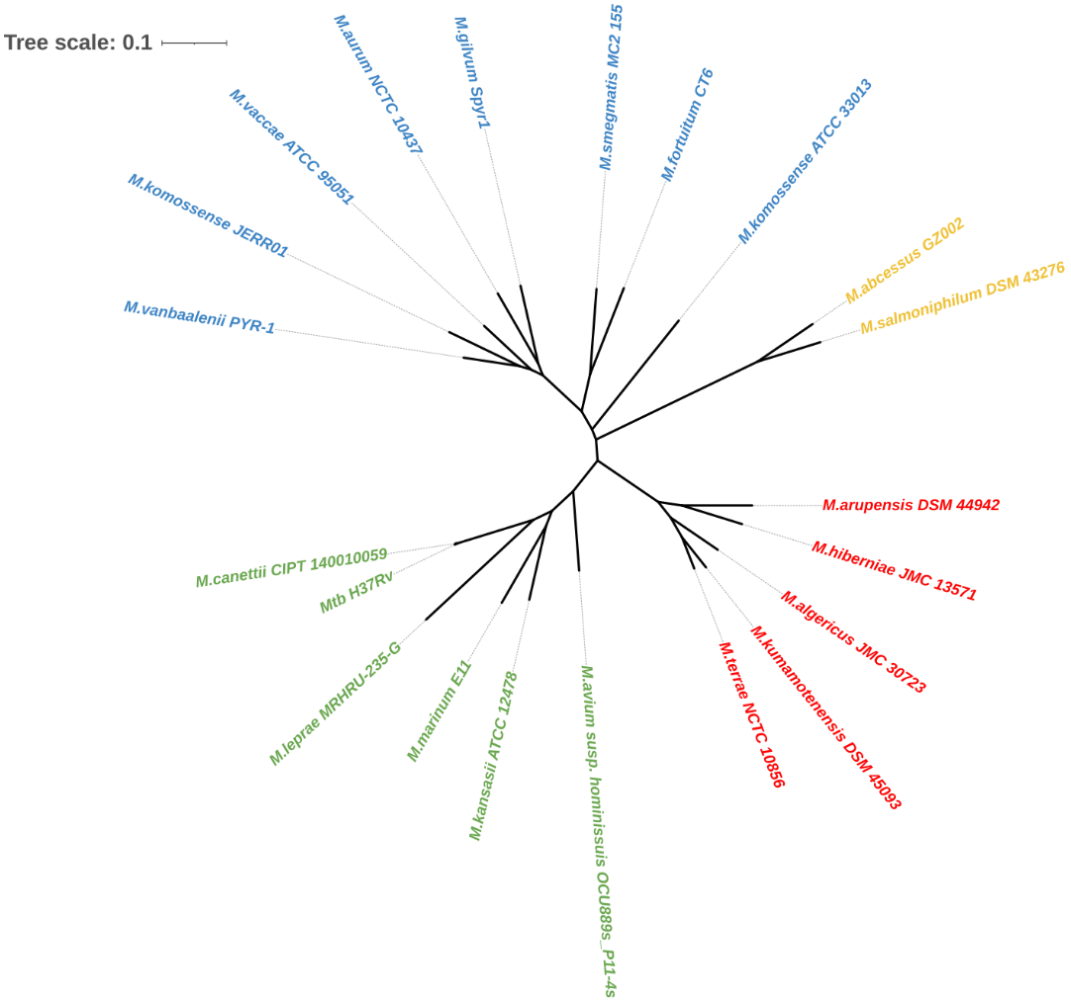


Figure 4-19. Core genome phylogeny of *M. komossense* JERR01 and representative *Mycobacterium* species.

The core genome alignment file generated using Panaroo was used to generate this phylogenetic tree. Panaroo used CLUSTAL for the alignments of the different genes using only the core genome and IQ-TREE was used to calculate phylogenetic the tree with 1,000 ultrafast bootstraps. The tree was mid-rooted and edited using Interactive Tree of Life iTOL v6.

4.3.6 Comparison between *M. komossense* JERR01 and *Mtb* H37Rv

The comparative genomic analysis undertaken facilitated a more detailed comparison between *M. komossense* JERR01 and *Mtb* H37Rv. In particular, it allowed for a detailed investigation of the conservation of loci linked to antibiotic resistance *Mtb* to help assess the suitability *M. komossense* JERR01 as a model organism studying drug resistance to anti-tuberculous antibiotics. Before focusing on this, it is relevant to reflect on the picture of the overall related that the whole genes provide.

As previously presented, at the DNA: DNA level *Mtb* H37Rv exhibit low level of identity (74.11%; Table 4-5) with *M. komossense* JERR01, and similarly a relatively low level of core gene conservation (2,802 shared proteins; Figure 4-20). This is indicative of the genus level taxonomic relatedness and sharing of some core functions, but also highlights the genetic divergence that exists between the two. Similarly, a comparison of genomes (Figure 4-21) demonstrates regions of difference across the chromosomes and chromosomal rearrangement, with weaker synteny across the chromosome, as was previously observed in the comparisons with the representatives of the *Fortuitum-Vaccae* clade.

Functionally, *M. komossense* JERR01 exhibits notable diversity in comparison to *Mtb* H37Rv (Figure 4-22). In the Signalling and cellular processes category *M. komossense* JERR01 has almost double the percentage (7.8%) of its genes within this category compared to *Mtb* H37Rv (4.4%). This highlights the broader regulatory capacity that *M. komossense* JERR01 possesses in signal transduction mechanisms and the regulation of cellular processes, which suggests that the strain has a greater capacity to detect and

respond to stimuli and stressors that may be found in diverse environments. By contrast, *Mtb*, is host restricted and therefore, its regulatory requirements are more constrained due to the close association with the host. Conversely, *Mtb* H37Rv dedicates a larger percentage (6.5%) of its genes to Metabolism of cofactors and vitamins compared to *M. komossense* JERR01 (4.8%). This might indicate variations in their requirements or utilization of specific organic compounds. In the subsystem of Protein families: metabolism, *M. komossense* JERR01 shows a lower percentage (3.8%) compared to *Mtb* H37Rv (5.7%). This could indicate the presence of unique metabolic pathways or variations in protein functions within *Mtb* H37Rv.

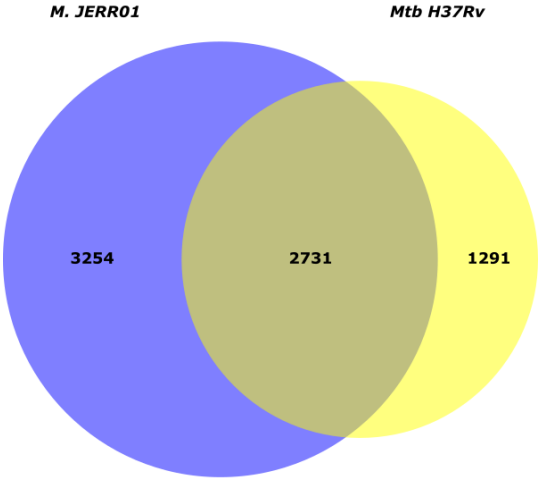


Figure 4-20. Shared and unique proteins comparison between of *M. komossense* JERR01 and *Mtb* H37Rv.
Proportional Venn diagram depicting the shared (2,731) and unique number (3,254 for *M. komossense* JERR01 and 1,291 for *Mtb* H37Rv) of proteins of *M. komossense* JERR01and *Mtb* H37Rv.

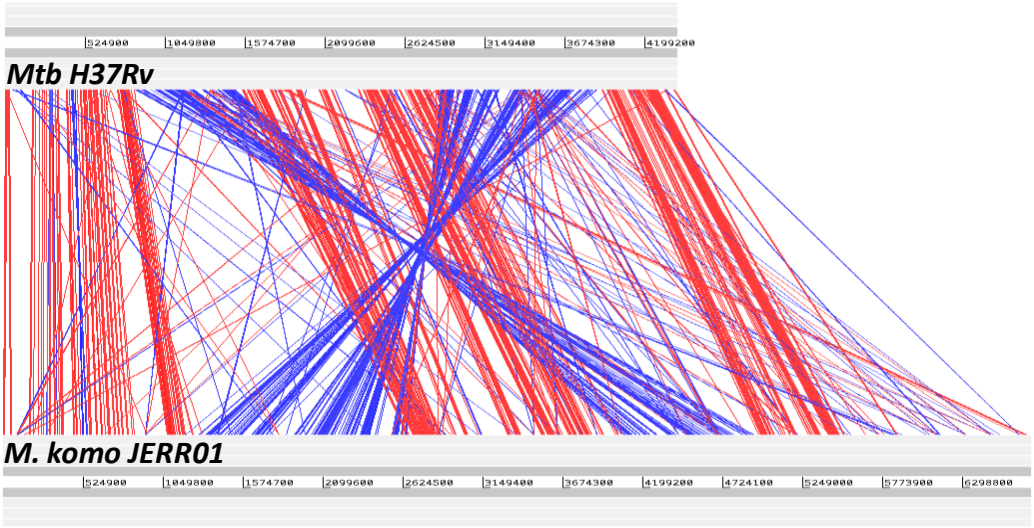


Figure 4-21. Pairwise genome comparison between *M. komossense* JERR01 and *Mtb* H37Rv.
The Artemis Comparison Tool (ACT) was used to display pairwise blastn comparisons of the genomes of *M. komossense* JERR01 against *Mtb* H37Rv. Genome sequences were aligned from the predicted *oriC* and visualised in ACT with a cut-off set to blast scores >600. Red and blue bars indicate matches in the same orientation (red) and reverse orientation (blue), and regions of uniqueness (white).

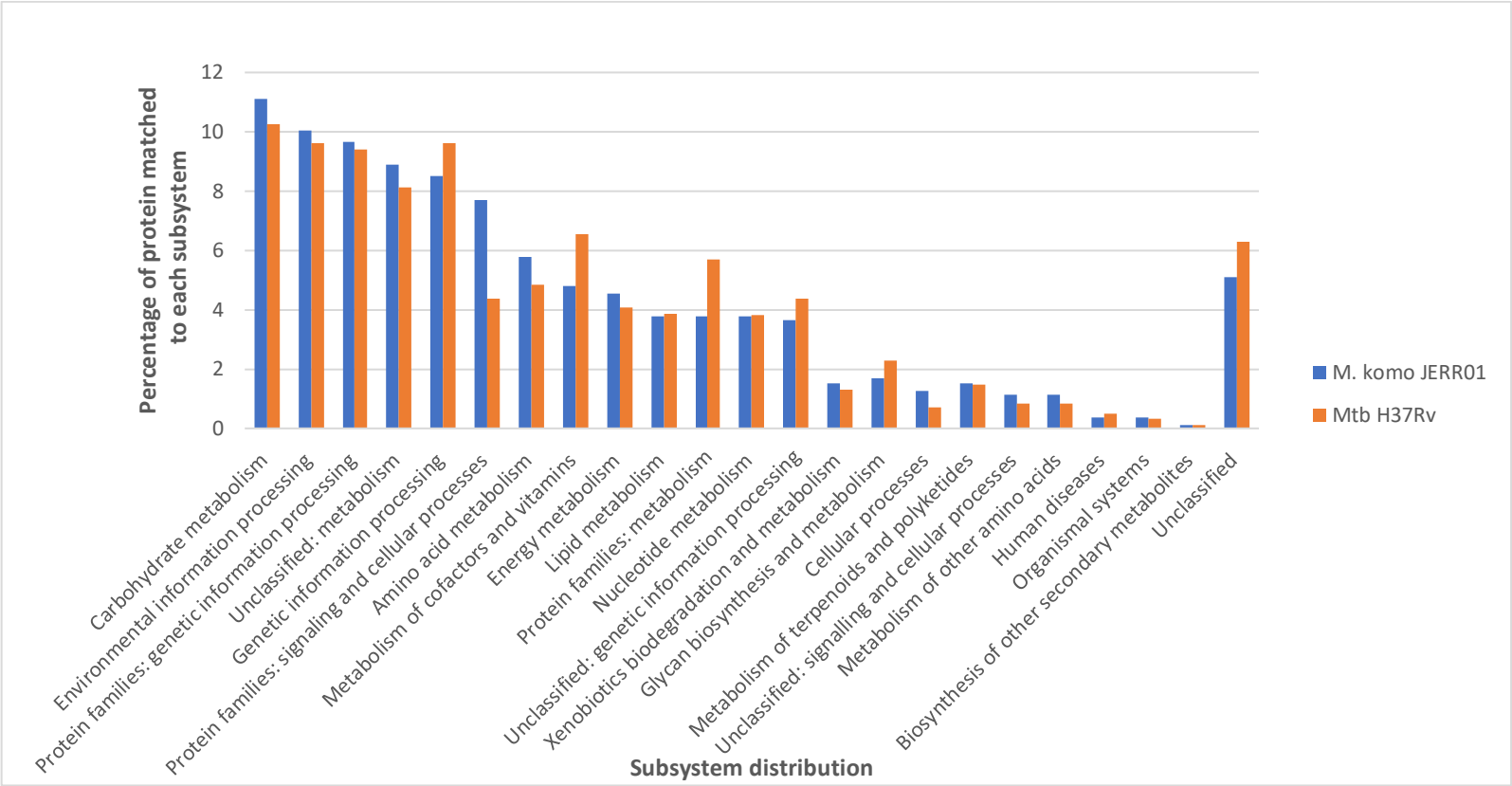


Figure 4-22. Comparative KEGG subsystem distribution of *M. komossense* JERR01 and *Mtb* H37Rv proteins.
This histogram demonstrates the comparative presence/absence of proteins within KEGG subsystem distribution for *M. komossense* JERR01 and *Mtb* H37Rv. KEGG IDs were assigned to protein sequences using GhostKOALA. The y-axis shows the percentage of proteins mapped to each subsystem, while the x-axis lists the individual subsystem. This comparative analysis reveals potential differences in the functional profiles of the two mycobacterial species based on their KEGG distribution.

4.3.6.1 Comparison between *M. komossense* JERR01 and *Mtb* H37Rv antimicrobial resistance (AMR) related genes.

As illustrated above, *M. komossense* JERR01 and *Mtb* H37Rv are divergent members of the genus, however, the phenotypic characterization of *M. komossense* JERR01 that was presented in the previous results chapter highlighted the common susceptibilities of antituberculosis antibiotics. This, therefore, raises the questions as to whether the two organisms share the same sensitives to these antibiotics, and if they also share the same routes to resistance via shared genetic determinants.

Comparative genomic analyses of *M. komossense* JERR01 and *Mtb* H37Rv were conducted looking for genes that have been demonstrated to confer susceptibility to clinically relevant antibiotics in *Mtb*. For this, the WHO catalogue of antibiotic resistance mutations was used as the reference database (World Health Organization, 2021). Homologs in *M. komossense* JERR01 of the WHO *Mtb* genes were identified by using blastp. From these, global alignments of the protein sequences were performed to calculate the percentage identity of the matches using the Needleman-Wunch algorithm. A high degree of sequence identity in the homologs suggests a shared function and therefore, a similar potential role in the mechanisms of AMR (Table 4.5).

In some cases, *M. komossense* JERR01 possessed more than one gene with similarity to *Mtb* AMR genes. In these cases, ACT comparison of the gene matches and the surrounding regions were used to distinguish orthologs from the paralogs. If an *M. komossense* JERR01 gene exhibited a high percentage of similarity identity to its *Mtb* H37Rv counterpart, and the surrounding genes presented a similar syntenic

organization, then, this gene would be suggested to have a strong orthologous relationship, indicating that these genes shared a common ancestor and possibly similar functions. Conversely, if the AMR gene in *M. komossense* JERR01 presented a lower percentage identity to its *Mtb* H37Rv gene counterpart, and the surrounding genes exhibited different structures and annotations, this suggested a paralogous relationship. The presence of these paralogous genes could be the result of gene duplication events within the *M. komossense* JERR01 genome or horizontal gene acquisition.

From the WHO catalogue, almost all of the AMR-related genes in *Mtb* H37Rv were found in *M. komossense* except for the gene *ahpC*. This gene encodes a subunit of the alkyl hydroperoxide reductase enzyme. This enzyme plays a crucial role in the bacterial defense mechanism against oxidative stress. In Chapter 3, the MIC_{90} testing of *M. komossense* JERR01 showed a high sensitivity of this organism towards 6 antitubercular drugs (RIF, EMB, MOXI, LEVO, BDQ, and LZD). In the case of INH, PZA, and PA, although genes targeted by these drugs were found, *M. komossense* JERR01 presented intrinsic resistance towards these antibiotics, suggesting the possibility of additional resistance mechanisms contributing to the observed resistance towards these antibiotics.

Table 4-5. Comparative analysis of homologues of *Mtb* antimicrobial resistance (AMR) associated genes in *M. komossense* JERR01

Drug	Locus tag in *M.* JERR01	Locus tag in *Mtb* H37Rv	Target (gene/protein function)	Abbreviation	Percentage ID (%)	Ortholog or Paralog?
Amikacin	HJANABMJ_03920	LFFBIPCO_01392	16S ribosomal RNA	Rrs	95	Ortholog
	HJANABMJ_05664	LFFBIPCO_02552	N-acetyltransferase Eis	Eis	32	Ortholog
Bedaquiline	HJANABMJ_00876	LFFBIPCO_00718	Siderophore exporter	MmpL5_1	63	Paralog
	HJANABMJ_01446	LFFBIPCO_00718	Siderophore exporter	MmpL5_2	66	Paralog
	HJANABMJ_02887	LFFBIPCO_00718	Siderophore exporter	MmpL5_3	74	Ortholog
	HJANABMJ_03930	LFFBIPCO_01381	ATP synthase subunit c	AtpH	96	Ortholog
	HJANABMJ_00642	LFFBIPCO_00720	HTH-type transcriptional regulator MmpR5	MmpR5	31	Ortholog
Delamanid	HJANABMJ_03142	LFFBIPCO_03735	Deazaflavin-dependent nitroreductase_1	Ddn_1	32	Paralog
	HJANABMJ_04822	LFFBIPCO_03735	Deazaflavin-dependent nitroreductase_2	Ddn_2	51	Ortholog
Ethambutol	HJANABMJ_05282	LFFBIPCO_03991	putative arabinosyltransferase A	EmbA	68	Ortholog
	HJANABMJ_05283	LFFBIPCO_03992	putative arabinosyltransferase B	EmbB	67	Ortholog
	HJANABMJ_05281	LFFBIPCO_03990	putative arabinosyltransferase C	EmbC	72	Ortholog
	HJANABMJ_01600	LFFBIPCO_01340	Transcriptional regulatory protein EmbR	EmbR	58	Ortholog
Ethionamide	HJANABMJ_02640	LFFBIPCO_01570	Enoyl-[acyl-carrier-protein] reductase [NADH]	InhA	86	Ortholog
	HJANABMJ_01813	LFFBIPCO_04057	FAD-containing monooxygenase EthA_1	EthA_1	45	Paralog
	HJANABMJ_05338	LFFBIPCO_04057	FAD-containing monooxygenase EthA_2	EthA_2	65	Ortholog
Isoniazid	HJANABMJ_02868	LFFBIPCO_02013	Catalase-peroxidase 2	KatG2	63	Ortholog
	HJANABMJ_02640	LFFBIPCO_01570	Enoyl-[acyl-carrier-protein] reductase [NADH]	InhA	86	Ortholog

Fluoroquinolones	HJANABMJ_00007	LFFBIPCO_00006	DNA gyrase subunit A	GyrA	89	Ortholog
	HJANABMJ_00006	LFFBIPCO_00005	DNA gyrase subunit B_1	GyrB_1	82	Ortholog
	HJANABMJ_00926	LFFBIPCO_00005	DNA gyrase subunit B_2	GyrB_2	75	Paralog
Linezolid	HJANABMJ_01107	LFFBIPCO_00743	50S ribosomal protein L3	RplC	90	Ortholog
Pyrazinamide	HJANABMJ_05375	LFFBIPCO_02160	Nicotinamidase/pyrazinamidase	PncA	67	Ortholog
	HJANABMJ_04935	LFFBIPCO_03788	Aspartate 1-decarboxylase	PanD	84	Ortholog
Rifampicin	HJANABMJ_01239	LFFBIPCO_03645	DNA-directed RNA polymerase subunit alpha	RpoA	93	Ortholog
	HJANABMJ_01059	LFFBIPCO_00708	DNA-directed RNA polymerase subunit beta	RpoB	90	Ortholog
	HJANABMJ_01060	LFFBIPCO_00709	DNA-directed RNA polymerase subunit beta'	RpoC	91	Ortholog
Streptomycin	HJANABMJ_03920	LFFBIPCO_01392	16S ribosomal RNA	Rrs	95	Ortholog
	HJANABMJ_01079	LFFBIPCO_00724	30S ribosomal protein S12	RpsL	96	Ortholog
	HJANABMJ_05694	LFFBIPCO_04125	Ribosomal RNA small subunit methyltransferase G	Gid	59	Ortholog

This table presents a side-by-side comparison of antimicrobial resistance elements in *M. komossense* JERR01 and *Mtb* H37Rv. Information on targeted drugs, associated resistance genes/proteins/RNAs, loci, percent identity between sequences, and orthology/paralogy relationships is provided for each resistance element.

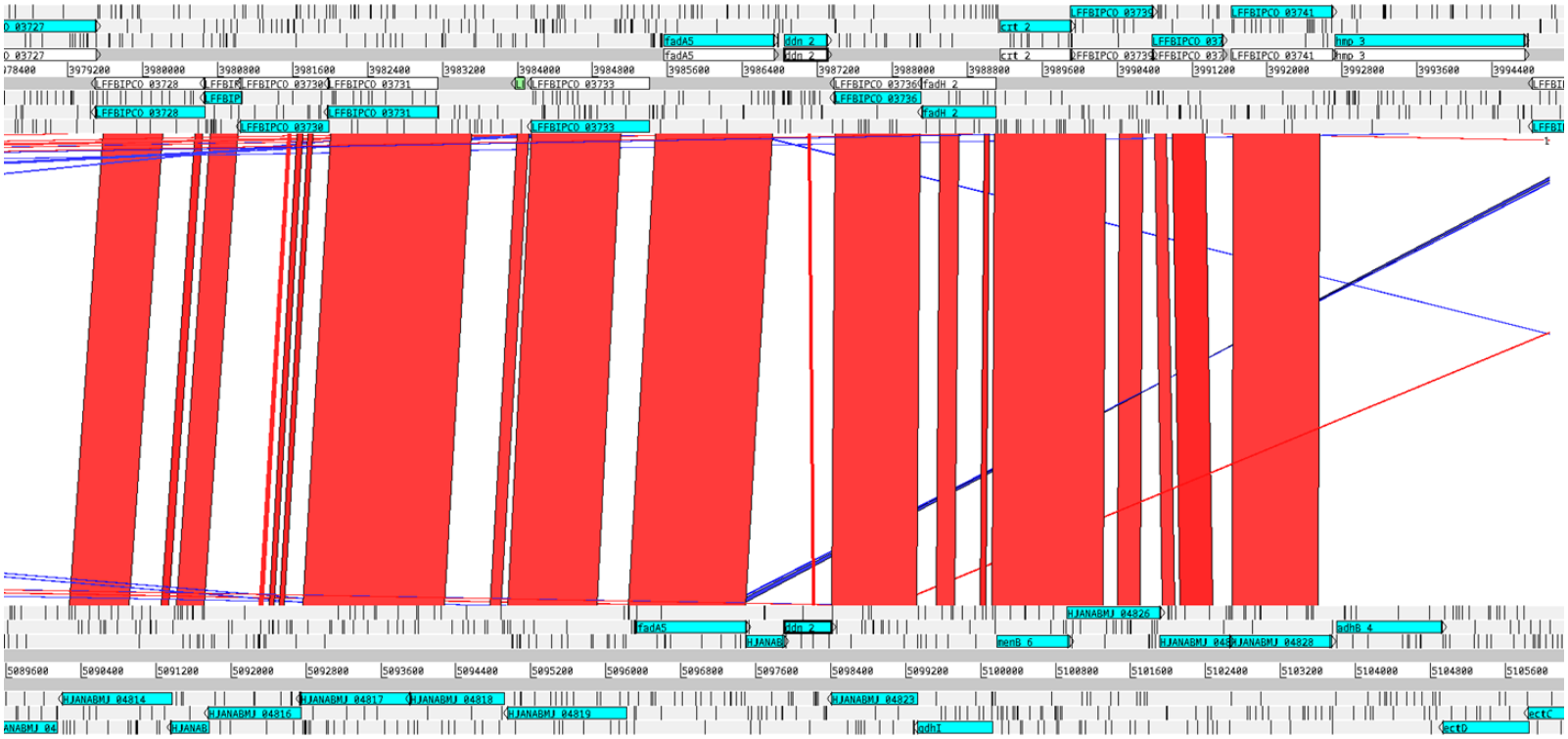


Figure 4-23. Comparison of Ddn orthologs in *Mtb* H37Rv and *M. komossense* JERR01.
This Artemis Comparison Tool (ACT) image shows a blastn pairwise alignment of Ddn genes from *Mtb* H37Rv and *M. komossense* JERR01. In picture above, Ddn_2 protein from *M. komossense* JERR01 was identified as the ortholog to *Mtb* H37Rv Ddn based on both high sequence identity and synteny conservation of surrounding genes. Red and blue bars indicate matches in the same orientation (red) and reverse orientation (blue), and regions of uniqueness (white).

4.4 Discussion

This chapter delves into the genomic characteristics of *M. komossense* JERR01, motivated by the phenotypic differences observed between *M. komossense* JERR01 and *M. komossense* ATCC 33013. In this chapter, the complete draft genome of *M. komossense* JERR01 was presented, and a targeted comparison between this strain and other members of the *Mycobacterium* genus, including *M. komossense* ATCC 33013 and *Mtb* H37Rv was performed.

In this thesis chapter, the closed chromosome of *M. komossense* JERR01 has been successfully obtained by using Oxford Nanopore sequencing data and Unicycler. One of the strengths of using Oxford Nanopore long read data in comparison to just using Illumina short read data is improving the completeness of the genome assembly. The quality of the assembly has also been assessed using other tools such as QUAST and BUSCO, and annotation via Prokka has utilized public databases for the annotation to provide functional predictions of genes found in the assembly. Although a closed chromosome was obtained with four other independent genetic elements, the quality analysis has revealed that this assembly is robust. It is important to mention that one of the weaknesses of Oxford Nanopore data is the higher rate of errors compared to Illumina short sequencing. This disadvantage of long read data requires additional error correction procedures that might not cover all the errors present in the sequencing. Nonetheless, it was possible to obtain a complete chromosome for this organism and identify plasmids, and this has never been done before for this representative of this species. In the past, it was thought that plasmids were a rare event in mycobacteria, however, their presence has been underestimated (Morgado and Vicente, 2021),

nowadays; due to advances in next-generation sequencing (NGS), it has been possible to characterize these plasmids and add them to public databases. While the annotation with Prokka has allowed to provide possible functionalities to the *M. komossense* JERR01 genes, since these annotations are based on genes already in the public databases, it is possible that novel genes in *M. komossense* JERR01 might have been overlooked by this annotation tool. Using GhostKoala and KEGG for function prediction, the overall functional distribution of the *M. komossense* proteins was explored, but 56.1% of the chromosome genome lacked KEGG annotations and this high percentage of proteins lacking annotation presents a limitation within this study. Although four of the contigs present in the genome were considered plasmids, this was based on the annotation of a limited number of proteins. A more in-depth analysis would require exploring the plasmid properties of these contigs, but due to time restrictions, these comparisons were not possible to be performed.

The comparison between *M. komossense* JERR01 and *M. komossense* ATCC 33013 genomes was performed using a combination of different methods (ANI, Panaroo, ACT and KEGG subsystem analysis), this *in silico* approach provided a picture of the taxonomic relationship between these two organisms. The results have provided evidence to conclude that *M. komossense* JERR01 and the type strain ATCC 33013 represent distinct species. The Pan-genome analysis reflected this; only 60.4% JERR01 CDSs were shared with ATCC 33013.

Functionally, 40.7% of *M. komossense* JERR01 proteins and 41.1% of proteins of the type strain ATCC 33013 were recognized by KEGG and used for the subsystem analysis, leaving

more than 50% of proteins for each of these organisms uncharacterized. The comparison analyses performed were computational, and therefore, additional experimental approaches like protein activity assays would be necessary to strengthen these conclusions. For example, in the previous chapter, it was observed that *M. komossense* JERR01 produced a yellow pigment. Notably amongst the unique genes in *M. komossense* JERR01 were genes for 15-cis-phytoene synthase (*crtB*) and Lycopene beta cyclase (*crtI*), genes responsible for producing yellow carotenoid pigments in mycobacteria (Ichiyama, Shimokata and Tsukamura., 1988; Robledo, Murillo and Rouzaud, 2011) and these genes were absent in *M. komossense* ATCC 33013. Due to time restrictions and challenges with growing *M. komossense* ATCC 33013 in the liquid culture, their potential functional divergence of these genes was not explored in these two strains.

Having identified that *M. komossense* JERR01 likely represented a distinct species, the comparison with other mycobacterial revealed a closer taxonomic relationship between *M. komossense* JERR01 and members of the *Fortuitum-Vaccae* clade. The ANI, ACT comparisons, and the Pan-genome analyses supported the taxonomic placement of *M. komossense* JERR01 within the *Fortuitum-Vaccae* clade of the genus *Mycobacterium*, with *M. vanbaalenii* PYR-1 being the closest taxonomic relative. Even with a close comparator, there was a high proportion of "Unclassified" proteins in the pairwise functional comparison which represents a gap in knowledge. Future studies could prioritize the functional characterization of these unclassified proteins for an understanding of potential metabolic adaptations specific to *M. komossense* JERR01 compared to other members of the *Fortuitum-Vaccae* clade. Only a proportion of the

proteins present in these genomes were used for the KEGG comparison; the proteins that were annotated as hypothetical proteins were not recognized by KEGG and, therefore, not used. A deeper look into these hypothetical proteins and a potential reannotation of these proteins could refine these results. The Pan-genome analysis allowed the identification of core genes (essential for basic functions) and accessory genes (specific to each genome and related to specific adaptations)(Costa et al., 2020). Future studies could focus on the accessory genes of *M. komossense* JERR01 to understand how this strain might have adapted to its specific environmental niche compared to other mycobacteria and also its pathogenic potential given its clinical origins. As mentioned earlier, only computational analyses were used to perform these comparisons; experimental approaches such as protein activity or targeted gene knockouts would be necessary to increase the robusticity of these results.

The comparative analysis employed for the identification of homologs in *M. komossense* JERR01 of genes associated with antibiotic resistance In *Mtb* H37Rv, which is a well characterized pathogen, provided information regarding the possible mechanisms of resistance in this *Mycobacterium* and its potential a model organism for *Mtb* resistance studies. The results derived from computational analysis successfully identified homologs of most genes associated with AMR from *Mtb*, with the exception of AhpC.

4.5 Summary

The genotypic characterization analysis performed in this chapter revealed significant differences between *M. komossense* JERR01 and the *M. komossense* type strain ATCC 33013 that suggest that it is potentially an unclassified species belonging to the

Fortuitum-Vaccae clade of the genus *Mycobacterium.* Therefore, for the rest of this thesis, the term *Mycobacterium* JERR01 (*M.* JERR01) will be used to describe the previously identified *M. komossense* strain JERR01. Despite the true taxonomic identity of *M.* JERR01 being unknown, the observed phenotypic sensitivities to a number of anti-tuberculosis drugs combined with the genomic identification of genes associated with resistance to them suggest that *M*. JERR01 may be suitable to act as a model organism for *Mtb* antibiotic resistance studies.

5 Evaluation of *M.* JERR01 as a model for the evolution of RIF and INH resistance.

5.1 Introduction

In the previous chapter, the genetic makeup *M.* JERR01 was characterised. An ANI comparison between *M.* JERR01 and *M. komossense* ATCC 33013, revealed an overall ANI of 75.13% and identified that both organisms were only distantly related. This lack of relatedness was also confirmed through a pairwise comparison and visualisation using ACT. Similarly, *M.* JERR01 showed limited ANI with representative species of the *Mycobacterium* genus, with the closest relative among the tested strains having an ANI of <85%, indicating that *M.* JERR01 was a distinct species of *Mycobacterium* that had not yet been described. Given its novel nature, a comparative analysis of genes associated with AMR in *Mtb* H37Rv and *M.* JERR01 was performed to identify commonalities and differences between pathogenic *Mtb* and the potential model organism for AMR studies. This analysis revealed that *M.* JERR01 shares almost all orthologs associated with ARM resistance in *Mtb* except for AhpC. This suggests that *M.* JERR01 could be a suitable model for evaluating the evolution/development of AMR towards certain drugs.

RIF and INH are the two most important components of first line of antitubercular treatment against drug-susceptible TB, such that resistance to them constitutes the formal definition of MDR-TB. RIF binds to the β subunit of the DNA-dependant RNA polymerase (or RpoB), inhibiting transcription. RIF resistance happens when the binding site of the β-subunit is altered, therefore, not allowing the antibiotic to bind and preventing the translation interference by RIF. These spontaneous mutations happen in a region inside the RpoB gene named Rifampicin Resistance Determining Region (RRDR).

Most of the mutations present in RIF-resistant *Mtb* strains happen within 4 sections of the RRDR, including the N-terminal cluster of the *Mtb* RpoB protein (residue 167-172), and clusters I (residues 423-457), II (residues 481-495) and III (residues 604-611) (Molodtsov *et al.*, 2017). Most of the mutations observed in RIF resistant strains are single amino acid substitutions (resulting from single nucleotide mutations) concentrated in cluster I, which is highly conserved among mycobacteria (Portelli *et al.*, 2020).

INH is a prodrug that gets activated by the catalase-peroxidase protein (KatG) and inhibits the synthesis of mycolic acid, which is a key component of the *Mtb* cell wall (Jena *et al.*, 2015). INH resistance in *Mtb* is associated with a range of mutations affecting one or more genes, such as the *katG* gene, the *inhA,* and AphC (Jena *et al.*, 2015), this last gene being absent in the *M.* JERR01 genome.

5.2 Aims and objectives

Having demonstrated in Chapters 3 and 4 that *M.* JERR01 strain is a fast-growing novel *Mycobacterium* that would facilitate experimental analysis, and would be susceptible to 6 antitubercular drugs, according to the clinical breakpoints and critical concentrations used for *Mtb* treatment, this chapter will continue to explore the possibility that *M.* JERR01 could be a suitable model for studying the emergence and evolution of AMR in *Mtb*.

This chapter aims to demonstrate that *M.* JER001 is a suitable model organism for studying resistance development against key antitubercular drugs through the following objectives:

1. Identify the conditions required for selection of antibiotic resistant mutants for RIF and INH in solid media.
2. Determine the mutation frequency of RIF and INH resistant mutants arising from a population given the identified conditions.
3. Characterize the level of AMR of the resistant *M.* JERR01 mutants towards the tested drugs.
4. Identify the genetic changes leading to resistance in a selection of resistant mutants and assess their potential functional impacts on resistance.

5.3 Results

5.3.1 Identification of conditions required for selection of antibiotics resistant mutants for RIF and INH

One of the key considerations in establishing whether *M.* JERR01 was a suitable model organism for studying AMR evolution in *Mtb* was to determine how *M.* JERR01 develops AMR during *in vitro* experimentation. To address this question, the two key frontline drugs, RIF and INH, were selected for investigation and comparison to studies performed in *Mtb* with these antibiotics. In chapter 3, the MIC_{90} of *M.* JERR01 for RIF and INH in liquid culture was determined (0.002 µg/ml and 32 µg/ml, respectively). However, liquid culture evolution experiments are prone to bottlenecking (Barrick and Lenski, 2013; Delaney, Letten and Engelstädter, 2023), allowing only the proliferation of the fittest clones and potentially limiting the type and number of mutants that can be selected via this method. By contrast, direct selection on solid medium supplemented with a suitable concentration of selective antibiotic can sample the population of resistant bacteria from the viable population in culture as well as drive the emergence of resistant clones through directed mutational selection upon antibiotic exposure. Based on these considerations, a plate-based approach for the selection of resistant mutants upon antibiotic exposure was chosen.

Since growth on a solid medium is different from the liquid culture growth used for the MIC_{90} determination, the first experiments performed were to assess whether resistance levels on agar plates supplemented with RIF or INH were different from those observed in liquid. To do this, 7H10 agar plates were supplemented with a range of RIF or INH concentrations starting from one dilution below the liquid MIC_{90} values of each antibiotic

(0.0008-0.4 µg/ml and 16-512 µg/ml for RIF and INH, respectively). A non-supplemented 7H10 plate served as a growth reference for each experiment.

Next, *M.* JERR01 was grown in 7H9 broth to stationary phase, serially diluted and spotted onto 7H10 plates supplemented without or with the respective antibiotics concentrations and incubated for 5 days at 30°C prior to growth evaluation (Figure 5-1 & Figure 5-2). No growth difference compared to the no antibiotic control was observed for RIF concentrations up to 0.0125 µg/ml Figure 5-1). By contrast, growth was reduced by approximately 100-fold at 0.025 µg/ml and 1000-fold at 0.05 µg/ml compared to the control as indicated by the growth of individual colonies at lower dilutions (Figure 5-1). At higher concentrations of RIF, the formation of individual colonies was no longer evident and only a confluent growth at the lowest dilution could be distinguished. The intensity of the confluent growth diminished with increasing RIF concentrations, being almost invisible at the highest RIF concentration (0.4 µg/ml) (Figure 5-1).

Similarly to the selective 7H10 Middlebrook plates containing RIF (Figure 5-1), no growth difference compared to the no antibiotic control was observed for concentrations up to 16 µg/ml on plates supplemented with INH (Figure 5-2). A reduction of ~100-fold can be seen at the MIC_{90} concentration of 32 µg/ml with overall colony sizes being reduced. At higher concentrations of INH (64 µg/ml - 512 µg/ml) (Figure 5-2), week confluent growth haloes were present at the lowest dilution indicating a strong antibiotic selection. Interestingly, single colonies could be identified on top of these haloes (Figure 5-2) that might reflect the selection of resistant clones.

These results showed that concentrations of RIF 0.1 µg/ml and above, and of INH 128 µg/ml above, could select against sensitive bacterial populations and would be suitable for the selection of resistant clones. However, increasing the selective pressure on the bacterial culture might result in a different population of resistant mutants being selected for compared to lower antibiotic concentrations. This was based on the rationale that there will be likely more mutations conferring low-level antibiotic resistance to the strain than those conferring high-level resistance. To accommodate this in the experimental design, three concentrations of each antibiotic to determine the evolution of AMR in *M.* JERR01 were selected, starting from the lowest concentration allowing selection (RIF: 0.1 µg/ml, 0.2 µg/ml and 0.4 µg/ml and for INH: 128 µg/ml, 256 µg/ml and 512 µg/ ml).

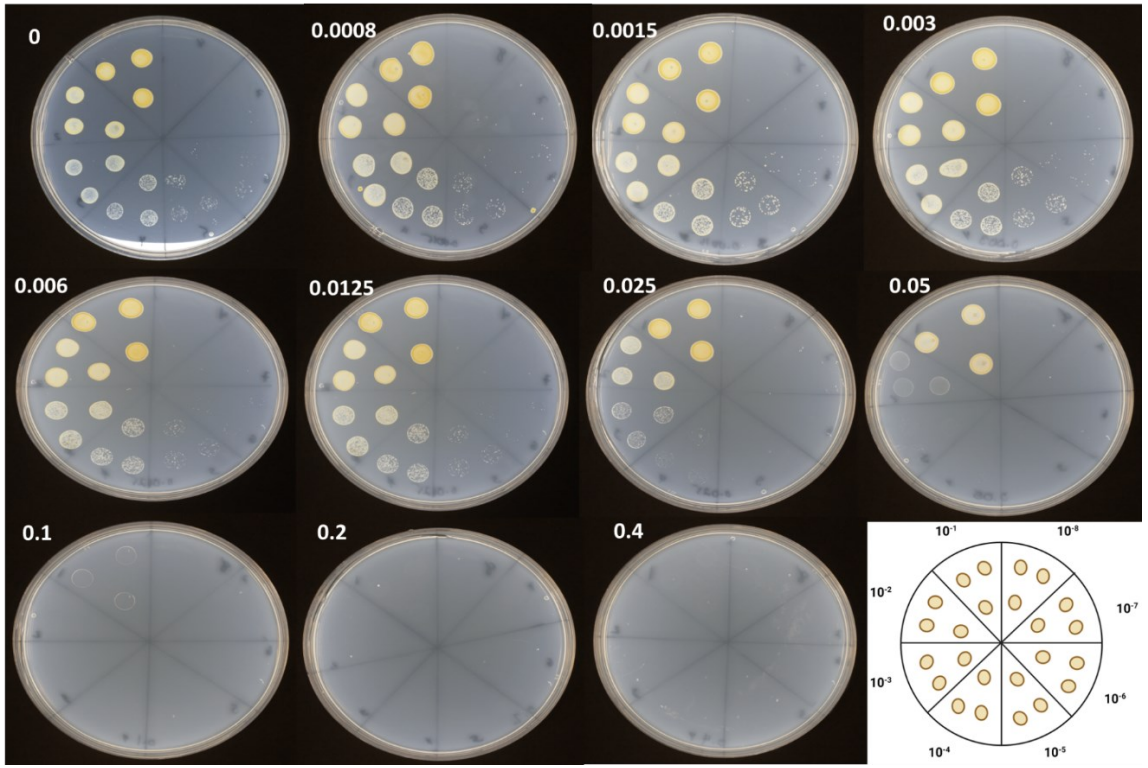


Figure 5-1. Growth of *M.* JERR01 on agar plate supplemented with different concentrations of RIF.
M. JERR01 was grown from a single colony in 7H9 at 30°C, 120 rpm to an OD_{600} of 3.0. 10-fold serial dilutions were prepared in PBS, and 10 µl were spotted onto 7H10 Middlebrook agar plates containing the indicated concentrations of RIF (µg/ml). The plates were incubated at 30°C for 5 days, bacterial growth was assessed, and pictures of each plate were taken. The bottom right panel shows the plate layout with dilutions for each segment indicated. The figure shows representative results of 3 biological repeats.

