UNIVERSITY OF ZULULAND

An *in silico* approach to understanding the role of P450s involved in secondary metabolites production in mycobacterial species

Ntokozo Minenhle Zondo

201418736

A dissertation submitted to the Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa, in fulfilment of the requirement for the degree of Master of Science

Supervisor: Prof Khajamohiddin Syed

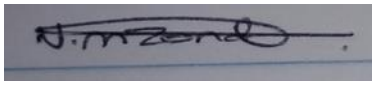
KwaDlangezwa

September 2022

DECLARATION

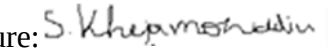
I, Ntokozo Minenhle Zondo, declare that this dissertation is entirely my work and has not been taken from the work of others, except where I have appropriately acknowledged and referenced the source. This dissertation has never been submitted for any degree for examination at any university. Considering that quality is more important than quantity (Gould, 2016), care has been taken to present the dissertation in a publication format (same as the International Journal of Molecular Sciences, impact factor 6.2) to enable the presentation of data in a concise manner and for easy understanding of the work. I state that the work presented in this dissertation has been under review as indicated in the research outputs section.

Student: Ntokozo Minenhle Zondo

Student signature: 

As the candidate's supervisor, I, Professor Khajamohiddin Syed, have read the thesis and have given approval for submission for examination.

Supervisor: Prof Khajamohiddin Syed

Supervisor signature: 

Reference: Gould, J., 2016. Future of the thesis. Nature, 535(7610), pp.26-29.

ABSTRACT

Cytochrome P450 monooxygenases (P450s/CYPs) are ubiquitous enzymes with unique regio- and stereo-selective oxidation activities. Due to these properties, P450s play a key role in the biosynthesis of natural metabolites. Mycobacterial species are well-known producers of complex metabolites that help them survive in diverse ecological niches, including in the host. In this study, a comprehensive analysis of P450s and their role in natural metabolite synthesis in 2666 mycobacterial species have been carried out. The study revealed the presence of 62815 P450s that can be grouped into 182 P450 families and 345 subfamilies. Blooming (presence of more than one copy of the same gene) and expansion (presence of the same gene in many species) were observed at the family and subfamily levels. CYP135 was the dominant family in mycobacterial species. The mycobacterial species have distinct P450 profiles, indicating lifestyle impacts P450 content in their genome vis a vis P450s play a key role in organisms' adaptation. Analysis of the P450 profile revealed a gradual loss of P450s from non-pathogenic to pathogenic mycobacteria. Pathogenic mycobacteria have more P450s in biosynthetic gene clusters that produce natural metabolites. This indicates that P450s are recruited for the biosynthesis of unique metabolites, thus helping these pathogens survive in their niches. This study is the first to analyze P450s and their role in natural metabolite synthesis in many mycobacterial species.

DEDICATION

I dedicate my dissertation work to my family, Nabane Eslina Zondo, my late father, Mzibeni Mapotwane Zondo, and my siblings, Sifiso Zondo and Khunjuliwe Tholiwe Zondo, and to My Child Melokuhle Nkosenhle Khanyile.

ACKNOWLEDGEMENTS

I would like to appreciate God for his wisdom and Mercy and for being by my side always.

Firstly, I would like to express my deepest thanks to Prof Syed for being my leader and research supervisor on this journey. I would also like to thank Prof David R Nelson from the University of Tennessee, Memphis, USA, for naming the P450s and assisting me in completing the study.

Secondly, I would like to thank the P450 group members for their continued support and help.

Lastly, I would like to extend my gratitude to my Mother for being my strength, my Brother for believing in me, and the Zondo family and friends for their encouragement, support, and support.

RESEARCH OUTPUTS

Articles

- **Ntokozo Minenhle Zondo** (co-author) 2019. Similarities, variations, and evolution of cytochrome P450s in *Streptomyces* versus *Mycobacterium*. Scientific reports, 9(1), pp.1-12.
- **Ntokozo Minenhle Zondo**, Tiara Padayachee, David R Nelson, Khajamohiddin Syed 2022. More P450s are involved in the biosynthesis of natural metabolites in pathogenic mycobacterial species. International Journal of Molecular Sciences (under review-ijms-1934237).

Conference attendance

- Conference on Genomics, Proteomics and Metabolomics: All in the Bioinformatics (CGPMB-2019), 27-28 July 2019, Umfolozi Hotel Casino and Convention Centre, Empangeni, KwaZulu-Natal.