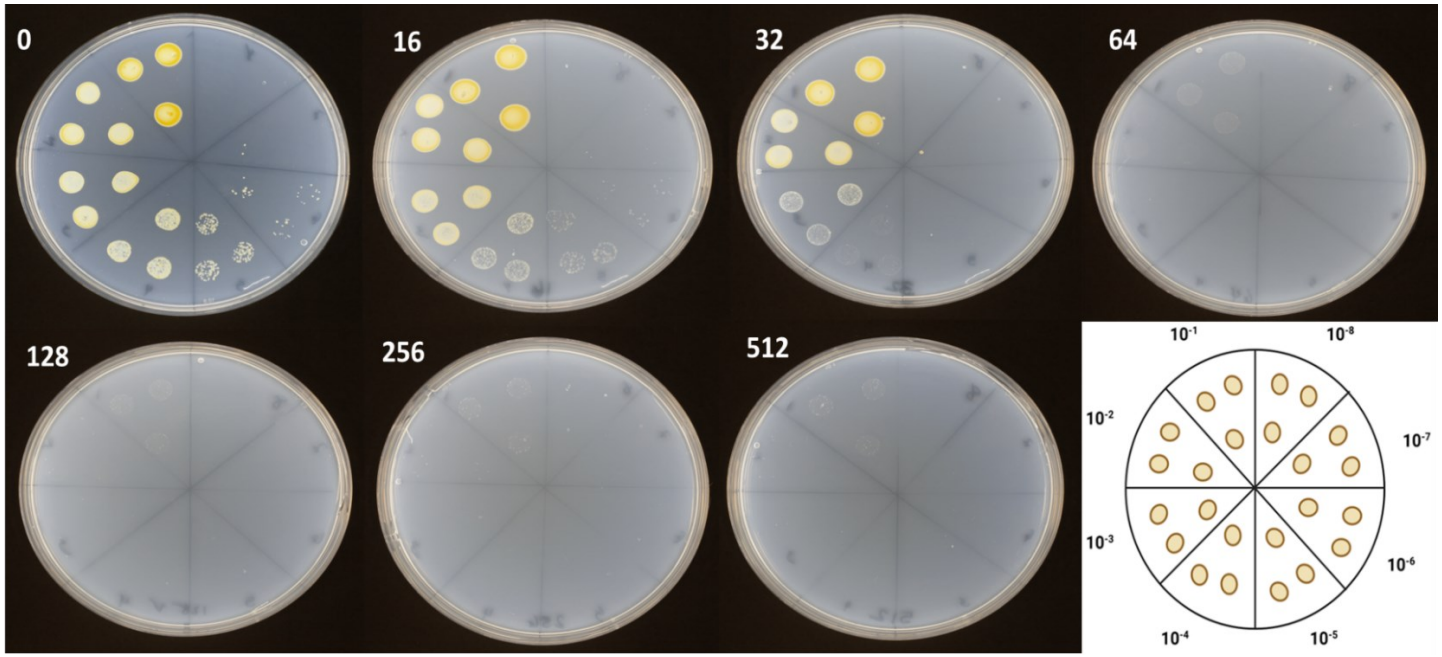


Figure 5-2. Growth of *M.* JERR01 on agar plate supplemented with different concentrations of INH.
M. JERR01 was grown from a single colony in 7H9 at 30°C, 120 rpm to an OD_{600} of 3.0. 10-fold serial dilutions were prepared in PBS, and 10 µl were spotted onto 7H10 Middlebrook agar plates containing the indicated concentrations of INH (µg/ml). The plates were incubated at 30°C for 5 days, bacterial growth was assessed, and pictures of each plate were taken. The bottom right panel shows the plate layout with dilutions for each segment indicated. The figure shows representative results of 3 biological repeats.

5.3.2 Mutation frequency and isolation of antibiotic resistant *M.* JERR01 mutant colonies

How likely a *bacterium* is to develop AMR, can be experimentally estimated by calculating the mutation frequency against specific antimicrobials. In order to compare how likely *M.* JERR01 was to develop resistance against RIF and INH in comparison to *Mtb*, the number of resistant bacteria selected from a given population of *M.* JERR01 using the conditions defined in the previous section was determined. Stationary phase *M.* JERR01 cultures (OD_{600}~3.0) were concentrated 100x by centrifugation into 1 ml PBS. This concentrated suspension was diluted another 100x in PBS, and a total of 2 ml from this suspension was spread evenly across 10-selective plates per antibiotic concentration, with four replicates per concentration. A serial dilution of this sample was prepared in PBS and plated onto nonselective 7H10 plates to determine the total number of bacteria plated in the experiment and allow the calculation of mutation frequencies. After 5-days of incubation at 30°C, the number of resistant colonies across all 10 plates of each experiment were collected, and mutation frequencies were determined (Figure 5-3).

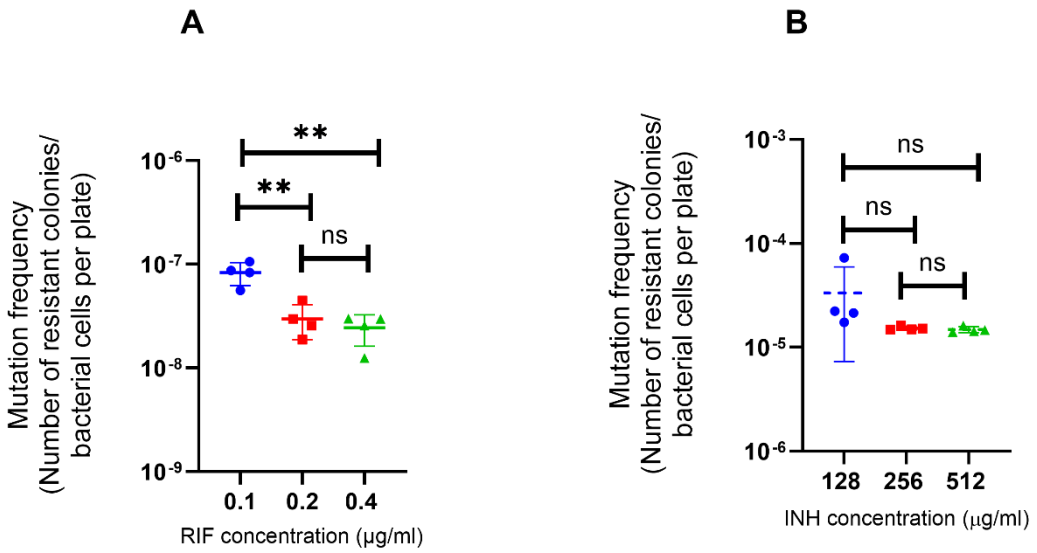


Figure 5-3. Mutation frequency of *M.* JERR01 to RIF and INH.
Mutation frequencies were determined by dividing the number of resistant colonies by the total number of bacterial cells in the experiment for **(A)** RIF and **(B)** INH at the defined concentrations of selective antibiotics. The coloured symbols represent the mean of the 10 antibiotic plates used to determine the mutation frequency. Central bars represent the estimated mutation frequency, and error bars representing the SD from the mean. Statistical analysis was performed using a One-way ANOVA followed by Tukey's post-test and significance levels for each comparison are indicated on top of the graphs: ns not significant; ** $p<0.01$.

Overall, mutation frequencies decreased with increasing concentrations of antibiotics for both RIF and INH (Figure 5-3 A and B, respectively). Mutation frequencies for RIF ranged from 8.32×10^{-8} ($\pm 2.064 \times 10^{-8}$) to 2.97×10^{-8} ($\pm 1.10 \times 10^{-8}$) and 2.44×10^{-8} ($\pm 8.21 \times 10^{-9}$) for 0.1, 0.2 and 0.4 µg/ml of the drug (Figure 5-3 A). The mutation frequencies between 0.1 and 0.2 µg/ml ($p = 0.0058$) and between 0.1 and 0.4 µg/ml ($p = 0.0019$) were statistically different from those observed between 0.2 and 0.4 µg/ml ($p = 0.7052$), in line with the hypothesis that increased selective pressure limits the possible mutations that can be selected for.

Mutation frequencies for INH (Figure 5-3 B) were higher compared to those observed for RIF (Figure 5-3 A). While there still was an overall trend for lower mutation frequencies at the two higher antibiotic concentrations, these were no longer statistically significant. The P-values for each of the groups were: 0.1135 (128 to 256 µg/ml), 0.0933 (128 to 512 µg/ml), and 0.9910 (256 to 512 µg/ml).

The observation of different mutation frequencies, therefore, suggested that the selected conditions could give rise to diverse types and levels of resistance. This hypothesis will be further evaluated in the next section.

5.3.3 Validation of antibiotic resistance profiles of *M.* JERR01 mutants

To further investigate the hypothesis that different selection conditions will select for different types of resistant mutants, for each antibiotic and biological replicate, 10 random colonies were purified from the following concentrations (0.1 µg/ml for RIF and 512 for INH µg/ml) and 40 random colonies were purified, from the following concentrations (0.4 µg/ml for RIF and 128 µg/ml of INH). These selective conditions were chosen, as mutation frequencies determined in section 5.3.2 indicated that there were significant (RIF) or indicative trends (INH) in mutation frequency differences, between the highest and all other conditions (Figure 5-3). Mutant colonies arising from the middle antibiotic concentrations were not further validated as there was not statistically significant difference in mutation frequencies between this and the highest concentration of each antibiotic tested (Figure 5-3).

To confirm whether the selected colonies represented stable, chromosomal mutations, these were firstly regrown on non-selective agar plates and subsequently a single colony

from these non-selective plates was regrown in selective agar plates at their respective selection concentrations, (0.1 μg/ml and 0.4 μg/ml) for RIF and (128 μg/ml and 512 μg/ml) for INH. All mutant colonies but not colonies from the parental strain were able to grow in the presence of the antibiotic concentrations used for selection, indicating that all selected mutants represented stable, genetic changes that led to resistance.

Having confirmed that all mutants were stably resistant towards the antibiotic used for evolution/selection, the level of antimicrobial resistance towards these drugs compared to the parental strain was measured. To do this, the susceptibility of each of these mutants towards the selective drugs was assessed using the broth microdilution method using the following concentration ranges for RIF (100 μg/ml - 0.002 μg/ml) and for INH (512 μg/ml - 4 μg/ml). As expected, all the RIF and INH mutants showed an MIC_{90} higher than that of the parent strain (0.002 μg/ml and 32 μg/ml, respectively for RIF and INH, Table 5-1). Interestingly, mutants selected for both antibiotics displayed a range of MIC_{90} values adding further support to the hypothesis that different genetic changes could account for the observed phenotypes.

Of the fifty RIF selected mutants, a total of 30 (60%) mutants presented an MIC_{90} of more than 100 μg/ml (the highest concentration assessed in the assay), while 11 (22%) mutants displayed an MIC_{90} of 0.78 μg/ml, 5 (10%) mutants an MIC_{90} of 12.5 μg/ml and 3 (6%) mutants an MIC_{90} of 1.56 μg/ml (Table 5-1). Regardless of the frequency distribution of MICs for the different mutants, even the mutants with the lowest MICs (0.78 μg/ml) were resistant to 400x the concentration of RIF relative to the parent strain,

while the most resistant mutants were able to grow at more than 50,000 higher RIF concentrations than the original strain.

The ten mutants selected at the highest INH concentration (512 µg/ml) all had an MIC_{90} to INH of ≥512 µg/ml on all plates (Table 5-1). By contrast, of the forty INH mutants incubated with 128 µg/ml, two were contaminated. Of the other 38, 31 (82%) had an MIC_{90} higher than 512 µg/ml INH, 5 (8%) had an MIC_{90} 256 µg/ml and 2 (5%) had an MIC_{90} of 128 µg/ml (Table 5-1). These observations added further evidence that a range of different genetic mutations that could give rise to distinct levels of AMR. The next section will determine the genetic changes likely to lead to these observed phenotypic differences.

Table 5-1. Distribution of MIC_{90} values for selected RIF and INH mutants.

Antibiotic used for selection	Concentration of antibiotic used for selection (µg/ml)	Number of isolates	MIC_{90} (µg/ml)
RIF	0.4 µg/ml	25	≥100
		1	50
		5	12.5
		3	1.56
		6	0.78
	0.1 µg/ml	5	≥100
		5	0.78
INH	512 µg/ml	10	≥512
	128 µg/ml	31	≥512
		5	256
		2	128

Data are representative of identical MICs observed in three biological replicates. MIC_{90} for the parent strain against RIF was 0.002 µg/ml and against INH 32 µg/ml.

5.3.4 Identification of genomic changes leading to RIF and INH resistance

MIC_{90} variations observed among the *M.* JERR01 mutants confirmed that varying degrees of resistance were selected during the experiment and suggested that different genetic alterations could be their cause. This observation provided us with the opportunity to study the molecular basis of how resistance evolves in *M.* JERR01 towards two important drugs used in the treatment of drug-susceptible TB.

To achieve this, single colonies from the characterized 50 *M.* JERR01 RIF mutants (10 colonies from 0.1 µg/ml and 40 colonies from 0.4 µg/ml) and 48 *M.* JERR01 INH mutants (10 colonies from 512 µg/ml and 38 colonies from 128 µg/ml), were chosen for WGS. Bacterial colonies of mutants were sent to MicrobesNG, where DNA was extracted for Oxford Nanopore sequencing and Illumina sequencing. The data was received (in the form of long reads, FASTA, and GenBank files), and their quality was assessed using QUAST and BUSCO. Variant calling analysis was performed using Medaka and Snippy, which uses deep learning to create consensus sequences and identify variants (SNPs and indels) in the data by comparing it to a reference genome, in this case, the parental strain *M.* JERR01.

5.3.4.1 Identification of mutations conferring RIF resistance

5.3.4.1.1 General comparison of *M.* JERR01 and *Mtb* RpoB proteins

Rifampicin resistance in *Mtb* is primarily driven by mutations in the *rpoB* gene altering specific amino acids in the RRDR region of the RpoB protein. To focus my investigation, I initially compared the RpoB protein of *M.* JERR01 and *Mtb* H37Rv. Both proteins were aligned using a Needleman-Wunsch Global Protein Alignment (Figure 5-4), showing an

overall protein identity of 90%, highlighting an overall high conservation of the protein. Interestingly, all RRDR regions of *Mtb* were fully conserved in *M.* JERR01 suggesting that similar mutational changes leading to RIF resistance might occur in this *Mycobacterium* (Figure 5-4).

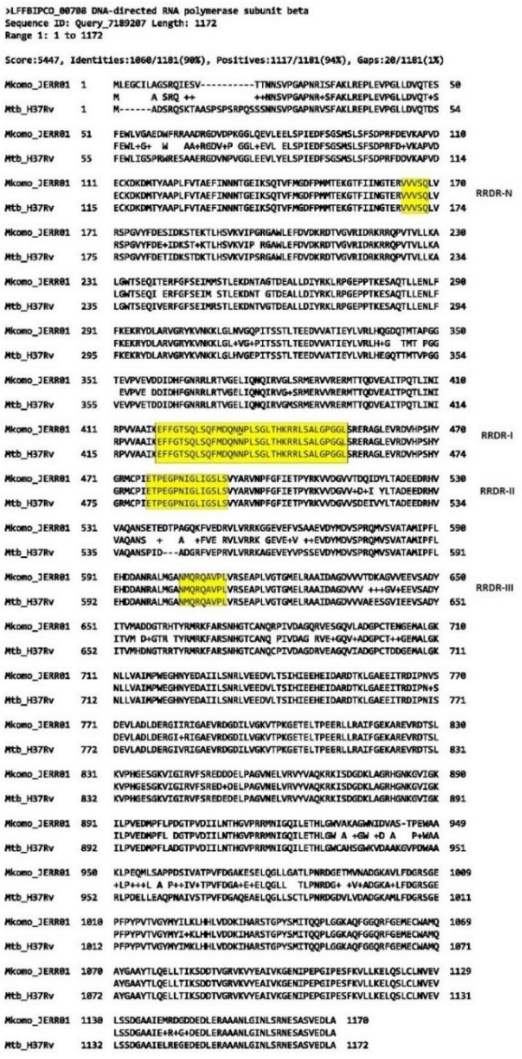


Figure 5-4. Global alignment analysis performed between *M.* JERR01 and *Mtb* H37Rv RpoB proteins.
M. JERR01 and *Mtb* H37Rv RpoB proteins were aligned using the Needleman–Wunsch algorithm available on the NCBI. RIF resistance-defining regions identified in *Mtb* and *M.* JERR01 are highlighted in yellow, and their region identifier is indicated on the right of the alignment.

5.3.4.1.2 Identification of *rpoB* mutants in RIF resistant colonies

The analysis in the above section indicated that RpoB and the RIF resistance defining regions were highly conserved between *Mtb* and *M.* JERR01, pointing towards a likely similar scenario of resistance development between the two species. I therefore focused my analysis on mutations targeting the *rpoB* gene. I identified single nucleotide polymorphisms (SNPs) in the *rpoB* gene of all RIF-resistant mutants (see Table 5-2 for details). Interestingly, all *rpoB* mutations resulted in amino acid changes in RRDR-I (Figure 5-4 and Figure 5-5) indicating the importance of this region of resistance development. A total of 12 different nucleotide substitutions were identified, affecting six amino acid residues in this region (Figure 5-1 and Table 5-2).

The most common single nucleotide mutation sites were codons: 437 at position 1,723,890 (10 mutants, 20%), 441 at position 1,723,878 (29 mutants, 58%), 446 at position 1,723,3863 (10 mutants, 20%), and one mutant reported at codon 431 position 1,723,3908 (1 mutant, 2%). Although all 50 mutants presented mutations on the *rpoB* gene, the following point mutations were identified: H441Y (11 mutants), S437F (10 mutants), H441D (7 mutants) and S446L (6 mutants), H441R (5 mutants), H441Q (4 mutants), S446W (3 mutants), H441P (2 mutants), D431V (1 mutant), and S446F (1 mutant)(Table 5.2).

Additionally, the following novel mutations were observed outside the RpoB protein of all *M.* JERR01 mutants: A500T in position 142,006; D142Y at position 5,211,331; D177E at position 5,432,385; A176V at position 68,933; and E136K at position 109,605. However, these mutations were found not to be associated with RIF resistance in *Mtb* (Hameed *et al.*, 2017), so they could not explain variations in phenotype. When

examining other genes associated with compensatory mutations in RIF resistant *Mtb,* such as *rpoA* or *rpoC*, no mutations were observed.

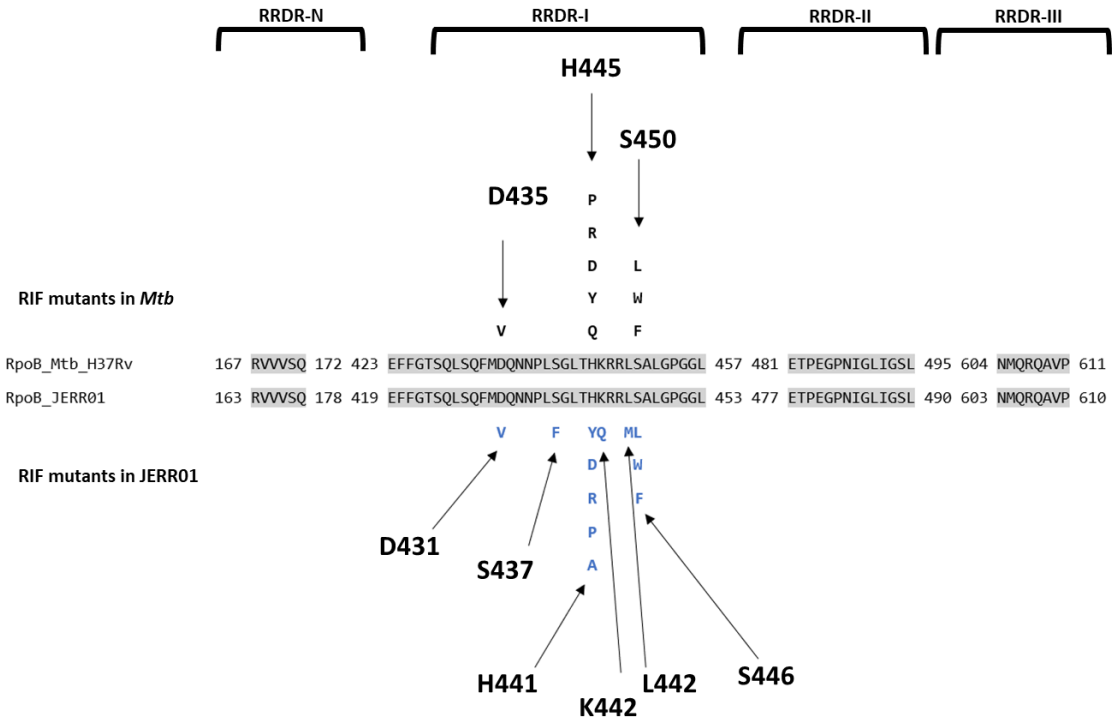


Figure 5-5. Alignment of the different RIF resistance-determining regions of *Mtb* H37Rv and *M.* JERR01 RpoB proteins and identified mutations.
RRDRs regions are indicated above the amino acid sequences. Mutations that confer RIF resistance in *Mtb* H37Rv are indicated above the *Mtb* RpoB sequence in black, while mutations in the *M.* JERR01 RpoB proteins are depicted below the *M.* JERR01 RpoB protein sequence in blue.

Table 5-2. Characterization of RIF mutations identified in *M.* JERR01.

Position	Reference	Mutant	Position	Amino Acid Change	Property Change	MIC (µg/ml)
431	GGT	GGA	1723908	D -> V	From acidic and hydrophilic amino acid to nonpolar hydrophobic amino acid.	0.78 (1 mutant)
437	CCG	CGA	1723890	S -> F	From polar, neutral, and hydrophilic amino acid to uncharged, nonpolar, aromatic, and hydrophobic amino acid	0.78 (6 mutants) 1.56 (3 mutants) >100 (1 mutant)
441/442	TGT/GGG	GGG/CGG	1723876/1723879	H -> Q K -> A	From polar, hydrophilic, and aromatic to a more hydrophilic, positively charged amino acid. From hydrophilic, non-polar to polar, hydrophobic, uncharged amino acid.	0.78 (4 mutants)
441	GTG	GCG	1723878	H -> R	Positively charged, hydrophilic to polar to another positively charged hydrophobic amino acid.	>100 (5 mutants)
441	GTG	GTA	1723879	H -> Y	Positively charged, hydrophilic amino acid to uncharged amino partially hydrophobic acid.	>100 (11 mutants)
441	GTG	GTC	1723879	H -> D	Positively charged hydrophilic amino acid to negatively charged hydrophilic amino acids.	>100 (7 mutants)
441	GTG	GGG	1723878	H -> P	Positively charged hydrophilic amino acid to partially hydrophilic uncharged amino acid that possesses a unique cyclic structure in its side chain.	>100 (2 mutants)

Position	Reference	Mutant	Position	Amino Acid Change	Property Change	MIC (µg/ml)
445/446	CGA/CAG	GGA/CAT	17233862/1723867	S -> F L -> M	From hydrophobic, nonpolar amino acid to a less hydrophobic, nonpolar amino acid with a sulphur-containing side chain. From a small, hydrophilic, and polar amino acid to a large, hydrophobic, nonpolar, aromatic amino acid.	50 (1 mutants)
446	CCG	CCA	17233863	S -> L	From polar, hydrophilic, neutral amino acid to uncharged, hydrophobic, and nonpolar amino acid.	12.5 (5 mutants) >100 (1 mutant)
446	GTG	GTC	1723386	S -> W	Positively charged, hydrophilic amino acid to a large, uncharged, nonpolar, hydrophobic aromatic amino acid.	>100 (3 mutants)

5.3.5 Characterization conferring INH resistance in *M.* JERR01

5.3.5.1.1 General comparison of *M.* JERR01 and *Mtb* catalase peroxidase proteins

INH resistance in *Mtb* is caused by mutations in KatG, InhA, and AhpC (Bakhtiyariniya *et al.*, 2022). Mutations in KatG prevent conversion of the INH prodrug into its active form, while mutations in the promoter region of InhA lead to the overexpression of this target, thus requiring higher concentrations of INH to achieve inhibition (Tseng *et al.*, 2015; Vilchèze and Kremer, 2017; Lempens *et al.*, 2018). The *ahpC* gene encodes an enzyme called alkyl hydroperoxidase reductase. This enzyme protects *Mtb* against oxidative stress, which is damage caused by reactive oxygen molecules. Mutations in *ahpC* gene have been associated with resistance towards INH, however, its impact on clinical isolates resistance remains still unclear (Cohen, Bishai and Pym, 2015).

5.3.5.1.2 Alignment and description of *Mtb* H37Rv and *M.* JERR01 KatG.

To determine whether the KatG proteins of *M.* JERR01 and *Mtb* H37Rv were conserved and could fulfil similar roles in INH resistance, a global alignment between the two proteins using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) was performed (Figure 5-6). This analysis showed that the two KatG proteins were only 63% identical and raised the possibility that the KatG protein of *M.* JERR01 might only be distantly related to the KatG of *Mtb*. While *M.* JERR01 only encodes one KatG protein, some members of the *Fortuitum-Vaccae* clade of *Mycobacteria,* such as *M. smegmatis* and *M. aurum* have more than one KatG protein (Phelan *et al.*, 2015). The KatG protein encoded in *M.* JERR01, therefore, might be more closely related to KatG proteins encoded in other members of the *Fortuitum-Vaccae* clade of *Mycobacterium*.

```
Score:2484, Identities:473/752(63%), Positives:567/752(75%), Gaps:17/752(2%)

Mkomo_JERR01  1    MTDTSDARPPHSDDKTSSRSESENPVIDSPEPKTHAPLTNKDWWPEQVDVSVLHKQNEKS  60
                   M +     PP ++  T + S    PV+   +        N+DWWP ++++ VLH+    +
Mtb_H37Rv     1    MPEQ---HPPITETTTGAASNG-CPVVGHMKYPVEGG-GNQDWWPNRLNLKVLHQNPAVA  55

Mkomo_JERR01  61   NPLGVDFDYATEFAKLDVEAFKRDVIDLINTSQDWWPADYGSYAGLFIRMSWHAAGTYRI  120
                   +P+G  FDYA E A +DV+A  RD+ +++ TSQ WWPADYG Y  LFIRM+WHAAGTYRI
Mtb_H37Rv     56   DPMGAAFDYAAEVATIDVDALTRDIEEVMTTSQPWWPADYGHYGPLFIRMAWHAAGTYRI  115

Mkomo_JERR01  121  FDGRGGAGQGSQRFAPLNSWPDNANLDKARRLLWPIKQKYGNKISWADLIAYAGNAALES  180
                    DGRGGAG G QRFAPLNSWPDNA+LDKARRLLWP+K+KYG K+SWADLI +AGN ALES
Mtb_H37Rv     116  HDGRGGAGGGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKKLSWADLIVFAGNCALES  175

Mkomo_JERR01  181  AGFQTFGFAFGREDIWEPEEMLWGQEDTWLGTDKRYGGTNDSDKRELAEPFGATTMGLIY  240
                    GF+TFGF FGR D WEP+E+ WG+E TWLG D+RY G     KR+L  P  A  MGLIY
Mtb_H37Rv     176  MGFKTFGFGFGRVDQWEPDEVYWGKEATWLG-DERYSG-----KRDLENPLAAVQMGLIY  229

Mkomo_JERR01  241  VNPEGPEGKPDPLAAAHDIRETFGRMAMNDEETAALIVGGHTLGKTHGAADVN-VGPEPE  299
                   VNPEGP G PDP+AAA DIRETF RMAMND ETAALIVGGHT GKTHGA   + VGPEPE
Mtb_H37Rv     230  VNPEGPNGNPDPMAAAVDIRETFRRMAMNDVETAALIVGGHTFGKTHGAGPADLVGPEPE  289

Mkomo_JERR01  300  GAPIEQQGLGWKCPFGTGNANDTVTSGLEVVWTTTPTKWSNAYLELLYGYEWELTKSPAG  359
                    AP+EQ GLGWK  +GTG   D +TSG+EVVWT TPTKW N++LE+LYGYEWELTKSPAG
Mtb_H37Rv     290  AAPLEQMGLGWKSSYGTGTGKDAITSGIEVVWTNTPTKWDNSFLEILYGYEWELTKSPAG  349

Mkomo_JERR01  360  AWQFEAKDAEAI--IPDPFGGPPRKPTMLVTDVSMRVDPIYGPITRRWLDHPEEMNEAFA  417
                   AWQ+ AKD      IPDPFGGP R PTML TD+S+RVDPIY  ITRRWL+HPEE+ + FA
Mtb_H37Rv     350  AWQYTAKDGAGAGTIPDPFGGPGRSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFA  409

Mkomo_JERR01  418  KAWYKLMHRDMGPVSRYLGPWVA-EAQLWQDPVPAVDHALIDESDVAALKTAVLQSGLSV  476
                   KAWYKL+HRDMGPV+RYLGP V  +  LWQDPVPAV H L+ E+++A+LK+ +  SGL+V
Mtb_H37Rv     410  KAWYKLIHRDMGPVARYLGPLVPKQTLLWQDPVPAVSHDLVGEAEIASLKSQIRASGLTV  469

Mkomo_JERR01  477  PQLVKTAWASASSFRGTDKRGGANGARLRLEPQRSWEANEPT-ELAKVLPALEKIQQDFN  535
                    QLV TAWA+ASSFRG+DKRGGANG R+RL+PQ  WE N+P  +L KV+  LE+IQ+ FN
Mtb_H37Rv     470  SQLVSTAWAAASSFRGSDKRGGANGGRIRLQPQVGWEVNDPDGDLRKVIRTLEEIQESFN  529

Mkomo_JERR01  536  ASATGGKKVSLADVIVLAGSAAIEKAAKDGGYEISVHFAPGRTDASQEQTDVESFAVLET  595
                   ++A G  KVS AD++VL G AAIEKAAK  G+ I+V F PGRTDASQEQTDVESFAVLE
Mtb_H37Rv     530  SAAPGNIKVSFADLVVLGGCAAIEKAAKAAGHNITVPFTPGRTDASQEQTDVESFAVLEP  589

Mkomo_JERR01  596  RADGFRNYARPGEKTPLEQLLIDKAYFLDLTAPELTALIGGLRTLNANHGGSKHGVFTDQ  655
                   +ADGFRNY   G   P E +L+DKA  L L+APE+T L+GGLR L AN+     GVFT+
Mtb_H37Rv     590  KADGFRNYLGKGNPLPAEYMLLDKANLLTLSAPEMTVLVGGLRVLGANYKRLPLGVFTEA  649

Mkomo_JERR01  656  PGVLSNDFFVNLLDMRTEWKPSELTENVYEGKDRATGAPKWTATAADLVFGSNSVLRAVA  715
                      L+NDFFVNLLDM   W+PS   +  Y+GKD  +G  KWT +  DLVFGSNS LRA+
Mtb_H37Rv     650  SESLTNDFFVNLLDMGITWEPSPADDGTYQGKD-GSGKVKWTGSRVDLVFGSNSELRALV  708

Mkomo_JERR01  716  EVYAQEDNKAKFVEDFVAAWVKVMNNDRFDLK  747
                   EVY  +D + KFV+DFVAAW KVMN DRFD++
Mtb_H37Rv     709  EVYGADDAQPKFVQDFVAAWDKVMNLDRFDVR  740
```

Figure 5-6. Global alignment analysis performed between *M.* JERR01 and *Mtb* H37Rv KatG proteins.

Global protein alignment between *M.* JERR01 and *Mtb* H37Rv KatG proteins using the Needleman–Wunsch algorithm presenting a percentage identity of 63%. Highlighted in yellow are key residues involved in INH binding.

To test if this was the case, the *M.* JERR01 KatG protein with KatG proteins of the environmental *M. smegmatis* MC2 155, which encodes three different KatG homologues (KatG1, KatG2 and KatG3) (Iwao and Nakata, 2018), were compared using the Needleman-Wunsch global alignment algorithm (Figure 5-7). This showed that the *M.* JERR01 KatG protein shared the highest sequence identity (78.24%) with the KatG2 protein of *M. smegmatis* MC2 155 (Figure 5-7). By contrast, comparisons of the *M.* JERR01 KatG protein to the *M. smegmatis* MC2 155 KatG1 and KatG3 proteins showed an identity of 56.7% and 62.3% respectively. Consequently, the *M.* JERR01 KatG protein was more closely related to the KatG2 protein of *M. smegmatis* MC2 than that of *Mtb.*

Conversely, comparison of the *Mtb* H37Rv KatG protein to those encoded in *M. smegmatis* MC2 155 revealed that the *Mtb* KatG protein was most closely related to the *M. smegmatis* KatG3 protein (71% identity) (Figure 5-8). Therefore, there was a substantial diversity in the number and sequences of KatG proteins present in the analysed species, which could explain their varying basal levels of resistance to INH.

Nevertheless, the alignment between *M.* JERR01 and *Mtb* H37Rv revealed conservation in the residues associated with INH resistance in clinical isolates of *Mtb* such as W300G, S315T, and W328R, as these mutations have been associated with loss of catalase-peroxidase activity and resistance to INH (Wengenack *et al.*, 1997; Kandler *et al.*, 2018) (Figure 5-6). These residues are important as mutations therein prevent INH interaction with KatG and prevent the conversion of the prodrug to its active form. For example, S315T accounts for 42%-95% of phenotypic INH resistance cases (Seifert *et al.*, 2015).

```
Score:3195, Identities:587/751(78%), Positives:654/751(87%), Gaps:7/751(0%)

Mkomo_JERR01     1    MT-DTSDARPPHSDDKTSSRSESENPVIDSPEPKTHAPLTNKDWWPEQVDVSVLHKQNEK  59
                      M+ DTSD+RPP+ D KT+S SESENP I SP+PK+ APL N+DWWP Q+DVS LH    +
M.smeg_MC2_155   1    MSSDTSDSRPPNPDTKTASTSESENPAIPSPKPKSGAPLRNQDWWPNQIDVSRLHPHPPQ  60

Mkomo_JERR01    60   SNPLGVDFDYATEFAKLDVEAFKRDVIDLINTSQDWWPADYGSYAGLFIRMSWHAAGTYR  119
                     NPLG DFDYA EFAKLDV A K D+  L+  SQDWWPADYG Y GLFIRMSWH+AGTYR
M.smeg_MC2_155  61   GNPLGEDFDYAEEFAKLDVNALKADLTALMTQSQDWWPADYGHYGGLFIRMSWHSAGTYR  120

Mkomo_JERR01   120  IFDGRGGAGQGSQRFAPLNSWPDNANLDKARRLLWPIKQKYGNKISWADLIAYAGNAALE  179
                    I DGRGG GQG+QRFAP+NSWPDN +LDKARRLLWPIKQKYGNKISWADL+ + GN ALE
M.smeg_MC2_155 121  IHDGRGGGGQGAQRFAPINSWPDNVSLDKARRLLWPIKQKYGNKISWADLLVFTGNVALE  180


Mkomo_JERR01   180  SAGFQTFGFAFGREDIWEPEEMLWGQEDTWLGTDKRYGGTNDSDKRELAEPFGATTMGLI  239
                    S GF+TFGF FGREDIWEPEE+L+G+ED WLGTDKRYGG    ++R+LAEP+GATTMGLI
M.smeg_MC2_155 181  SMGFKTFGFGFGREDIWEPEEILFGEEDEWLGTDKRYGG---GEQRQLAEPYGATTMGLI  237

Mkomo_JERR01   240  YVNPEGPEGKPDPLAAAHDIRETFGRMAMNDEETAALIVGGHTLGKTHGAADVN-VGPEP  298
                    YVNPEGPEG+PDPLAAAHDIRETFGRMAMNDEETAALIVGGHT GKTHGA D + VGPEP
M.smeg_MC2_155 238  YVNPEGPEGQPDPLAAAHDIRETFGRMAMNDEETAALIVGGHTFGKTHGAGDASLVGPEP  297

Mkomo_JERR01   299  EGAPIEQQGLGWKCPFGTGNANDTVTSGLEVVWTTTPTKWSNAYLELLYGYEWELTKSPA  358
                    E APIEQQGLGWK  +GTG   DT+TSGLEVVWT TPTKW N++LE+LYGYEWELTKSPA
M.smeg_MC2_155 298  EAAPIEQQGLGWKSSYGTGKGPDTITSGLEVVWTNTPTKWDNSFLEILYGYEWELTKSPA  357

Mkomo_JERR01   359  GAWQFEAKDAEAI--IPDPFGGPPRKPTMLVTDVSMRVDPIYGPITRRWLDHPEEMNEAF  416
                    GAWQF AKD      IPDPFGGP R PTMLVTD+SMRVDPIYG ITRRWLDHPEE++EAF
M.smeg_MC2_155 358  GAWQFTAKDGAGAGTIPDPFGGPGRNPTMLVTDISMRVDPIYGKITRRWLDHPEELSEAF  417

Mkomo_JERR01   417  AKAWYKLMHRDMGPVSRYLGPWVAEAQLWQDPVPAVDHALIDESDVAALKTAVLQSGLSV  476
                    AKAWYKL+HRDMGP+SRYLGPWVAE QLWQDPVPAVDH L+D+ D+AALK+ VL SGLS
M.smeg_MC2_155 418  AKAWYKLLHRDMGPISRYLGPWVAEPQLWQDPVPAVDHPLVDDQDIAALKSTVLDSGLST  477

Mkomo_JERR01   477  PQLVKTAWASASSFRGTDKRGGANGARLRLEPQRSWEANEPTELAKVLPALEKIQQDFNA  536
                    QL+KTAWASA+S+R TDKRGGANGAR+RLEPQ++W+ NEP ELA VLP LE+IQQDFNA
M.smeg_MC2_155 478  GQLIKTAWASAASYRNTDKRGGANGARVRLEPQKNWDVNEPAELATVLPVLERIQQDFNA  537

Mkomo_JERR01   537  SATGGKKVSLADVIVLAGSAAIEKAAKDGGYEISVHFAPGRTDASQEQTDVESFAVLETR  596
                    SA+GGKKVSLAD+IVLAGSAAIEKAAKDGGY ++V FAPGRTDASQE TDVESFAVLE R
M.smeg_MC2_155 538  SASGGKKVSLADLIVLAGSAAIEKAAKDGGYNVTVPFAPGRTDASQENTDVESFAVLEPR  597

Mkomo_JERR01   597  ADGFRNYARPGEKTPLEQLLIDKAYFLDLTAPELTALIGGLRTLNANHGGSKHGVFTDQP  656
                    ADGFRNY RPGEK  LE++L+++AYFL +TAP+LTAL+GGLR L+ NHGG+KHGVFTD+P
M.smeg_MC2_155 598  ADGFRNYVRPGEKVQLEKMLLERAYFLGVTAPQLTALVGGLRALDVNHGGTKHGVFTDRP  657

Mkomo_JERR01   657  GVLSNDFFVNLLDMRTEWKPSELTENVYEGKDRATGAPKWTATAADLVFGSNSVLRAVAE  716
                    G L+NDFFVNLLDM TEWK SE TENVYEG DR TG  KWTATA DLVFGS+SVLRAVAE
M.smeg_MC2_155 658  GALTNDFFVNLLDMGTEWKTSETTENVYEGVDRKTGQLKWTATANDLVFGSHSVLRAVAE  717

Mkomo_JERR01   717  VYAQEDNKAKFVEDFVAAWVKVMNNDRFDLK  747
                    VYAQ DN  +FV DFV AWVKVMNNDRFDLK
M.smeg_MC2_155 718  VYAQSDNGERFVNDFVKAWVKVMNNDRFDLK  748
```

Figure 5-7. Global protein alignment between *M.* JERR01 KatG and *M. smegmatis* MC2 155 KatG2.

Global protein alignment between *M.* JERR01 and *M. smegmatis MC2 155* KatG proteins using the Needleman–Wunsch algorithm presenting a percentage identity of 78%. Highlighted in yellow is the key residue involved in INH binding in *Mtb*. The alignment was performed using the Needleman–Wunsch algorithm on NCBI (Needleman and Wunsch, 1970).

```
Range 1: 1 to 739
NW Score (2776), Identities 531/748 (71%), Positives 602/748 (80%), Gaps 17/748 (2%)

Mtb_H37Rv        1    MPEQHP-----PITETTTGAASNGCPV-VGHMKYPVEGGGNQDWWPNRLNLKVLHQNPAV  54
                      MPE  P     PI E  T A + GCP   G +K PV GG N+DWWPN+LNLK+L +NP V
Msmeg_MC2_155    1    MPEDRPIEDSPPIGEAQTDAPAGGCPAGFGRIKPPVAGGSNRDWWPNQLNLKILQKNPDV  60

Mtb_H37Rv       55   ADPMGAAFDYAAEVATIDVDALTRDIEEVMTTSQPWWPADYGHYGPLFIRMAWHAAGTYR  114
                      +P+   FDY + V  +DVDAL  DI EVM TSQ WWPAD+GHYGPLFIRMAWHAAGTYR
Msmeg_MC2_155    61   INPLDEDFDYRSAVQNLDVDALRADIVEVMHTSQDWWPADFGHYGPLFIRMAWHAAGTYR  120

Mtb_H37Rv       115  IHDGRGGAGGGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKKLSWADLIVFAGNCALE  174
                      + DGRGGAG GMQRFAPLNSWPDNASLDKARRLLWPVKKKYGK LSWADLIV+AGN ALE
Msmeg_MC2_155    121  VSDGRGGAGAGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKNLSWADLIVYAGNVALE  180

Mtb_H37Rv       175  SMGFKTFGFGFGRVDQWEPDE-VYWGKEATWLGDERYSGKRDLENPLAAVQMGLIYVNPE  233
                      MGF+T GF FGR D+WEP+E VYWG E  WL D+RY+G+RDLENPLAAVQMGLIYVNPE
Msmeg_MC2_155    181  DMGFRTAGFAFGREDRWEPEEDVYWGPEQEWLDDKRYTGERDLENPLAAVQMGLIYVNPE  240

Mtb_H37Rv       234  GPNGNPDPMAAAVDIRETFRRMAMNDVETAALIVGGHTFGKTHGAGPADLVGPEPEAAPL  293
                      GPNGNPDP A+A+DIRETF RMAMNDVETAALIVGGHTFGKTHG G A LVGPEPEAAPL
Msmeg_MC2_155    241  GPNGNPDPQASAIDIRETFGRMAMNDVETAALIVGGHTFGKTHGNGDASLVGPEPEAAPL  300

Mtb_H37Rv       294  EQMGLGWKSSYGTGTGKDAITSGIEVVWTNTPTKWDNSFLEILYGYEWELTKSPAGAWQY  353
                      E++GLGW++  GTG GKDAITSG+EV WT+TPTKWDNSFLEILYG EWELTKSPAGA Q+
Msmeg_MC2_155    301  EEVGLGWRNPQGTGVGKDAITSGLEVTWTHTPTKWDNSFLEILYGNEWELTKSPAGANQW  360

Mtb_H37Rv       354  TAKDGAGAGTIPDPFGGPGRSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFAKAWY  413
                        KD   A ++P         P+ML +DL+LRVDPIYE+ITRRWL+HPEELA+EFAKAW+
Msmeg_MC2_155    361  KPKDNGWANSVPLAHEDGKTHPSMLTSDLALRVDPIYEQITRRWLDHPEELAEEFAKAWF  420

Mtb_H37Rv       414  KLIHRDMGPVARYLGPLVPKQTLLWQDPVPAVSHDLVGEAEIASLKSQIRASGLTVSQLV  473
                      KL+HRDMGPV RYLGP VPK T LWQD +PA  +DL  + E+A LK  I  SGLTVSQLV
Msmeg_MC2_155    421  KLLHRDMGPVTRYLGPEVPKDTWLWQDNIPA-GNDL-SDDEVAKLKELIADSGLTVSQLV  478

Mtb_H37Rv       474  STAWAAASSFRGSDKRGGANGGRIRLQPQVGWEVNDPDGDLRKVIRTLEEIQESFNSAAP  533
                      STAW AAS+FR SD RGGANGGRIRLQPQ+GWE N+PD +L +V+R  EEIQ++
Msmeg_MC2_155    479  STAWKAASTFRSSDLRGGANGGRIRLQPQLGWEANEPD-ELAQVVRKYEEIQKA------  531

Mtb_H37Rv       534  GNIKVSFADLVVLGGCAAIEKAAKAAGHNITVPFTPGRTDASQEQTDVESFAVLEPKADG  593
                        I VSFADLVVLGG   +EKAAKAAG ++TVPFTPGR DA+QE+TDV+SFA LEPKADG
Msmeg_MC2_155    532  SGINVSFADLVVLGGNVGVEKAAKAAGFDVTVPFTPGRGDATQEETDVDSFAYLEPKADG  591

Mtb_H37Rv       594  FRNYLGKGNPLPAEYMLLDKANLLTLSAPEMTVLVGGLRVLGANYKRLPLGVFTEASESL  653
                      FRNYLGKG+ LPAE+ L+D+ANLL LSAPEMT LVGGLRVL  N+     GV T+   +L
Msmeg_MC2_155    592  FRNYLGKGSDLPAEFKLIDRANLLGLSAPEMTTLVGGLRVLDVNHGGTKHGVLTDKPGAL  651

Mtb_H37Rv       654  TNDFFVNLLDMGITWEPSPADDGTYQGKD-GSGKVKWTGSRVDLVFGSNSELRALVEVYG  712
                      T DFFVNLLDM   W+PSPADDGTY G D  +G  KWTG+RVDLVF SNS+LRAL EVY
Msmeg_MC2_155    652  TTDFFVNLLDMSTAWKPSPADDGTYIGTDRATGSPKWTGTRVDLVFASNSQLRALAEVYA  711

Mtb_H37Rv       713  ADDAQPKFVQDFVAAWDKVMNLDRFDVR  740
                      DD++ KFV+DFVAAW KVM+ DRFDV
Msmeg_MC2_155    712  EDDSKEKFVKDFVAAWTKVMDADRFDVA  739
```

Figure 5-8. Global alignment between *Mtb* H37Rv KatG and *M.smegmatis* MC2 155 KatG3.

Global protein alignment between *M. smegmatis* MC2 155 and *Mtb* H37Rv KatG proteins using the Needleman–Wunsch algorithm presenting a percentage identity of 71%. Highlighted in yellow is the key residue involved in INH binding in *Mtb*. The alignment was performed using the Needleman–Wunsch algorithm on NCBI (Needleman and Wunsch, 1970).

5.3.5.1.3 General description of the INH *inhA* resistance locus

InhA is the target of the activated INH drug and mutations in InhA result in an inability of INH to bind to InhA and block mycolic acid biosynthesis (Unissa *et al.*, 2016b). To identify whether the InhA proteins of *Mtb* and *M.* JERRR01 were conserved, the InhA protein of *M.* JERR01 was aligned to the InhA protein of *Mtb* H37Rv (Figure 5-9). This alignment showed that the two proteins were 86% identical and that regions associated with resistance in clinical isolates of *Mtb,* such I21T, S94A, and I194T (Kandler *et al.*, 2018) were highly conserved between both species (Figure 5-9). The overall conservation of InhA suggested that similar mutations as in *Mtb* might lead to INH resistance in *M.* JERR01.

```
Score:1190, Identities:230/269(86%), Positives:253/269(94%), Gaps:1/269(0%)

Mkomo_JERR01  1    MT-LLQGKRILVTGIITDSSIAFYIAKVAQEAGAEVICTGFNRLRLIERILERLPSKPPL  59
                   MT LL GKRILV+GIITDSSIAF+IA+VAQE GA+++ TGF+RLRLI+RI +RLP+K PL
Mtb_H37Rv     1    MTGLLDGKRILVSGIITDSSIAFHIARVAQEQGAQLVLTGFDRLRLIQRITDRLPAKAPL  60

Mkomo_JERR01  60   LELDVQDDKHLDTLADRVTEVIGEGNKLDGVVHSIGFMPQTGMGINPFFDAPYEDVAKGI  119
                   LELDVQ+++HL +LA RVTE IG GNKLDGVVHSIGFMPQTGMGINPFFDAPY DV+KGI
Mtb_H37Rv     61   LELDVQNEEHLASLAGRVTEAIGAGNKLDGVVHSIGFMPQTGMGINPFFDAPYADVSKGI  120

Mkomo_JERR01  120  HISAYSYASLAKAVLPVMNPGGGIVGMDFDPTRAMPAYNWMTVAKSALESVNRFVAREAG  179
                   HISAYSYAS+AKA+LP+MNPGG IVGMDFDP+RAMPAYNWMTVAKSALESVNRFVAREAG
Mtb_H37Rv     121  HISAYSYASMAKALLPIMNPGGSIVGMDFDPSRAMPAYNWMTVAKSALESVNRFVAREAG  180

Mkomo_JERR01  180  KVGVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQMQLLEEGWDQRAPLGWNMKDPTPVAK  239
                   K GVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQ+QLLEEGWDQRAP+GWNMKD TPVAK
Mtb_H37Rv     181  KYGVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQIQLLEEGWDQRAPIGWNMKDATPVAK  240

Mkomo_JERR01  240  TVCALLSDWLPATTGTVIFADGGASTQLL  268
                   TVCALLSDWLPATTG +I+ADGGA TQLL
Mtb_H37Rv     241  TVCALLSDWLPATTGDIIYADGGAHTQLL  269
```

Figure 5-9. Global alignment between *M.* JERR01 and *Mtb* H37Rv InhA proteins.
Global protein alignment between *M.* JERR01 and *Mtb* H37Rv InhA proteins using the Needleman–Wunsch algorithm presenting a percentage identity of 86%. Highlighted in yellow are key residues involved in INH binding in *Mtb*. The alignment was performed using the Needleman–Wunsch algorithm on NCBI (Needleman and Wunsch, 1970). The alignment was performed using the Needleman–Wunsch algorithm on NCBI (Needleman and Wunsch, 1970).

In addition to the protein sequence of InhA, resistance mutations to INH have also been identified in the intergenic region between *fabG* and *inhA* of *Mtb* H37Rv. This region encodes for the promoter of *inhA,* and promoter mutations resulting in increased *inhA* expression are known to increase resistance to INH, accounting for 6% to 43% of INH resistant cases in *Mtb* (Seifert *et al.*, 2015).

5.3.5.1.4 Identification of *katG and inhA* mutants in INH resistant colonies

Unlike RIF mutants, the mutations observed in INH mutants were more diverse, with deletions, insertions (indels), and SNPs present (Table 5-3). Thirty-two (32) of INH mutants (62.5%), showed deletions of either part of *katG* gene, the full *katG* gene, or larger deletions comprising *katG* and the surrounding genes (Table 5-3). In 16 mutants (33.3%) of INH mutants, large insertions in *katG* or its negative regulator-encoding gene *furA* (Table 5-3) were observed. Out of the 48 mutants, only one INH mutant (mutant 8) had mutations in its *inhA* gene resulting in an amino acid change (I24S). However, this is not a mutation associated with INH resistance in *Mtb* (Bakhtiyariniya *et al.*, 2022). To confirm whether this InhA mutation was likely to result in high-level INH resistance or whether this was caused by a separate mutation in INH resistance-related genes, the *katG* region of this mutant was analysed in more detail. Visualising the coverage and mapping of sequencing reads using the IGV browser, it was observed that the read coverage was interrupted abruptly before the beginning of the *katG* sequence. Since sequencing was performed using paired end reads, the location of the matched pair for each sequencing read was identified, which was interrupted on either side of the coverage break. Neither of the reads mapped to the other side of the interruption but rather to a long insertion sequence, indicating disruption of the transcriptional and

regulatory context of *katG* and *furA*. The insertion likely disrupts the ribosomal binding site (RBS) of *katG*, resulting in an absence of protein expression. Furthermore, the length of the insertion sequence could also impact the level and regulation of transcription of the *katG* mRNA as the FurA binding site was likely separated from *katG* and since transcription of *katG* is predicted to originate from the *furA* promoter. Therefore, INH resistance of mutant 8 was likely a consequence of loss of KatG expression resulting in an inability to convert the prodrug into its active form.

Seven mutants (14.6%) were observed to have SNPs present, when looking at their sequence on IGV, no deletions nor insertions were observed in INH resistant related genes such as *katG, inhA* nor *furA*, suggesting that these SNPs are leading to resistance in these mutants (Table 5-3).

The conclusion from this was that the genetic basis INH resistance in *M.* JERR01 is more diverse than RIF resistance, including, insertions, SNPs, and deletions. With deletions of the *katG* being the most common cause for INH resistant (60.4% of mutants) in *M.* JERR01. Insertions within *katG* and the regulator *furA* (18.8% mutants), and even disruptions can lead to INH resistance, although this last example was only observed in one case, suggesting being an uncommon occurrence. While mutations in *inhA* were rare, SNPs in INH resistance-related genes were observed in a small portion (14.6%) of mutants, suggesting they might play a role but require further investigation. For some mutants (INH mutants 7, 9, 12, 13, 18, and 19), the mutation responsible for INH resistance was not determined. The mutation was not called by the variant calling analysis programs and when analysing the mutants' genome in IGV, no insertions or

deletions were noted in the key genes *katG* and *inhA*. It is possible that mutations in other parts of the genome, that were not detected by the analysis undertaken, could contribute to the resistance. This study highlights the diverse mechanisms of INH resistance in *M.* JERR01 and the importance of analysing gene expression and regulatory regions for a complete understanding of drug resistance.

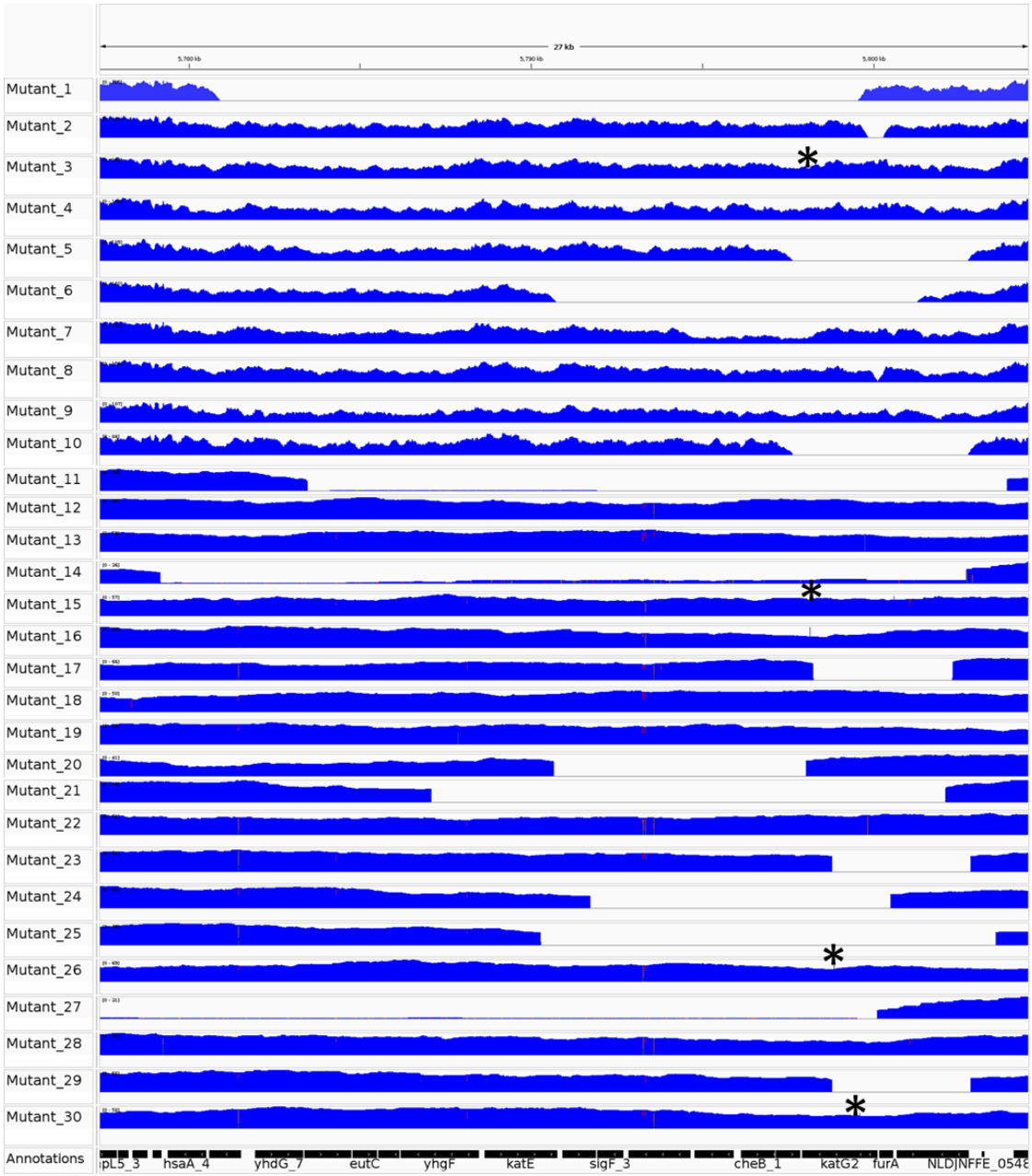
Table 5-3. Mutations observed in 48 INH resistant mutants.

Selective pressure	INH mutants	MIC µg/ml	Start	End	Gene	Mutation Positive strand
512 µg/ml	mutant 1	>512 µg/ml	5,780,914	5,799,550	*katG*	18,636 bp deletion
	mutant 2	>512 µg/ml	5,799,858	5,800,298	*katG*	440 bp deletion
	mutant 3	>512 µg/ml	5,800,105	5,800,107	*katG*	2 bp insertion
	mutant 4	>512 µg/ml	5,800,105	5,800,107	*katG*	2 bp deletion
	mutant 5	>512 µg/ml	5,797,636	5,802,765	*katG*	5,129 bp deletion
	mutant 6	>512 µg/ml	5,790,727	5,801,288	*katG*	10,561 bp deletion
	mutant 7	>512 µg/ml				Undefined
	mutant 8	>512 µg/ml	72,505	72,505	*inhA*	I24A (CGA -> CGC)
	mutant 8	>512 µg/ml	5,800,091	5,800,141	*katG/furA*	Deletion of starting codon on *katG* and deletion of RBS
	mutant 9	>512 µg/ml				Undefined
128 µg/ml	mutant 10	>512 µg/ml	5,797,644	5,802,757	*katG*	5,113 bp deletion
	mutant 11	>512 µg/ml	5,783,467	5,803,900	*katG/furA*	20,433 bp deletion
	mutant 12	>512 µg/ml				Undefined
	mutant 13	>512 µg/ml				Undefined
	mutant 14	256 µg/ml	5,779,176	5,802,707	*katG*	23,531 bp deletion
	mutant 15	256 µg/ml	5,800,595	5,801,970	*furA*	1,420 bp insertion
	mutant 16	>512 µg/ml	5,789,049	5,798,153	*katG*	9,104 bp deletions
	mutant 16	>512 µg/ml	5,789,155	5,800,567	*furA*	11,412 bp deletion

mutant 17	>512 µg/ml	5,798,239	5,802,312	*katG/furA*	4,072 bp deletion
mutant 18	>512 µg/ml				Undefined
mutant 19	>512 µg/ml				Undefined
mutant 20	>512 µg/ml	5,790,661	5,798,031	*katG*	7,370 bp deletion
mutant 21	>512 µg/ml	5,787,085	5,802,107	*katG*	15,022 bp deletion
mutant 22	>512 µg/ml	5,799,840	5,799,840	*katG*	D99H (GAC -> CAC)
mutant 23	>512 µg/ml	5,798,790	5,802,837	*katG*	4,047 bp deletion
mutant 24	>512 µg/ml	5,791,726	5,800,509	*katG/furA*	8,783 bp deletion
mutant 25	>512 µg/ml	5,790,278	5,803,571	*katG/furA*	13,293 bp deletion
mutant 26	>512 µg/ml	5,798,848	5,800,170	*katG*	1,322 bp insertion
mutant 27	>512 µg/ml	5,774,130	5,804,298	*katG*	30,168 bp deletion
mutant 28	>512 µg/ml	5,796,752	5,800,863	*furA/katG*	4,111 bp deletion
mutant 29	>512 µg/ml	5,798,790	5,802,837	*katG/furA*	4,047 bp deletion
mutant 30	>512 µg/ml	5,799,605	5,801,019	*katG*	1,414 bp insertion
mutant 31	256 µg/ml	5,800,509	5,801,876	*furA/katG*	1,367 bp insertion
mutant 32	256 µg/ml	5,779,976	5,803,544	*katG*	23,568 bp deletion
mutant 33	>512 µg/ml	5,798,663	5,798,663	*katG*	A491L (GCG -> GAG)
mutant 34	>512 µg/ml	5,799,287	5,799,287	*katG*	L283P (CAG -> CGG)
mutant 35	256 µg/ml	5,7943,43	5,803,705	*katG*	9,362 bp deletion
mutant 36	256 µg/ml	5,800,145	5,800,145	intergenic region *katG/furA*	(CTT -> CTG)
mutant 37	128 µg/ml	5,799,526	5,799,526		W330C (CCA -> TCA)
mutant 37	128 µg/ml	5,790,662	5,798,030	*katG*	7,369 bp deletion
mutant 38	>512 µg/ml	5,790,487	5,803,109	*katG*	12,622 bp deletion

mutant 39	128 µg/ml	5,779,657	5,803,530	*katG*	23,873 bp deletion
mutant 40	512 µg/ml	5,798,885	5,800,111	*katG*	1,226 bp deletion
mutant 41	>512 µg/ml	5,797,766	5,807,572	*katG*	9,806 bp deletion
mutant 42	>512 µg/ml	5,800,573	5,801,976	*furA*	1,403 bp insertion
mutant 43	>512 µg/ml	5,799,686	5,800,675	*katG*	989 bp deletion
mutant 44	>512 µg/ml	5,798,217	5,799,638	*katG*	1,421 bp insertion
mutant 45	>512 µg/ml	5,798,502	5,799,797	*katG*	1,295 bp insertion
mutant 46	>512 µg/ml	5,781,667	5,799,434	*katG*	17,767 bp deletion
mutant 47	>512 µg/ml	5,799,585	5,799,585	*katG*	T330P (CTG -> CTA)
mutant 48	>512 µg/ml	5,7998,46	5,801,181	*katG*	1,335 bp insertion

This table details the selective pressures applied to generate *M.* JERR01 INH mutants, alongside the minimum inhibitory concentration (MIC_{90}) of INH for each mutant, the identified mutations, including deletions, insertions, and single nucleotide polymorphisms (SNPs).


27 kb
5,700 kb
5,790 kb
5,800 kb
Mutant_1
Mutant_2
Mutant_3
Mutant_4
Mutant_5
Mutant_6
Mutant_7
Mutant_8
Mutant_9
Mutant_10
Mutant_11
Mutant_12
Mutant_13
Mutant_14
Mutant_15
Mutant_16
Mutant_17
Mutant_18
Mutant_19
Mutant_20
Mutant_21
Mutant_22
Mutant_23
Mutant_24
Mutant_25
Mutant_26
Mutant_27
Mutant_28
Mutant_29
Mutant_30
Annotations
hsaA_4
yhdG_7
eutC
yhgF
katE
sigF_3
cheB_1
katG2
furA

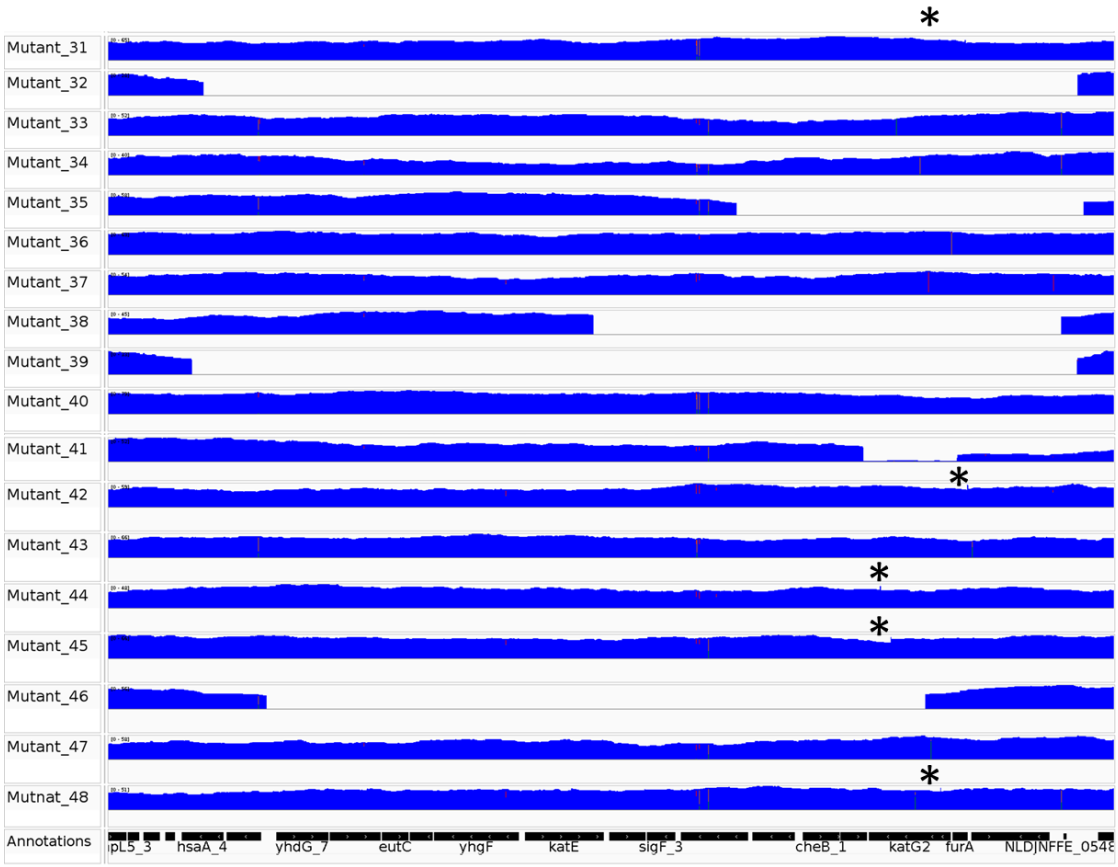


Figure 5-10. IGV coverage profile of 48 *M.* JERR01 INH mutants (deletions, insertions, and SNPs).

This figure illustrates the coverage profile across a set of 48 *M.* JERR01 mutants containing several types of mutations: insertions, deletions, and single nucleotide polymorphisms (SNPs) with the asterisk representing insertions in the different mutants.

5.4 Discussion

In this chapter, mutants of *M.* JERR01 were isolated with respect to the two most important antitubercular drugs used in the treatment of drug-susceptible TB, RIF and INH. Furthermore, a comparison of the mutation frequency reported of *M.* JERR01 for RIF and INH was conducted, and this has not been previously investigated using this organism. The main target for RIF is the beta subunit of the RNA polymerase, which plays an essential role in bacterial transcription (Monama, Olotu and Tastan Bishop, 2023). These mutations involved specific changes in specific amino acids that alter the binding interactions between the *M.* JERR01 RpoB protein and RIF. Since specific changes are required to happen to develop resistance in *M.* JERR01 towards RIF, these events can be considered as rare events, and thus, producing a low RIF mutation frequency.

Among the mutations observed in *M.* JERR01, the most common were: H441Y, S437P, H441A, and S446L, respectively (Figure 5-11), which are also common mutation sites for RIF resistant *Mtb*. The mutation S446L is the most common mutation in clinical practice, followed by the mutation H441Y in *Mtb* (Jagielski *et al.*, 2018; Nontuthuko E. Maningi *et al.*, 2018) and these mutations were observed in *M.* JERR01 RIF mutants. The mutation S450L is also associated with higher levels of resistance to RIF in *Mtb* clinical isolates, while in our study, the mutation S446L was associated with only one strain being resistant to ≥100 µg/ml and 5 strains resistant to 12.5 µg/ml of RIF. Even with the same mutation, variations on phenotypic expressions are possible, this could be due to background mutations happening in other places (even outside the *rpoB* gene) that can result in variations in gene expression and protein folding, thus leading to MIC_{90} variations. In *Mtb* it is known that compensatory mutations in the *rpoA* and *rpoC* can

lead to variations in RIF MIC_{90}, however, when examining these genes in *M.* JERR01, no mutations were observed in these genes, suggesting that additional changes in the *M.* JERR01 genome could be responsible for the observed resistance level variations. When examining the genome to find alternative mutations that could explain the difference in phenotypes, different SNPs were identified outside the *rpoB* gene, however, these same SNPs were also present in the reference genome used for variant calling. Since Oxford Nanopore sequencing is also prone to errors on raw sequences, some of these reported SNPs might not be products of mutations and instead just programmed errors of the sequencing. Thus, identifying additional factors that contributed to different phenotypes was not possible to be performed at this moment. A future approach to identify compensatory mutations in *M.* JERR01 would be to sequence key mutants (mutants that contained same mutations but different phenotypes) using short read Illumina sequencing (higher sequencing accuracy) (Delahaye and Nicolas, 2021; Linde *et al.*, 2023).

INH was used for evolution experiments despite the high MIC_{90} reported for INH, as the prevalence of INH-resistance among TB patients is higher than the prevalence of RIF resistance (Dean *et al.*, 2020). Since *M.* JERR01 had a high MIC_{90} towards INH and its intrinsic resistance, a higher mutation frequency compared to what it was observed in RIF was expected. This was indeed observed (10^{-5} mutation/total bacterial cells) at a concentration of 512 µg/ml of INH per selective plate. The difference observed in the two mutation frequencies for these two antibiotics can be due to their targets and their different mutation requirements. In the case of INH, this antibiotic depends on the actions of KatG, which is the protein responsible for the conversion of the prodrug INH

into its active NAD+ and NADP+ adducts, which in turn inhibit the actions of InhA (essential for mycolic acid synthesis)(Lee *et al.*, 2018). Since resistance towards INH can arise from different mechanisms such as SNPs, deletions, and insertions events within the *katG* gene, a specific amino acid change in a specific region is not required for developing resistance, thus, leading to more frequent mutation events, therefore, contributing to a higher mutation frequency when compared to RIF.

The deletion of *katG* observed with *M.* JERR01 was observed *in vitro* studies using *Mtb* H37Rv. In the study performed by Bergval and colleagues (2009), *Mtb* H37Rv was exposed to high concentrations of INH *in vitro* 0.1 µg/ml, 0.4 µg/ml, and 20 µg/ml (which corresponds to 0x, 4x and 200x the MIC for *Mtb* H37Rv respectively) (Namouchi *et al.*, 2017), to obtain an INH mutation rate using a laboratory strain. This resulted in a mutation rate of 3.2×10^{-7} mutations per cell division for INH (Bergval *et al.*, 2009). It was also noted in this experiment that *katG* was partially or fully deleted, and neither of the clinically relevant mutations observed in *Mtb in vivo* were found in *Mtb* H37Rv *in vitro* (S315T nor mutations in the promoter of InhA)(Bergval *et al.*, 2009). These results are interesting considering that *katG* gene deletions are a rare occurrence in clinical isolates (De Maio *et al.*, 2021). Therefore, *M.* JERR01 INH resistance *in vitro* model is not a good representation of what happens in *Mtb in vivo*.

5.5 Summary

This chapter investigated the potential of *M.* JERR01 as a model organism for studying drug resistance development against antitubercular drugs RIF and INH. The specific aims of this chapter were to:

1. Identify the conditions required for the selection of antibiotic resistant mutants in solid media.
2. Determine the mutation frequency of RIF and INH resistant mutants.
3. Characterize the level of AMR of the resistant *M.* JERR01 mutants.
4. Identify the genetic changes leading to resistance in a selection of resistant mutants and assess their potential functional impact.

Both *M.* JERR01 and *Mtb* display a low mutation frequency towards RIF drug due to the specific mutation requirements for the antibiotic. Mutations in the *rpoB* gene, specifically in the RIF Resistance Determining Region (RRDR-I), were identified in RIF-resistant *M*. JERR01 mutants. The most common mutations observed in *M.* JERR01 (H441Y, S437P, H441A, and S446L) are also reported in RIF-resistant *Mtb* isolates. However, compensatory mutations in *rpoA* and *rpoC* genes, known to influence RIF resistance levels in *Mtb*, were not observed in *M.* JERR01 mutants. This suggests that other genetic factors might be responsible for the observed variations in resistance levels among *M.* JERR01 mutants, however due to time restrictions, these factors were not addressed. Future experiments using more accurate technologies, such as Illumina sequencing, could help identify these additional compensatory mutations.

A high mutation frequency towards INH was observed compared to RIF, likely due to the different mechanisms of action and mutation requirements for this antibiotic. Sixty percent (60%) of *M.* JERR01 mutants displayed partial/full deletions in the *katG* gene, which is not a common resistance mechanism observed in clinical isolates of *Mtb*. This

suggests that the *M.* JERR01 INH resistance model might not accurately reflect INH resistance in *Mtb* isolates.

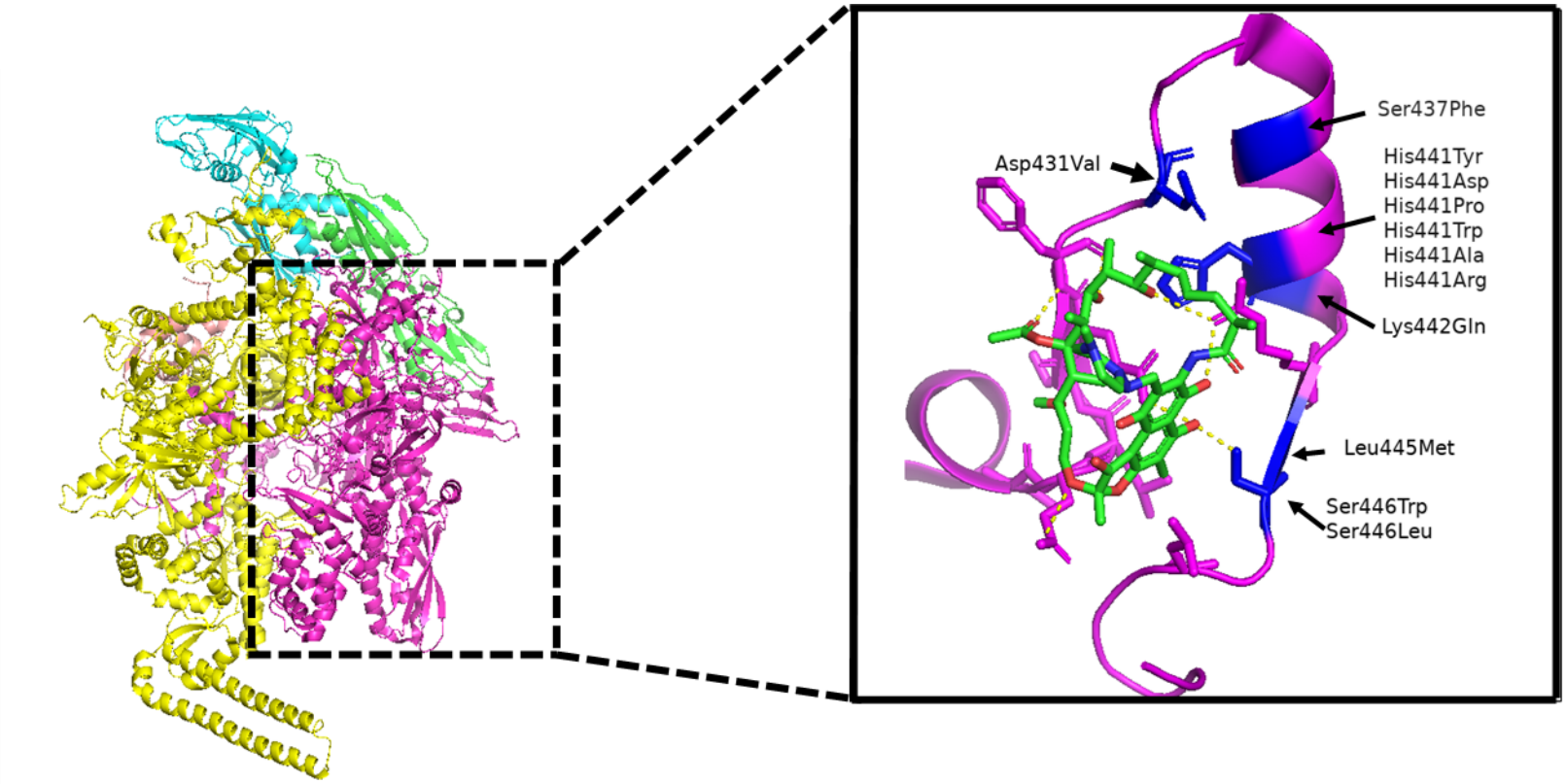


Figure 5-11. The predicted structure of the RIF binding pocket in *M.* JERR01 RpoB RRDR-I region (residues 419-453).
Picture designed by using Pymol and the RpoB protein structure of *Mtb* H37Rv to identify the beta subunit (PDB ID:5UHC) (Borukhov and Nudler, 2003; Wei *et al.*, 2017). To this model the RRDR-I cluster depicted on pink of *M.* JERR01 was added. The RRDR region of *M.* JERR01 is shown with the RIF stick model (shown in green) bound to the RRDR-I by hydrogen bonding interactions. In blue and labelled, are all the mutations observed in *M.* JERR01 RIF mutants.

6 General Discussion

6.1 General Discussion

Mycobacteria models used for *Mtb* studies are important to improve understanding of the biology of *Mtb,* including its resistance mechanisms against antitubercular drugs. The rising incidence, and public health threat caused by drug resistant TB increases the urgency to find appropriate model organisms for this purpose. The overarching aim of this thesis was to examine whether a clinically isolated mycobacterial strain, previously characterised as *M. komossense* obtained from SMRL (strain JERR01), exhibited properties that would make it a suitable model organism for *Mtb* drug resistance studies.

The thesis focused on 3 main areas:

1. Phenotypic characterization of *M.* JERR01 and comparison with the organism initially believed to be its closest relative, *M. komossense* ATCC 33013

2. Genotypic characterization of *M.* JERR01 and genotypic comparison with other mycobacteria species, including other non-tuberculous bacteria and members of the MTBC

3. Selection experiments using RIF and INH to demonstrate the evolution of antibiotic resistance in *M*. JERR01 in comparison with prior observations in *Mtb.*

6.2 Phenotypic characterization of *M*. JERR01

The phenotypic characterization of this novel organism was an important first step as the physical appearance and growth kinetics of this organism had not previously been described or related to a reference strain of *M. komossense*, the species which it was

initially believed to represent. The main findings from Chapter 3 demonstrate substantial differences between *M*. JERR01 and the type strain, *M. komossense* ATCC 33013. *M.* JERR01 produced sharp-edged colonies and a velvety bright yellow pigment when grown at 30°C and 37°C while *M. komossens*e ATCC 33013 produced more rounded colonies with no pigment and did not grow at 37°C. Overall, these phenotypic observations suggested that *M.* JERR01 and ATCC 33013 might only be distantly related and paved the way for more detailed genomic studies to carefully interrogate the speciation of *M.* JERR01.

As the main goal of this thesis was to determine whether *M.* JERR01 was a suitable model organism for studying the development of antimicrobial resistance to antitubercular drugs used in the treatment of *Mtb*, it was particularly important to delineate the AST profile of the new organism. As a starting point, MIC_{90} profiles were calculated for nine drugs of interest. The specific drugs chosen were selected based on their current importance in the management of DS-TB and DR-TB. Both well-established and recently introduced antibiotics were considered. For six drugs (RIF, EMB, MOXI, LEVO, BDQ, and LZD), the MIC_{90} was below the currently established clinical breakpoints or critical concentration for *Mtb.* This finding was encouraging, as model organisms are more reliable if molecular targets for antibiotic susceptibility and mechanisms of resistance are similar to the pathogen of interest. Closely matched AST profiles suggest that this is more likely to be the case.

Notably, for other important drugs (INH, PZA, and PA), the MIC_{90} profile of *M.* JERR01 suggested intrinsic resistance when compared to clinical breakpoints or critical

concentrations for *Mtb*. This suggests that *M.* JERR01 may be less useful as a model organism for antibiotic resistance studies of these drugs. However, particularly in relation to INH, it should be noted that *M. smegmatis* has been successfully used to improve understanding of the effect of antibiotic exposure on mycobacteria, even though it is also intrinsically resistant to INH (Aldridge *et al.*, 2012; Wakamoto *et al.*, 2013). Therefore, model organisms can still sometimes be effectively used provided that their biological properties and limitations are well understood. Additionally, at least 8% of clinical TB cases worldwide are caused by INH mono-resistant *Mtb* (Worawat Chumpangern, 2018), and having a rapidly growing *in vitro* model organism that replicates this situation may be valuable.