ABBREVIATIONS

BGCs	Biosynthetic gene clusters
CYP/P450	Cytochrome P450 monooxygenase
MTBC	*Mycobacterium tuberculosis* complex
MCAC	Mycobacterium chelonae abscessus complex
MCL	Mycobacteria causing leprosy
NTM	Nontuberculous mycobacteria
SAP	Saprophytes

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Background and problem statement

Cytochrome P450 monooxygenases (CYPs/P450s) are undoubtedly one of the most extensively studied enzymes due to their role in organisms' primary and secondary metabolism (Yamazaki, 2014, McLean et al., 2015, Nelson, 2018). The word “cytochrome” indicates the presence of a chromophore such as heme as a prosthetic group, and “P450” denotes their characteristic absorbance spectrum of 450 nm wavelength when they are reduced and complexed with carbon monoxide (Garfinkel, 1958, Klingenberg, 1958, Omura and Sato, 1962, Omura, 2011). The word “monooxygenase” indicates the incorporation of one oxygen into substrates by these enzymes (Garfinkel, 1958, Klingenberg, 1958, Omura and Sato, 1962, White and Coon, 1980, Omura, 2011). However, since their discovery, it has been identified that P450s catalyze diverse enzymatic reactions with regio- and stereo-selectivity (Sono et al., 1996, Bernhardt, 2006, Yan et al., 2022). Due to their unique enzymatic properties, P450s were explored for their biomedical and biotechnological applications (Kelly and Kelly, 2013, Jawallapersand et al., 2014, Girvan and Munro, 2016, Lepesheva et al., 2018, Urlacher and Girhard, 2019, Li et al., 2020).

P450s are ubiquitous as they were found in living and non-living entities such as viruses (Nelson, 2018, Lamb et al., 2019). The International P450 Nomenclature Committee developed a nomenclature and annotation method for correctly identifying P450s in species across all the biological kingdoms (Nelson, 1998, Nelson, 2006, Nelson, 2009). The nomenclature system begins with the prefix “CYP” for cytochrome P450 monooxygenase, followed by an Arabic numeral designating the family, a capital letter representing the subfamily, and an Arabic digit specifying the individual P450 in a family. The annotation criteria include assigning family and subfamily with > 40% identity belong to the same family and all P450s with > 55% identity belong to the same subfamily (Nelson et al., 1993, Nelson, 1998, Nelson, 2006, Nelson, 2009). All researchers have universally accepted this nomenclature and annotation method; thus, any P450 identified should be subjected to this criterion for its proper identification.

P450 application in synthesizing natural metabolites has gained momentum (Podust and Sherman, 2012, Rudolf et al., 2017, Greule et al., 2018). Natural products are metabolites, either primary or secondary, produced by organisms. Primary metabolites are involved in organisms' physiology. In contrast, the secondary metabolites, although they have no direct role in organisms' physiology, they tend to play a role indirectly helping organisms in their

survival. The ongoing genome sequencing rush revealed many P450s from the species across the biological kingdoms (Nelson, 2018). It is impossible to clone, express, and characterize many P450s. Due to this hurdle, *in silico* analysis of P450s will help understand their role in different biological processes. One such *in silico* tool is identifying P450s that possibly play in the biosynthesis of natural products.

In silico studies based on genome-wide analysis of P450s and their role in the biosynthesis of natural products revealed the presence of many P450s in biosynthetic gene clusters (BGCs) in bacterial species. BGCs are none other than a physical grouping of two or more genes in a genome that is responsible for the biosynthesis of a metabolite (Medema et al., 2015). Analysis of BGCs across different bacterial species revealed that *Salinispora* species have the highest percentage of P450s, and *Bacteroidetes* species and cyanobacterial species have the lowest percentage of P450s as part of BGCs (Parvez et al., 2016, Mthethwa et al., 2018, Senate et al., 2019, Khumalo et al., 2020, Padayachee et al., 2020, Mnguni et al., 2020, Msomi et al., 2021, Malinga et al., 2022, Msweli et al., 2022, Nkosi et al., 2022, Nzuza et al., 2021). The percentage of P450s part of BGCs from highest to lowest is as follows: *Salinispora* species (47%) > *Streptomyces* species (23% > *Firmicutes* species (18%) > mycobacterial species (15%) > proteobacterial species (12%) > *Bacteroidetes* species = cyanobacterial species (8%).