My phenotypic characterization was targeted rather than comprehensive because of limited time and resources. Had the opportunity been available, I would have sought to extend this work by performing a detailed biochemical investigation of *M.* JERR01, including the study of the optimal energy sources (e.g., carbon or nitrogen) and comparison of metabolic pathways between this organism and *Mtb*.

6.3 Genomic characteristics of *M*. JERR01

M. JERR01 was initially characterized as *M. komossense* after sequencing its *hsp65* gene in 2010 when it was isolated from a patient sample. However, the *in silico* analysis performed in Chapter 4 on the whole *M.* JERR01 genome demonstrated that the original species designation was imprecise. This likely reflects the limited sequence data available to compare the *hsp65* gene to at the time of its initial isolation and characterization. Since 2010, there has been a massive expansion in the amount of genomic data in the public sequence databases, which has also been matched by broader taxonomic

coverage. Reappraisal of the taxonomic status of *M.* JERR01 fourteen years later, with the benefits of this increased context and genome data, has provided a detailed characterization of the genetic makeup of the strain, but its species identification remains elusive.

Whilst *M*. JERR01 is an unclassified mycobacterial species, the characteristics of the genome suggest that it belongs to the *Fortuitum-Vaccae* clade of environmental mycobacteria. *M.* JERR01 possesses homologs of most AMR-associated genes from *Mtb*, suggesting its potential as a model organism for resistance studies for different *Mtb* drugs. In this analysis, it was also noted that *M.* JERR01 plasmids contained genes coding for ESX-1, a secretion system that is important for virulence and immune modulation in *Mtb* (Samten et al., 2009). After running bioinformatic programs to detect possible virulence factors in JERR01, none were found. However, the presence of annotated ESX-1, 2, 3, 5, and PE/PPE family proteins in the main chromosome and plasmids suggests a deeper look into the function of these proteins is required, as in *Mtb* these proteins have been linked to virulence and pathogenicity (Ates *et al.*, 2016; Newton-Foot *et al.*, 2016; Qian *et al.*, 2020). Although initially considered a BSL-1 organism due to its presumed identity as *M. komossense* (non-pathogenic *Mycobacterium* classified as BSL-1 organism)(Kazda and Muller, 1979), the presence of the ESX secretion system and PE/PPE proteins in *M.* JERR01 necessitates a more cautious approach. Although their functions in *M.* JERR01 remain unknown, their presence introduces some uncertainty regarding the organism's potential virulence. Furthermore, no pathogenicity tests have been performed on this organism yet. Such tests would be crucial in definitively assessing its

ability to cause disease. In the absence of definitive data and due to the organism's origin from a human host, I have decided to classify *M.* JERR01 as a BSL-2 organism as a precautionary measure. This current classification could be revisited based on the results of forthcoming pathogenicity tests and any additional characterization of *M.* JERR01.

Due to time restrictions, accessory genes of *M*. JERR01 were not studied, but in future experiments, the comparison between accessory genes between mycobacteria could shed light on how this organism adapted to specific environments, such as the human host from which it was originally isolated. Allied to this, experimental validation through protein activity assays or targeted gene knockouts would strengthen the conclusions drawn from computational analyses on antibiotic resistance mechanisms.

6.4 Evolution of *M*. JERR01 for rifampicin and isoniazid mutants.

Following the genomic characterisation of *M.* JERR01, the final area of experimentation turned to the exploration of the evolution of resistance to RIF and INH. Mutant selection experiments, based on the study of *M.* JERR01 colonies grown under RIF selective pressure, showed that the mutation frequency of *M*. JERR01 was close to the mutation frequency previously reported for *Mtb* (Billington, Mchugh, and Gillespie, 1999), and that mutations observed in the RRDR-I region of the *rpoB* gene in *M.* JERR01, were similar to mutations previously observed in *Mtb* (Taniguchi *et al.*, 1996a; Cavusoglu, Karaca-Derici and Bilgic, 2004; Koch, Mizrahi and Warner, 2014; Hameed *et al.*, 2017; Monama, Olotu and Tastan Bishop, 2023). These results provided further support for the argument that *M.* JERR01 could be used for RIF resistance studies. As RIF is still considered as the single most important drug for the effective treatment of DS-TB, a new

model organism to facilitate a better understanding of how its resistance mechanisms evolve could be very advantageous.

Since some RIF mutants with similar *rpoB* sequences presented different anti-microbial resistance phenotypes (differences in MIC_{90}), I would have liked to identify potential compensatory mutations in other regions of the genome and map them to these phenotypic differences. However, due to time restrictions, this analysis could not be performed but would be a key area of future work.

Conversely, mutant selection experiments based on the study of *M.* JERR01 colonies grown under INH selective pressure showed a different mutation frequency and different emergence of mutations compared to observations previously reported from *in vivo* and clinical studies of *Mtb*. This may be due to biological differences between the organisms or differences between *in vitro* and *in vivo* experimental conditions. *In vivo,* KatG is a key determinant for the survival of *Mtb* in macrophages. Deletions of KatG, and inactivation of its function or expression are therefore unlikely to be selected for during *in vivo* conditions (Master *et al.*, 2001). Since an immune response cannot be replicated *in vitro* and, therefore, was not a limiting selective pressure, this may have led to the deletion of the *katG* gene in most of the INH mutants. In other words, KatG is not essential for *in vitro* growth. The fact that *M.* JERR01 is intrinsically resistant to INH meant that to select for further resistance, experiments in this area required much higher INH exposure than would be used clinically, and this may also have influenced the range of mutations observed. Rather than selecting for mutations of the INH drug target, most of the mutants identified were unable to process the prodrug to its active form. This high initial

selective pressure could have prevented the identification of functional mutations in INH target genes. Overall, the evidence that *M.* JERR01 could be used as a model for INH resistance studies is less strong than for RIF.

It would have been useful to repeat the mutation selection experiments done for RIF and INH with the other anti-tubercular antibiotics which were evaluated in Chapter 3 (e.g., the fluoroquinolones and new drugs such as BDQ). Once again, time and resource limitations prevented me from undertaking this so far, but it is an important area for future work.

Overall, this thesis has made an important contribution to our understanding of how mycobacteria are characterised and speciated, and considers how non-tuberculous mycobacteria may be used as model organisms for antibiotic resistance studies in *Mtb*. The phenotypic and genotypic descriptive work showed that the *bacterium* initially identified as *M. komossense* JERR01 is likely to belong to another (perhaps entirely new) species, highlighting the extent to which advances in WGS in the last fourteen years have enabled us to examine and describe micro-organisms more precisely. The examination of the antibiotic susceptibility profile of *M.* JERR01 and the mutations which underpin that has illustrated that *M.* JERR01 could be a useful model for *Mtb* but that there are some limitations.

Finally, it is essential that work on novel non-pathogenic, rapidly growing, non-tuberculous mycobacteria and their similarity to *Mtb* continues, so that pre-clinical research to develop and optimise new treatments can be conducted quickly and

effectively. This endeavour is an often neglected but essential activity in the global effort to tackle TB.

7 References

Adékambi, T., Drancourt, M. and Raoult, D. (2009) 'The *rpoB*', *Trends in Microbiology*, 17(1), pp. 37–45. Available at: https://doi.org/10.1016/j.tim.2008.09.008.

Aldridge, B.B. *et al.* (2012) 'Asymmetry and Aging of Mycobacterial Cells Lead to Variable Growth and Antibiotic Susceptibility', *Science (New York, N.Y.)*, 335(2006), pp. 100–104.

Van Alen, I. *et al.* (2021) 'The G132S Mutation Enhances the Resistance of *Mycobacterium tuberculosis* β-Lactamase against Sulbactam', *Biochemistry*, 60, pp. 2236–2245. Available at: https://doi.org/10.1021/acs.biochem.1c00168.

Andries, K. *et al.* (2005) 'A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*', *Science*, 307(5707), pp. 223–227. Available at: https://doi.org/10.1126/science.1106753.

Ang, M.L.T. *et al.* (2017) 'EthA/R-independent killing of *Mycobacterium tuberculosis* by ethionamide', *Frontiers in Microbiology*, 8(APR), pp. 1–12. Available at: https://doi.org/10.3389/fmicb.2017.00710.

Anton, V., Rougé, P. and Daffé, M. (1996) 'Identification of the sugars involved in mycobacterial cell aggregation', *FEMS Microbiology Letters*, 144(2–3), pp. 167–170. Available at: https://doi.org/10.1016/0378-1097(96)00356-4.

Arbex, M.A. *et al.* (2010) 'Antituberculosis drugs: Drug interactions, adverse effects, and use in special situations. Part 1: First-line drugs', *J Bras Pneumol.*, 36(June), pp. 626–640. Available at: https://doi.org/S1806-37132010000500016 [pii].

Armocida, E. and Martini, M. (2020) 'Tuberculosis: A timeless challenge for medicine', *Journal of Preventive Medicine and Hygiene*, 61(2), pp. E143–E147. Available at: https://doi.org/10.15167/2421-4248/jpmh2020.61.2.1402.

Aronson, J.D. (1926) 'SPONTANEOUS TUBERCULOSIS IN SALT WATER FISH', *The Journal of Infectious Diseases*, 39(4), pp. 315–320.

Asadian, M. *et al.* (2022) 'Genomic characteristics of two most widely used BCG vaccine strains: Danish 1331 and Pasteur 1173P2', *BMC Genomics*, 23(1), pp. 1–13. Available at: https://doi.org/10.1186/s12864-022-08826-9.

Ates, L.S. *et al.* (2016) 'The ESX-5 System of Pathogenic Mycobacteria Is Involved In Capsule Integrity and Virulence through Its Substrate PPE10', *PLoS Pathogens*, 12(6), pp. 1–26. Available at: https://doi.org/10.1371/journal.ppat.1005696.

Ayukekbong, J.A., Ntemgwa, M. and Atabe, A.N. (2017) 'The threat of antimicrobial resistance in developing countries: Causes and control strategies', *Antimicrobial Resistance and Infection Control*, 6(1), pp. 1–8. Available at: https://doi.org/10.1186/s13756-017-0208-x.

Bachmann, N.L. *et al.* (2020) 'Key Transitions in the Evolution of Rapid and Slow Growing Mycobacteria Identified by Comparative Genomics', *Frontiers in Microbiology*, 10(January), pp. 1–12. Available at: https://doi.org/10.3389/fmicb.2019.03019.

Bakhtiyariniya, P. *et al.* (2022) 'Detection and characterization of mutations in genes related to isoniazid resistance in *Mycobacterium tuberculosis* clinical isolates from Iran', *Molecular Biology Reports*, 49(7), pp. 6135–6143. Available at: https://doi.org/10.1007/s11033-022-07404-2.

Barberis, I. *et al.* (2017) 'The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus', *Journal of Preventive Medicine and Hygiene*, 58(1), pp. E9–E12.

Barrick, J.E. and Lenski, R.E. (2013) 'Genome dynamics during experimental evolution', *Nature Reviews Genetics*, 14(12), pp. 827–839. Available at: https://doi.org/10.1038/nrg3564.

Bergval, I.L. *et al.* (2009) 'Resistant mutants of *Mycobacterium tuberculosis* selected in vitro do not reflect the in vivo mechanism of isoniazid resistance', *Journal of Antimicrobial Chemotherapy*, 64(3), pp. 515–523. Available at: https://doi.org/10.1093/jac/dkp237.

Biadglegne, F. *et al.* (2014) 'Drug resistance of *Mycobacterium tuberculosis* isolates from tuberculosis lymphadenitis patients in Ethiopia', *Indian Journal of Medical Research*, 140(JUL), pp. 116–122.

Bian, S. *et al.* (2021) 'Genetic Structure, Function, and Evolution of Capsule Biosynthesis Loci in Vibrio parahaemolyticus', *Frontiers in Microbiology*, 11(January), pp. 1–16. Available at: https://doi.org/10.3389/fmicb.2020.546150.

Billington, O.J., Mchugh, T.D. and Gillespie, S.H. (1999) 'Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*', *Antimicrobial Agents and Chemotherapy*, 43(8), pp. 1866–1869. Available at: https://doi.org/10.1128/aac.43.8.1866.

Blondeau, J.M. (2004) 'Fluoroquinolones: Mechanism of action, classification, and development of resistance', *Survey of Ophthalmology*, 49(2 SUPPL. 2), pp. 1–6. Available at: https://doi.org/10.1016/j.survophthal.2004.01.005.

Boddinghaus, B. *et al.* (1990) 'Detection and identification of mycobacteria by amplification of rRNA', *Journal of Clinical Microbiology*, 28(8), pp. 1751–1759. Available at: https://doi.org/10.1128/jcm.28.8.1751-1759.1990.

Bonnet, M. *et al.* (2020) 'Bacterial culture through selective and non-selective conditions: the evolution of culture media in clinical microbiology', *New Microbes and New Infections*, 34. Available at: https://doi.org/10.1016/j.nmni.2019.100622.

Borukhov, S. and Nudler, E. (2003) 'RNA polymerase holoenzyme: Structure, function and biological implications', *Current Opinion in Microbiology*, 6(2), pp. 93–100. Available at: https://doi.org/10.1016/S1369-5274(03)00036-5.

Bottaro, Larsen, B. (2010) 'To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis?', *Current Opinion in Microbiology*, 13(1), pp. 86–92. Available at: https://doi.org/10.1016/j.mib.2009.11.006.To.

Briken, V. *et al.* (2018) '*Mycobacterium tuberculosis*: An Adaptable Pathogen Associated With Multiple Human Diseases', *Frontiers in Cellular and Infection Microbiology*, 8(158). Available at: https://doi.org/10.3389/fcimb.2018.00158.

Brites, D. and Gagneux, S. (2015) 'Co-evolution of *Mycobacterium tuberculosis* and Homo sapiens', *Immunological Reviews*, 264(1), pp. 6–24. Available at: https://doi.org/10.1111/imr.12264.

Brown-Elliott, B.A. and Philley, J. V. (2017) 'Rapidly growing mycobacteria', *Tuberculosis and Nontuberculous Mycobacterial Infections*, pp. 703–724. Available at: https://doi.org/10.1128/9781555819866.ch41.

Bush, S.J. *et al.* (2020) 'Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines', *GigaScience*, 9(2), pp. 1–21. Available at: https://doi.org/10.1093/GIGASCIENCE/GIAA007.

C Reygaert, W. (2018) 'An overview of the antimicrobial resistance mechanisms of bacteria', *AIMS Microbiology*, 4(3), pp. 482–501. Available at: https://doi.org/10.3934/microbiol.2018.3.482.

Cambau, E. *et al.* (2019) 'Rationale for defining a reference method for the determination of minimum inhibitory concentrations (MIC) of anti-tuberculous agents for *Mycobacterium tuberculosis* complex', *European Journal of Clinical Microbiology and Infectious Diseases*, (Mic), pp. 1–2.

Cambier, C.J., Falkow, S. and Ramakrishnan, L. (2014) 'Host evasion and exploitation schemes of Mycobacterium tuberculosis', *Cell*, 159(7), pp. 1497–1509. Available at: https://doi.org/10.1016/j.cell.2014.11.024.

Canetti, D. *et al.* (2022) '*Mycobacterium marinum*: A brief update for clinical purposes', *European Journal of Internal Medicine*, 105(July), pp. 15–19. Available at: https://doi.org/10.1016/j.ejim.2022.07.013.

Canetti, G; Froman, S; Grosset, J; Hauduroy, P; Langerová, M; Mahler, H T; Meissner, G; Mitchison, D A; Sula, L. *et al.* (1963) 'Mycobacteria: Laboratory Methods for Testing Drug Sensitivity and Resistance.', *Bull. Wld Hlth Org.*, 29(December 1961), pp. 565–578.

Carter, A.P. *et al.* (2000) 'Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics', *Nature*, 407(6802), pp. 340–348. Available at: https://doi.org/10.1038/35030019.

Carver, T.J. *et al.* (2005) 'ACT: The Artemis comparison tool', *Bioinformatics*, 21(16), pp. 3422–3423. Available at: https://doi.org/10.1093/bioinformatics/bti553.

Case, R.J. *et al.* (2007) 'Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies', *Applied and Environmental Microbiology*, 73(1), pp. 278–288. Available at: https://doi.org/10.1128/AEM.01177-06.

Cavusoglu, C., Karaca-Derici, Y. and Bilgic, A. (2004) 'In-vitro activity of rifabutin against rifampicin-resistant *Mycobacterium tuberculosis* isolates with known rpoB mutations', *Clinical Microbiology and Infection*, 10(7), pp. 662–665. Available at: https://doi.org/10.1111/j.1469-0691.2004.00917.x.

Clary, G. *et al.* (2018) '*Mycobacterium abscessus* smooth and rough morphotypes form antimicrobial-tolerant biofilm phenotypes but are killed by acetic acid', *Antimicrobial Agents and Chemotherapy*, 62(3), pp. 1–17. Available at: https://doi.org/10.1128/AAC.01782-17.

Cohen, K.A., Bishai, W.R. and Pym, A.S. (2015) 'Molecular basis of drug resistance in *Mycobacterium tuberculosis*', *Molecular Genetics of Mycobacteria*, pp. 411–429. Available at: https://doi.org/10.1128/9781555818845.ch21.

Cole, S.T. *et al.* (1998) 'Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence', *Nature*, 393(NOVEMBER), pp. 537–544.

Comella-del-Barrio, P. *et al.* (2021) 'Impact of COVID-19 on Tuberculosis Control', *Archivos de Bronconeumologia*, 57, pp. 5–6. Available at: https://doi.org/10.1016/j.arbres.2020.11.016.

Costa, S.S. *et al.* (2020) 'First Steps in the Analysis of Prokaryotic Pan-Genomes', *Bioinformatics and Biology Insights*, 14, pp. 1–9. Available at: https://doi.org/10.1177/1177932220938064.

Covarrubias, A.J. *et al.* (2021) 'NAD+ metabolism and its roles in cellular processes during ageing', 22(2), pp. 119–141. Available at: https://doi.org/10.1038/s41580-020-00313-x.NAD.

Van Crevel, R. *et al.* (2011) 'Innate immune recognition of *mycobacterium tuberculosis*', *Clinical and Developmental Immunology*, 2011. Available at: https://doi.org/10.1155/2011/405310.

Daniel, T.M. (2006) 'The history of tuberculosis', *Respiratory Medicine*, 100(11), pp. 1862–1870. Available at: https://doi.org/10.1016/j.rmed.2006.08.006.

D'Arcy Hart, P. and Armstrong, J.A. (1974) 'Strain virulence and the lysosomal response in macrophages infected with Mycobacterium tuberculosis', *Infection and Immunity*, 10(4), pp. 742–746. Available at: https://doi.org/10.1128/iai.10.4.742-746.1974.

Datta, A. (2021) 'Determination of Viable Microbial Count Present in Tap Water', *International Journal of Innovative Science and Research Technology*, 6(4), pp. 19–28. Available at: https://doi.org/10.1520/STP36000S.

Daugelat, S. *et al.* (2003) 'The RD1 proteins of *Mycobacterium tuberculosis*: Expression in *Mycobacterium smegmatis* and biochemical characterization', *Microbes and Infection*, 5(12), pp. 1082–1095. Available at: https://doi.org/10.1016/S1286-4579(03)00205-3.

Dean, A.S. *et al.* (2020) 'Prevalence and genetic profiles of isoniazid resistance in tuberculosis patients: A multicountry analysis of cross-sectional data', *PLoS Medicine*, 17(1), pp. 1–13. Available at: https://doi.org/10.1371/JOURNAL.PMED.1003008.

Delahaye, C. and Nicolas, J. (2021) 'Sequencing DNA with nanopores: Troubles and biases', *PLoS ONE*, 16(10 October). Available at: https://doi.org/10.1371/journal.pone.0257521.

Delaney, O., Letten, A.D. and Engelstädter, J. (2023) 'Frequent, infinitesimal bottlenecks maximize the rate of microbial adaptation', *Genetics*, 225(4), pp. 1–6. Available at: https://doi.org/10.1093/genetics/iyad185.

Delogu, G., Sali, M. and Fadda, G. (2013) 'The Biology of *Mycobacterium Tuberculosis* Infection', *Citation: Mediterr J Hematol Infect Dis*, 2013(1), p. 2013070. Available at: https://doi.org/10.4084/MJHID.2013.070.

DeVito, J.A. and Morris, S. (2003) 'Exploring the structure and function of the mycobacterial KatG protein using trans-dominant mutants', *Antimicrobial Agents and Chemotherapy*, 47(1), pp. 188–195. Available at: https://doi.org/10.1128/AAC.47.1.188-195.2003.

Dias, L.M. *et al.* (2018) 'Genomic architecture of the two cold-adapted genera *Exiguobacterium* and *Psychrobacter*: Evidence of functional reduction in the *Exiguobacterium antarcticum* B7 genome', *Genome Biology and Evolution*, 10(3), pp. 731–741. Available at: https://doi.org/10.1093/gbe/evy029.

Dong, M.J., Luo, H. and Gao, F. (2022) 'Ori-Finder 2022: A Comprehensive Web Server for Prediction and Analysis of Bacterial Replication Origins', *Genomics, Proteomics and Bioinformatics*, 20(6), pp. 1207–1213. Available at: https://doi.org/10.1016/j.gpb.2022.10.002.

Dorbniewski, F. (2002) 'Antimicrobial susceptibility testing of *Mycobacterium tuberculosis*', *Clinical Microbiology and Infection*, 8(10), pp. 1–10. Available at: https://doi.org/10.1111/j.1469-0691.2002.00478.x.

EUCAST (2019) 'SOP for calibrating surrogate MIC methods for *M. tuberculosis* against the EUCAST reference MIC method', (July), pp. 2–4.

EUCAST (2024) '"The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 14.0, 2024.', *Http://Www.Eucast.Org*, pp. 0–77. Available at: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_5.0_Breakpoint_Table_01.pdf.

European Centre for Disease Control (2018) *Handbook on tuberculosis laboratory diagnostic methods in the European Union*, *Technical Report*. Available at: www.ecdc.europa.eu.

Falconer, K. *et al.* (2024) 'Rapid determination of antimicrobial susceptibility of Gram-negative bacteria from clinical blood cultures using a scattered light-integrated collection device', *Journal of medical microbiology*, 73(2), pp. 1–11. Available at: https://doi.org/10.1099/jmm.0.001812.

Fleres, G. *et al.* (2018) 'Detection of *Legionella Anisa* in water from hospital dental chair units and molecular characterization by whole-genome sequencing', *Microorganisms*, 6(3), pp. 1–15. Available at: https://doi.org/10.3390/microorganisms6030071.

Flynn, J.L., Chan, J. and Lin, P.L. (2011) 'Macrophages and control of granulomatous inflammation in tuberculosis', *Nature*, 4(3), pp. 271–278. Available at: https://doi.org/10.1038/mi.2011.14.

Fontes, F.L. *et al.* (2020) 'The Acid-Base Equilibrium of Pyrazinoic Acid Drives the pH Dependence of Pyrazinamide-Induced *Mycobacterium tuberculosis* Growth Inhibition', *ACS Infectious Diseases*, 6(11), pp. 3004–3014. Available at: https://doi.org/10.1021/acsinfecdis.0c00507.

Gautam, P. and Meena, L.S. (2020) 'Revelation of point mutations effect in *Mycobacterium tuberculosis* MfpA protein that involved in mycobacterial DNA supercoiling and fluoroquinolone resistance', *Biochemistry and Molecular Biology, Inc*, 68(6), pp. 1357–1371. Available at: https://doi.org/10.1002/bab.2058.

Getahun, H. *et al.* (2015) 'Management of latent *Mycobacterium tuberculosis* infection: WHO guidelines for low tuberculosis burden countries', *European Respiratory Journal*, 46(6), pp. 1563–1576. Available at: https://doi.org/10.1183/13993003.01245-2015.

Ghazy, R.M. *et al.* (2022) 'A systematic review and meta-analysis of the catastrophic costs incurred by tuberculosis patients', *Scientific Reports* |, 12(588). Available at: https://doi.org/10.1038/s41598-021-04345-x.

Gideon, H.P. and Flynn, J.L. (2011) 'Latent tuberculosis: what the host "'sees'"?', *Immunological Reviews*, 50, pp. 202–212. Available at: https://doi.org/10.1007/s12026-011-8229-7.

Glaziou, P., Floyd, K. and Raviglione, M. (2018) 'Trends in tuberculosis in the UK', *Thorax*, 73(8), pp. 702–703. Available at: https://doi.org/10.1136/thoraxjnl-2018-211537.

Goldstein, B.P. (2014a) 'Resistance to rifampicin: a review', *The Journal of Antibiotics*, 67, pp. 625–630. Available at: https://doi.org/10.1038/ja.2014.107.

Goldstein, B.P. (2014b) 'Resistance to rifampicin: a review', *The Journal of Antibiotics*, 67, pp. 625–630. Available at: https://doi.org/10.1038/ja.2014.107.

Gómez-González, P.J. *et al.* (2021) 'Genetic diversity of candidate loci linked to *Mycobacterium tuberculosis* resistance to bedaquiline, delamanid and pretomanid', *Scientific Reports*, 11(1), pp. 1–13. Available at: https://doi.org/10.1038/s41598-021-98862-4.

Grant, J.R. *et al.* (2023) 'Proksee: In-depth characterization and visualization of bacterial genomes', *Nucleic Acids Research*, 51(W1), pp. W484–W492. Available at: https://doi.org/10.1093/nar/gkad326.

Gupta, A. and Bhakta, S. (2012) 'An integrated surrogate model for screening of drugs against *mycobacterium tuberculosis*', *Journal of Antimicrobial Chemotherapy*, 67(6), pp. 1380–1391. Available at: https://doi.org/10.1093/jac/dks056.

Gupta, R.S., Lo, B. and Son, J. (2018) 'Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera', *Frontiers in Microbiology*, 9(FEB), pp. 1–41. Available at: https://doi.org/10.3389/fmicb.2018.00067.

Gurevich, A. *et al.* (2013) 'QUAST: Quality assessment tool for genome assemblies', *Bioinformatics*, 29(8), pp. 1072–1075. Available at: https://doi.org/10.1093/bioinformatics/btt086.

Gurumurthy, M. *et al.* (2013) 'A novel F420-dependent anti-oxidant mechanism protects *Mycobacterium tuberculosis* against oxidative stress and bactericidal agents', *Molecular Microbiology*, 87(4), pp. 744–755. Available at: https://doi.org/10.1111/mmi.12127.

Gygli, S.M. *et al.* (2017) 'Antimicrobial resistance in *Mycobacterium tuberculosis*: Mechanistic and evolutionary perspectives', *FEMS Microbiology Reviews*, 41(3), pp. 354–373. Available at: https://doi.org/10.1093/femsre/fux011.

Haberecht, H.B., Cull, D.B. and Wieland, C.N. (2024) '*Mycobacterium marinum*', *Mayo Clinic Proceedings*, 99(2), pp. 196–197. Available at: https://doi.org/10.1016/j.mayocp.2023.08.005.

Hameed, S. *et al.* (2017) '*Sequence analysis of the rifampicin resistance determining region (RRDR) of rpoB gene in multidrug resistance confirmed and newly diagnosed tuberculosis patients of Punjab, Pakistan*', *PLoS ONE*, 12(8). Available at: https://doi.org/10.1371/JOURNAL.PONE.0183363.

Hammond, R. *et al.* (2017) 'Scattered Light Integrating Collector (SLIC) – Point-Of-Care (Near Patient) Clinical Breakpoint Analysis', p. 2017.

Hammond, R.J.H. *et al.* (2022) 'A simple label-free method reveals bacterial growth dynamics and antibiotic action in real-time', *Scientific Reports*, 12(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-022-22671-6.

Han, S.J. *et al.* (2015) 'Complete genome sequence of *Mycobacterium tuberculosis* K from a Korean high school outbreak, belonging to the Beijing family', *Standards in Genomic Sciences*, 10(1), pp. 1–8. Available at: https://doi.org/10.1186/s40793-015-0071-4.

He, Z.L., Du, F.W. and Du, X.Z. (2013) 'The viable *Mycobacterium tuberculosis* H37Ra strain induces a stronger mouse macrophage response compared to the heat-inactivated H37Rv strain', *Molecular Medicine Reports*, 7(5), pp. 1597–1602. Available at: https://doi.org/10.3892/mmr.2013.1363.

Heinrichs, M.T. *et al.* (2017) '*Mycobacterium tuberculosis* Strains H37ra and H37rv have Equivalent Minimum Inhibitory Concentrations to Most Antituberculosis Drugs', *International Journal of Mycobacteriology*, 6(3), pp. 239–245. Available at: https://doi.org/10.4103/ijmy.ijmy.

Holzheimer, M., Buter, J. and Minnaard, A.J. (2021) 'Chemical Synthesis of Cell Wall Constituents of *Mycobacterium tuberculosis*', *Chemical reviews*, 121, pp. 9554–9643. Available at: https://doi.org/10.1021/acs.chemrev.1c00043.

Horwitz, M.A. *et al.* (2010) 'Induce Comparable Protective Immunity Against Tuberculosis', 27(3), pp. 441–445. Available at: https://doi.org/10.1016/j.vaccine.2008.10.058.Commonly.

Ichiyama, S., Shimokata, K. and Tsukamura., M. (1988) 'Relationship between Mycobacterial Species and Their Carotenoids Pigments.', *Microbiology and Immunology*, 32(5), pp. 473–479.

Isa, M.A. *et al.* (2021) 'Identification of potent inhibitors of ATP synthase subunit c (AtpE) from Mycobacterium tuberculosis using in silico approach', *Heliyon*, 7(12), p. e08482. Available at: https://doi.org/10.1016/j.heliyon.2021.e08482.

Iseman, M.D. (2002) 'Tuberculosis therapy: Past, present and future', *European Respiratory Journal, Supplement*, 20(36), pp. 87–94. Available at: https://doi.org/10.1183/09031936.02.00309102.

Jagielski, T. *et al.* (2018) 'Characterization of Mutations Conferring Resistance to Rifampin in *Mycobacterium tuberculosis* Clinical Strains', *American Society For Microbiology*, 62(10).

Jang, J.G. and Chung, J.H. (2020) 'Diagnosis and treatment of multidrug-resistant tuberculosis', *Reichman and Hershfield's Tuberculosis: A Comprehensive, International Approach, Third Edition*, 37(4), pp. 417–457. Available at: https://doi.org/10.3904/kjm.2015.88.5.509.

Jarher, V. and Nlkaldo, H. (1994) *Mycobacterial cell wall: Structure and role in natural resistance to antibiotics*, *FEMS Microbiology Letters*. Available at: https://doi.org/10.1111/j.1574-6968.1994.tb07194.x.

Jena, L. *et al.* (2015) 'Study of mechanism of interaction of truncated isoniazid-nicotinamide adenine dinucleotide adduct against multiple enzymes of *Mycobacterium tuberculosis* by a computational approach', *International Journal of Mycobacteriology*, 4(4), pp. 276–283. Available at: https://doi.org/10.1016/j.ijmyco.2015.06.006.

Kandler, J.L. *et al.* (2018) 'Validation of novel *Mycobacterium tuberculosis* isoniazid resistance mutations not detectable by common molecular tests', *bioR*, pp. 1–43.

Kanehisa, M. and Sato, Y. (2000) 'KEGG Mapper for inferring cellular functions from protein sequences', *Nucleic Acids Research*, 28(1), pp. 28–35. Available at: https://doi.org/10.1002/pro.3711.

Kanehisa, M., Sato, Y. and Morishima, K. (2016) 'BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences', *Journal of Molecular Biology*, 428(4), pp. 726–731. Available at: https://doi.org/10.1016/j.jmb.2015.11.006.

Kazda, J. and Muller, K. (1979) '*Mycobacterium komossense*', *International journal of Systematic Bacteriology*, (ll), pp. 361–365. Available at: https://doi.org/10.1007/s00894-009-0480-7.

Kerantzas, C.A. and Jacobs, W.R. (2017) 'Origins of combination therapy for tuberculosis: Lessons for future antimicrobial development and application', *American Society For Microbiology*, 8(2). Available at: https://doi.org/10.1128/mBio.01586-16.

Keshavjee, S. and Farmer, P.E. (2012) 'Tuberculosis, Drug Resistance, and the History of Modern Medicine', *New England Journal of Medicine*, 367(10), pp. 931–936. Available at: https://doi.org/10.1056/nejmra1205429.

Kestler, B. and Tyler, S.K. (2022) 'Latent tuberculosis testing through the ages: the search for a sleeping killer', *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 322(3), pp. L412–L419. Available at: https://doi.org/10.1152/ajplung.00217.2021.

Khan, S.R., Manialawy, Y. and Siraki, A.G. (2019) 'Isoniazid and host immune system interactions: A proposal for a novel comprehensive mode of action', *British Journal of Pharmacology*, 176(24), pp. 4599–4608. Available at: https://doi.org/10.1111/bph.14867.

Kim, H. and Ryoo, S. (2011) 'Exploitation of Culture Medium for *Mycobacterium tuberculosis*', *jbv*, 41(4), pp. 237–244. Available at: https://doi.org/10.4167/jbv.2011.41.4.237.

Kim, M.J. *et al.* (2010) 'Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism', *EMBO Molecular Medicine*, 2(7), pp. 258–274. Available at: https://doi.org/10.1002/emmm.201000079.

Koch, A., Mizrahi, V. and Warner, D.F. (2014) 'The impact of drug resistance on *Mycobacterium tuberculosis* physiology: What can we learn from rifampicin?', *Emerging Microbes and Infections*, 3. Available at: https://doi.org/10.1038/emi.2014.17.

Köser, C.U., Ellington, M.J. and Peacock, S.J. (2014) 'Whole-genome sequencing to control antimicrobial resistance', *Trends in Genetics*, 30(9), pp. 401–407. Available at: https://doi.org/10.1016/j.tig.2014.07.003.

Krishnan Hyderabad, V. (2019) 'Tuberculosis: experts question evidence and safety data used to approve latest drug', *The BMJ*, 367. Available at: https://doi.org/10.1136/bmj.l6832.

Krugman, J. and Terence, C. (2022) 'When a Touch of Gold Was Used to Heal the King's Evil', *Emerging Infectious Diseases*, 11(1), pp. 765–767. Available at: https://doi.org/10.1057/s41280-020-00164-x.

Kurz, S.G. and Bonomo, R.A. (2012) 'Reappraising the use of β-lactams to treat tuberculosis', *Expert Rev Infect Ther.*, 10(9), pp. 999–1006. Available at: https://doi.org/10.1586/eri.12.96.

Lambert, S.M., Walker, S.L. and Harnisch, J.P. (2013) *CHAPTER 40 Leprosy (Hansen ' s Disease)*. Fifth Edit, *The Travel and Tropical Medicine Manual*. Fifth Edit. Elsevier Inc. Available at: https://doi.org/10.1016/B978-0-323-37506-1.00040-4.

Laws, M., Jin, P. and Rahman, K.M. (2022) 'Efflux pumps in *Mycobacterium tuberculosis* and their inhibition to tackle antimicrobial resistance', *Trends in Microbiology*, 30(1), pp. 57–68. Available at: https://doi.org/10.1016/j.tim.2021.05.001.

Lee, H.-N. *et al.* (2018) 'Roles of three FurA paralogs in the regulation of genes pertaining to peroxide defense in *Mycobacterium smegmatis* mc 2 155', *Molecular Microbiology*, 108(6), pp. 661–682. Available at: https://doi.org/10.1111/mmi.13956.

Lee, I. *et al.* (2016) 'OrthoANI: An improved algorithm and software for calculating average nucleotide identity', *International Journal of Systematic and Evolutionary Microbiology*, 66(2), pp. 1100–1103. Available at: https://doi.org/10.1099/ijsem.0.000760.

Lee, J.Y. (2015) 'Diagnosis and Treatment of Extrapulmonary Tuberculosis', *Tuberculosis and Respiratory Diseases*, 78, pp. 47–55. Available at: https://doi.org/10.4046/trd.2015.78.2.47.

Lelovic, N. *et al.* (2020) 'Application of *Mycobacterium smegmatis* as a surrogate to evaluate drug leads against *Mycobacterium tuberculosis*', *Journal of Antibiotics*, 73(11), pp. 780–789. Available at: https://doi.org/10.1038/s41429-020-0320-7.

Lempens, P. *et al.* (2018) 'Isoniazid resistance levels of *Mycobacterium tuberculosis* can largely be predicted by high-confidence resistance-conferring mutations', *Scientific Reports*, 8(1), pp. 1–9. Available at: https://doi.org/10.1038/s41598-018-21378-x.

Letunic, I. and Bork, P. (2021) 'Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation', *Nucleic Acids Research*, 49(W1), pp. W293–W296. Available at: https://doi.org/10.1093/nar/gkab301.

Levin, R., Grinstein, S. and Canton, J. (2016) 'The life cycle of phagosomes: formation, maturation, and resolution', *Immunological Reviews*, 273, pp. 156–179.

Linde, J. *et al.* (2023) 'Comparison of Illumina and Oxford Nanopore Technology for genome analysis of *Francisella tularensis*, *Bacillus anthracis*, and *Brucella suis*', *BMC Genomics*, 24(1), pp. 1–15. Available at: https://doi.org/10.1186/s12864-023-09343-z.

Liu, J. *et al.* (2019) 'Mutations in efflux pump Rv1258c (Tap) cause resistance to pyrazinamide, isoniazid, and streptomycin in *M. tuberculosis*', *Frontiers in Microbiology*, 10(FEB), pp. 1–7. Available at: https://doi.org/10.3389/fmicb.2019.00216.

Long, K.S. and Vester, B. (2012) 'Resistance to linezolid caused by modifications at its binding site on the ribosome', *Antimicrobial Agents and Chemotherapy*, 56(2), pp. 603–612. Available at: https://doi.org/10.1128/AAC.05702-11.

Luca, S. and Mihaescu, T. (2013) *History of BCG Vaccine*, *Maedica A Journal of Clinical Medicine*.

Mahomed, S. *et al.* (2017) 'Whole genome sequencing for the management of drug-resistant TB in low income high TB burden settings: Challenges and implications', *Tuberculosis*, 107, pp. 137–143. Available at: https://doi.org/10.1016/j.tube.2017.09.005.

De Maio, F. *et al.* (2021) 'First description of the katG gene deletion in a *Mycobacterium tuberculosis* clinical isolate and its impact on the mycobacterial fitness', *International Journal of Medical Microbiology*, 311(4), p. 151506. Available at: https://doi.org/10.1016/j.ijmm.2021.151506.

Maladan, Y. *et al.* (2021) 'The whole-genome sequencing in predicting *Mycobacterium tuberculosis* drug susceptibility and resistance in Papua, Indonesia', *BMC Genomics*, 22(1), pp. 1–11. Available at: https://doi.org/10.1186/s12864-021-08139-3.

Manjunatha, U., Boshoff, H.I.M. and Barry, C.E. (2009) 'The mechanism of action of PA-824', *Communicative and Integrative Biology*, 2(3), pp. 215–218. Available at: https://doi.org/10.4161/cib.2.3.7926.

Maphasa, R.E., Meyer, M. and Dube, A. (2021) 'The Macrophage Response to *Mycobacterium tuberculosis* and Opportunities for Autophagy Inducing Nanomedicines for Tuberculosis Therapy', *Frontiers in Microbiology*, 10(618414). Available at: https://doi.org/10.3389/fcimb.2020.618414.

Marchese, A. and Debbia, E.A. (2016a) 'The role of *gyrA*, *gyrB*, and *dnaA* functions in bacterial conjugation', *Annals of Microbiology*, 66(1), pp. 223–228. Available at: https://doi.org/10.1007/s13213-015-1098-x.

Marchese, A. and Debbia, E.A. (2016b) 'The role of gyrA, gyrB, and dnaA functions in bacterial conjugation', *Annals of Microbiology*, 66(1), pp. 223–228. Available at: https://doi.org/10.1007/s13213-015-1098-x.

Master, S. *et al.* (2001) 'Mapping of *Mycobacterium tuberculosis katG* promoters and their differential expression in infected macrophages', *Journal of Bacteriology*, 183(13), pp. 4033–4039. Available at: https://doi.org/10.1128/JB.183.13.4033-4039.2001.

Master, S.S. *et al.* (2002) 'Oxidative stress response genes in *Mycobacterium tuberculosis*: Role of ahpC in resistance to peroxynitrite and stage-specific survival in macrophages', *Microbiology*, 148(10), pp. 3139–3144. Available at: https://doi.org/10.1099/00221287-148-10-3139.

Meehan, C.J. *et al.* (2021) 'Reconstituting the genus *Mycobacterium*', *International Journal of Systematic and Evolutionary Microbiology*, 71(9). Available at: https://doi.org/10.1099/ijsem.0.004922.

Migliori, G.B. *et al.* (2007) '125 years after Robert Koch's discovery of the tubercle bacillus: The new XDR-TB threat. Is "science" enough to tackle the epidemic?', *European Respiratory Journal*, 29(3), pp. 423–427. Available at: https://doi.org/10.1183/09031936.00001307.

Miles, A.A., Misra, S.S. and Irwin, J.O. (1938) 'The estimation of the bactericidal power of the blood', *Journal of Hygiene*, 38(6), pp. 732–749. Available at: https://doi.org/10.1017/S002217240001158X.

Molodtsov, V. *et al.* (2017) 'Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*', *Molecular Microbiology*, 103(6), pp. 1034–1045. Available at: https://doi.org/10.1111/mmi.13606.

Monama, M.Z., Olotu, F. and Tastan Bishop, Ö. (2023) 'Investigation of Multi-Subunit *Mycobacterium tuberculosis* DNA-Directed RNA Polymerase and Its Rifampicin Resistant Mutants', *International Journal of Molecular Sciences*, 24(4). Available at: https://doi.org/10.3390/ijms24043313.

Morgado, S.M. and Vicente, A.C.P. (2021) 'Comprehensive in silico survey of the *Mycolicibacterium* mobilome reveals an as yet underexplored diversity', *Microbial Genomics*, 7(3). Available at: https://doi.org/10.1099/mgen.0.000533.

Moure, R. *et al.* (2014) 'Characterization of the embB gene in Mycobacterium tuberculosis isolates from barcelona and rapid detection of main mutations related to ethambutol resistance using a low-density DNA array', *Journal of Antimicrobial Chemotherapy*, 69(4), pp. 947–954. Available at: https://doi.org/10.1093/jac/dkt448.

Nahid, P. *et al.* (2016) 'Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis', *Clinical Infection diseases*, 63, pp. 853–867. Available at: https://doi.org/10.1093/cid/ciw566.

Nair, S.S. *et al.* (2015) 'In vitro antimycobacterial activity of acetone extract of *Glycyrrhiza glabra*', *Journal of Pharmacy and Pharmacognosy Research*, 3(4), pp. 80–86. Available at: https://doi.org/10.56499/jppres15.062_3.4.80.

Namouchi, A. *et al.* (2017) 'Phenotypic and genomic comparison of *Mycobacterium aurum* and surrogate model species to *Mycobacterium tuberculosis*: Implications for drug discovery', *BMC Genomics*, 18(1), pp. 25–28. Available at: https://doi.org/10.1186/s12864-017-3924-y.

Nasiri, M.J. *et al.* (2017) 'New Insights in to the Intrinsic and Acquired Drug Resistance Mechanisms in Mycobacteria', *Frontiers in Microbiology | www.frontiersin.org*, 8(681). Available at: https://doi.org/10.3389/fmicb.2017.00681.

Ndlovu, H. and Marakalala, M.J. (2016) 'Granulomas and inflammation: Host-Directed Therapies for Tuberculosis', *Frontiers in Microbiology*, 7(434). Available at: https://doi.org/10.3389/fimmu.2016.00434.

Needleman, S.B. and Wunsch, C.D. (1970) 'A general method applicable to the search for similarities in the amino acid sequence of two proteins', *Journal of Molecular Biology*, 48(3), pp. 443–453. Available at: https://doi.org/10.1016/0022-2836(70)90057-4.

Negrete-Paz, A.M. *et al.* (2023) 'Pangenome Reconstruction of *Mycobacterium tuberculosis* as a Guide to Reveal Genomic Features Associated with Strain Clinical Phenotype', *Microorganisms*, 11(6). Available at: https://doi.org/10.3390/microorganisms11061495.

Newton-Foot, M. *et al.* (2016) 'The plasmid-mediated evolution of the mycobacterial ESX (Type VII) secretion systems', *BMC Evolutionary Biology*, 16(1), pp. 1–12. Available at: https://doi.org/10.1186/s12862-016-0631-2.

Niederweis, M. (2003) 'Mycobacterial porins-new channel proteins in unique outer membranes', *Molecular Microbiology*, 49(5), pp. 1167–1177. Available at: https://doi.org/10.1046/j.1365-2958.2003.03662.x.

Nimmo, C. *et al.* (2022) 'Evolution of Mycobacterium tuberculosis drug resistance in the genomic era', *Frontiers in Cellular and Infection Microbiology*, 12(October), pp. 1–12. Available at: https://doi.org/10.3389/fcimb.2022.954074.

Niward, K. *et al.* (2016) 'Susceptibility testing breakpoints for Mycobacterium tuberculosis categorize isolates with resistance mutations in *gyrA* as susceptible to fluoroquinolones: Implications for MDR-TB treatment and the definition of XDR-TB', *Journal of Antimicrobial Chemotherapy*, 71(2), pp. 333–338. Available at: https://doi.org/10.1093/jac/dkv353.

Nontuthuko E. Maningi, L.T.D. *et al.* (2018) 'Multi- and Extensively Drug Resistant *Mycobacterium tuberculosis* in South Africa : a Molecular Analysis of Historical', *Journal of Clinical Microbiology*, 56(5).

Oelofse, S. *et al.* (2021) 'Pretomanid with bedaquiline and linezolid for drug-resistant TB: a comparison of prospective cohorts', *INT J TUBERC LUNG DIS*, 25(6). Available at: https://doi.org/10.5588/ijtld.21.0035.

Ogier, J.C. *et al.* (2019) 'RpoB, a promising marker for analyzing the diversity of bacterial communities by amplicon sequencing', *BMC Microbiology*, 19(1), pp. 1–16. Available at: https://doi.org/10.1186/s12866-019-1546-z.

O'neill, M.B. *et al.* (2019) 'Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia', *Molecular Ecology*, 28, pp. 3241–3256. Available at: https://doi.org/10.1111/mec.15120.

Palaci, M. *et al.* (2007) 'Cavitary disease and quantitative sputum bacillary load in cases of pulmonary tuberculosis', *Journal of Clinical Microbiology*, 45(12), pp. 4064–4066. Available at: https://doi.org/10.1128/JCM.01780-07.

Palomino, J.C. and Martin, A. (2014) 'Drug resistance mechanisms in *Mycobacterium tuberculosis*', *Antibiotics*, 3(3), pp. 317–340. Available at: https://doi.org/10.3390/antibiotics3030317.

Pan, Q., Zhao, F.L. and Ye, B.C. (2018) 'Eis, a novel family of arylalkylamine N-acetyltransferase (EC 2.3.1.87)', *Scientific Reports*, 8(1), pp. 1–8. Available at: https://doi.org/10.1038/s41598-018-20802-6.

Passiflora Sarathy, J. *et al.* (2012) 'The Role of Transport Mechanisms in *Mycobacterium Tuberculosis* Drug Resistance and Tolerance', *Pharmaceuticals*, 5, pp. 1210–1235. Available at: https://doi.org/10.3390/ph5111210.

Percival, S.L. and Williams, D.W. (2013) Mycobacterium. Second Edi, *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks: Second Edition*. Second Edi. Elsevier. Available at: https://doi.org/10.1016/B978-0-12-415846-7.00009-3.

Pereira, A.C. *et al.* (2020) 'Non-tuberculous mycobacteria: Molecular and physiological bases of virulence and adaptation to ecological niches', *Microorganisms*, 8(9), pp. 1–49. Available at: https://doi.org/10.3390/microorganisms8091380.

Petrini, B. (2006) '*Mycobacterium marinum*: ubiquitous agent of waterborne granulomatous skin infections', *European Journal of Clinical Microbiology and Infectious Diseases*, 25(10), pp. 609–613. Available at: https://doi.org/10.1007/s10096-006-0201-4.

Phelan, J. *et al.* (2015) 'The draft genome of *Mycobacterium aurum*, a potential model organism for investigating drugs against *Mycobacterium tuberculosis* and *Mycobacterium leprae*', *International Journal of Mycobacteriology*, 4(3), pp. 207–216. Available at: https://doi.org/10.1016/j.ijmyco.2015.05.001.

Pietersen, R.D. *et al.* (2020) 'Tween 80 induces a carbon flux rerouting in *Mycobacterium tuberculosis*', *Journal of Microbiological Methods*, 170(November 2019), p. 105795. Available at: https://doi.org/10.1016/j.mimet.2019.105795.

Portelli, S. *et al.* (2020) 'Prediction of rifampicin resistance beyond the RRDR using structure-based machine learning approaches', *Scientific Reports*, 10(1), pp. 1–13. Available at: https://doi.org/10.1038/s41598-020-74648-y.

Pranger, A.D. *et al.* (2019) 'The Role of Fluoroquinolones in the Treatment of Tuberculosis in 2019', *Drugs*, 79(2), pp. 161–171. Available at: https://doi.org/10.1007/s40265-018-1043-y.

Preece, C.L. *et al.* (2016) 'A novel culture medium for isolation of rapidly-growing mycobacteria from the sputum of patients with cystic fibrosis', *Journal of Cystic Fibrosis*, 15(2), pp. 186–191. Available at: https://doi.org/10.1016/j.jcf.2015.05.002.

Qian, J. *et al.* (2020) 'Role of the PE/PPE Family in Host–Pathogen Interactions and Prospects for Anti-Tuberculosis Vaccine and Diagnostic Tool Design', *Frontiers in Cellular and Infection Microbiology*, 10(November), pp. 1–8. Available at: https://doi.org/10.3389/fcimb.2020.594288.

Queval, C.J., Brosch, R. and Simeone, R. (2017) 'The Macrophage: A Disputed Fortress in the Battle against *Mycobacterium tuberculosis*', *Frontiers in Microbiology*, 8(23). Available at: https://doi.org/10.3389/fmicb.2017.02284.

Quinn, G.A. *et al.* (2020) '*Streptomyces* from traditional medicine: sources of new innovations in antibiotic discovery', *Journal of Medical Microbiology*, 69(8), pp. 1040–1048. Available at: https://doi.org/10.1099/jmm.0.001232.

Ramakrishnan, L. (2004) 'Using *Mycobacterium marinum* and its hosts to study tuberculosis', *Current Science*, 86(1), pp. 82–92.

Raynaud, C. *et al.* (2002) *The functions of OmpATb, a pore-forming protein of Mycobacterium tuberculosis*, *Molecular Microbiology*. Available at: http://www.stoptb.org/tuberculosis/.

Rekha, S.R., Kulandhaivel, M. and Hridhya, K. V. (2018) 'Antibacterial efficacy and minimum inhibitory concentrations of medicinal plants against wound pathogens', *Biomedical and Pharmacology Journal*, 11(1), pp. 237–246. Available at: https://doi.org/10.13005/bpj/1368.

Reyrat, J.M. and Kahn, D. (2001) '*Mycobacterium smegmatis*: an absurd model for tuberculosis?', *Trends in Microbiology*, 9(10), pp. 472–473. Available at: https://doi.org/10.1016/s0966-842x(01)02168-0.

Robinson, J.T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*, 29(1), pp. 24–26. Available at: https://doi.org/10.1038/nbt.1754.

Robledo, J.A., Murillo, A.M. and Rouzaud, F. (2011) 'Physiological role and potential clinical interest of mycobacterial pigments', *IUBMB Life*, 63(2), pp. 71–78. Available at: https://doi.org/10.1002/iub.424.

Samten, B., Wang, X. and Barnes, P.F. (2009) '*Mycobacterium tuberculosis* ESX-1 system-secreted protein ESAT-6 but not CFP10 inhibits human T-cell immune responses', *Tuberculosis*, 89(SUPPL.1), pp. S74–S76. Available at: https://doi.org/10.1016/S1472-9792(09)70017-4.

Sanderson, N.D. *et al.* (2020) 'High precision *Neisseria gonorrhoeae* variant and antimicrobial resistance calling from metagenomic Nanopore sequencing', *Genome Research*, 30(9), pp. 1354–1363. Available at: https://doi.org/10.1101/GR.262865.120.

Santos-López, A., Rodríguez-Beltrán, J. and Millán, A.L.S. (2021) 'The bacterial capsule is a gatekeeper for mobile DNA', *PLoS Biology*, 19(7), pp. 4–7. Available at: https://doi.org/10.1371/journal.pbio.3001308.

Sarathy, J.P., Gruber, G. and Dick, T. (2019) 'Re-understanding the mechanisms of action of the anti-mycobacterial drug bedaquiline', *Antibiotics*, 8(4). Available at: https://doi.org/10.3390/antibiotics8040261.

Sawyer, J. *et al.* (2023) '*Mycobacterium bovis* and its impact on human and animal tuberculosis', *Journal of Medical Microbiology*, 72(11), pp. 1–7. Available at: https://doi.org/10.1099/jmm.0.001769.

Schön, T. *et al.* (2017) '*Mycobacterium tuberculosis* drug-resistance testing: challenges, recent developments and perspectives', *Clinical Microbiology and Infection*, 23(3), pp. 154–160. Available at: https://doi.org/10.1016/j.cmi.2016.10.022.

Schön, T. *et al.* (2020) 'Antimicrobial susceptibility testing of *Mycobacterium tuberculosis* complex isolates – the EUCAST broth microdilution reference method for MIC determination', *Clinical Microbiology and Infection*, 26(11), pp. 1488–1492. Available at: https://doi.org/10.1016/j.cmi.2020.07.036.

Seifert, M. *et al.* (2015) 'Genetic mutations associated with isoniazid resistance in *Mycobacterium tuberculosis*: A systematic review', *PLoS ONE*, 10(3), pp. 1–13. Available at: https://doi.org/10.1371/journal.pone.0119628.

Selengut, J.D. and Haft, D.H. (2010) 'Unexpected abundance of coenzyme F420-dependent enzymes in Mycobacterium tuberculosis and other actinobacteria', *Journal of Bacteriology*, 192(21), pp. 5788–5798. Available at: https://doi.org/10.1128/JB.00425-10.

Senghore, M. *et al.* (2020) 'Evolution of *Mycobacterium tuberculosis* complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali', *Nature Research*, 10(327). Available at: https://doi.org/10.1038/s41598-019-56001-0.

Shaku, M., Ealand, C. and Kana, B.D. (2020) 'Cell Surface Biosynthesis and Remodeling Pathways in Mycobacteria Reveal New Drug Targets', *Frontiers in Microbiology*, 10(603382). Available at: https://doi.org/10.3389/fcimb.2020.603382.

Silver, G.A. (2014) *A History of Public Health*, *American Journal of Public Health and the Nations Health*. Available at: https://doi.org/10.2105/ajph.48.7.944-b.

Simão, F.A. *et al.* (2015) 'BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs', *Bioinformatics*, 31(19), pp. 3210–3212. Available at: https://doi.org/10.1093/bioinformatics/btv351.

Singh, R. *et al.* (2020) 'Recent updates on drug resistance in *Mycobacterium tuberculosis*.', *Journal of Applied Microbiology*, 128(6), pp. 1547–1567. Available at: https://doi.org/10.1111/jam.14478.

Smeulders, M.J. *et al.* (2017) 'To catch a killer. What can mycobacterial models teach us about *Mycobacterium tuberculosis* pathogenesis?', *Frontiers in Microbiology*, 10(1), pp. 1–7. Available at: https://doi.org/10.1086/647952.

Smith, I. (2003) '*Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence', *CLINICAL MICROBIOLOGY REVIEWS*, 16(3), pp. 463–496. Available at: https://doi.org/10.1128/CMR.16.3.463-496.2003.

Smith, Tasha *et al.* (2012) 'Molecular Biology of Drug Resistance in *Mycobacterium tuberculosis*', *Current Topics in Microbiology and Immunology*, 374, pp. 53–80. Available at: https://doi.org/10.1007/82_2012_279.

Songane, M. *et al.* (2012) 'The role of autophagy in host defence against *Mycobacterium tuberculosis* infection', *Tuberculosis*, 92(5), pp. 388–396. Available at: https://doi.org/10.1016/j.tube.2012.05.004.

Sparks, I.L. *et al.* (2023) '*Mycobacterium smegmatis*: The Vanguard of Mycobacterial Research', *Journal of Bacteriology*, 205(1). Available at: https://doi.org/10.1128/jb.00337-22.

Squeglia, F., Ruggiero, A. and Berisio, R. (2018) 'Chemistry of Peptidoglycan in *Mycobacterium tuberculosis* Life Cycle: An off-the-wall Balance of Synthesis and Degradation', *Chemistry - A European Journal*, 24(11), pp. 2533–2546. Available at: https://doi.org/10.1002/chem.201702973.

Ssengooba, W. *et al.* (2015) 'Feasibility of establishing a biosafety level 3 tuberculosis culture laboratory of acceptable quality standards in a resource-limited setting: An experience from Uganda', *Health Research Policy and Systems*, 13(1), pp. 1–10. Available at: https://doi.org/10.1186/1478-4505-13-4.

Stinear, T.P. *et al.* (2008) 'Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*', *Genome Research*, 18(5), pp. 729–741. Available at: https://doi.org/10.1101/gr.075069.107.

Su, M., Satola, S.W. and Read, T.D. (2019) 'Genome-based prediction of bacterial antibiotic resistance', *Journal of Clinical Microbiology*, 57(3). Available at: https://doi.org/10.1128/JCM.01405-18.

Subramani, E. *et al.* (2017) 'Mycobacterial heat shock protein 65 mediated metabolic shift in decidualization of human endometrial stromal cells', *Scientific Reports*, 7(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-017-04024-w.

Swaim, L.E. *et al.* (2006) '*Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity', *Infection and Immunity*, 74(11), pp. 6108–6117. Available at: https://doi.org/10.1128/IAI.00887-06.

Szumowski, J.D. *et al.* (2012) 'Antimicrobial Efflux Pumps and *Mycobacterium tuberculosis* Drug Tolerance: Evolutionary Considerations', *Current Topics in Microbiology and Immunology*, 374, pp. 81–108. Available at: https://doi.org/10.1007/82_2012_300.

T, J.A.S. *et al.* (2020) 'Features of the biochemistry of *Mycobacterium smegmatis*, as a possible model for *Mycobacterium tuberculosis*', *Journal of Infection and Public Health*, 13(9), pp. 1255–1264. Available at: https://doi.org/10.1016/j.jiph.2020.06.023.

Talaat, A.M. *et al.* (1998) 'Goldfish, Carassius auratus, a novel animal model for the study of *Mycobacterium marinum* pathogenesis', *Infection and Immunity*, 66(6), pp. 2938–2942. Available at: https://doi.org/10.1128/iai.66.6.2938-2942.1998.

Taniguchi, H. *et al.* (1996a) *Rifampicin resistance and mutation of the <i>rpoB gene in* Mycobacterium tuberculosis, *FEMS Microbiology Letters*. Available at: https://doi.org/10.1111/j.1574-6968.1996.tb08515.x.

Taniguchi, H. *et al.* (1996b) *Rifampicin resistance and mutation of the rpoB gene in Mycobacterium tuberculosis*, *FEMS Microbiology Letters*. Available at: https://doi.org/10.1111/j.1574-6968.1996.tb08515.x.

Tao, J. *et al.* (2013) '*Mycobacterium* fluoroquinolone resistance protein B, a novel small GTPase, is involved in the regulation of DNA gyrase and drug resistance', *Nucleic Acids Research*, 41(4), pp. 2370–2381. Available at: https://doi.org/10.1093/nar/gks1351.

Tapiero, H., D.M. Townsend and Tew, K.D. (2018) 'HHS Public Access', *Physiology & behavior*, 176(1), pp. 139–148. Available at: https://doi.org/10.1016/j.biopha.2003.12.006.The.

Tay, S.T.L. *et al.* (1998) 'Two new *Mycobacterium* strains and their role in toluene degradation in a contaminated stream', *Applied and Environmental Microbiology*, 64(5), pp. 1715–1720. Available at: https://doi.org/10.1128/aem.64.5.1715-1720.1998.

Tindall, B.J. (1999) 'Misunderstanding the Bacteriological Code', *International Journal of Systematic Bacteriology*, 49(3), pp. 1313–1316. Available at: https://doi.org/10.1099/00207713-49-3-1313.

Tobin, D.M. and Ramakrishnan, L. (2008) 'Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*', *Cellular Microbiology*, 10(5), pp. 1027–1039. Available at: https://doi.org/10.1111/j.1462-5822.2008.01133.x.

Toh, S.M. *et al.* (2008) 'The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA', *Rna*, 14(1), pp. 98–106. Available at: https://doi.org/10.1261/rna.814408.

Tonkin-Hill, G. *et al.* (2020) 'Producing polished prokaryotic pangenomes with the Panaroo pipeline', *Genome Biology*, 21(1), pp. 1–21. Available at: https://doi.org/10.1186/s13059-020-02090-4.

Tortoli, E. *et al.* (2019) 'Same meat, different gravy: Ignore the new names of mycobacteria', *European Respiratory Journal*, 54(1), pp. 19–21. Available at: https://doi.org/10.1183/13993003.00795-2019.

Triebel, S. *et al.* (2023) 'De novo genome assembly resolving repetitive structures enables genomic analysis of 35 European *Mycoplasmopsis bovis* strains', *BMC Genomics*, 24(1), pp. 1–14. Available at: https://doi.org/10.1186/s12864-023-09618-5.

Tseng, S.T. *et al.* (2015) 'The mutations of *katG* and *inhA* genes of isoniazid-resistant *Mycobacterium tuberculosis* isolates in Taiwan', *Journal of Microbiology, Immunology and Infection*, 48(3), pp. 249–255. Available at: https://doi.org/10.1016/j.jmii.2013.08.018.

Unissa, A.N. *et al.* (2016a) 'Overview on mechanisms of isoniazid action and resistance in *Mycobacterium tuberculosis*', *Infection, Genetics and Evolution*, 45(1), pp. 474–492. Available at: https://doi.org/10.1016/j.meegid.2016.09.004.

Unissa, A.N. *et al.* (2016b) 'Overview on mechanisms of isoniazid action and resistance in *Mycobacterium tuberculosis*', *Infection, Genetics and Evolution*, 45(1), pp. 474–492. Available at: https://doi.org/10.1016/j.meegid.2016.09.004.

Urbanowski, M.E. *et al.* (2020) 'Cavitary tuberculosis: the gateway of disease transmission', *The Lancet Infectious Diseases*, 20(6), pp. e117–e128. Available at: https://doi.org/10.1016/S1473-3099(20)30148-1.

Vergne, I. *et al.* (2004) 'CELL BIOLOGY OF *MYCOBACTERIUM TUBERCULOSIS* PHAGOSOME', *Annu. Rev. Cell Dev. Biol*, 20, pp. 367–94. Available at: https://doi.org/10.1146/annurev.cellbio.20.010403.114015.

Vilchèze, C. and Kremer, L. (2017) 'Acid-fast positive and acid-fast negative *Mycobacterium tuberculosis*: The Koch paradox', *Tuberculosis and the Tubercle Bacillus: Second Edition*, pp. 519–532. Available at: https://doi.org/10.1128/9781555819569.ch23.

Vuorio, R. *et al.* (1999) 'A new rapidly growing mycobacterial species, *Mycobacterium murale* sp. nov., isolated from the indoor walls of a children's day care centre', *International Journal of Systematic Bacteriology*, 49(1), pp. 25–35. Available at: https://doi.org/10.1099/00207713-49-1-25.

Waddington, C. *et al.* (2022) 'Exploiting genomics to mitigate the public health impact of antimicrobial resistance', *Genome Medicine*, 14(1), pp. 1–14. Available at: https://doi.org/10.1186/s13073-022-01020-2.

Wakamoto, Y. *et al.* (2013) 'Dynamic Persistence of Antibiotic-Stressed Mycobacteria', *Science*, 339, pp. 91–95.

Wei, L. *et al.* (2017) 'Structural basis of *Mycobacterium tuberculosis* transcription and transcription inhibition', *Physiology & behavior*, 176(1), pp. 139–148. Available at: https://doi.org/10.1016/j.molcel.2017.03.001.Structural.

Wengenack, N.L. *et al.* (1997) 'Recombinant *Mycobacterium tuberculosis* KatG(S315t) is a competent catalase-peroxidase with reduced activity toward isoniazid', *Journal of Infectious Diseases*, 176(3), pp. 722–727. Available at: https://doi.org/10.1086/514096.

Wick, R.R. *et al.* (2015) 'Bandage: Interactive visualization of de novo genome assemblies', *Bioinformatics*, 31(20), pp. 3350–3352. Available at: https://doi.org/10.1093/bioinformatics/btv383.

Wick, R.R. *et al.* (2017) 'Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads', *PLoS Computational Biology*, 13(6), pp. 1–22. Available at: https://doi.org/10.1371/journal.pcbi.1005595.

Wiegand, I., Hilpert, K. and Hancock, R.E.W. (2008) 'Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances', *Nature Protocols*, 3(2), pp. 163–175. Available at: https://doi.org/10.1038/nprot.2007.521.

Woodman, M., Haeusler, I.L. and Grandjean, L. (2019) 'Tuberculosis Genetic Epidemiology: A Latin American Perspective', *Genes*, 10(53). Available at: https://doi.org/10.3390/genes10010053.

Woods, G.L. *et al.* (2018) 'Performance Standards for Susceptibility Testing of Mycobacteria, Nocardia spp., and Other Aerobic Actinomycetes'.

Worawat Chumpangern, W.R. (2018) *WHO treatment guidelines for isoniazid- resistant tuberculosis*, *Supplement to the WHO treatment guidelines for drug-resistant tuberculosis*.

World Health Organization (2020) *WHO consolidated guidelines on tuberculosis*.

World Health Organization (2021) *Catalogue of mutations in* Mycobacterium tuberculosis *complex and their association with drug resistance*. Available at: https://www.who.int/publications/i/item/9789240028173.

World Health Organization 2018 (2018) *Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis*, *Biomass Chem Eng*. Available at: http://journal.stainkudus.ac.id/index.php/equilibrium/article/view/1268/1127%0Ahttp://publicacoes.cardiol.br/portal/ijcs/portugues/2018/v3103/pdf/3103009.pdf%0Ahttp://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-75772018000200067&lng=en&tlng=.

World Health Organization (WHO) (2021) *Global Tuberculosis Report 2021*.

World Health Organization (WHO) (2023) *Global tuberculosis report 2023*, *World Health Organization*.

Xu, Y., Wang, G. and Xu, M. (2020) 'Biohazard levels and biosafety protection for *Mycobacterium tuberculosis* strains with different virulence', *Biosafety and Health*, 2(3), pp. 135–141. Available at: https://doi.org/10.1016/j.bsheal.2020.04.001.

Yamori, S. *et al.* (1992) 'Bacteriostatic and Bactericidal Activity of Antituberculosis Drugs against *Mycobacterium tuberculosis*, *Mycobacterium avium-Mycobacterium intracellulare* Complex and *Mycobacterium kansasii* in Different Growth Phases', *Microbiology and Immunology*, 36(4), pp. 361–368. Available at: https://doi.org/10.1111/j.1348-0421.1992.tb02035.x.

Yang, R. *et al.* (2014) 'The RD1 locus in the *Mycobacterium tuberculosis* genome contributes to the maturation and secretion of IL-1α from infected macrophages through the elevation of cytoplasmic calcium levels and calpain activation', *Pathogens and Disease*, 70(1), pp. 51–60. Available at: https://doi.org/10.1111/2049-632X.12075.

Zhang, L. *et al.* (2020) 'Structures of cell wall arabinosyltransferases with the anti-tuberculosis drug ethambutol', *Science*, 368(6496), pp. 1211–1219. Available at: https://doi.org/10.1126/science.aba9102.

Zhang, Q. *et al.* (2019) 'Uncovering the Resistance Mechanism of *Mycobacterium tuberculosis* to Rifampicin Due to RNA Polymerase H451D/Y/R Mutations From Computational Perspective', *Frontiers in Chemistry | www.frontiersin.org*, 7(819). Available at: https://doi.org/10.3389/fchem.2019.00819.

Zhang, Y. *et al.* (2003) 'Mode of action of pyrazinamide: disruption of Mycobacterium tuberculosis membrane transport and energetics by pyrazinoic acid', *Journal of Antimicrobial Chemotherapy*, 52, pp. 790–795. Available at: https://doi.org/10.1093/jac/dkg446.

Zheng, H. *et al.* (2008) 'Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv', *PLoS ONE*, 3(6). Available at: https://doi.org/10.1371/journal.pone.0002375.

Zheng, S. *et al.* (2021) 'Implication of Surface Properties, Bacterial Motility, and Hydrodynamic Conditions on Bacterial Surface Sensing and Their Initial Adhesion', *Frontiers in Bioengineering and Biotechnology*, 9(February), pp. 1–22. Available at: https://doi.org/10.3389/fbioe.2021.643722.

Zhu, M. and Dai, X. (2018) 'On the intrinsic constraint of bacterial growth rate: *M. tuberculosis's* view of the protein translation capacity', *Critical Reviews in Microbiology*, 44(4), pp. 455–464. Available at: https://doi.org/10.1080/1040841X.2018.1425672.

Zwick, E.D. and Pepperell, C.S. (2020) 'Tuberculosis sanatorium treatment at the advent of the chemotherapy era', *BMC Infectious Diseases*, 20(1), pp. 1–11. Available at: https://doi.org/10.1186/s12879-020-05539-w.

8 Appendix A: Chapter 4 Supplementary data

AMR related proteins global alignments

```
Query: HJANABMJ_03920 16S ribosomal RNA Query ID: lcl|Query_5201301 Length: 1519


>LFFBIPCO_01392 16S ribosomal RNA
Sequence ID: Query_5201303 Length: 1529
Range 1: 1 to 1529

NW Score:2632,  Identities:1452/1535(95%),  Gaps:22/1535(1%), Strand: Plus/Plus

Mkomo_JERR01  1     TTGGAGAGTTTGATTCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTC  60
                    ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1     TTGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTC  60

Mkomo_JERR01  61    GAACGGAAAGGCC-CTTCGGGG-TGCTCGAGTGGCGAACGGGTGAGTAACACGTGGGTGA  118
                    |||||||||||| | |||||| | | ||||||||||||||||||||||||||||||||||
Mtb_H37Rv     61    GAACGGAAAGGTCTCTTCGGAGATACTCGAGTGGCGAACGGGTGAGTAACACGTGGGTGA  120

Mkomo_JERR01  119   TCTGCCCTGCACTTTGGGATAAGCCTGGGAAACTGGGTCTAATACCGAATATACCCTTCT  178
                    |||||||||||||| ||||||||||||||||||||||||||||||| |||   ||  |
Mtb_H37Rv     121   TCTGCCCTGCACTTCGGGATAAGCCTGGGAAACTGGGTCTAATACCGGATAGGACCA-CG  179

Mkomo_JERR01  179   GGTTGCATGGCCTGGGAGGGGAAAGC--TTTTGCGGTGTGGGATGGGCCCGCGGCCTATC  236
                    || |||||| | || | | |||||||  ||| |||||||||||||| ||||||||||||||
Mtb_H37Rv     180   GGATGCATGTCTTGTG-GTGGAAAGCGCTTTAGCGGTGTGGGATGAGCCCGCGGCCTATC  238

Mkomo_JERR01  237   AGCTTGTTGGTGAGGTTACGGCTCACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTG  296
                    |||||||||||| ||| |||||  ||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     239   AGCTTGTTGGTGGGGTGACGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTG  298

Mkomo_JERR01  297   ACCGGCCACACTGGGACTGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT  356
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     299   TCCGGCCACACTGGGACTGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAAT  358

Mkomo_JERR01  357   ATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGAGGGATGACGGCCTTCGGG  416
                    |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Mtb_H37Rv     359   ATTGCACAATGGGCGCAAGCCTGATGCAGCGACGCCGCGTGGGGGATGACGGCCTTCGGG  418

Mkomo_JERR01  417   TTGTAAACCTCTTTCGCTCACGACGAAGCGCAAG----------TGACGGTAGTGAGA  464
                    |||||||||||||| |   ||||||||  |  |           |||||||||   ||
Mtb_H37Rv     419   TTGTAAACCTCTTTCACCATCGACGAAGGTCCGGGTTCTCTCGGATTGACGGTAGGTGGA  478

Mkomo_JERR01  465   GAAGAAGGACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTCCGAGCGTTG  524
                    ||||||| ||||||||||||||||||||||||||||||||||||||| |||||||||
Mtb_H37Rv     479   GAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCGAGCGTTG  538

Mkomo_JERR01  525   TCCGGAATTACTGGGCGTAAAGAGCTCGTAGGTGGTTTGTCGCGTTGTCCGTGAAAACTC  584
                    |||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||
Mtb_H37Rv     539   TCCGGAATTACTGGGCGTAAAGAGCTCGTAGGTGGTTTGTCGCGTTGTTCGTGAAATCTC  598

Mkomo_JERR01  585   ACAGCTCAACTGTGGGCGTGCGGGCGATACGGGCAGACTGGAGTACTGCAGGGGAGACTG  644
                    || ||| ||||||| |||||||||||||||||||||| ||||||||||||||||||||
Mtb_H37Rv     599   ACGGCTTAACTGTGAGCGTGCGGGCGATACGGGCAGACTAGAGTACTGCAGGGGAGACTG  658

Mkomo_JERR01  645   GAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCG  704
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     659   GAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGGTGGCGAAGGCG  718

Mkomo_JERR01  705   GGTCTCTGGGCAGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGAT  764
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     719   GGTCTCTGGGCAGTAACTGACGCTGAGGAGCGAAAGCGTGGGGAGCGAACAGGATTAGAT  778

Mkomo_JERR01  765   ACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATC  824
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     779   ACCCTGGTAGTCCACGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATC  838

Mkomo_JERR01  825   CGTGCCGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACT  884
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     839   CGTGCCGTAGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAAACT  898

Mkomo_JERR01  885   CAAAGAAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACG  944
                    ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Mtb_H37Rv     899   CAAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGATGCAACG  958


Mkomo_JERR01  945   CGAAGAACCTTACCTGGGTTTGACATGCACAGGACGCCGGCAGAGATGTCGGTTCCCTTG  1004
                    |||||||||||||||||||||||||||||||||||||||    ||||||    |||||||||
Mtb_H37Rv     959   CGAAGAACCTTACCTGGGTTTGACATGCACAGGACGCGTCTAGAGATAGGCGTTCCCTTG  1018

Mkomo_JERR01  1005  TGGTCTGTGTGCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAA  1064
                    ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1019  TGGCCTGTGTGCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAA  1078

Mkomo_JERR01  1065  GTCCCGCAACGAGCGCAACCCTTGTCTCATGTTGCCAGCACGTTATGGTGGGGACTCGTG  1124
                    |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Mtb_H37Rv     1079  GTCCCGCAACGAGCGCAACCCTTGTCTCATGTTGCCAGCACGTAATGGTGGGGACTCGTG  1138

Mkomo_JERR01  1125  AGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTT  1184
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1139  AGAGACTGCCGGGGTCAACTCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGCCCCTT  1198

Mkomo_JERR01  1185  ATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGCCGTGAGGT  1244
                    |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Mtb_H37Rv     1199  ATGTCCAGGGCTTCACACATGCTACAATGGCCGGTACAAAGGGCTGCGATGCCGCGAGGT  1258

Mkomo_JERR01  1245  GGAGCGAATCCTTTCAAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACTCGACCCCGTG  1304
                    ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1259  TAAGCGAATCCTTA-AAAGCCGGTCTCAGTTCGGATCGGGGTCTGCAACTCGACCCCGTG  1317

Mkomo_JERR01  1305  AAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCT  1364
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1318  AAGTCGGAGTCGCTAGTAATCGCAGATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCT  1377

Mkomo_JERR01  1365  TGTACACACCGCCCGTCACGTCATGAAAGTCGGTAACACCCGAAGCCGGTGGCCTAACCC  1424
                    |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Mtb_H37Rv     1378  TGTACACACCGCCCGTCACGTCATGAAAGTCGGTAACACCCGAAGCCAGTGGCCTAACCC  1437

Mkomo_JERR01  1425  CTTGTGGGAGGGAGCCGTCGAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTA  1484
                     | |  ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1438  -TCG--GGAGGGAGCTGTCGAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTA  1494

Mkomo_JERR01  1485  GCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTT  1519
                    |||||||||||||||||||||||||||||||||||
Mtb_H37Rv     1495  GCCGTACCGGAAGGTGCGGCTGGATCACCTCCTTT  1529
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv rrs.*

```
Query: HJANABMJ_03930 ATP synthase subunit c Query ID: lcl|Query_6353975 Length: 81

>LFFBIPCO_01381 ATP synthase subunit c
Sequence ID: Query_6353977 Length: 81
Range 1: 1 to 81

Score:390, Identities:78/81(96%), Positives:80/81(98%), Gaps:0/81(0%)

Mkomo_JERR01  1   MDPTIAAGALIGGGLIMAGGAIGAGIGNGVAGNALISGIARQPEAQGRLFTPFFITVGLV  60
                  MDPTIAAGALIGGGLIMAGGAIGAGIG+GVAGNALISG+ARQPEAQGRLFTPFFITVGLV
Mtb_H37Rv     1   MDPTIAAGALIGGGLIMAGGAIGAGIGDGVAGNALISGVARQPEAQGRLFTPFFITVGLV  60

Mkomo_JERR01  61  EAAYFINLAFMALFVFATPVG  81
                  EAAYFINLAFMALFVFATPV
Mtb_H37Rv     61  EAAYFINLAFMALFVFATPVK  81
```

Global protein alignment between *M.* JERR01 and *Mtbs*_H37Rv AtpH.