Mycobacterial P450s gained attention due to their potential as drug targets (Ortiz de Montellano, 2018) and the production of valuable human metabolites (van Beilen et al., 2005). CYP121A1 and CYP128A1 from *Mycobacterium tuberculosis* H37Rv are involved in the biosynthesis of natural metabolites mycocyclosin (McLean et al., 2008, Belin et al., 2009) and sulfomenaquinone (Sogi et al., 2016). A point to be noted is that *CYP121A1* is essential for the survival of *M. tuberculosis,* indicating that the metabolite it synthesizes is a vital primary metabolite (McLean et al., 2008). Menaquinone is a primary metabolite involved in electron transport in *M. tuberculosis* (Sogi et al., 2016). CYP139 family members were found to be part of different BGC types indicating their involvement in the biosynthesis of diverse natural compounds (Syed et al., 2019). Based on the results authors proposed that these metabolites might provide mycobacterial species with advantageous traits in diverse niches competing with other microbial or viral agents and helping these microbes infect hosts by interfering with the host's metabolism and immune system (Syed et al., 2019). Genome-wide analysis of P450s in 60 mycobacterial species revealed that 15% of P450s belonging to 31 different families are part of BGCs (Parvez et al., 2016). It is well-known that mycobacterial species have complex

metabolites as part of their cellular structure, and some of these metabolites are well-known for their virulence (Quadri, 2014, Bansal-Mutalik and Nikaido, 2014, Ghazaei, 2018). Based on P450s and their BGCs, the authors concluded that these P450s of BGCs possibly play a role in the synthesis of complex metabolites (Senate et al., 2019).

Studies as mentioned earlier, which are limited to a family (CYP121, CYP128, and CYP139) and a few species (60 species), to date, a comprehensive analysis of P450s and their role in natural metabolites biosynthesis in a large number of mycobacterial species has not been carried out. This study aims to address this research gap to understand the P450s role in natural product biosynthesis concerning different mycobacterial categories.

1.2. Aim and Objective

1.2.1. Aim

To annotate and understand the P450s role in secondary metabolites synthesis in mycobacterial species.

1.2.2. Objectives

- To perform genome-wide identification of P450s in 2666 mycobacterial species
- To annotate identified P450s from 2666 mycobacterial species
- To identify the P450s part of biosynthetic gene-clusters
- To perform a comparative analysis of P450 features between mycobacterial species with other bacterial species

1.3. Thesis layout

This dissertation is divided into five chapters apart from the supplementary information listed below:

Chapter 1: Introduction

Chapter 2: Literature review

Chapter 3: Results and discussion

Chapter 4: Methodology

Chapter 5: Conclusions and future perspectives

Supplementary information: Due to the nature of the data, a separate excel file with six tables and a word file with a dataset are provided.

All the information provided in this dissertation is the same as the article submitted to the International Journal of Molecular Sciences (see research outputs). The dissertation layout is designed in the same format as the journal article.

CHAPTER 2: LITERATURE REVIEW

2.1. The genus *Mycobacterium*

The genus *Mycobacterium* contains gram-positive, rod-shaped, and acid-fast species (Ingenbleek, 2017, Ventura et al., 2007). These species adapted to diverse ecological niches and can be found in water bodies, soil, and metalworking fluids (Ingenbleek, 2017). Due to their diverse lifestyles, species belonging to this genus were categorized into six categories (Parvez et al., 2016). The six categories are *Mycobacterium tuberculosis* complex (MTBC), *M. chelonae-abscessus* complex (MCAC), *M. avium* complex (MAV), Mycobacteria causing leprosy (MCL), Nontuberculous mycobacteria (NTM) and Saprophytes (SAP).

2.2. *Mycobacterium tuberculosis* complex (MTBC)

2.2.1. *Mycobacterium tuberculosis* H37Rv

M. tuberculosis H37Rv is a pathogenic, aerobic, chemoorganotrophic, rod-shaped, non-motile bacterium that causes Tuberculosis (TB) in humans (Cole et al., 1998). *M. tuberculosis* is typically a slow-growing, dormant, intracellular pathogen with a complex cell envelope and genetic homogeneity (Cole et al., 1998). *M. tuberculosis*, although primarily affecting the lungs, can affect almost every organ in the body, producing symptoms imitating both inflammatory and malignant disease (Gillespie and Bamford, 2012). The presence of *M. tuberculosis* in a clinical specimen is almost always associated with infection (Menzies et al., 2007). The genome sequence of *M. tuberculosis* H37Rv is probably the best-characterized strain. It has been investigated and analyzed to improve our understanding of its biology and help discover new prophylactic and therapeutic interventions (Cole et al., 1998, Camus et al., 2002).