```
Query: HJANABMJ_03142 Deazaflavin-dependent nitroreductase Query ID: lcl|Query_7832193 Length:
139

>LFFBIPCO_03735 Deazaflavin-dependent nitroreductase
Sequence ID: Query_7832195 Length: 151
Range 1: 1 to 151

Score:163, Identities:49/153(32%), Positives:72/153(47%), Gaps:16/153(10%)

Mkomo_JERR01  1    M--TDPEDMNTQLIQEF------------RANDGKVASFGNGEMPLLVLHTTGAKSGRKR  46
                   M  + P  +N+ L   F            R NDG+       ++P+ +L TTG K+G+ R
Mtb_H37Rv     1    MPKSPPRFLNSPLSDFFIKWMSRINTWMYRRNDGEGLGGTFQKIPVALLTTTGRKTGQPR  60

Mkomo_JERR01  47   VSLLIYQDLGGSWAVFASNGGASTHPAWYHNIIANPHVEIEVGSATHRALARVADMVERQ  106
                   V+ L +   GG   V AS GGA  +P WY N+ ANP V++++        AR A   ER
Mtb_H37Rv     61   VNPLYFLRDGGRVIVAASKGGAEKNPMWYLNLKANPKVQVQIKKEVLDLTARDATDEERA  120

Mkomo_JERR01  107  SIWDAQRAVMPEIDDIAANSSRLIPVVVLDPAP  139
                     W     + P   D  + + R IP+VV +P
Mtb_H37Rv     121  EYWPQLVTMYPSYQDYQSWTDRTIPIVVCEP--  151
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv Ddn_1*

```
Query: HJANABMJ_04822 Deazaflavin-dependent nitroreductase Query ID: lcl|Query_808081 Length:
166


>LFFBIPCO_03735 Deazaflavin-dependent nitroreductase
Sequence ID: Query_808083 Length: 151
Range 1: 1 to 151

Score:409, Identities:85/167(51%), Positives:109/167(65%), Gaps:17/167(10%)

Mkomo_JERR01  1    MAK-PPRPLSAKQVDGLNSPAVGVGIKWMSKINTWVYKATGGRVGYNWRLGSNRFSAPPP  59
                   M K PPR         LNSP     IKWMS+INTW+Y+   G        G        P
Mtb_H37Rv     1    MPKSPPR--------FLNSPLSDFFIKWMSRINTWMYRRNDGE-------GLGGTFQKIP  45

Mkomo_JERR01  60   IGILTTIGRKSGQPRESPLLFLREGDRIVLVASQGGRAANPMWYLNLKADPQVTFQIKNE  119
                   + +LTT GRK+GQPR +PL FLR+G R+++ AS+GG   NPMWYLNLKA+P+V  QIK E
Mtb_H37Rv     46   VALLTTTGRKTGQPRVNPLYFLRDGGRVIVAASKGGAEKNPMWYLNLKANPKVQVQIKKE  105

Mkomo_JERR01  120  KLRLVAREATDAERDEYWPRLDAIYPDFANYRSYTDRKIPILICDPV  166
                    L L AR+ATD ER EYWP+L  +YP + +Y+S+TDR IPI++C+P
Mtb_H37Rv     106  VLDLTARDATDEERAEYWPQLVTMYPSYQDYQSWTDRTIPIVVCEP-  151
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv Ddn_2*

```
Query: HJANABMJ_05664 N-acetyltransferase Eis Query ID: lcl|Query_6124285 Length: 250


>LFFBIPCO_02552 N-acetyltransferase Eis
Sequence ID: Query_6124287 Length: 400
Range 1: 1 to 400

Score:446, Identities:132/411(32%), Positives:162/411(39%), Gaps:172/411(41%)

Mkomo_JERR01      M-----------------------------------------------------------
                  M
Mtb_H37Rv       1 MTLCSPTEDDWPGMFLLAAASFTDFIGPESATAWRTLVPTDGAVVVRDGAGPGSEVVGMA  60

Mkomo_JERR01      ------------------------------------------------------------

Mtb_H37Rv      61 LYMDLRLTVPGEVVLPTAGLSFVAVAPTHRRRGLLRAMCAELHRRIADSGYPVAALHASE  120

Mkomo_JERR01    2 ---------------------------------------GVRVVAPAAHRARLEDIEQRW  22
                                                          VR+V P  HR   E I +RW
Mtb_H37Rv      121 GGIYGRFGYGPATTLHELTVDRRFARFHADAPGGGLGGSSVRLVRPTEHRGEFEAIYERW  180

Mkomo_JERR01   23 RLQTPGGLHTPAALWDEVLADREALRDGGSALYCLLHTDGYAMYRVHGGDETDDGKGRAE  82
                  R Q PGGL  P  LWDE+LA+ +A   G    + LLH DGYA+YRV   D TD     A
Mtb_H37Rv      181 RQQVPGGLLRPQVLWDELLAECKAAPGGDRESFALLHPDGYALYRV---DRTD--LKLAR  235

Mkomo_JERR01   83 VTHLAAVTPEAFIALWRALLGLDLMGSVNVVTHPDHLLPYLLTDPRLVRHRGSHDGLWLR  142
                  V+ L AVT +A  ALWRAL+GLD M  ++++THP   LP+LLTD RL R     DGLWLR
Mtb_H37Rv      236 VSELRAVTADAHCALWRALIGLDSMERISIITHPQDPLPHLLTDTRLARTTWRQDGLWLR  295

Mkomo_JERR01  143 LLDIPRALQARSYG---ADLSLVLDVTDGALGGGGGRFALDVRDGRAHCAATDADADVEL  199
                  ++++P AL+AR Y     + S VL+V+DG      GRFAL + DGRA C  TDA A++E+
Mtb_H37Rv      296 IMNVPAALEARGYAHEVGEFSTVLEVSDG------GRFALKIGDGRARCTPTDAAAEIEM  349

Mkomo_JERR01  200 DLSVLGSLYLGAHRASAFATAGRLRYAKPDHLVTLDSAFASDVPAELSFGF  250
                  D  VLGSLYLGAHRAS  A A RLR      L  LD+AFASDVP + +F F
Mtb_H37Rv      350 DRDVLGSLYLGAHRASTLAAANRLRTKDSQLLRRLDAAFASDVPVQTAFEF  400
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv* Eis

```
Query: HJANABMJ_05282 putative arabinosyltransferase A Query ID: lcl|Query_104055 Length: 1082


>LFFBIPCO_03991 putative arabinosyltransferase A
Sequence ID: Query_104057 Length: 1094
Range 1: 1 to 1094

Score:3831, Identities:743/1097(68%), Positives:862/1097(78%), Gaps:18/1097(1%)

Mkomo_JERR01  1     MPAPS-------ARLVAVIAGLLGMVLCGLAPLLPVTQSTATIVWPQAADEDGFVGDITA  53
                    MP          ARL AV++G+ G++LCG+ PLLPV Q+TATI WPQ +  DG +  ITA
Mtb_H37Rv     1     MPHDGNERSHRIARLAAVVSGIAGLLLCGIVPLLPVNQTTATIFWPQGSTADGNITQITA  60

Mkomo_JERR01  54    PLVSGAPRALEVTIPCAAVATLPDDDGVVFSTIPAGGIDAGRNGLFVRANADVVYVAFRD  113
                    PLVSGAPRAL+++IPC+A+ATLP + G+V ST+PAGG+D G+ GLFVRAN D V VAFRD
Mtb_H37Rv     61    PLVSGAPRALDISIPCSAIATLPANGGLVLSTLPAGGVDTGKAGLFVRANQDTVVVAFRD  120

Mkomo_JERR01  114   TVAAVAPRADVESGACSELRIWADVGAVGADFVGIPGAAGTLAADKRPQVAGVFTELEAT  173
                    +VAAVA R+ + +G CS L IWAD G  GADF+GIPG AGTL  +K+PQV G+FT+L+
Mtb_H37Rv     121   SVAAVAARSTIAAGGCSALHIWADTGGAGADFMGIPGGAGTLPPEKKPQVGGIFTDLKVG  180

Mkomo_JERR01  174   PDSGVRARIDVDTRFITAPTTLKLAVMVLGVLCVLAAVVALAMLDR-AWGRRL----PRG  228
                       G+ AR+D+DTRFIT P  LK AVM+LGVL VL A+V LA LDR + GR L     R
Mtb_H37Rv     181   AQPGLSARVDIDTRFITTPGALKKAVMLLGVLAVLVAMVGLAALDRLSRGRTLRDWLTRY  240

Mkomo_JERR01  229   ARRHRAGLWTWLTDAAVIGGLLIWHLVGALSSDDGYNTTIARVSGEAGYITNYYRYFGAS  288
                      R R G  + L DAAVI  LL+WH++GA SSDDGY  T+ARV+ +AGY+ NYYRYFG +
Mtb_H37Rv     241   RPRVRVGFASRLADAAVIATLLLWHVIGATSSDDGYLLTVARVAPKAGYVANYYRYFGTT  300

Mkomo_JERR01  289   EAPFDWYQSVLAQLASVSTASVWLRLPATAAAIGTWLIISRCVLPRLGRR---VAANRVA  345
                    EAPFDWY SVLAQLA+VSTA VW+RLPAT A I  WLI+SR VL RLG     +A+NRVA
Mtb_H37Rv     301   EAPFDWYTSVLAQLAAVSTAGVWMRLPATLAGIACWLIVSRFVLRRLGPGPGGLASNRVA  360

Mkomo_JERR01  346   VLTAGAVFLAAWLPFNNGLRPEPLIAFGVVAAWMLVENALGTRRLWPYAAAIVVAVFSVT  405
                    V TAGAVFL+AWLPFNNGLRPEPLIA GV+  W+LVE ++   RL P A AI+VA  + T
Mtb_H37Rv     361   VFTAGAVFLSAWLPFNNGLRPEPLIALGVLVTWVLVERSIALGRLAPAAVAIIVATLTAT  420

Mkomo_JERR01  406   LAPQGLVAVAPLLVGARGVTRIVAARRPAAGLPAQLAPLAAAASLVFVVVFRDQTLATVA  465
                    LAPQGL+A+APLL GAR + + +  RR   GL A LA LAAA SL+ VVVFRDQTLATVA
Mtb_H37Rv     421   LAPQGLIALAPLLTGARAIAQRIRRRRATDGLLAPLAVLAAALSLITVVVFRDQTLATVA  480

Mkomo_JERR01  466   ESVRIKYVVGPTIPWYQEFLRYYFLTVEESVDGSLTRRFSVLILLLCLFGLLAVLLRRGS  525
                    ES RIKY VGPTI WYQ+FLRYYFLTVE +V+GS++RRF+VL+LL CLFG+L VLLRRG
Mtb_H37Rv     481   ESARIKYKVGPTIAWYQDFLRYYFLTVESNVEGSMSRRFAVLVLLFCLFGVLFVLLRRGR  540

Mkomo_JERR01  526   LPGAVNGPVWRLAGTTAIGLLLLTFTPTKWAVQFGVFAGLAGALGAVTAFAFARVGLHSR  585
                    + G  +GP WRL GTTA+GLLLLTFTPTKWAVQFG FAGLAG LGAVTAF FAR+GLHSR
Mtb_H37Rv     541   VAGLASGPAWRLIGTTAVGLLLLTFTPTKWAVQFGAFAGLAGVLGAVTAFTFARIGLHSR  600

Mkomo_JERR01  586   RNLALYVTALLFVLAWATSGINGWFYTGNYGVPWFDKQPVIVGYPVTTIFLVLAIVGGLL  645
                    RNL LYVTALLFVLAWATSGINGWFY GNYGVPW+D QPVI  +PVT++FL L+I+ GLL
Mtb_H37Rv     601   RNLTLYVTALLFVLAWATSGINGWFYVGNYGVPWYDIQPVIASHPVTSMFLTLSILTGLL  660

Mkomo_JERR01  646   AGWLHFRMDYAGHTEVANTGRNRALASTPLLVVAVIMVVLQLGSMVKATVGRYPVYTTGA  705
                    A W HFRMDYAGHTEV +  RNR LASTPLLVVAVIMV  ++GSM KA V RYP+YTT
Mtb_H37Rv     661   AAWYHFRMDYAGHTEVKDNRRNRILASTPLLVVAVIMVAGEVGSMAKAAVFRYPLYTTAK  720

Mkomo_JERR01  706   ANIAALRSGLDGTSCAMADDVLVEADTNAGMLTPVPGQRFGEYGPLGGENPVGFTPNGVS  765
                    AN+ AL +GL  +SCAMADDVL E D NAGML PVPGQ FG  GPLGG +PVGF P GV
Mtb_H37Rv     721   ANLTALSTGL--SSCAMADDVLAEPDPNAGMLQPVPGQAFGPDGPLGGISPVGFKPEGVG  778

Mkomo_JERR01  766   DTLEPAEPVAANPGTVNSDGPVDKPNIGVGYAAGTGGGYGPEGVNGSRVFLPFGLDPGTT  825
                    + L+ ++PV + PG VNSD   +KPN  +  +AGT GG GP G+NGS   LPFGLDP  T
Mtb_H37Rv     779   EDLK-SDPVVSKPGLVNSDASPNKPNAAITDSAGTAGGKGPVGINGSHAALPFGLDPART  837

Mkomo_JERR01  826   PVMGSFGENTVAAKATSAWYQLPDRTPDRPLVSVAAAGAIWYYNEDGSFNYGQSLKLQWG  885
                    PVMGS+GEN +AA ATSAWYQLP R+PDRPLV V+AAGAIW Y EDG F YGQSLKLQWG
Mtb_H37Rv     838   PVMGSYGENNLAATATSAWYQLPPRSPDRPLVVVSAAGAIWSYKEDGDFIYGQSLKLQWG  897
```

```
Mkomo_JERR01  886   VHRPDGSYEELGEVQPIDIFVQKAWRNLRFPLAWAPPEANVARIVADDPNLSEDQWFAFT  945
                    V  PDG  + LG+V PIDI  Q AWRNLRFPLAWAPPEA+VARIVA DPNLS +QWFAFT
Mtb_H37Rv     898   VTGPDGRIQPLGQVFPIDIGPQPAWRNLRFPLAWAPPEADVARIVAYDPNLSPEQWFAFT  957

Mkomo_JERR01  946   PPRVPVLQSAQDYLGRDTPVLMDIATAANFPCQRPFAERLGIAELPEYRIMPNFKQIVVS  1005
                    PPRVPVL+S Q  +G  TPVLMDIATAANFPCQRPF+E LGIAELP+YRI+P+ KQ   S
Mtb_H37Rv     958   PPRVPVLESLQRLIGSATPVLMDIATAANFPCQRPFSEHLGIAELPQYRILPDHKQTAAS  1017

Mkomo_JERR01  1006  SNQWQAAQDGGPFLFIQALLRTESIPTYLRGDWYRDWGSLERYLRVVPEDEAPDAVIEEG  1065
                    SN WQ++  GGPFLF QALLRT +I TYLRGDWYRDWGS+E+Y R+VP D+APDAV+EEG
Mtb_H37Rv     1018  SNLWQSSSTGGPFLFTQALLRTSTIATYLRGDWYRDWGSVEQYHRLVPADQAPDAVVEEG  1077

Mkomo_JERR01  1066  STRVFGWNRGGPIRALP  1082
                       V GW R GPIRALP
Mtb_H37Rv     1078  VITVPGWGRPGPIRALP  1094
```

Global protein alignment between *M.* JERR01 and *Mtb* H37Rv EmbA

```
Query: HJANABMJ_05283 putative arabinosyltransferase B Query ID: lcl|Query_851067  Length: 1108


>LFFBIPCO_03992 putative arabinosyltransferase B
Sequence ID: Query_851069 Length: 1098
Range 1: 1 to 1098

Score:3916, Identities:757/1130(67%), Positives:887/1130(78%), Gaps:54/1130(4%)

Mkomo_JERR01  1     MSGTSGDAVSGTSGDAVNGTSDNAVNGTSDKAVNGTSVRDDVRIARWVATIAGLLGFVMA  60
                    M+  +    S         T + A+ G    A          R  RWVATIAGL+GFV++
Mtb_H37Rv     1     MTQCASRRKS---------TPNRAILGAFASA----------RGTRWVATIAGLIGFVLS  41

Mkomo_JERR01  61    VAVPLLPVTQTTATLNWPQQGELANVTAPLISQAPVNLTATVPCEVIAGMPADGGLVLGT  120
                    VA PLLPV QTTA L+WPQ+G+L +VTAPLIS  PV+ TATVPC+V+  MP  GG+VLGT
Mtb_H37Rv     42    VATPLLPVVQTTAMLDWPQRGQLGSVTAPLISLTPVDFTATVPCDVVRAMPPAGGVVLGT  101

Mkomo_JERR01  121   APAEGRDAALNAMLVNVTGGGTQRTETGGGRVDVIVRNVVVASVERDRVTGTGCERIEIT  180
                    AP +G+DA L A+ V V+            RVDV  RNVV+ SV R++VT   C+RIE+T
Mtb_H37Rv     102   APKQGKDANLQALFVVVSAQ----------RVDVTDRNVVILSVPREQVTSPQCQRIEVT  151

Mkomo_JERR01  181   SSLDGTFAEFVGLTDDDGAPLRSGYADPNLRPAIVGVFTDLTGPAPPGMSFSADIDTRFT  240
                    S+  GTFA FVGL D  GAPLRSG+ DPNLRP IVGVFTDLTGPAPPG++ SA IDTRF+
Mtb_H37Rv     152   STHAGTFANFVGLKDPSGAPLRSGFPDPNLRPQIVGVFTDLTGPAPPGLAVSATIDTRFS  211

Mkomo_JERR01  241   TEPTALKLAAIVLAIVCTVIALLALWRLDRLDGRR------------------MHRLIPT  282
                    T PT LKL AI+ AIV TV+AL+ALWRLD+LDGR                    M RLIP
Mtb_H37Rv     212   TRPTTLKLLAIIGAIVATVVALIALWRLDQLDGRGSIAQLLLRPFRPASSPGGMRRLIPA  271

Mkomo_JERR01  283   RWRTVTAVDGVVISGFAIWYVIGANSSDDGYILQMARVAEHAGYMSNYFRWFGSPEDPFG  342
                     WRT T  D VVI GF +W+VIGANSSDDGYIL MARVA+HAGYMSNYFRWFGSPEDPFG
Mtb_H37Rv     272   SWRTFTLTDAVVIFGFLLWHVIGANSSDDGYILGMARVADHAGYMSNYFRWFGSPEDPFG  331

Mkomo_JERR01  343   WYYNVLALMTQISTASIWIRLPDLICALVCWLLLSREVLPRLGPAVSGSRAALWAAGLVL  402
                    WYYN+LALMT +S AS+W+RLPDL   LVCWLLLSREVLPRLGPAV  S+ A WAA +VL
Mtb_H37Rv     332   WYYNLLALMTHVSDASLWMRLPDLAAGLVCWLLLSREVLPRLGPAVEASKPAYWAAAMVL  391

Mkomo_JERR01  403   LAAWMPFNNGLRPEGQIATGALITYVLIERAVTSGRLTPAALAVTTAAFTLGIQPTGLIA  462
                    L AWMPFNNGLRPEG IA G+L+TYVLIER++   RLTPAALAV TAAFTLG+QPTGLIA
Mtb_H37Rv     392   LTAWMPFNNGLRPEGIIALGSLVTYVLIERSMRYSRLTPAALAVVTAAFTLGVQPTGLIA  451

Mkomo_JERR01  463   VAALVAGGRPILRILMRRRQTVGLWPLLAPLLAAGTVILAVVFADQTLATVLEATKIRTA  522
                    VAALVAGGRP+LRIL+RR + VG  PL++P+LAAGTVIL VVFADQTL+TVLEAT++R
Mtb_H37Rv     452   VAALVAGGRPMLRILVRRHRLVGTLPLVSPMLAAGTVILTVVFADQTLSTVLEATRVRAK  511

Mkomo_JERR01  523   IGPSQEWYTENLRYYYLILPTTDGAIARRVAFLFTALCLFSSLFIMLRRKRVPGVARGPA  582
                    IGPSQ WYTENLRYYYLILPT DG+++RR  FL TALCLF+++FIMLRRKR+P VARGPA
Mtb_H37Rv     512   IGPSQAWYTENLRYYYLILPTVDGSLSRRFGFLITALCLFTAVFIMLRRKRIPSVARGPA  571

Mkomo_JERR01  583   WRLMGVIFATIFFLMFTPTKWVHHFGLFAAAGGAMAALATVLVSPVVLRSARNRMTFLAA  642
                    WRLMGVIF T+FFLMFTPTKWVHHFGLFAA G AMAAL TVLVSP VLR +RNRM FLAA
Mtb_H37Rv     572   WRLMGVIFGTMFFLMFTPTKWVHHFGLFAAVGAAMAALTTVLVSPSVLRWSRNRMAFLAA  631

Mkomo_JERR01  643   VLFVLALSFASTNGWWYVSNFGVPYNTSVPQLGGVTVSAVFFALFGIAALWAFLLHVGNK  702
                    + F+LAL +A+TNGWWYVS++GVP+N+++P++ G+TVS +FFALF IAA +A  LH   +
Mtb_H37Rv     632   LFFLLALCWATTNGWWYVSSYGVPFNSAMPKIDGITVSTIFFALFAIAAGYAAWLHFAPR  691

Mkomo_JERR01  703   Q--ESRLVHTLTAAPIPLAAGLMVVFMMASLAIGVVRQYPTYSNGWSNIRAFAGGCGLAD  760
                       E RL+  LT AP+P+ AG M    +AS+  G+VRQYPTYSNGWSN+RAF GGCGLAD
Mtb_H37Rv     692   GAGEGRLIRALTTAPVPIVAGFMAAVFVASMVAGIVRQYPTYSNGWSNVRAFVGGCGLAD  751

Mkomo_JERR01  761   DVLVEPNSNDGFLQPIPPAPGQPFWGPLGPLGGVDATGFSPNGVPDRIIAEAIRLNNPQP  820
                    DVLVEP++N GF++P+    G   WGPLGPLGGV+  GF+PNGVP+  +AEAI +   QP
Mtb_H37Rv     752   DVLVEPDTNAGFMKPLDGDSGS--WGPLGPLGGVNPVGFTPNGVPEHTVAEAIVMKPNQP  809

Mkomo_JERR01  821   GTDYDWNRPIALSRPGVNGSTVPLPYGLDPARVPVAGTYSTGPQQESRLTSAWYALPAAD  880
                    GTDYDW+ P  L+ PG+NGSTVPLPYGLDPARVP+AGTY+TG QQ+S L SAWY LP  D
Mtb_H37Rv     810   GTDYDWDAPTKLTSPGINGSTVPLPYGLDPARVPLAGTYTTGAQQQSTLVSAWYLLPKPD  869
```

```
Mkomo_JERR01  881   DAHPLVVITAAGTIAGTSVANAFTSGQTVELEYATAGPDGAPVLAGRVSPYDI-GPTP-S  938
                    D HPLVV+TAAG IAG SV + +T GQTV LEYA  GP GA V AGR+ P D+ G  P +
Mtb_H37Rv     870   DGHPLVVVTAAGKIAGNSVLHGYTPGQTVVLEYAMPGP-GALVPAGRMVPDDLYGEQPKA  928

Mkomo_JERR01  939   WRNLRFPRAQIADDAVAVRVIAEDLSLGQGDWVAVTPPRVPEVRSVQEYVGSQQPVLMDW  998
                    WRNLRF RA++  DAVAVRV+AEDLSL   DW+AVTPPRVP++RS+QEYVGS QPVL+DW
Mtb_H37Rv     929   WRNLRFARAKMPADAVAVRVVAEDLSLTPEDWIAVTPPRVPDLRSLQEYVGSTQPVLLDW  988

Mkomo_JERR01  999   AVGLAFPCQQPMLHANGVTEIPRFRISPDYFAKLQSTDTWQDGINGGLLGITDLLLRASV  1058
                    AVGLAFPCQQPMLHANG+ EIP+FRI+PDY AK   TDTW+DG NGGLLGITDLLLRA V
Mtb_H37Rv     989   AVGLAFPCQQPMLHANGIAEIPKFRITPDYSAKKLDTDTWEDGTNGGLLGITDLLLRAHV  1048

Mkomo_JERR01  1059  MSTYLSNDWGQDWGSLRRFDTIVDAEPAEIELGSTTHSGLWSPGPMRFQP  1108
                    M+TYLS DW +DWGSLR+FDT+VDA PA++ELG+ T SGLWSPG +R  P
Mtb_H37Rv     1049  MATYLSRDWARDWGSLRKFDTLVDAPPAQLELGTATRSGLWSPGKIRIGP  1098
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv* EmbB

```
Query: HJANABMJ_05281 putative arabinosyltransferase C Query ID: lcl|Query_1837519 Length: 1079


>LFFBIPCO_03990 putative arabinosyltransferase C
Sequence ID: Query_1837521 Length: 1094
Range 1: 1 to 1094

Score:4097, Identities:791/1100(72%), Positives:897/1100(81%), Gaps:27/1100(2%)

Mkomo_JERR01  1     MVQQAA---------SVEPTRAVANHRTVRLIAIVAGLLGTMLAVATPFLPVNQTTAELN  51
                    M  +AA         S     A AN+R  R +A+VAGLLG +LA+ATP LPVNQTTA+LN
Mtb_H37Rv     1     MATEAAPPRIAVRLPSTSVRDAGANYRIARYVAVVAGLLGAVLAIATPLLPVNQTTAQLN  60

Mkomo_JERR01  52    WPQEGVLGSVDAPLIGYVATDLEITVPCSAAAGL---DRPGRTVLLSTVPKQAPNAVDRG  108
                    WPQ G   SV+APLIGYVATDL ITVPC AAAGL      G+TVLLSTVPKQAP AVDRG
Mtb_H37Rv     61    WPQNGTFASVEAPLIGYVATDLNITVPCQAAAGLAGSQNTGKTVLLSTVPKQAPKAVDRG  120

Mkomo_JERR01  109   LLIERVNDDLLVIVRNTPVVSAPLDQVLSPACDQLRFTAHADRVTGEFVGLQTEPEGAGQ  168
                    LL++R NDDL+++VRN P+V+APL QVL P C +L FTAHADRV  EFVGL    +G
Mtb_H37Rv     121   LLLQRANDDLVLVVRNVPLVTAPLSQVLGPTCQRLTFTAHADRVAAEFVGLV---QGPNA  177

Mkomo_JERR01  169   EPSAEPLRGERSGYDFRPQIVGVFTDLSGPAPPGLEFSATIDTRYSTSPTLLKLLAMIVG  228
                    E    PLRGERSGYDFRPQIVGVFTDL+GPAPPGL FSA++DTRYS+SPT LK+ AMI+G
Mtb_H37Rv     178   EHPGAPLRGERSGYDFRPQIVGVFTDLAGPAPPGLSFSASVDTRYSSSPTPLKMAAMILG  237

Mkomo_JERR01  229   VAMTVIALGALHVLDRADGRRHKRFLPPRWWSLTPLDGLVAAVLVWWHFVGANTADDGYI  288
                    VA+T  AL ALH+LD ADG RH+RFLP RWWS   LD LV AVLVWWHFVGANT+DDGYI
Mtb_H37Rv     238   VALTGAALVALHILDTADGMRHRRFLPARWWSTGGLDTLVIAVLVWWHFVGANTSDDGYI  297

Mkomo_JERR01  289   LTMARVSEGAGYMANYYRWFGTPEAPFGWYYDLLALWANVSTTSVWMRLPTLLMGLACWW  348
                    LTMARVSE AGYMANYYRWFGTPEAPFGWYYDLLALWA+VST S+WMRLPTL M L CWW
Mtb_H37Rv     298   LTMARVSEHAGYMANYYRWFGTPEAPFGWYYDLLALWAHVSTASIWMRLPTLAMALTCWW  357

Mkomo_JERR01  349   VISREVLPRLGTAVKHSQAAAWTAAGLFLAFWLPLNNGLRPEPIIALGILLTWCSVERGV  408
                    VISREV+PRLG AVK S+AAAWTAAG+FLA WLPL+NGLRPEPIIALGILLTWCSVER V
Mtb_H37Rv     358   VISREVIPRLGHAVKTSRAAAWTAAGMFLAVWLPLDNGLRPEPIIALGILLTWCSVERAV  417

Mkomo_JERR01  409   ATSRLLPVAVAIIIGALTLFSGPTGIAAIGALLVAIGPLKTIVAAHVSRFGYLALLAPIA  468
                    ATSRLLPVA+A IIGALTLFSGPTGIA+IGALLVAIGPL+TI+     RFG L L+API
Mtb_H37Rv     418   ATSRLLPVAIACIIGALTLFSGPTGIASIGALLVAIGPLRTILHRRSRRFGVLPLVAPIL  477

Mkomo_JERR01  469   AAGTVTIFLIFRDQTLASELQASSFKSAVGPSLAWFDEHIRYSRLFTTSPDGSVARRFAV  528
                    AA TVT   IFRDQT A E+QA+  K AVGPSL WFDEHIRY RLF  SPDGS+ARRFAV
Mtb_H37Rv     478   AAATVTAIPIFRDQTFAGEIQANLLKRAVGPSLKWFDEHIRYERLFMASPDGSIARRFAV  537

Mkomo_JERR01  529   LTLLLALAVAVAMTLRKGRIPGTATGPTLRIIGISVISFLAMMFTPTKWTHHFGVFAGLA  588
                    L L+LALAV+VAM+LRKGRIPGTA GP+ RIIGI++ISFLAMMFTPTKWTHHFGVFAGLA
Mtb_H37Rv     538   LALVLALAVSVAMSLRKGRIPGTAAGPSRRIIGITIISFLAMMFTPTKWTHHFGVFAGLA  597

Mkomo_JERR01  589   GSLGALAAVAISAAAMRSPRNRAMFTAAVLFLTALSFATVNGWWYVSNFGVPWSNAFPEW  648
                    GSLGALAAVA++ AAMRS RNR +F A V+F+ ALSFA+VNGWWYVSNFGVPWSN+FP+W
Mtb_H37Rv     598   GSLGALAAVAVTGAAMRSRRNRTVFAAVVVFVLALSFASVNGWWYVSNFGVPWSNSFPKW  657

Mkomo_JERR01  649   KFGFTTMLLGLSLVALLVAAWLHF-SGRDVPSDVAPPRWKR----ITQAPLAIVTWALVI  703
                    ++  TT LL L+++ LL+AAW HF +  D      P R++     I Q+PLAI TW LV+
Mtb_H37Rv     658   RWSLTTALLELTVLVLLLAAWFHFVANGDGRRTARPTRFRARLAGIVQSPLAIATWLLVL  717

Mkomo_JERR01  704   FQVFSLTVAMLGQYPAWTVGRSNLQALTGKTCGMAEDVLVEPDANAGMLTPIGVAIGDAL  763
                    F+V SLT AM+ QYPAW+VGRSNLQAL GKTCG+AEDVLVE D NAGML P+   + DAL
Mtb_H37Rv     718   FEVVSLTQAMISQYPAWSVGRSNLQALAGKTCGLAEDVLVELDPNAGMLAPVTAPLADAL  777

Mkomo_JERR01  764   GEVTSDGFRPNGIPSDVSADPVMEQPGSDNFADSDGSGETGNEAGTEGGTTAAAGVNGSR  823
                    G   S+ F PNGIP+DV+ADPVME+PG  +F + DG   TG+E GTEGGTTAA G+NGSR
Mtb_H37Rv     778   GAGLSEAFTPNGIPADVTADPVMERPGDRSFLNDDGL-ITGSEPGTEGGTTAAPGINGSR  836

Mkomo_JERR01  824   ARLPYGLDPATTPVLGSWRSGTQQPAFLRSAWYQLPPGWNAQDRSESLLVISAAGRFDAS  883
                    ARLPY LDPA TPVLGSWR+G Q PA LRS WY+LP   N Q     LLV++AAGRFD+
Mtb_H37Rv     837   ARLPYNLDPARTPVLGSWRAGVQVPAMLRSGWYRLPT--NEQRDRAPLLVVTAAGRFDSR  894
```

```
Mkomo_JERR01  884   EVVVQWAGPD----GEPAGSIEFGDVGAAPAWRNLRAPLSAIPAEATRIRLVASDDDLSP  939
                    EV +QWA  +    G   GS+EF DVGAAPAWRNLRAPLSAIP+ AT++RLVA D DL+P
Mtb_H37Rv     895   EVRLQWATDEQAAAGHHGGSMEFADVGAAPAWRNLRAPLSAIPSTATQVRLVADDQDLAP  954

Mkomo_JERR01  940   EHWIAVTPPRIPELRTLQDVVGSSDPVLLDWLVGLAFPCQRPFGHRNGVIEVPQWRILPD  999
                    +HWIA+TPPRIP +RTLQ+VVG++DPV LDWLVGLAFPCQRPFGH+ GV E P+WRILPD
Mtb_H37Rv     955   QHWIALTPPRIPRVRTLQNVVGAADPVFLDWLVGLAFPCQRPFGHQYGVDETPKWRILPD  1014

Mkomo_JERR01  1000  RFGAEANSPVMDYLGGGPLGITELLFRAMTVPTYLKDDWFRDWGALQRLVPFYPDAEPAR  1059
                    RFGAEANSPVMD+ GGGPLGITELL RA TV +YLKDDWFRDWGALQRL P+YPDA+PA
Mtb_H37Rv     1015  RFGAEANSPVMDHNGGGPLGITELLMRATTVASYLKDDWFRDWGALQRLTPYYPDAQPAD  1074

Mkomo_JERR01  1060  LDLGTEQRSGLWSPAPLRLS  1079
                    L+LGT  RSGLWSPAPLR
Mtb_H37Rv     1075  LNLGTVTRSGLWSPAPLRRG  1094
```

Global protein alignment between *M.* JERR01 and *Mtb* H37Rv EmbC

```
Query: HJANABMJ_01600 Transcriptional regulatory protein EmbR Query ID: lcl|Query_4186263
Length: 384


>LFFBIPCO_01340 Transcriptional regulatory protein EmbR
Sequence ID: Query_4186265 Length: 388
Range 1: 1 to 388

Score:1106, Identities:232/398(58%), Positives:279/398(70%), Gaps:24/398(6%)

Mkomo_JERR01  1    MGA-------LEFAVLGPLKLTVDGADTAVGTPKLRALLAMLVINRNRPVAVDSLIDALW  53
                   M         L+F +LGPL++T+DG     GTPK RA+LAMLVINRNRPV VD+LI ALW
Mtb_H37Rv     1    MAGSATVEKRLDFGLLGPLQMTIDGTPVPSGTPKQRAVLAMLVINRNRPVGVDALITALW  60

Mkomo_JERR01  54   GESPPAGARTSLHSYVSNLRRLL-----TPHHVLDSVPPGYRLKVADTACDLDRFAADKA  108
                    E PP+GAR S+HSYVSNLR+LL      P  VL + PPGYRL + D  CDL RF A+K
Mtb_H37Rv     61   EEWPPSGARASIHSYVSNLRKLLGGAGIDPRVVLAAAPPGYRLSIPDNTCDLGRFVAEKT  120

Mkomo_JERR01  109  AGVQAAAVGDFETARRRLSTALRHWRGPVLEDLRAFSFVGPFAAALDEERLTVATHHAEA  168
                   AGV AAA G FE A R LS ALR WRGPVL+DLR F FV PFA AL E+++   T  AEA
Mtb_H37Rv     121  AGVHAAAAGRFEQASRHLSAALREWRGPVLDDLRDFQFVEPFATALVEDKVLAHTAKAEA  180

Mkomo_JERR01  169  ELACGRAADVIGELEQLAVENPYREPIWVQLISAYYLTERQSEALAAYQRLKTTLADELG  228
                   E+ACGRA+ VI ELE L  E+PYREP+W QLI+AYYL++RQS+AL AY+R+KTTLAD+LG
Mtb_H37Rv      181 EIACGRASAVIAELEALTFEHPYREPLWTQLITAYYLSDRQSDALGAYRRVKTTLADDLG  240

Mkomo_JERR01  229  IDPGPTVQDLHGKILRQLPLDIRGPARA-AAGAETVLEQRTQIDSGAAVAWLVE-ATGIR  286
                   IDPGPT++ L+ +ILRQ PLD +  A+  AAG  TVL+QRT      AVA+L + A+G
Mtb_H37Rv     241  IDPGPTLRALNERILRQQPLDAKKSAKTTAAGTVTVLDQRTMASGQQAVAYLHDIASGRG  300

Mkomo_JERR01  287  HPLNSLATRIGRLPDNDIVLDDAKVSRHHAAIIDTGTGFAINDLRSANGVEVGARRIQGS  346
                   +PL + ATRIGRL DNDIVLD A VSRHHA I+DTGT + INDLRS+NGV V   RI+ +
Mtb_H37Rv     301  YPLQAAATRIGRLHDNDIVLDSANVSRHHAVIVDTGTNYVINDLRSSNGVHVQHERIRSA  360

Mkomo_JERR01  347  APLTHGDQIRICGHVFTYEHSTAGHPDTEHPTGSAADA  384
                     L  GD IRIC H FT++ S   H  T
Mtb_H37Rv     361  VTLNDGDHIRICDHEFTFQISAGTHGGT----------  388
```

Global protein alignment between *M.* JERR01 and *Mtb H37Rv* EmbR

```
Query: HJANABMJ_01813 FAD-containing monooxygenase EthA Query ID: lcl|Query_4600907 Length: 513


>LFFBIPCO_04057 FAD-containing monooxygenase EthA
Sequence ID: Query_4600909 Length: 489
Range 1: 1 to 489

Score:1150, Identities:230/515(45%), Positives:308/515(59%), Gaps:28/515(5%)

Mkomo_JERR01  1    MDQVYPETEFSDVLIIGAGISGIGAAYRIQERNPALKYTILERRSRIGGTWDLHRYPGIR  60
                   M      TE  DV+I+GAGISG+ AA+ +Q+R P   Y ILE+R  +GGTWDL RYPGIR
Mtb_H37Rv     1    M------TEHLDVVIVGAGISGVSAAWHLQDRCPTKSYAILEKRESMGGTWDLFRYPGIR  54

Mkomo_JERR01  61   SDSDIFTLSFPWEPWTRPENVADGADIRTYLTETAHKHGIDRHIRFDTRVTSANWDSNTD  120
                   SDSD++TL F + PWT  + +ADG  I  Y+  TA  +GIDRHIRF  +V SA+W +  +
Mtb_H37Rv     55   SDSDMYTLGFRFRPWTGRQAIADGKPILEYVKSTAAMYGIDRHIRFHHKVISADWSTAEN  114

Mkomo_JERR01  121  TWTVQAEQNGASRVHRARFLFFATGYYDYDHPYRPDFPGLENFTGEVVHPQHWPEDLDYS  180
                    WTV  + +G        FLF  +GYY+YD  Y P F G E+F G ++HPQHWPEDLDY
Mtb_H37Rv     115  RWTVHIQSHGTLSALTCEFLFLCSGYYNYDEGYSPRFAGSEDFVGPIIHPQHWPEDLDYD  174

Mkomo_JERR01  181  GKNVVVIGSGATAISLIPSLTERAA-HVTMVQRSPTYLLSGQRINPIVNLLRKVPPRRLG  239
                    KN+VVIGSGATA++L+P+L +  A HVTM+QRSPTY++S    + I   L +  P  +
Mtb_H37Rv     175  AKNIVVIGSGATAVTLVPALADSGAKHVTMLQRSPTYIVSQPDRDGIAEKLNRWLPETMA  234

Mkomo_JERR01  240  YRLAWLYNVLFIVAVYAIARKAPRFSRRAIRAVAKHYLPAGYPVDTHFNPSYDPWDQRLC  299
                   Y      NVL   AVY+  +K PR  R+   ++ +  LP GY V  HF P Y+PWDQRLC
Mtb_H37Rv     235  YTAVRWKNVLRQAAVYSACQKWPRRMRKMFLSLIQRQLPEGYDVRKHFGPHYNPWDQRLC  294

Mkomo_JERR01  300  LILSGDFYDAISAGRAEVVTDRIDHLDAHGVVLQSGRRIDADVLVTATGIQLQALGGIVV  359
                   L+ +GD + AI  G+ EVVTD I+   A G+ L SGR + AD+++TATG+ LQ  GG
Mtb_H37Rv     295  LVPNGDLFRAIRHGKVEVVTDTIERFTATGIRLNSGRELPADIIITATGLNLQLFGGATA  354

Mkomo_JERR01  360  SVDGEQVKPADRFVYKEHLLEDVPNMAWCVGYINASWTLRADLTARAVAKLLAYMDAHGY  419
                   ++DG+QV       YK  +L  +PNMA+ VGY NASWTL+ADL +  V +LL YMD +G+
Mtb_H37Rv     355  TIDGQQVDITTTMAYKGMMLSGIPNMAYTVGYTNASWTLKADLVSEFVCRLLNYMDDNGF  414

Mkomo_JERR01  420  THAYPHLGQTPMPEKPAWNINAGYVHRAAHVLPKSGTHRPWNVRHNYVLDAIDHRLDRIE  479
                            + + E+P      GYV R+   LPK G+  PW +  NY+ D    R  +I+
Mtb_H37Rv     415  DTVVVERPGSDVEERPFMEFTPGYVLRSLDELPKQGSRTPWRLNQNYLRDIRLIRRGKID  474

Mkomo_JERR01  480  -ESMVFGRTPPKLYGGSELPDGAVLETSLERPVGG  513
                    E + F + P                     PVG
Mtb_H37Rv     475  DEGLRFAKRPA--------------------PVGV  489
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 EthA_1.

```
Query: HJANABMJ_05338 FAD-containing monooxygenase EthA Query ID: lcl|Query_3721637 Length: 491


>LFFBIPCO_04057 FAD-containing monooxygenase EthA
Sequence ID: Query_3721639 Length: 489
Range 1: 1 to 489

Score:1761, Identities:318/491(65%), Positives:384/491(78%), Gaps:2/491(0%)

Mkomo_JERR01  1    MTEFVDVVIVGAGISGISAAWHLQDRCPGKSYVILERRDNLGGTWDLFKYPGIRSDSDMF  60
                   MTE +DVVIVGAGISG+SAAWHLQDRCP KSY ILE+R+++GGTWDLF+YPGIRSDSDM+
Mtb_H37Rv     1    MTEHLDVVIVGAGISGVSAAWHLQDRCPTKSYAILEKRESMGGTWDLFRYPGIRSDSDMY  60

Mkomo_JERR01  61   TLGFRFKPWTSEKAIADGPDIMSYLKETVAEYGIDKHIRYGQNVLSADWSDADGRWTVRV  120
                   TLGFRF+PWT  +AIADG  I+ Y+K T A YGID+HIR+   V+SADWS A+ RWTV +
Mtb_H37Rv     61   TLGFRFRPWTGRQAIADGKPILEYVKSTAAMYGIDRHIRFHHKVISADWSTAENRWTVHI  120

Mkomo_JERR01  121  DRGGEQIEISTRYLMACSGYYNYDEGYSPEFPGAGNFAGTVIHPQHWPEDLDYAGKRIVV  180
                      G    ++  +L  CSGYYNYDEGYSP F G+ +F G +IHPQHWPEDLDY  K IVV
Mtb_H37Rv     121  QSHGTLSALTCEFLFLCSGYYNYDEGYSPRFAGSEDFVGPIIHPQHWPEDLDYDAKNIVV  180

Mkomo_JERR01  181  IGSGATAVTLIPALANSGAAHVTMLQRSPTFIAALPDVDPFTVATNKRLPAKAAYVLNRW  240
                   IGSGATAVTL+PALA+SGA HVTMLQRSPT+I + PD D      N+ LP   AY   RW
Mtb_H37Rv     181  IGSGATAVTLVPALADSGAKHVTMLQRSPTYIVSQPDRDGIAEKLNRWLPETMAYTAVRW  240

Mkomo_JERR01  241  KSILAQSAQYQVARKFPDLFRKQLLTMAKHRLPEGYDVHRHFNPRYNPWDERVCLAPNGD  300
                   K++L Q+A Y   +K+P   RK  L++ + +LPEGYDV +HF P YNPWD+R+CL PNGD
Mtb_H37Rv     241  KNVLRQAAVYSACQKWPRRMRKMFLSLIQRQLPEGYDVRKHFGPHYNPWDQRLCLVPNGD  300

Mkomo_JERR01  301  LFKTIRAGKADVVTDTIDTFTETGIQLASGAHLDADIIVTATGLNLQLFGGATISRNGEA  360
                   LF+ IR GK +VVTDTI+ FT TGI+L SG  L ADII+TATGLNLQLFGGAT + +G+
Mtb_H37Rv     301  LFRAIRHGKVEVVTDTIERFTATGIRLNSGRELPADIIITATGLNLQLFGGATATIDGQQ  360

Mkomo_JERR01  361  IELNDTMAYKGMMLTNMPNMAFTIGYTNASWTLKADLVSEFFCRVINYMDAHSYDRVEPR  420
                   +++  TMAYKGMML+ +PNMA+T+GYTNASWTLKADLVSEF CR++NYMD + +D V
Mtb_H37Rv     361  VDITTTMAYKGMMLSGIPNMAYTVGYTNASWTLKADLVSEFVCRLLNYMDDNGFDTVVVE  420

Mkomo_JERR01  421  HPGRSVDERPLMDFTPGYVLRALDYLPKAGSRAPWRLKQNYLLDLQLIRRGKVDDEALAF  480
                    PG  V+ERP M+FTPGYVLR+LD LPK GSR PWRL QNYL D++LIRRGK+DDE L F
Mtb_H37Rv     421  RPGSDVEERPFMEFTPGYVLRSLDELPKQGSRTPWRLNQNYLRDIRLIRRGKIDDEGLRF  480

Mkomo_JERR01  481  SKHPAPVTASA  491
                   +K PAPV
Mtb_H37Rv     481  AKRPAPV--GV  489
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 EthA_2