2.2.2. *Mycobacterium africanum* GM041182

M. africanum was first isolated from a non-infected, HIV-positive patient in 2004 with a positive smear for pulmonary tuberculosis (Bentley et al., 2012). In contrast to other species in this category, *M. africanum* possesses all the biochemical properties usually used to differentiate between the different species, especially *M tuberculosis* and *M. bovis* (Mostowy et al., 2004). Due to this trait, *M. africanum* is an unspecified classification and is generally applied to isolates in the category that differs in their phenotypic presentation and genomic content (Mostowy et al., 2004).

M. africanum is also known for causing TB in humans, especially in West African countries (Vasconcellos et al., 2010). Symptoms of infection are very similar to those caused

by *M. tuberculosis*, and predisposing factors include immune-compromised states, such as in patients with HIV (Galagan et al., 2010). Although the host range of *M. africanum* is believed to be human, the organism is isolated with a much lower frequency than *M tuberculosis* (Mostowy et al., 2004). *M. africanum* is found commonly in Northern, Western, and Central Africa (Galagan et al., 2010).

2.2.3. *Mycobacterium bovis*

M. bovis causes bovine, and human TB and avirulent strains of the organism are used for the Bacille Calmette-Guérin (BCG) vaccine (Bentley et al., 2012). Although the organism mainly causes TB in cattle, it may infect other animals, including dogs, cats, swine, rabbits, antelope, lions, buffalo, and sometimes certain birds of prey (Vissa and Brennan, 2001). Bovine TB is still common in less developed countries, causing severe economic losses from livestock deaths, chronic illness, and trade restrictions. The organism may also be a threat to endangered animals.

2.3. *Mycobacterium chelonae-abscessus* complex (MCAC)

2.3.1. *Mycobacterium abscessus*

M. abscessus ATCC 19977 this strain was first isolated in 1953 in a subcutaneous lesion from a knee infection in a human (Choo et al., 2012), and *M. abscessus* subspecies *bolletii* 50594 was isolated from a Korean patient with a pulmonary infection (Kim et al., 2017). *M. abscessus* is one of the most clinically important of the rapidly growing mycobacteria (Ripoll et al., 2009). It is a common water and soil contaminant (Ripoll et al., 2009). *M. abscessus* commonly causes skin infection, often evolving into draining subcutaneous abscesses, particularly in patients with compromised immunity (Yano et al., 2006). Although present in many water sources, the organism is also an important causative agent of pulmonary disease in patients with chronic respiratory disease (Ripoll et al., 2009). Other predisposing factors for infection include organ transplantation, rheumatoid arthritis, trauma, and invasive medical procedures (Yano et al., 2006). Lung infections with *M. abscessus* in patients with cystic fibrosis are becoming increasingly frequent (Ripoll et al., 2009).

Pulmonary infection with *M. abscessus* is considered very serious, with a much higher fatality rate than any other rapidly growing *mycobacterium* (Ripoll et al., 2009). Although considered an environmental contaminant, this organism is frequently found on hospital equipment. It has been associated with nosocomial and community-acquired infections, ranging from superficial skin and soft tissue infection to disseminated disease (Choo et al.,

2012). The frequent observation of the organism in healthcare-associated disease is because the organism is resistant to most disinfectants and biocides and can survive in harsh environments (Ripoll et al., 2009).

2.4. *Mycobacterium avium* complex (MAC)

2.4.1. *Mycobacterium avium*

M. avium is a natural pathogen of birds (Wu et al., 2006, Gillespie and Bamford, 2012) and an opportunistic pathogen of humans. However, very little epidemiologic information is available regarding human disease (Horan et al., 2006). It is a gram-positive, non-spore-forming, slightly curved, aerobic, slow-growing bacteria (Galagan et al., 2010). In the childrens, they cause mycobacterial lymphadenitis, osteomyelitis in immune-compromised patients, and chronic pulmonary infection in the elderly (Gillespie and Bamford, 2012). In the advanced stages of AIDS, the organism can cause a disseminated infection or bacterial sepsis (Vissa and Brennan, 2001, Gillespie and Bamford, 2012).