```
Query: HJANABMJ_05694 Ribosomal RNA small subunit methyltransferase G Query ID:
lcl|Query_1934251 Length: 228


>LFFBIPCO_04125 Ribosomal RNA small subunit methyltransferase G
Sequence ID: Query_1934253 Length: 224
Range 1: 1 to 224

Score:657, Identities:135/230(59%), Positives:168/230(73%), Gaps:8/230(3%)

Mkomo_JERR01  1    MKHADIPSPPSAADSVFGATLPAAEKYAQILAGAGVERGLLGPREVDRLWDRHVLNCAVV  60
                   M     P  P AA ++FG  L  A +YA+ LAG GVERGL+GPREV RLWDRH+LNCAV+
Mtb_H37Rv     1    MS----PIEP-AASAIFGPRLGLARRYAEALAGPGVERGLVGPREVGRLWDRHLLNCAVI  55

Mkomo_JERR01  61   GELLNTGERIADIGSGAGLPGIPLALARPDVHVILVEPLLRRSEFLREVIEELGIVCEVV  120
                   GELL  G+R+ DIGSGAGLPG+PLA+ARPD+ V+L+EPLLRR+E LRE++ +LG+  E+V
Mtb_H37Rv     56   GELLERGDRVVDIGSGAGLPGVPLAIARPDLQVVLLEPLLRRTESLREMVTDLGVAVEIV  115

Mkomo_JERR01  121  RGRAEDRAVREEMGGSMDVVVSRAVASLDKLTKWSAPLLRPGGRMLAIKGERADEEVEQH  180
                   RGRAE+  V++++GGS D  VSRAVA+LDKLTKWS PL+RP GRMLAIKGERA +EV +H
Mtb_H37Rv     116  RGRAEESWVQDQLGGS-DAAVSRAVAALDKLTKWSMPLIRPNGRMLAIKGERAHDEVREH  174

Mkomo_JERR01  181  RRAMAALGLSKVKVERCGGQYVQPPATVVVGLM--ETASPSRPRPDRRHK  228
                   RR M A G   V+V  CG  Y++PPATVV      + A  S
Mtb_H37Rv     175  RRVMIASGAVDVRVVTCGANYLRPPATVVFARRGKQIARGSARMASGGTA  224
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 Gid

```
Query: HJANABMJ_00007 DNA gyrase subunit A Query ID: lcl|Query_5061089 Length: 841


>LFFBIPCO_00006 DNA gyrase subunit A
Sequence ID: Query_5061091 Length: 818
Range 1: 1 to 818

Score:3806, Identities:752/841(89%), Positives:791/841(94%), Gaps:23/841(2%)

Mkomo_JERR01  1    MTDTTLPPGGEAGDRIEPVDIQQEMQRSYIDYAMSVIVGRALPEVRDGLKPVHRRVLYAM  60
                   MTDTTLPP  ++ DRIEPVDI+QEMQRSYIDYAMSVIVGRALPEVRDGLKPVHRRVLYAM
Mtb_H37Rv     1    MTDTTLPPD-DSLDRIEPVDIEQEMQRSYIDYAMSVIVGRALPEVRDGLKPVHRRVLYAM  59

Mkomo_JERR01  61   FDSGFRPDRGHAKSARSVAETMGNYHPHGDSSIYDTLVRMAQPWSLRYPLVDGQGNFGSP  120
                   FDSGFRPDR HAKSARSVAETMGNYHPHGD+SIYD+LVRMAQPWSLRYPLVDGQGNFGSP
Mtb_H37Rv     60   FDSGFRPDRSHAKSARSVAETMGNYHPHGDASIYDSLVRMAQPWSLRYPLVDGQGNFGSP  119

Mkomo_JERR01  121  GNDPPAAMRYTEARLTPLAMEMLREIDEETVDFVPNYDGRVQEPTVLPSRFPNLLANGSG  180
                   GNDPPAAMRYTEARLTPLAMEMLREIDEETVDF+PNYDGRVQEPTVLPSRFPNLLANGSG
Mtb_H37Rv     120  GNDPPAAMRYTEARLTPLAMEMLREIDEETVDFIPNYDGRVQEPTVLPSRFPNLLANGSG  179

Mkomo_JERR01  181  GIAVGMATNIPPHNLRELAEAVYWCLDNHEADEEATCAAVMERVKGPDFPTRGLIVGSQG  240
                   GIAVGMATNIPPHNLRELA+AV+W L+NH+ADEE T AAVM RVKGPDFPT GLIVGSQG
Mtb_H37Rv     180  GIAVGMATNIPPHNLRELADAVFWALENHDADEEETLAAVMGRVKGPDFPTAGLIVGSQG  239

Mkomo_JERR01  241  IEEAYKTGRGSIRMRGVVEIEEDSRGRTSLVITELPYQVNHDNFITSIAEQVRDGKLSGI  300
                     +AYKTGRGSIRMRGVVE+EEDSRGRTSLVITELPYQVNHDNFITSIAEQVRDGKL+GI
Mtb_H37Rv     240  TADAYKTGRGSIRMRGVVEVEEDSRGRTSLVITELPYQVNHDNFITSIAEQVRDGKLAGI  299

Mkomo_JERR01  301  SNIEDQSSDRVGLRIVVEIKRDAVAKVVLNNLYKHTQLQTSFGANMLSIVDGVPRTLRLD  360
                   SNIEDQSSDRVGLRIV+EIKRDAVAKVV+NNLYKHTQLQTSFGANML+IVDGVPRTLRLD
Mtb_H37Rv     300  SNIEDQSSDRVGLRIVIEIKRDAVAKVVINNLYKHTQLQTSFGANMLAIVDGVPRTLRLD  359

Mkomo_JERR01  361  QMIRHYVDHQLEVIVRRTRYRLRKANERAHILRGLVKALDALDEVIALIRASQTVEIARS  420
                   Q+IR+YVDHQL+VIVRRT YRLRKANERAHILRGLVKALDALDEVIALIRAS+TV+IAR+
Mtb_H37Rv     360  QLIRYYVDHQLDVIVRRTTYRLRKANERAHILRGLVKALDALDEVIALIRASETVDIARA  419

Mkomo_JERR01  421  GLIELLDIDEIQAQAILDMQLRRLAALERQRIVDDLAKIEAEIADLEDILAKPERQRAIV  480
                   GLIELLDIDEIQAQAILDMQLRRLAALERQRI+DDLAKIEAEIADLEDILAKPERQR IV
Mtb_H37Rv     420  GLIELLDIDEIQAQAILDMQLRRLAALERQRIIDDLAKIEAEIADLEDILAKPERQRGIV  479

Mkomo_JERR01  481  HDELKEIVDKYGDDRRTRIIPADGEVADEDLIAREDVVVTITETGYAKRTKTDLYRSQKR  540
                    DEL EIVD++GDDRRTRII ADG+V+DEDLIAREDVVVTITETGYAKRTKTDLYRSQKR
Mtb_H37Rv     480  RDELAEIVDRHGDDRRTRIIAADGDVSDEDLIAREDVVVTITETGYAKRTKTDLYRSQKR  539

Mkomo_JERR01  541  GGKGVQGAGLKQDDIVNHFFVCSTHDWILFFTTQGRVYRAKAYDLPEASRTARGQHVANL  600
                   GGKGVQGAGLKQDDIV HFFVCSTHD ILFFTTQGRVYRAKAYDLPEASRTARGQHVANL
Mtb_H37Rv     540  GGKGVQGAGLKQDDIVAHFFVCSTHDLILFFTTQGRVYRAKAYDLPEASRTARGQHVANL  599

Mkomo_JERR01  601  LAFQPEERIAQVIQIKSYEDAPYLVLATRNGLVKKSRLVDFDSNRSGGIVAVNLRDGDEL  660
                   LAFQPEERIAQVIQI+ Y DAPYLVLATRNGLVKKS+L DFDSNRSGGIVAVNLRD DEL
Mtb_H37Rv     600  LAFQPEERIAQVIQIRGYTDAPYLVLATRNGLVKKSKLTDFDSNRSGGIVAVNLRDNDEL  659

Mkomo_JERR01  661  VGAVLCSSDDDLLLVSANGQSIRFSATDEALRPMGRATSGVQGMRFNAEDALLSLNVIRE  720
                   VGAVLCS+ DDLLLVSANGQSIRFSATDEALRPMGRATSGVQGMRFN +D LLSLNV+RE
Mtb_H37Rv     660  VGAVLCSAGDDLLLVSANGQSIRFSATDEALRPMGRATSGVQGMRFNIDDRLLSLNVVRE  719

Mkomo_JERR01  721  DTYLLVATSGGYAKRTAIEEYPVQGRGGKGVLTIQFDKRRGTLVGALIVDDDTELYAITS  780
                    TYLLVATSGGYAKRTAIEEYPVQGRGGKGVLT+ +D+RRG LVGALIVDDD+ELYA+TS
Mtb_H37Rv     720  GTYLLVATSGGYAKRTAIEEYPVQGRGGKGVLTVMYDRRRGRLVGALIVDDDSELYAVTS  779

Mkomo_JERR01  781  GGGVIRTAARQVRKAGRQTKGVRLMNLGDGDTLIAIARNAEAGDGTDEVNTDPEATTSEE  840
                   GGGVIRTAARQVRKAGRQTKGVRLMNLG+GDTL+AIARN
Mtb_H37Rv     780  GGGVIRTAARQVRKAGRQTKGVRLMNLGEGDTLLAIARN----------------------  818

Mkomo_JERR01  841  S  841
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 GyrA

```
Query: HJANABMJ_00006 DNA gyrase subunit B Query ID: lcl|Query_6986909 Length: 675


>LFFBIPCO_00005 DNA gyrase subunit B
Sequence ID: Query_6986911 Length: 714
Range 1: 1 to 714

Score:3018, Identities:588/714(82%), Positives:630/714(88%), Gaps:39/714(5%)

Mkomo_JERR01  1    M-------------------------------------AAQKKNAPAEYGADSIKVLE  21
                   M                                     AAQKK A  EYGA SI +LE
Mtb_H37Rv     1    MGKNEARRSALAPDHGTVVCDPLRRLNRMHATPEESIRIVAAQKKKAQDEYGAASITILE  60

Mkomo_JERR01  22   GLEAVRKRPGMYIGSTGERGLHHLIWEVVDNAVDEAMAGFATKVDVRILEDGGVQVTDDG  81
                   GLEAVRKRPGMYIGSTGERGLHHLIWEVVDNAVDEAMAG+AT V+V +LEDGGV+V DDG
Mtb_H37Rv     61   GLEAVRKRPGMYIGSTGERGLHHLIWEVVDNAVDEAMAGYATTVNVVLLEDGGVEVADDG  120

Mkomo_JERR01  82   RGIPVAMHATGIPTVDVVMTVLHAGGKFEEGAYQVSGGLHGVGVSVVNALSTRLEADIRK  141
                   RGIPVA HA+GIPTVDVVMT LHAGGKF+  AY +SGGLHGVGVSVVNALSTRLE +I++
Mtb_H37Rv     121  RGIPVATHASGIPTVDVVMTQLHAGGKFDSDAYAISGGLHGVGVSVVNALSTRLEVEIKR  180

Mkomo_JERR01  142  DGHEWFQTYDRSVPGTLRQGGQTKETGTTIRFWADPDIFETTNYDFETIARRLQEMAFLN  201
                   DG+EW Q Y++S P  L+QG  TK+TG+T+RFWADP +FETT YDFET+ARRLQEMAFLN
Mtb_H37Rv     181  DGYEWSQVYEKSEPLGLKQGAPTKKTGSTVRFWADPAVFETTEYDFETVARRLQEMAFLN  240

Mkomo_JERR01  202  KGLTIELTDERVAPEEVVDDIVSDHAEAPKSAEEKAAEAAAPQKVKHRVFHYPGGLVDFV  261
                   KGLTI LTDERV  +EVVD++VSD AEAPKSA E+AAE+ AP KVK R FHYPGGLVDFV
Mtb_H37Rv     241  KGLTINLTDERVTQDEVVDEVVSDVAEAPKSASERAAESTAPHKVKSRTFHYPGGLVDFV  300

Mkomo_JERR01  262  KHINRTKTPIQPSVIDFDGKGEGHEVEVAMQWNAGYSESVHTFANTINTHEGGTHEEGFR  321
                   KHINRTK  I  S++DF GKG GHEVE+AMQWNAGYSESVHTFANTINTHEGGTHEEGFR
Mtb_H37Rv     301  KHINRTKNAIHSSIVDFSGKGTGHEVEIAMQWNAGYSESVHTFANTINTHEGGTHEEGFR  360

Mkomo_JERR01  322  AALTSVVNKYAKDKKLLKDKDPNLTGDDIREGLAAVISVKVSQPQFEGQTKTKLGNTEVK  381
                   +ALTSVVNKYAKD+KLLKDKDPNLTGDDIREGLAAVISVKVS+PQFEGQTKTKLGNTEVK
Mtb_H37Rv     361  SALTSVVNKYAKDRKLLKDKDPNLTGDDIREGLAAVISVKVSEPQFEGQTKTKLGNTEVK  420

Mkomo_JERR01  382  SFVQKICNEQLSHWFEANPAEAKTVVNKAVSSAQARMAARKARELVRRKSATDIGGLPGK  441
                   SFVQK+CNEQL+HWFEANP +AK VVNKAVSSAQAR+AARKARELVRRKSATDIGGLPGK
Mtb_H37Rv     421  SFVQKVCNEQLTHWFEANPTDAKVVVNKAVSSAQARIAARKARELVRRKSATDIGGLPGK  480

Mkomo_JERR01  442  LADCRSTDPTKSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE  501
                   LADCRSTDP KSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE
Mtb_H37Rv     481  LADCRSTDPRKSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE  540

Mkomo_JERR01  502  VQAIITALGTGIHDEFDIEKLRYHKIVLMADADVDGQHISTLLLTLLFRFMKPLVENGHI  561
                   VQAIITALGTGIHDEFDI KLRYHKIVLMADADVDGQHISTLLLTLLFRFM+PL+ENGH+
Mtb_H37Rv     541  VQAIITALGTGIHDEFDIGKLRYHKIVLMADADVDGQHISTLLLTLLFRFMRPLIENGHV  600

Mkomo_JERR01  562  FLAQPPLYKLKWQRADPEFAYSDRERDGLLEAGRAAGKRINVDDGIQRYKGLGEMDAKEL  621
                   FLAQPPLYKLKWQR+DPEFAYSDRERDGLLEAG  AGK+IN +DGIQRYKGLGEMDAKEL
Mtb_H37Rv     601  FLAQPPLYKLKWQRSDPEFAYSDRERDGLLEAGLKAGKKINKEDGIQRYKGLGEMDAKEL  660

Mkomo_JERR01  622  WETTMDPSVRVLRQVTLDDAAAADELFSILMGEDVEARRSFITRNAKDVRFLDV  675
                   WETTMDPSVRVLRQVTLDDAAAADELFSILMGEDV+ARRSFITRNAKDVRFLDV
Mtb_H37Rv     661  WETTMDPSVRVLRQVTLDDAAAADELFSILMGEDVDARRSFITRNAKDVRFLDV  714
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 GyrB_1

```
Query: HJANABMJ_00926 DNA gyrase subunit B Query ID: lcl|Query_92815 Length: 640


>LFFBIPCO_00005 DNA gyrase subunit B
Sequence ID: Query_92817 Length: 714
Range 1: 1 to 714


Mkomo_JERR01  1    M-----------------------------------------TSSNSYGAESITILE  16
                   M                                          + + YGA SITILE
Mtb_H37Rv     1    MGKNEARRSALAPDHGTVVCDPLRRLNRMHATPEESIRIVAAQKKKAQDEYGAASITILE  60

Mkomo_JERR0   17   GLDVVRKRPGMYIGSTGERGLHHMIWEVVDNGVDEAMAGHASRVEVTLLADGGVEVVDDG  76
                   GL+ VRKRPGMYIGSTGERGLHH+IWEVVDN VDEAMAG+A+ V V LL DGGVEV DDG
Mtb_H37Rv     61   GLEAVRKRPGMYIGSTGERGLHHLIWEVVDNAVDEAMAGYATTVNVVLLEDGGVEVADDG  120

Mkomo_JERR0   77   RGIPVAQHVTGIPTIDVVMTQLHAGGKFDSDSYSVSGGLHGVGVSVVNALSTRVEVEIRR  136
                   RGIPVA H +GIPT+DVVMTQLHAGGKFDSD+Y++SGGLHGVGVSVVNALSTR+EVEI+R
Mtb_H37Rv     121  RGIPVATHASGIPTVDVVMTQLHAGGKFDSDAYAISGGLHGVGVSVVNALSTRLEVEIKR  180

Mkomo_JERR0   137  DGYVWSQSYDRSLPGVLTRGPRTPDTGTTVRFWADPDIFETVTYSVEKVARRLQEMAFLN  196
                   DGY WSQ Y++S P  L +G  T  TG+TVRFWADP +FET  Y  E VARRLQEMAFLN
Mtb_H37Rv     181  DGYEWSQVYEKSEPLGLKQGAPTKKTGSTVRFWADPAVFETTEYDFETVARRLQEMAFLN  240

Mkomo_JERR0   197  RGLTLTLDDRR------------DADAGP------------------RTFHYADGLVDFV  226
                   +GLT+ L D R            D    P                  RTFHY  GLVDFV
Mtb_H37Rv     241  KGLTINLTDERVTQDEVVDEVVSDVAEAPKSASERAAESTAPHKVKSRTFHYPGGLVDFV  300

Mkomo_JERR0   227  THLNRTREPIQPSIVGFTGAATGIEVEVAMQWNAGYTESVHTFANTINTHEGGTHEEGFR  286
                    H+NRT+  I  SIV F+G  TG EVE+AMQWNAGY+ESVHTFANTINTHEGGTHEEGFR
Mtb_H37Rv     301  KHINRTKNAIHSSIVDFSGKGTGHEVEIAMQWNAGYSESVHTFANTINTHEGGTHEEGFR  360

Mkomo_JERR0   287  AALTSVVNKYARDKKLLKDKDPNLTGDDIREGLAAVIAVKIAEPQFEGQTKTRLGNAEVR  346
                   +ALTSVVNKYA+D+KLLKDKDPNLTGDDIREGLAAVI+VK++EPQFEGQTKT+LGN EV+
Mtb_H37Rv     361  SALTSVVNKYAKDRKLLKDKDPNLTGDDIREGLAAVISVKVSEPQFEGQTKTKLGNTEVK  420

Mkomo_JERR0   347  SCVQKVCHERLTHWFEANPAEAKTVITKVTSSAHARLAARKARDLVRRKSATDIGGLPGK  406
                   S VQKVC+E+LTHWFEANP +AK V+ K  SSA AR+AARKAR+LVRRKSATDIGGLPGK
Mtb_H37Rv     421  SFVQKVCNEQLTHWFEANPTDAKVVVNKAVSSAQARIAARKARELVRRKSATDIGGLPGK  480

Mkomo_JERR0   407  LADCRSTDPTKSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE  466
                   LADCRSTDP KSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE
Mtb_H37Rv     481  LADCRSTDPRKSELYVVEGDSAGGSAKSGRDSMFQAILPLRGKIINVEKARIDRVLKNTE  540

Mkomo_JERR0   467  VQAIITALGTGIHDEFDIEKLRYHKIVLMADADVDGQHISTLLLTLLFRFMKPLVENGHI  526
                   VQAIITALGTGIHDEFDI KLRYHKIVLMADADVDGQHISTLLLTLLFRFM+PL+ENGH+
Mtb_H37Rv     541  VQAIITALGTGIHDEFDIGKLRYHKIVLMADADVDGQHISTLLLTLLFRFMRPLIENGHV  600

Mkomo_JERR0   527  FLAQPPLYKLKWQRSDPEFAYSDRERDGLLEAGRAAGKRINVDDGIQRYKGLGEMDAKEL  586
                   FLAQPPLYKLKWQRSDPEFAYSDRERDGLLEAG  AGK+IN +DGIQRYKGLGEMDAKEL
Mtb_H37Rv     601  FLAQPPLYKLKWQRSDPEFAYSDRERDGLLEAGLKAGKKINKEDGIQRYKGLGEMDAKEL  660

Mkomo_JERR0   587  WETTMDPSVRVLRQVTLDDAAAADELFSILMGDDVDARRSFITRNAKDIRFLDI  640
                   WETTMDPSVRVLRQVTLDDAAAADELFSILMG+DVDARRSFITRNAKD+RFLD+
Mtb_H37Rv     661  WETTMDPSVRVLRQVTLDDAAAADELFSILMGEDVDARRSFITRNAKDVRFLDV  714
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 GyrB_2

```
Query: HJANABMJ_02640 Enoyl-[acyl-carrier-protein] reductase [NADH] Query ID: lcl|Query_4242007
Length: 268


>LFFBIPCO_01570 Enoyl-[acyl-carrier-protein] reductase [NADH]
Sequence ID: Query_4242009 Length: 269
Range 1: 1 to 269

Score:1190, Identities:230/269(86%), Positives:253/269(94%), Gaps:1/269(0%)

Mkomo_JERR01  1    MT-LLQGKRILVTGIITDSSIAFYIAKVAQEAGAEVICTGFNRLRLIERILERLPSKPPL  59
                   MT LL GKRILV+GIITDSSIAF+IA+VAQE GA+++ TGF+RLRLI+RI +RLP+K PL
Mtb_H37Rv     1    MTGLLDGKRILVSGIITDSSIAFHIARVAQEQGAQLVLTGFDRLRLIQRITDRLPAKAPL  60

Mkomo_JERR01  60   LELDVQDDKHLDTLADRVTEVIGEGNKLDGVVHSIGFMPQTGMGINPFFDAPYEDVAKGI  119
                   LELDVQ+++HL +LA RVTE IG GNKLDGVVHSIGFMPQTGMGINPFFDAPY DV+KGI
Mtb_H37Rv     61   LELDVQNEEHLASLAGRVTEAIGAGNKLDGVVHSIGFMPQTGMGINPFFDAPYADVSKGI  120

Mkomo_JERR01  120  HISAYSYASLAKAVLPVMNPGGGIVGMDFDPTRAMPAYNWMTVAKSALESVNRFVAREAG  179
                   HISAYSYAS+AKA+LP+MNPGG IVGMDFDP+RAMPAYNWMTVAKSALESVNRFVAREAG
Mtb_H37Rv     121  HISAYSYASMAKALLPIMNPGGSIVGMDFDPSRAMPAYNWMTVAKSALESVNRFVAREAG  180

Mkomo_JERR01  180  KVGVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQMQLLEEGWDQRAPLGWNMKDPTPVAK  239
                   K GVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQ+QLLEEGWDQRAP+GWNMKD TPVAK
Mtb_H37Rv     181  KYGVRSNLVAAGPIRTLAMSAIVGGALGEEAGAQIQLLEEGWDQRAPIGWNMKDATPVAK  240

Mkomo_JERR01  240  TVCALLSDWLPATTGTVIFADGGASTQLL  268
                   TVCALLSDWLPATTG +I+ADGGA TQLL
Mtb_H37Rv     241  TVCALLSDWLPATTGDIIYADGGAHTQLL  269
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 InhA

```
Query: HJANABMJ_02868 Catalase-peroxidase 2 Query ID: lcl|Query_5940237 Length: 747


>LFFBIPCO_02013 Catalase-peroxidase
Sequence ID: Query_5940239 Length: 740
Range 1: 1 to 740

Score:2484, Identities:473/752(63%), Positives:567/752(75%), Gaps:17/752(2%)

Mkomo_JERR01  1    MTDTSDARPPHSDDKTSSRSESENPVIDSPEPKTHAPLTNKDWWPEQVDVSVLHKQNEKS  60
                   M +     PP ++  T + S    PV+   +        N+DWWP ++++ VLH+    +
Mtb_H37Rv     1    MPEQ---HPPITETTTGAASNG-CPVVGHMKYPVEGG-GNQDWWPNRLNLKVLHQNPAVA  55

Mkomo_JERR01  61   NPLGVDFDYATEFAKLDVEAFKRDVIDLINTSQDWWPADYGSYAGLFIRMSWHAAGTYRI  120
                   +P+G  FDYA E A +DV+A  RD+ +++ TSQ WWPADYG Y  LFIRM+WHAAGTYRI
Mtb_H37Rv     56   DPMGAAFDYAAEVATIDVDALTRDIEEVMTTSQPWWPADYGHYGPLFIRMAWHAAGTYRI  115

Mkomo_JERR01  121  FDGRGGAGQGSQRFAPLNSWPDNANLDKARRLLWPIKQKYGNKISWADLIAYAGNAALES  180
                    DGRGGAG G QRFAPLNSWPDNA+LDKARRLLWP+K+KYG K+SWADLI +AGN ALES
Mtb_H37Rv     116  HDGRGGAGGGMQRFAPLNSWPDNASLDKARRLLWPVKKKYGKKLSWADLIVFAGNCALES  175

Mkomo_JERR01  181  AGFQTFGFAFGREDIWEPEEMLWGQEDTWLGTDKRYGGTNDSDKRELAEPFGATTMGLIY  240
                    GF+TFGF FGR D WEP+E+ WG+E TWLG D+RY G     KR+L  P  A  MGLIY
Mtb_H37Rv     176  MGFKTFGFGFGRVDQWEPDEVYWGKEATWLG-DERYSG-----KRDLENPLAAVQMGLIY  229

Mkomo_JERR01  241  VNPEGPEGKPDPLAAAHDIRETFGRMAMNDEETAALIVGGHTLGKTHGAADVN-VGPEPE  299
                   VNPEGP G PDP+AAA DIRETF RMAMND ETAALIVGGHT GKTHGA   + VGPEPE
Mtb_H37Rv     230  VNPEGPNGNPDPMAAAVDIRETFRRMAMNDVETAALIVGGHTFGKTHGAGPADLVGPEPE  289

Mkomo_JERR01  300  GAPIEQQGLGWKCPFGTGNANDTVTSGLEVVWTTTPTKWSNAYLELLYGYEWELTKSPAG  359
                    AP+EQ GLGWK  +GTG   D +TSG+EVVWT TPTKW N++LE+LYGYEWELTKSPAG
Mtb_H37Rv     290  AAPLEQMGLGWKSSYGTGTGKDAITSGIEVVWTNTPTKWDNSFLEILYGYEWELTKSPAG  349

Mkomo_JERR01  360  AWQFEAKDAEAI--IPDPFGGPPRKPTMLVTDVSMRVDPIYGPITRRWLDHPEEMNEAFA  417
                   AWQ+ AKD      IPDPFGGP R PTML TD+S+RVDPIY  ITRRWL+HPEE+ + FA
Mtb_H37Rv     350  AWQYTAKDGAGAGTIPDPFGGPGRSPTMLATDLSLRVDPIYERITRRWLEHPEELADEFA  409

Mkomo_JERR01  418  KAWYKLMHRDMGPVSRYLGPWVA-EAQLWQDPVPAVDHALIDESDVAALKTAVLQSGLSV  476
                   KAWYKL+HRDMGPV+RYLGP V  +  LWQDPVPAV H L+ E+++A+LK+ +  SGL+V
Mtb_H37Rv     410  KAWYKLIHRDMGPVARYLGPLVPKQTLLWQDPVPAVSHDLVGEAEIASLKSQIRASGLTV  469

Mkomo_JERR01  477  PQLVKTAWASASSFRGTDKRGGANGARLRLEPQRSWEANEPT-ELAKVLPALEKIQQDFN  535
                    QLV TAWA+ASSFRG+DKRGGANG R+RL+PQ  WE N+P  +L KV+  LE+IQ+ FN
Mtb_H37Rv     470  SQLVSTAWAAASSFRGSDKRGGANGGRIRLQPQVGWEVNDPDGDLRKVIRTLEEIQESFN  529

Mkomo_JERR01  536  ASATGGKKVSLADVIVLAGSAAIEKAAKDGGYEISVHFAPGRTDASQEQTDVESFAVLET  595
                   ++A G  KVS AD++VL G AAIEKAAK  G+ I+V F PGRTDASQEQTDVESFAVLE
Mtb_H37Rv     530  SAAPGNIKVSFADLVVLGGCAAIEKAAKAAGHNITVPFTPGRTDASQEQTDVESFAVLEP  589

Mkomo_JERR01  596  RADGFRNYARPGEKTPLEQLLIDKAYFLDLTAPELTALIGGLRTLNANHGGSKHGVFTDQ  655
                   +ADGFRNY   G   P E +L+DKA  L L+APE+T L+GGLR L AN+     GVFT+
Mtb_H37Rv     590  KADGFRNYLGKGNPLPAEYMLLDKANLLTLSAPEMTVLVGGLRVLGANYKRLPLGVFTEA  649

Mkomo_JERR01  656  PGVLSNDFFVNLLDMRTEWKPSELTENVYEGKDRATGAPKWTATAADLVFGSNSVLRAVA  715
                      L+NDFFVNLLDM   W+PS   +  Y+GKD  +G  KWT +  DLVFGSNS LRA+
Mtb_H37Rv     650  SESLTNDFFVNLLDMGITWEPSPADDGTYQGKD-GSGKVKWTGSRVDLVFGSNSELRALV  708

Mkomo_JERR01  716  EVYAQEDNKAKFVEDFVAAWVKVMNNDRFDLK  747
                   EVY  +D + KFV+DFVAAW KVMN DRFD++
Mtb_H37Rv     709  EVYGADDAQPKFVQDFVAAWDKVMNLDRFDVR  740
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 KatG

```
Query: HJANABMJ_00876 Siderophore exporter MmpL5 Query ID: lcl|Query_6489543 Length: 959


>LFFBIPCO_00718 Siderophore exporter MmpL5
Sequence ID: Query_6489545 Length: 964
Range 1: 1 to 964

Score:3204, Identities:606/966(63%), Positives:760/966(78%), Gaps:9/966(0%)

Mkomo_JERR01  1    MTTF-----TDALPPARHAAPARPKLPRFIRTFAVPIVLAWIAIVALLNTVVPQLEEVGK  55
                   M        T ++PP RHAA  RP +PR IRTFAVPI+L W+  +A+LN  VPQLE VG+
Mtb_H37Rv     1    MIVQRTAAPTGSVPPDRHAA--RPFIPRMIRTFAVPIILGWLVTIAVLNVTVPQLETVGQ  58

Mkomo_JERR01  56   LRAVSMSPNDAPALIATKHVGEKFEEYDTSSSVMIVVEGEEPLGPDAHAFYDEVVRQLNA  115
                   ++AVSMSP+ AP++I+ KH+G+ FEE D+ S+ MIV+EG+ PLG  AHAFYD+++ +L A
Mtb_H37Rv     59   IQAVSMSPDAAPSMISMKHIGKVFEEGDSDSAAMIVLEGQRPLGDAAHAFYDQMIGRLQA  118

Mkomo_JERR01  116  DTEHVQHVQDFWGDTLTAAGAQSIDGKAAYVQVYIAGDQGETLANESVHAVRAIVDRNQA  175
                   DT HVQ +QDFWGD LTA GAQS DGKAAYVQV +AG+QGE+LANESV AV+ IV+R
Mtb_H37Rv     119  DTTHVQSLQDFWGDPLTATGAQSSDGKAAYVQVKLAGNQGESLANESVEAVKTIVERLAP  178

Mkomo_JERR01  176  PAGVSAYVTGPAALTTDQNIVGDASMRTIESVTIGIIIVMLLIIYRSVVTTVVTMAMVFV  235
                   P GV  YVTG AAL  DQ   GD S++ IE+VT  +IIVMLL++YRS++T+ + + MV +
Mtb_H37Rv     179  PPGVKVYVTGSAALVADQQQAGDRSLQVIEAVTFTVIIVMLLLVYRSIITSAIMLTMVVL  238

Mkomo_JERR01  236  GLLSARGIVSFLGYYQVFGLTTFATSMVVTLAIAAATDYAIFLIGRYQGARRSGMDRESA  295
                   GLL+ RG V+FLG++++ GL+TFAT+++V LAIAAATDYAIFLIGRYQ AR  G DRESA
Mtb_H37Rv     239  GLLATRGGVAFLGFHRIIGLSTFATNLLVVLAIAAATDYAIFLIGRYQEARGLGQDRESA  298

Mkomo_JERR01  296  YYDMFHGTTHVVLASGLTIAGATACLHFTRLPYFQSMGFPLAIGMIIVVAAALTLGPALI  355
                   YY MF GT HVVL SGLTIAGAT CL FTRLPYFQ++G PLAIGM+IVVAAALTLGPA+I
Mtb_H37Rv     299  YYTMFGGTAHVVLGSGLTIAGATFCLSFTRLPYFQTLGVPLAIGMVIVVAAALTLGPAII  358

Mkomo_JERR01  356  SIVTRFGNVLEPKGNGRARGWRKLGSATVRWPGAVLVMATVLCLVGLLALPGYHTNYNDR  415
                   ++ +RFG +LEPK   R RGWRK+G+A VRWPG +LV A  L LVGLL LPGY TNYNDR
Mtb_H37Rv     359  AVTSRFGKLLEPKRMARVRGWRKVGAAIVRWPGPILVGAVALALVGLLTLPGYRTNYNDR  418

Mkomo_JERR01  416  IYLPDGVPANVGYAAADRHFSDAKMNPDLVMVESDHDMRNPADFLVIEKIAKALTRVHGI  475
                    YLP  +PAN GYAAA+RHFS A+MNP+++MVESDHDMRN ADFLVI KIAKA+  V GI
Mtb_H37Rv     419  NYLPADLPANEGYAAAERHFSQARMNPEVLMVESDHDMRNSADFLVINKIAKAIFAVEGI  478

Mkomo_JERR01  476  ASVTTITRPDGKPIKHASLAYTISQSGNGQIMNNDFQQTVLENTLQQANEMQVTIDSMEE  535
                   + V  ITRPDGKPI+H S+ + IS  G  Q +   + Q +    L+Q N++Q  ID ME
Mtb_H37Rv     479  SRVQAITRPDGKPIEHTSIPFLISMQGTSQKLTEKYNQDLTARMLEQVNDIQSNIDQMER  538

Mkomo_JERR01  536  MQRITLELSEVTREMADKMKDTSANLNEVRDHLADFDDQFRPLRNYFYWEPHCYNIPMCW  595
                   M  +T ++++VT EM  +M     ++ E+R+H+ADFDD FRP+R+YFYWE HCY+IP+CW
Mtb_H37Rv     539  MHSLTQQMADVTHEMVIQMTGMVVDVEELRNHIADFDDFFRPIRSYFYWEKHCYDIPVCW  598

Mkomo_JERR01  596  ALRSVFDSLDGISTMSDDFAELVPSIERMAQLTPQMAALMPAMIQTMKNQKQIMLNQYQA  655
                   +LRSVFD+LDGI  M++D   L+P ++R+  L PQ+ A+MP MIQTMK+ K  ML+ +
Mtb_H37Rv     599  SLRSVFDTLDGIDVMTEDINNLLPLMQRLDTLMPQLTAMMPEMIQTMKSMKAQMLSMHST  658

Mkomo_JERR01  656  QKMQQDQNIAMQEDTTAMGEAFDTARNDDTFYLPPEAFQTADFQRGIKLFMSPDGKAVRF  715
                   Q+  QDQ  AMQED+ AMGEAFD +RNDD+FYLPPE F   DFQRG++ F+SPDG AVRF
Mtb_H37Rv     659  QEGLQDQMAAMQEDSAAMGEAFDASRNDDSFYLPPEVFDNPDFQRGLEQFLSPDGHAVRF  718

Mkomo_JERR01  716  TVFHQGDPLTEAGTARIDPLRVAAADAIKGTPLEGSTVYVGGSAAMYKDMQQGADYDLLI  775
                    + H+GDP+++AG ARI  ++ AA +AIKGTPLEGS +Y+GG+AAM+KD+  G  YDL+I
Mtb_H37Rv     719  IISHEGDPMSQAGIARIAKIKTAAKEAIKGTPLEGSAIYLGGTAAMFKDLSDGNTYDLMI  778

Mkomo_JERR01  776  AAVASLILIFLIMVILTRAVAAAAVIVGTVVLSLGTSFGLSVLLWQHIIGIPLGWMVLPM  835
                   A +++L LIF+IM+I TR+V AAAVIVGTVVLSLG SFGLSVL+WQHI+GI L W+VL M
Mtb_H37Rv     779  AGISALCLIFIIMLITTRSVVAAAVIVGTVVLSLGASFGLSVLIWQHILGIELHWLVLAM  838

Mkomo_JERR01  836  SVIVLLAVGADYNLLLVSRMKEELHAGVNTGIIRSMAGTGSVVTAAGFVFAFTMIGMIVS  895
                   +VI+LLAVGADYNLLLV+R+KEE+HAG+NTGIIR+M G+GSVVTAAG VFAFTM+   VS
Mtb_H37Rv     839  AVIILLAVGADYNLLLVARLKEEIHAGINTGIIRAMGGSGSVVTAAGLVFAFTMMSFAVS  898
```

```
Mkomo_JERR01  896  DMIVIGQVGTTIGLGLLFDTLVVRSLMTPSIAALMGKWFWWPMHVRQRPKPQPWPRGVAA  955
                   ++ V+ QVGTTIG+GLLFDTL+VRS MTPSIAAL+GKWFWWP  VRQRP PQPWP   +A
Mtb_H37Rv     899  ELTVMAQVGTTIGMGLLFDTLIVRSFMTPSIAALLGKWFWWPQVVRQRPIPQPWPSPASA  958