Patients at risk for *M. avium* include those suffering or affected by chronic obstructive pulmonary disease, bronchiectasis, chronic aspiration or recurrent pneumonia, TB, pneumoconiosis, and bronchogenic carcinoma (Zink et al., 2003). There has been some correlation between infection with *M. avium* and cystic fibrosis (Zink et al., 2003). In non-immune compromised patients, the disease presents like an infection with *M. tuberculosis* (Zink et al., 2003), and occasionally patients are asymptomatic. *M. avium* 104 is widely distributed in the environment and is an opportunistic pathogen known to cause infection in immune-compromised patients (Wu et al., 2006). The first strain was isolated in 1983 in South California from an adult patient infected with AIDS.

M. avium subspecies *paratuberculosis* K10 is an obligate pathogen of animals such as cattle, sheep, antelopes, deer, and giraffes, causing a disease characterized by enteritis (Wu et al., 2006). Infection in animals causes significant economic losses, especially in the dairy industry (Wu et al., 2006). *M. avium* subspecies *paratuberculosis* has also been associated with Chron's disease in humans, with symptoms like intestinal symptoms caused by an intestinal infection in cattle (Wu et al., 2006). This strain can only be treated with a combination of antibiotics such as Rifabutin and Clarithromycin. It is not susceptible to many anti-TB drugs, making treatment very difficult (Wynne et al., 2010).

M. avium subspecies *paratuberculosis* MAP4 this strain is the causative agent for Johne's disease in animals like cattle, sheep, goats, deer, giraffes, antelopes, and camels (You et al.,

2012, Bannantine et al., 2014). It is associated with Chron's disease in humans (Bannantine et al., 2014). Chron's disease is a chronic granulomatous infection in the small intestine of these animals with a worldwide distribution (Bannantine et al., 2014). Chron's disease takes a tremendous economic toll on the livestock industry (Bannantine et al., 2014). This strain was isolated in 2000 from the breast milk of a patient with Chron's disease. The organism has a complex cell wall structure with mycolic acids and several lipids very similar to other strains in this genus, but it is the most slow-growing of these organisms (Bannantine et al., 2014). The organism requires 8 to 16 weeks before colonies are visible, and a siderophore is needed in the laboratory medium for growth (Bannantine et al., 2014).

2.5. *Mycobacterium* causing Leprosy

2.5.1. *Mycobacterium leprae*

M. leprae is gram-positive, aerobic, and surrounded by the unique mycobacterial waxy coating. The organism was identified in 1873 before *M. tuberculosis*, making it the first human pathogenic *mycobacterium* to be identified (Vissa and Brennan, 2001). Up to date humans are the only known host (Monot et al., 2009).

This *bacterium* causes leprosy, or Hansen's disease, in humans. Leprosy is a chronic infectious and primarily granulomatous disease of the peripheral nerves, mucosa, and upper respiratory tract producing skin lesions. The organism initially attacks peripheral nerves, followed by digital destruction and deformity, often leaving the patient severely disabled (Monot et al., 2009, Gillespie and Bamford, 2012). Patients at risk are immune-compromised (Gillespie and Bamford, 2012). *M. leprae* Br4923 strain was first isolated from a patient in Brazil, the country with the second highest incidence of leprosy (Monot et al., 2009). *M. leprae* TN strain was first isolated from an armadillo in Tamil Nadu, India (Cole et al., 2001).

2.6. Nontuberculosis mycobacteria

2.6.1. *Mycobacterium marinum*

M. marinum is a facultatively anaerobic, fast-growing organism with a growth rate of 4 to 6 hours (Stinear et al., 2008). Unlike *M. tuberculosis*, *M. marinum* cannot reduce nitrate and produces a bright yellow carotenoid pigment when exposed to light, which protects it from UV damage (Stinear et al., 2008). This organism can replicate in both single-celled organisms, such as amoeba, and cultured mammalian macrophages, meaning the organism can survive in both intracellular and unpredictable extracellular environments (Stinear et al., 2008).