Mkomo_JERR01  956  AM--RT  959

Mtb_H37Rv     959  RTFALV  964
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 MmpL5_1

```
Query: HJANABMJ_01446 Siderophore exporter MmpL5 Query ID: lcl|Query_7668639 Length: 954


>LFFBIPCO_00718 Siderophore exporter MmpL5
Sequence ID: Query_7668641 Length: 964
Range 1: 1 to 964

Score:3365, Identities:641/964(66%), Positives:794/964(82%), Gaps:10/964(1%)

Mkomo_JERR01  1    M--SRSGEDTQSKP---HFFG--IARFIRRASIPIILIWIAIAAFLNISVPQLETVGKLR  53
                   M   R+   T S P   H     I R IR  ++PIIL W+   A LN++VPQLETVG+++
Mtb_H37Rv     1    MIVQRTAAPTGSVPPDRHAARPFIPRMIRTFAVPIILGWLVTIAVLNVTVPQLETVGQIQ  60

Mkomo_JERR01  54   SVSMSPDEAPAVISMQRIGEVFEEYESNSSVMIVLEGQEPLGDDTRSYYADLISRLDADT  113
                   +VSMSPD AP++ISM+ IG+VFEE +S+S+ MIVLEGQ PLGD   ++Y  +I RL ADT
Mtb_H37Rv     61   AVSMSPDAAPSMISMKHIGKVFEEGDSDSAAMIVLEGQRPLGDAAHAFYDQMIGRLQADT  120

Mkomo_JERR01  114  THVEHIQDLWSDPLTASGAQSADGKATYFQVYLAGNQGEALANESVKAVQDVVAESQPPD  173
                   THV+ +QD W DPLTA+GAQS+DGKA Y QV LAGNQGE+LANESV+AV+ +V    PP
Mtb_H37Rv     121  THVQSLQDFWGDPLTATGAQSSDGKAAYVQVKLAGNQGESLANESVEAVKTIVERLAPPP  180

Mkomo_JERR01  174  GVRAYVTGASALAAEQENAGHESMRVVEALTFLVITVMLVLFFRSIVTTLLILVMVGLSL  233
                   GV+ YVTG++AL A+Q+ AG  S++V+EA+TF VI VML+L +RSI+T+ ++L MV L L
Mtb_H37Rv     181  GVKVYVTGSAALVADQQQAGDRSLQVIEAVTFTVIIVMLLLVYRSIITSAIMLTMVVLGL  240

Mkomo_JERR01  234  MTVRGAVAFLGFHEIIGLSTFATSLLVTLAIAIAVDYAIFLIGRYQEARGTGATAESAYY  293
                   +  RG VAFLGFH IIGLSTFAT+LLV LAIA A DYAIFLIGRYQEARG G   ESAYY
Mtb_H37Rv     241  LATRGGVAFLGFHRIIGLSTFATNLLVVLAIAAATDYAIFLIGRYQEARGLGQDRESAYY  300

Mkomo_JERR01  294  TMFGGTAHVIVGSGLTIAGATFCLSFTRLPYFQTLGVPLALGMIVLIMVAMTFGPAVVTV  353
                   TMFGGTAHV++GSGLTIAGATFCLSFTRLPYFQTLGVPLA+GM++++  A+T GPA++ V
Mtb_H37Rv     301  TMFGGTAHVVLGSGLTIAGATFCLSFTRLPYFQTLGVPLAIGMVIVVAAALTLGPAIIAV  360

Mkomo_JERR01  354  ASRFG-LLDPKRAMRVRGWRKIGAATVRWPAAILVASVAVSLIGLLALPGYQTSYNDRNY  412
                    SRFG LL+PKR  RVRGWRK+GAA VRWP  ILV +VA++L+GLL LPGY+T+YNDRNY
Mtb_H37Rv     361  TSRFGKLLEPKRMARVRGWRKVGAAIVRWPGPILVGAVALALVGLLTLPGYRTNYNDRNY  420

Mkomo_JERR01  413  LPADIQSNVGYAAAERHFSPARLNPEVLMVETDHDLRNSTDFIVIDKIAKALFRVEGIGR  472
                   LPAD+ +N GYAAAERHFS AR+NPEVLMVE+DHD+RNS DF+VI+KIAKA+F VEGI R
Mtb_H37Rv     421  LPADLPANEGYAAAERHFSQARMNPEVLMVESDHDMRNSADFLVINKIAKAIFAVEGISR  480

Mkomo_JERR01  473  VQTITRPDGTPISSTTIPYIMSRQGTTQQLNEKYMLDRMDDMLVQADALQTTINTMEKMQ  532
                   VQ ITRPDG PI  T+IP+++S QGT+Q+L EKY  D    ML Q + +Q+ I+ ME+M
Mtb_H37Rv     481  VQAITRPDGKPIEHTSIPFLISMQGTSQKLTEKYNQDLTARMLEQVNDIQSNIDQMERMH  540

Mkomo_JERR01  533  ALMGQMAEITQSMVTKTKTMAADITELRDHIANLDDFFRPVRNYFYWEPHCYNIPVCWSM  592
                   +L  QMA++T  MV +   M  D+ ELR+HIA+ DDFFRP+R+YFYWE HCY+IPVCWS+
Mtb_H37Rv     541  SLTQQMADVTHEMVIQMTGMVVDVEELRNHIADFDDFFRPIRSYFYWEKHCYDIPVCWSL  600

Mkomo_JERR01  593  RSLFDLIDGTNLLTDNIQDLVPELERLAALMPQLVALMPSQIESMKSQREMMLTMYQTQN  652
                   RS+FD +DG +++T++I +L+P ++RL  LMPQL A+MP  I++MKS +  ML+M+ TQ
Mtb_H37Rv     601  RSVFDTLDGIDVMTEDINNLLPLMQRLDTLMPQLTAMMPEMIQTMKSMKAQMLSMHSTQE  660

Mkomo_JERR01  653  GQLEQAAAMGVDATAMGDAFNDAMNDDTFYLPPEAFDNADFQRGIEQFISPNGHAVRFII  712
                   G  +Q AAM  D+ AMG+AF+ + NDD+FYLPPE FDN DFQRG+EQF+SP+GHAVRFII
Mtb_H37Rv     661  GLQDQMAAMQEDSAAMGEAFDASRNDDSFYLPPEVFDNPDFQRGLEQFLSPDGHAVRFII  720

Mkomo_JERR01  713  AHEGDPLSPAGIEKIDAIKTAAKEAIKGTPLEGSTIYLGGTAATFKDMADGTRYDLLIAG  772
                   +HEGDP+S AGI +I  IKTAAKEAIKGTPLEGS IYLGGTAA FKD++DG  YDL+IAG
Mtb_H37Rv     721  SHEGDPMSQAGIARIAKIKTAAKEAIKGTPLEGSAIYLGGTAAMFKDLSDGNTYDLMIAG  780

Mkomo_JERR01  773  IAAIGLIFIIMLILTRAIVAAAVIVGTVVLSLGASFGLSVLVWQHLVGIELHWMVLPMAV  832
                   I+A+ LIFIIMLI TR++VAAAVIVGTVVLSLGASFGLSVL+WQH++GIELHW+VL MAV
Mtb_H37Rv     781  ISALCLIFIIMLITTRSVVAAAVIVGTVVLSLGASFGLSVLIWQHILGIELHWLVLAMAV  840

Mkomo_JERR01  833  IILLAVGADYNLLVVSRLKEEVHAGLHTGLIRTMGGSGSVVTAAGMVFALTMMTMAVSDL  892
                   IILLAVGADYNLL+V+RLKEE+HAG++TG+IR MGGSGSVVTAAG+VFA TMM+ AVS+L
Mtb_H37Rv     841  IILLAVGADYNLLLVARLKEEIHAGINTGIIRAMGGSGSVVTAAGLVFAFTMMSFAVSEL  900
```

```
Mkomo_JERR01  893  TIIGQVGTTIGMGLIFDTLVIRSFMTPSIAALMGRWFWWPQRIRTRPRPSPWPSPATPVS  952
                   T++ QVGTTIGMGL+FDTL++RSFMTPSIAAL+G+WFWWPQ +R RP P PWPSPA+  +
Mtb_H37Rv     901  TVMAQVGTTIGMGLLFDTLIVRSFMTPSIAALLGKWFWWPQVVRQRPIPQPWPSPASART  960

Mkomo_JERR01  953  PS--  954
                   +
Mtb_H37Rv     961  FALV  964
```

Global alignment between *Mtb* H37Rv and *M*. JERR01 MmpL5_2

Query: HJANABMJ_02887 Siderophore exporter MmpL5 Query ID: lcl|Query_7131697 Length: 968

>LFFBIPCO_00718 Siderophore exporter MmpL5
Sequence ID: Query_7131699 Length: 964
Range 1: 1 to 964

Score:3745, Identities:721/968(74%), Positives:827/968(85%), Gaps:4/968(0%)

```
Mkomo_JERR01  1    MSAPAPDARTDEFPAARPPHRPFIPRVIRLFALPIILGWVALIVILNLTVPQLEAVGEMR  60
                   M      A T   P  R   RPFIPR+IR FA+PIILGW+  I +LN+TVPQLE VG+++
Mtb_H37Rv     1    MIVQRTAAPTGSVPPDRHAARPFIPRMIRTFAVPIILGWLVTIAVLNVTVPQLETVGQIQ  60

Mkomo_JERR01  61   AVSMSPNDAPSMISMKKVGELFEEGDSDSSVMVVFEGDAPLGDEAHAWYDELVERLRADT  120
                   AVSMSP+ APSMISMK +G++FEEGDSDS+ M+V EG  PLGD AHA+YD+++ RL+ADT
Mtb_H37Rv     61   AVSMSPDAAPSMISMKHIGKVFEEGDSDSAAMIVLEGQRPLGDAAHAFYDQMIGRLQADT  120

Mkomo_JERR01  121  VHVQSVQDFWSDPLTASGSQSNDGLAAYVQVKLAGNQGEALANESVKAVQGIVSSLEPPP  180
                    HVQS+QDFW DPLTA+G+QS+DG AAYVQVKLAGNQGE+LANESV+AV+ IV  L PPP
Mtb_H37Rv     121  THVQSLQDFWGDPLTATGAQSSDGKAAYVQVKLAGNQGESLANESVEAVKTIVERLAPPP  180

Mkomo_JERR01  181  GVKAYVTGPAALAADQHIAGDRSVRTIETVTFAVIIVMLLLVYRSIVTVLITLVMVVLSL  240
                   GVK YVTG AAL ADQ  AGDRS++ IE VTF VIIVMLLLVYRSI+T  I L MVVL L
Mtb_H37Rv     181  GVKVYVTGSAALVADQQQAGDRSLQVIEAVTFTVIIVMLLLVYRSIITSAIMLTMVVLGL  240

Mkomo_JERR01  241  TTARGVVAFLGYHEIIGLSTFATNLLVTLAIAAATDYAIFLIGRYQEARTMGEDKESAFY  300
                      RG VAFLG+H IIGLSTFATNLLV LAIAAATDYAIFLIGRYQEAR +G+D+ESA+Y
Mtb_H37Rv     241  LATRGGVAFLGFHRIIGLSTFATNLLVVLAIAAATDYAIFLIGRYQEARGLGQDRESAYY  300

Mkomo_JERR01  301  TMFHGTAHVVLGSGLTIAGATFCLSFTRLPYFQTLGVPLAIGMFVVMMAGVVGMVAVISV  360
                   TMF GTAHVVLGSGLTIAGATFCLSFTRLPYFQTLGVPLAIGM +V+ A +    A+I+V
Mtb_H37Rv     301  TMFGGTAHVVLGSGLTIAGATFCLSFTRLPYFQTLGVPLAIGMVIVVAAALTLGPAIIAV  360

Mkomo_JERR01  361  VTRFGKLLEPKRAMRIRGWRKVGAAVVRWPGPILVATVALALVGLLALPGYRTNYNDRAY  420
                    +RFGKLLEPKR  R+RGWRKVGAA+VRWPGPILV  VALALVGLL LPGYRTNYNDR Y
Mtb_H37Rv     361  TSRFGKLLEPKRMARVRGWRKVGAAIVRWPGPILVGAVALALVGLLTLPGYRTNYNDRNY  420

Mkomo_JERR01  421  LPADLPANEGYAVADEHFDPARMNPELLMVESDHDLRNSADFLVIDKIAKALFRVEGIAR  480
                   LPADLPANEGYA A+ HF  ARMNPE+LMVESDHD+RNSADFLVI+KIAKA+F VEGI+R
Mtb_H37Rv     421  LPADLPANEGYAAAERHFSQARMNPEVLMVESDHDMRNSADFLVINKIAKAIFAVEGISR  480

Mkomo_JERR01  481  VQAITRPDGKPIKHTSIPFQMSMQGTTQRLNEKYMQDRMADMLVQADEMADTIATMEKMS  540
                   VQAITRPDGKPI+HTSIPF +SMQGT+Q+L EKY QD  A ML Q +++   I  ME+M
Mtb_H37Rv     481  VQAITRPDGKPIEHTSIPFLISMQGTSQKLTEKYNQDLTARMLEQVNDIQSNIDQMERMH  540

Mkomo_JERR01  541  NLTAQMASITNDMVSKMENMLTDIEDLRDSIANFDDFFRPIRNYFYWEPHCYNIPVCWAL  600
                   +LT QMA +T++MV +M  M+ D+E+LR+ IA+FDDFFRPIR+YFYWE HCY+IPVCW+L
Mtb_H37Rv     541  SLTQQMADVTHEMVIQMTGMVVDVEELRNHIADFDDFFRPIRSYFYWEKHCYDIPVCWSL  600

Mkomo_JERR01  601  RSVFDTLDGINVMTDDFREILPDMRRLNTLMPQMVALMPEMIETMKTMRTMMLTMYQSQK  660
                   RSVFDTLDGI+VMT+D   +LP M+RL+TLMPQ+ A+MPEMI+TMK+M+  ML+M+ +Q+
Mtb_H37Rv     601  RSVFDTLDGIDVMTEDINNLLPLMQRLDTLMPQLTAMMPEMIQTMKSMKAQMLSMHSTQE  660

Mkomo_JERR01  661  GQQDQMAAMSEDADAMGEAFDNSMNDDSFYLPPEIFENKDFQRGLEQFLSPDGHAVRFII  720
                   G QDQMAAM ED+ AMGEAFD S NDDSFYLPPE+F+N DFQRGLEQFLSPDGHAVRFII
Mtb_H37Rv     661  GLQDQMAAMQEDSAAMGEAFDASRNDDSFYLPPEVFDNPDFQRGLEQFLSPDGHAVRFII  720

Mkomo_JERR01  721  SHEGDPLSPEGVAKIEKIKTAAKEAIKGTPLEGSKIYLGGTAATFKDMRDGNNYDLLIAG  780
                   SHEGDP+S  G+A+I KIKTAAKEAIKGTPLEGS IYLGGTAA FKD+ DGN YDL+IAG
Mtb_H37Rv     721  SHEGDPMSQAGIARIAKIKTAAKEAIKGTPLEGSAIYLGGTAAMFKDLSDGNTYDLMIAG  780

Mkomo_JERR01  781  ISALCLIFIIMLILTRAVVAAAVIVGTVAMSLGASFGLSILLWQHILGIELHWMVFAMAV  840
                   ISALCLIFIIMLI TR+VVAAAVIVGTV +SLGASFGLS+L+WQHILGIELHW+V AMAV
Mtb_H37Rv     781  ISALCLIFIIMLITTRSVVAAAVIVGTVVLSLGASFGLSVLIWQHILGIELHWLVLAMAV  840

Mkomo_JERR01  841  IVLLGVGADYNLLLVSRLKEEIHAGVGTGIIRAMGGSGSVVTAAGLVFALTMMSMAISEL  900
                   I+LL VGADYNLLLV+RLKEEIHAG+ TGIIRAMGGSGSVVTAAGLVFA TMMS A+SEL
Mtb_H37Rv     841  IILLAVGADYNLLLVARLKEEIHAGINTGIIRAMGGSGSVVTAAGLVFAFTMMSFAVSEL  900
```

```
Mkomo_JERR01  901  TVIAQVGTTIGLGLLFDTLVVRSFMTPSIAALLGPWFWWPQRVRTRPVPSPWPRPGGLQS  960
                   TV+AQVGTTIG+GLLFDTL+VRSFMTPSIAALLG WFWWPQ VR RP+P PWP P   ++
Mtb_H37Rv     901  TVMAQVGTTIGMGLLFDTLIVRSFMTPSIAALLGKWFWWPQVVRQRPIPQPWPSPASART  960

Mkomo_JERR01  961  DPSEGVKA  968

Mtb_H37Rv     961  F----ALV  964
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 MmpL5_3

```
Query: HJANABMJ_00642 HTH-type transcriptional regulator MmpR5 Query ID: lcl|Query_7595579 Length:
156


>LFFBIPCO_00720 HTH-type transcriptional regulator MmpR5
Sequence ID: Query_7595581 Length: 165
Range 1: 1 to 165

Score:160, Identities:52/166(31%), Positives:80/166(48%), Gaps:11/166(6%)

Mkomo_JERR01  1    MTPS---------PKLHEAAERLAIVLAGHGLQRMTARVLAALLFSEQPSVTMAELADDL  51
                   M+ +         P + E   E++        L R+  R+L  LL  +    +  ELA  L
Mtb_H37Rv     1    MSVNDGVDQMGAEPDIMEFVEQMGGYFESRSLTRLAGRLLGWLLVCDPERQSSEELATAL  60

Mkomo_JERR01  52   HASAGSISGALKTLTAVGLAERVPARSSRRDHFRLRPDAWAMLYTGQNETIAALLDAADA  111
                    AS+G IS   + L   G  ER+     RR +FRLRP+A+A     +   +A L D AD
Mtb_H37Rv     61   AASSGGISTNARMLIQFGFIERLAVAGDRRTYFRLRPNAFAAGERERIRAMAELQDLADV  120

Mkomo_JERR01  112  GIVATNVSSPARAR-LTEMRDFYAFLMAEIPAVLERWHARRSLGND  156
                   G+ A   + P R+R L EMRD  A++   +   L R+ ++R+  +D
Mtb_H37Rv     121  GLRALGDAPPQRSRRLREMRDLLAYMENVVSDALGRY-SQRTGEDD  165
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 MmpR5

```
Query: HJANABMJ_04935 Aspartate 1-decarboxylase Query ID: lcl|Query_6983069 Length: 135

>LFFBIPCO_03788 Aspartate 1-decarboxylase
Sequence ID: Query_6983071 Length: 139
Range 1: 1 to 139

Score:583, Identities:117/139(84%), Positives:124/139(89%), Gaps:4/139(2%)

Mkomo_JERR01  1    MLRTMLKSKIHRATVTQADLHYVGSVTVDADLMDAADLLEGEQVTIVDVDNGARLVTYVI  60
                   MLRTMLKSKIHRATVT ADLHYVGSVT+DADLMDAADLLEGEQVTIVD+DNGARLVTY I
Mtb_H37Rv     1    MLRTMLKSKIHRATVTCADLHYVGSVTIDADLMDAADLLEGEQVTIVDIDNGARLVTYAI  60

Mkomo_JERR01  61   TGERGSGVIGINGAAAHLVHPGDLVILIAYATVDDAEARTYRPRVVFVDAENRVIDLGAD  120
                   TGERGSGVIGINGAAAHLVHPGDLVILIAYAT+DDA ARTY+PR+VFVDA N+ ID+G D
Mtb_H37Rv     61   TGERGSGVIGINGAAAHLVHPGDLVILIAYATMDDARARTYQPRIVFVDAYNKPIDMGHD  120

Mkomo_JERR01  121  PAFAPAGAAELLSPR----  135
                   PAF P   A ELL PR
Mtb_H37Rv     121  PAFVPENAGELLDPRLGVG  139
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 PanD

```
Query: HJANABMJ_05375 Nicotinamidase/pyrazinamidase Query ID: lcl|Query_6851757 Length: 186


>LFFBIPCO_02160 Nicotinamidase/pyrazinamidase
Sequence ID: Query_6851759 Length: 186
Range 1: 1 to 186

Score:616, Identities:127/189(67%), Positives:141/189(74%), Gaps:6/189(3%)

Mkomo_JERR01  1    MTSNRALIIVDVQNDFCEGGSLAVAGGAAVARAINDLLATAPGYDHVVATKDFHIDPGPH  60
                   M   RALIIVDVQNDFCEGGSLAV GGAA+ARAI+D LA A  Y HVVATKDFHIDPG H
Mtb_H37Rv     1    M---RALIIVDVQNDFCEGGSLAVTGGAALARAISDYLAEAADYHHVVATKDFHIDPGDH  57

Mkomo_JERR01  61   FSDHPDFVDSWPVHCVAQTAGAEFHPDLDTTPIEAVFKKGHYSAAYSGFEGTLDAGGSLA  120
                   FS  PD+  SWP HCV+ T GA+FHP LDT+ IEAVF KG Y+ AYSGFEG  + G  L
Mtb_H37Rv     58   FSGTPDYSSSWPPHCVSGTPGADFHPSLDTSAIEAVFYKGAYTGAYSGFEGVDENGTPLL  117

Mkomo_JERR01  121  EWLRDRDVTAVDVVGIATDHCVKATAADAVAAGFTTRVLLDLTAGVSLATTTAAVGELRD  180
                    WLR R V  VDVVGIATDHCV+ TA DAV  G  TRVL+DLTAGVS  TT AA+ E+R
Mtb_H37Rv     118  NWLRQRGVDEVDVVGIATDHCVRQTAEDAVRNGLATRVLVDLTAGVSADTTVAALEEMRT  177

Mkomo_JERR01  181  AGVEIL---  186
                   A VE++
Mtb_H37Rv     178  ASVELVCSS  186
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 PncA

```
Query: HJANABMJ_01107 50S ribosomal protein L3 Query ID: lcl|Query_2073091 Length: 218


>LFFBIPCO_00743 50S ribosomal protein L3
Sequence ID: Query_2073093 Length: 217
Range 1: 1 to 217

Score:1011, Identities:197/218(90%), Positives:208/218(95%), Gaps:1/218(0%)

Mkomo_JERR01  1    MARKGILGTKLGMTQVFDENNKVVPVTVVKAGPNVVTRIRTPERDGYSAVQLAYGEISPR  60
                   MARKGILGTKLGMTQVFDE+N+VVPVTVVKAGPNVVTRIRTPERDGYSAVQLAYGEISPR
Mtb_H37Rv     1    MARKGILGTKLGMTQVFDESNRVVPVTVVKAGPNVVTRIRTPERDGYSAVQLAYGEISPR  60

Mkomo_JERR01  61   KVNKPVTGQFAAAGVNPRRHLAELRLDNEAAAAEYEVGQELTAEIFADGAYVDVTGTSKG  120
                   KVNKP+TGQ+ AAGVNPRR+LAELRLD+  AA EY+VGQELTAEIFADG+YVDVTGTSKG
Mtb_H37Rv     61   KVNKPLTGQYTAAGVNPRRYLAELRLDDSDAATEYQVGQELTAEIFADGSYVDVTGTSKG  120

Mkomo_JERR01  121  KGFAGTMKRHGFRGQGASHGAQAVHRRPGSIGGCATPGRVFKGTRMSGRMGNDRVTTQNL  180
                   KGFAGTMKRHGFRGQGASHGAQAVHRRPGSIGGCATP RVFKGTRM+GRMGNDRVT  NL
Mtb_H37Rv     121  KGFAGTMKRHGFRGQGASHGAQAVHRRPGSIGGCATPARVFKGTRMAGRMGNDRVTVLNL  180

Mkomo_JERR01  181  KVHKVDAENGVLLIKGAIPGRNGGLVVVRSAIKRGEKA  218
                    VHKVDAENGVLLIKGA+PGR GGLV+VRSAIKRGEK
Mtb_H37Rv     181  LVHKVDAENGVLLIKGAVPGRTGGLVMVRSAIKRGEK-  217
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 RplC

```
Query: HJANABMJ_01239 DNA-directed RNA polymerase subunit alpha Query ID: lcl|Query_262873
Length: 354


>LFFBIPCO_03645 DNA-directed RNA polymerase subunit alpha
Sequence ID: Query_262875 Length: 347
Range 1: 1 to 347

Score:1648, Identities:328/354(93%), Positives:338/354(95%), Gaps:7/354(1%)

Mkomo_JERR01  1    MLISQRPTLSEEVIAENRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVL  60
                   MLISQRPTLSE+V+ +NRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVL
Mtb_H37Rv     1    MLISQRPTLSEDVLTDNRSQFVIEPLEPGFGYTLGNSLRRTLLSSIPGAAVTSIRIDGVL  60

Mkomo_JERR01  61   HEFTTVPGVKEDVTDIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHN  120
                   HEFTTVPGVKEDVT+IILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHN
Mtb_H37Rv     61   HEFTTVPGVKEDVTEIILNLKSLVVSSEEDEPVTMYLRKQGPGEVTAGDIVPPAGVTVHN  120

Mkomo_JERR01  121  PEMHIATLNDKGKLEVELVVERGRGYVPAVQNKASGAEIGRIPVDSIYSPVLKVTYKVEA  180
                   P MHIATLNDKGKLEVELVVERGRGYVPAVQN+ASGAEIGRIPVDSIYSPVLKVTYKV+A
Mtb_H37Rv     121  PGMHIATLNDKGKLEVELVVERGRGYVPAVQNRASGAEIGRIPVDSIYSPVLKVTYKVDA  180

Mkomo_JERR01  181  TRVEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNIEAEGIEIGPSPAE  240
                   TRVEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELN+EAEGIEIGPSPAE
Mtb_H37Rv     181  TRVEQRTDFDKLILDVETKNSISPRDALASAGKTLVELFGLARELNVEAEGIEIGPSPAE  240

Mkomo_JERR01  241  ADQAAHFALPIDDLDLTVRSYNCLKREGVHTVGELVARTESDLLDIRNFGQKSIDEVKIK  300
                   AD  A FALPIDDLDLTVRSYNCLKREGVHTVGELVARTESDLLDIRNFGQKSIDEVKIK
Mtb_H37Rv     241  ADHIASFALPIDDLDLTVRSYNCLKREGVHTVGELVARTESDLLDIRNFGQKSIDEVKIK  300

Mkomo_JERR01  301  LHQLGLSLKDSPASFDPSEVAGYDVATGTWNSDAGYDSGYEADDNQDYAETEQL  354
                   LHQLGLSLKDSP SFDPSEVAGYDVATGTW+++  YD        QDYAETEQL
Mtb_H37Rv     301  LHQLGLSLKDSPPSFDPSEVAGYDVATGTWSTEGAYDE-------QDYAETEQL  347
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 RpoA

```
Query: HJANABMJ_01059 DNA-directed RNA polymerase subunit beta Query ID: lcl|Query_7189205
Length: 1170


>LFFBIPCO_00708 DNA-directed RNA polymerase subunit beta
Sequence ID: Query_7189207 Length: 1172
Range 1: 1 to 1172

Score:5447, Identities:1060/1181(90%), Positives:1117/1181(94%), Gaps:20/1181(1%)

Mkomo_JERR01  1     MLEGCILAGSRQIESV----------TTNNSVPGAPNRISFAKLREPLEVPGLLDVQTES  50
                    M      A SRQ ++           ++NNSVPGAPNR+SFAKLREPLEVPGLLDVQT+S
Mtb_H37Rv     1     M------ADSRQSKTAASPSPSRPQSSSNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDS  54

Mkomo_JERR01  51    FEWLVGAEDWFRRAADRGDVDPKGGLQEVLEELSPIEDFSGSMSLSFSDPRFDEVKAPVD  110
                    FEWL+G+  W   AA+RGDV+P GGL+EVL ELSPIEDFSGSMSLSFSDPRFD+VKAPVD
Mtb_H37Rv     55    FEWLIGSPRWRESAAERGDVNPVGGLEEVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVD  114

Mkomo_JERR01  111   ECKDKDMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLV  170
                    ECKDKDMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLV
Mtb_H37Rv     115   ECKDKDMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLV  174

Mkomo_JERR01  171   RSPGVYFDESIDKSTEKTLHSVKVIPGRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKA  230
                    RSPGVYFDE+IDKST+KTLHSVKVIP RGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKA
Mtb_H37Rv     175   RSPGVYFDETIDKSTDKTLHSVKVIPSRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKA  234

Mkomo_JERR01  231   LGWTSEQITERFGFSEIMMSTLEKDNTAGTDEALLDIYRKLRPGEPPTKESAQTLLENLF  290
                    LGWTSEQI ERFGFSEIM STLEKDNT GTDEALLDIYRKLRPGEPPTKESAQTLLENLF
Mtb_H37Rv     235   LGWTSEQIVERFGFSEIMRSTLEKDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLF  294

Mkomo_JERR01  291   FKEKRYDLARVGRYKVNKKLGLNVGQPITSSTLTEEDVVATIEYLVRLHQGDQTMTAPGG  350
                    FKEKRYDLARVGRYKVNKKLGL+VG+PITSSTLTEEDVVATIEYLVRLH+G  TMT PGG
Mtb_H37Rv     295   FKEKRYDLARVGRYKVNKKLGLHVGEPITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGG  354

Mkomo_JERR01  351   TEVPVEVDDIDHFGNRRLRTVGELIQNQIRVGLSRMERVVRERMTTQDVEAITPQTLINI  410
                     EVPVE DDIDHFGNRRLRTVGELIQNQIRVG+SRMERVVRERMTTQDVEAITPQTLINI
Mtb_H37Rv     355   VEVPVETDDIDHFGNRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINI  414

Mkomo_JERR01  411   RPVVAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHY  470
                    RPVVAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHY
Mtb_H37Rv     415   RPVVAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHY  474

Mkomo_JERR01  471   GRMCPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVVTDQIDYLTADEEDRHV  530
                    GRMCPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVV+D+I YLTADEEDRHV
Mtb_H37Rv     475   GRMCPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHV  534

Mkomo_JERR01  531   VAQANSETEDTPAGQKFVEDRVLVRRKGGEVEFVSAAEVDYMDVSPRQMVSVATAMIPFL  590
                    VAQANS  +   A  +FVE RVLVRRK GEVE+V ++EVDYMDVSPRQMVSVATAMIPFL
Mtb_H37Rv     535   VAQANSPID---ADGRFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFL  591

Mkomo_JERR01  591   EHDDANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVTDKAGVVEEVSADY  650
                    EHDDANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVV +++GV+EEVSADY
Mtb_H37Rv     592   EHDDANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADY  651

Mkomo_JERR01  651   ITVMADDGTRHTYRMRKFARSNHGTCANQRPIVDAGQRVESGQVLADGPCTENGEMALGK  710
                    ITVM D+GTR TYRMRKFARSNHGTCANQ PIVDAG RVE+GQV+ADGPCT++GEMALGK
Mtb_H37Rv     652   ITVMHDNGTRRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGK  711

Mkomo_JERR01  711   NLLVAIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNVS  770
                    NLLVAIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPN+S
Mtb_H37Rv     712   NLLVAIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNIS  771

Mkomo_JERR01  771   DEVLADLDERGIIRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSL  830
                    DEVLADLDERGI+RIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSL
Mtb_H37Rv     772   DEVLADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSL  831

Mkomo_JERR01  831   KVPHGESGKVIGIRVFSREDDDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGK  890
                    KVPHGESGKVIGIRVFSRED+DELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGK
Mtb_H37Rv     832   KVPHGESGKVIGIRVFSREDEDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGK  891
```

```
Mkomo_JERR01  891   ILPVEDMPFLPDGTPVDIILNTHGVPRRMNIGQILETHLGWVAKAGWNIDVAS-TPEWAA  949
                    ILPVEDMPFL DGTPVDIILNTHGVPRRMNIGQILETHLGW A +GW +D A   P+WAA
Mtb_H37Rv     892   ILPVEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKGVPDWAA  951

Mkomo_JERR01  950   KLPEQMLSAPPDSIVATPVFDGAKESELQGLLGATLPNRDGETMVNADGKAVLFDGRSGE  1009
                    +LP+++L A P++IV+TPVFDGA+E+ELQGLL  TLPNRDG+ +V+ADGKA+LFDGRSGE
Mtb_H37Rv     952   RLPDELLEAQPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLFDGRSGE  1011

Mkomo_JERR01  1010  PFPYPVTVGYMYILKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQ  1069
                    PFPYPVTVGYMYI+KLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQ
Mtb_H37Rv     1012  PFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQ  1071

Mkomo_JERR01  1070  AYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEV  1129
                    AYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEV
Mtb_H37Rv     1072  AYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEV  1131

Mkomo_JERR01  1130  LSSDGAAIEMRDGDDEDLERAAANLGINLSRNESASVEDLA  1170
                    LSSDGAAIE+R+G+DEDLERAAANLGINLSRNESASVEDLA
Mtb_H37Rv     1132  LSSDGAAIELREGEDEDLERAAANLGINLSRNESASVEDLA  1172
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 RpoB

```
Query: HJANABMJ_01060 DNA-directed RNA polymerase subunit beta' Query ID: lcl|Query_483421
Length: 1316


>LFFBIPCO_00709 DNA-directed RNA polymerase subunit beta'
Sequence ID: Query_483423 Length: 1316
Range 1: 1 to 1316


Mkomo_JERR01  1     MLDVNFFDELRIGLATADDIRNWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWEC  60
                    MLDVNFFDELRIGLATA+DIR WSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWEC
Mtb_H37Rv     1     MLDVNFFDELRIGLATAEDIRQWSYGEVKKPETINYRTLKPEKDGLFCEKIFGPTRDWEC  60

Mkomo_JERR01  61    YCGKYKRVRFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDL  120
                    YCGKYKRVRFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDL
Mtb_H37Rv     61    YCGKYKRVRFKGIICERCGVEVTRAKVRRERMGHIELAAPVTHIWYFKGVPSRLGYLLDL  120

Mkomo_JERR01  121   APKDLEKIIYFAAYVITAVDDEMRHNELSTLEAEMAVERKAIEDQRDADLEARAQKLEAD  180
                    APKDLEKIIYFAAYVIT+VD+EMRHNELSTLEAEMAVERKA+EDQRD +LEARAQKLEAD
Mtb_H37Rv     121   APKDLEKIIYFAAYVITSVDEEMRHNELSTLEAEMAVERKAVEDQRDGELEARAQKLEAD  180

Mkomo_JERR01  181   MKELEDEGAKSDVKRKVRDGGEREMRQLRDRAQRELDRLEEIWTTFTKLAPKQLIVDELL  240
                    + ELE EGAK+D +RKVRDGGEREMRQ+RDRAQRELDRLE+IW+TFTKLAPKQLIVDE L
Mtb_H37Rv     181   LAELEAEGAKADARRKVRDGGEREMRQIRDRAQRELDRLEDIWSTFTKLAPKQLIVDENL  240

Mkomo_JERR01  241   YRELQDRYGEYFEGAMGAESIKKLIENFDIEAEAESLRDTIKNGKGQKKLRALKRLKVVA  300
                    YREL DRYGEYF GAMGAESI+KLIENFDI+AEAESLRD I+NGKGQKKLRALKRLKVVA
Mtb_H37Rv     241   YRELVDRYGEYFTGAMGAESIQKLIENFDIDAEAESLRDVIRNGKGQKKLRALKRLKVVA  300

Mkomo_JERR01  301   AFQTNRNSPMGMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLIDL  360
                    AFQ + NSPMGMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLIDL
Mtb_H37Rv     301   AFQQSGNSPMGMVLDAVPVIPPELRPMVQLDGGRFATSDLNDLYRRVINRNNRLKRLIDL  360

Mkomo_JERR01  361   GAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLLGK  420
                    GAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLLGK
Mtb_H37Rv     361   GAPEIIVNNEKRMLQESVDALFDNGRRGRPVTGPGNRPLKSLSDLLKGKQGRFRQNLLGK  420

Mkomo_JERR01  421   RVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAKRMVERQR  480
                    RVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAKRMVERQR
Mtb_H37Rv     421   RVDYSGRSVIVVGPQLKLHQCGLPKLMALELFKPFVMKRLVDLNHAQNIKSAKRMVERQR  480

Mkomo_JERR01  481   PQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPQLVEGKAIQLHPLVCEAFNADFDGDQ  540
                    PQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEP LVEGKAIQLHPLVCEAFNADFDGDQ
Mtb_H37Rv     481   PQVWDVLEEVIAEHPVLLNRAPTLHRLGIQAFEPMLVEGKAIQLHPLVCEAFNADFDGDQ  540

Mkomo_JERR01  541   MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEIEGDKGEFV  600
                    MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTE+ GD GE+
Mtb_H37Rv     541   MAVHLPLSAEAQAEARILMLSSNNILSPASGRPLAMPRLDMVTGLYYLTTEVPGDTGEYQ  600

Mkomo_JERR01  601   PAATDQPESGVYSSPAEAIMAMDRGALSPRARIRVRLTQLRPSAEVEAEQFP-DGWTMGD  659
                    PA+ D PE+GVYSSPAEAIMA DRG LS RA+I+VRLTQLRP  E+EAE F   GW  GD
Mtb_H37Rv     601   PASGDHPETGVYSSPAEAIMAADRGVLSVRAKIKVRLTQLRPPVEIEAELFGHSGWQPGD  660

Mkomo_JERR01  660   AWTAETTLGRVLFNELLPKGYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLKDAG  719
                    AW AETTLGRV+FNELLP GYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLKDAG
Mtb_H37Rv     661   AWMAETTLGRVMFNELLPLGYPFVNKQMHKKVQAAIINDLAERYPMIVVAQTVDKLKDAG  720

Mkomo_JERR01  720   FYWATRSGVTVSMADVLVPPEKQEILERYEAEADSIEKQYQRGKLNKDERNEALVKIWQD  779
                    FYWATRSGVTVSMADVLVPP K+EIL+ YE  AD +EKQ+QRG LN DERNEALV+IW++
Mtb_H37Rv     721   FYWATRSGVTVSMADVLVPPRKKEILDHYEERADKVEKQFQRGALNHDERNEALVEIWKE  780

Mkomo_JERR01  780   ATEEVGQALRSHYPKDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPIKSSF  839
                    AT+EVGQALR HYP DNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRP+KSSF
Mtb_H37Rv     781   ATDEVGQALREHYPDDNPIITIVDSGATGNFTQTRTLAGMKGLVTNPKGEFIPRPVKSSF  840

Mkomo_JERR01  840   REGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVRETDCETERGITVT  899
                    REGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVRE DC+TERGI V
Mtb_H37Rv     841   REGLTVLEYFINTHGARKGLADTALRTADSGYLTRRLVDVSQDVIVREHDCQTERGIVVE  900
```

```
Mkomo_JERR01  900   LAELQGE-QLVRDQHIETSAYARTLATDAVDANGNVIVERGHDLGDPAIDALLAAGITEV  958
                    LAE   +  L+RD +IETSAYARTL TDAVD  GNVIVERG DLGDP IDALLAAGIT+V
Mtb_H37Rv     901   LAERAPDGTLIRDPYIETSAYARTLGTDAVDEAGNVIVERGQDLGDPEIDALLAAGITQV  960

Mkomo_JERR01  959   KVRSVLTCATGTGVCAMCYGRSMATGKLVDIGEAVGIVAAQSIGEPGTQLTMRTFHQGGV  1018
                    KVRSVLTCAT TGVCA CYGRSMATGKLVDIGEAVGIVAAQSIGEPGTQLTMRTFHQGGV
Mtb_H37Rv     961   KVRSVLTCATSTGVCATCYGRSMATGKLVDIGEAVGIVAAQSIGEPGTQLTMRTFHQGGV  1020

Mkomo_JERR01  1019  TGGADIVGGLPRVQELFEARIPRNRAPIADVSGRVRLEESDKFYKITIVPDDGGEEVVYD  1078
                      G DI GGLPRVQELFEAR+PR +APIADV+GRVRLE+ ++FYKITIVPDDGGEEVVYD
Mtb_H37Rv     1021  --GEDITGGLPRVQELFEARVPRGKAPIADVTGRVRLEDGERFYKITIVPDDGGEEVVYD  1078

Mkomo_JERR01  1079  KLSRRQRLKVFKHDDGSERLLTDGDHVEVGQQLLEGSADPHEVLRVQGPREVQIHLVKEV  1138
                    K+S+RQRL+VFKH+DGSER+L+DGDHVEVGQQL+EGSADPHEVLRVQGPREVQIHLV+EV
Mtb_H37Rv     1079  KISKRQRLRVFKHEDGSERVLSDGDHVEVGQQLMEGSADPHEVLRVQGPREVQIHLVREV  1138

Mkomo_JERR01  1139  QEVYRAQGVSIHDKHIEVIVRQMLRRVTIIDSGATEFLPGSLTERGEFETENRRVVAEGG  1198
                    QEVYRAQGVSIHDKHIEVIVRQMLRRVTIIDSG+TEFLPGSL +R EFE ENRRVVAEGG
Mtb_H37Rv     1139  QEVYRAQGVSIHDKHIEVIVRQMLRRVTIIDSGSTEFLPGSLIDRAEFEAENRRVVAEGG  1198

Mkomo_JERR01  1199  EPAAGRPVLMGITKASLATDSWLSAASFQETTRVLTDAAINCRSDKLQGLKENVIIGKLI  1258
                    EPAAGRPVLMGITKASLATDSWLSAASFQETTRVLTDAAINCRSDKL GLKENVIIGKLI
Mtb_H37Rv     1199  EPAAGRPVLMGITKASLATDSWLSAASFQETTRVLTDAAINCRSDKLNGLKENVIIGKLI  1258

Mkomo_JERR01  1259  PAGTGINRYRNIQVQPTEEARAAAYTIPSYEDQYYSPDFGQATGAAVPLDDYGYSDYR  1316
                    PAGTGINRYRNI VQPTEEARAAAYTIPSYEDQYYSPDFG ATGAAVPLDDYGYSDYR
Mtb_H37Rv     1259  PAGTGINRYRNIAVQPTEEARAAAYTIPSYEDQYYSPDFGAATGAAVPLDDYGYSDYR  1316
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 RpoC

```
Query: HJANABMJ_01079 30S ribosomal protein S12 Query ID: lcl|Query_585747 Length: 124


>LFFBIPCO_00724 30S ribosomal protein S12
Sequence ID: Query_585749 Length: 124
Range 1: 1 to 124

Score:605, Identities:119/124(96%), Positives:121/124(97%), Gaps:0/124(0%)

Mkomo_JERR01  1    MPTINQLVRKGRRDKIAKVKTAALKGSPQRRGVCTRVYTTTPKKPNSALRKVARVRLTSA  60
                   MPTI QLVRKGRRDKI+KVKTAALKGSPQRRGVCTRVYTTTPKKPNSALRKVARV+LTS
Mtb_H37Rv     1    MPTIQQLVRKGRRDKISKVKTAALKGSPQRRGVCTRVYTTTPKKPNSALRKVARVKLTSQ  60

Mkomo_JERR01  61   VEVTAYIPGEGHNLQEHSMVLVRGGRVKDLPGVRYKIIRGSLDTQGVKNRKQARSRYGAK  120
                   VEVTAYIPGEGHNLQEHSMVLVRGGRVKDLPGVRYKIIRGSLDTQGVKNRKQARSRYGAK
Mtb_H37Rv     61   VEVTAYIPGEGHNLQEHSMVLVRGGRVKDLPGVRYKIIRGSLDTQGVKNRKQARSRYGAK  120

Mkomo_JERR01  121  KEKS  124
                   KEK
Mtb_H37Rv     121  KEKG  124
```

Global alignment between *Mtb* H37Rv and *M.* JERR01 RpsL