It is mainly a pathogen of fish and amphibians (Stinear et al., 2008) and is found in aquatic environments such as swimming pools and drinking water. The systemic granulomatous infection in amphibians and fish has histologic similarities to lesions of TB in humans (Stinear et al., 2008). In humans, it causes granulomatous skin infections, mainly acquired from contaminated rivers, poorly maintained swimming pools, and fish tanks through cuts and scratches in the skin (Stinear et al., 2008). Infection with *M. marinum*, commonly referred to as fish tank or aquarium tank granuloma (Stinear et al., 2008), has been reported in lifeguards, fishermen, and people working in the aquatic industry (Yano et al., 2006). The granulomatous infection is generally limited to the skin and soft tissue extremities, and it resembles like dermal disease caused by *M. tuberculosis* (Stinear et al., 2008). *M. marinum* has an optimal growth temperature at 35ºC in Middlebrook 7H9 medium and grows poorly at 37ºC, which explains its tendency to infect poikilotherms and a superficial disease restricted to cooler extremities of the human body (Stinear et al., 2008).

2.6.2. *Mycobacterium ulcerans* Agy99

This organism causes the devastating Buruli ulcer in humans and is believed to have derived from *M. marinum (*Stinear et al., 2007, Stinear et al., 2008). Buruli ulcer is a necrotic disease of subcutaneous tissue and is widespread in West and Central Africa (Stinear et al., 2007). The transmission mode is still uncertain (Stinear et al., 2008). This is the only pathogen that does not have an intracellular lifestyle in the granulomas, as it can lyse cells using its polyketide toxin (Stinear et al., 2008). The toxin produced by the organism, mycolactone, is described as a macrolide toxin (Stinear et al., 2007). The slow growth, ability to stimulate immune suppressors, cell wall remodeling and modification and biofilm-forming capability give this organism its survival ability (Stinear et al., 2007).

2.7. Saprophytes

2.7.1. *Mycobacterium vanbaalenii* PYR-1

M. vanbaalenii is a rod shaped, non-motile, non-spore forming bacteria. *M. vanbaalenii* plays a role in bioremediation. This organism can degrade a wide range of environmentally toxic chemicals, including high-molecular-weight polycyclic aromatic hydrocarbons (PAHs), like pyrene, as a sole source of carbon and energy (Ripoll et al., 2009). This strain was first isolated at a site close to the Harbour Island oil tank farm in the watershed of Redfish Bay. The metabolic diversity of *M. vanbaalenii* makes it a potential source for bioremediation of PAH-contaminated areas (Kim et al., 2017). The genome sequence of *M. vanbaalenii* contains 6012

protein-coding sequences, many of which are involved in the catabolism of aromatic compounds (Kim et al., 2017).

2.7.2. *M. smegmatis* JS623

M. smegmatis JS623, an ethane oxidising organism, can grow on vinyl chloride as a carbon and energy source (Jin et al., 2010). Vinyl chloride is a toxic ground water pollutant associated with plastic manufacture and chlorinated solvent used is a known human carcinogen (Jin et al., 2010). This trait makes it potentially useful in the bioremediation of such pollutants (Jin et al., 2010). This strain is also a useful model system for studying microbial enzyme evolution in response to xenobiotic compounds (Jin et al., 2010).

2.7.3. *M. gilvum* PYR-GCK

This strain was isolated from the sediment of the Grand Calumet River in North-western Indiana and can utilize the toxic polycyclic hydrocarbon pyrene for growth (Badejo et al., 2014). It has been studied to try and glean some insight into mechanisms related to its exceptional bioremediation abilities (Badejo et al., 2014).

2.8. *M. tuberculosis* P450s

Cytochrome P450 monooxygenases (CYPs/P450s) are heme-thiolate proteins well known for their diverse enzymatic reactions, particularly in a stereo- and regio-specific fashion (Bernhardt, 2006, Sono et al., 1996). As a result of this function, P450s have become vital for organismal survival, playing a critical role in primary and secondary metabolism (Kelly and Kelly, 2013). P450s are multifunctional oxidoreductases found in organisms across all biological kingdoms (Nelson, 2018). P450s are divided into mainly two divisions by their functionality; some play a role in biological processes in organisms, such as fatty acid metabolism and sterol and cholesterol synthesis, and those that break down xenobiotic compounds in organisms (Kelly and Kelly, 2013).

Genome sequence analysis of *M. tuberculosis* H37Rv shows the existence of 20 P450s in its genome (Cole et al., 1998) (Figure 2.1). Among 20 P450s, only a few P450s have been functionally characterized (Figure 2.1). The characterized P450s appear to have the less specific physiological function of P450s rather than being devoted to removing xenobiotics (Figure 2.1 and Table 2.1).

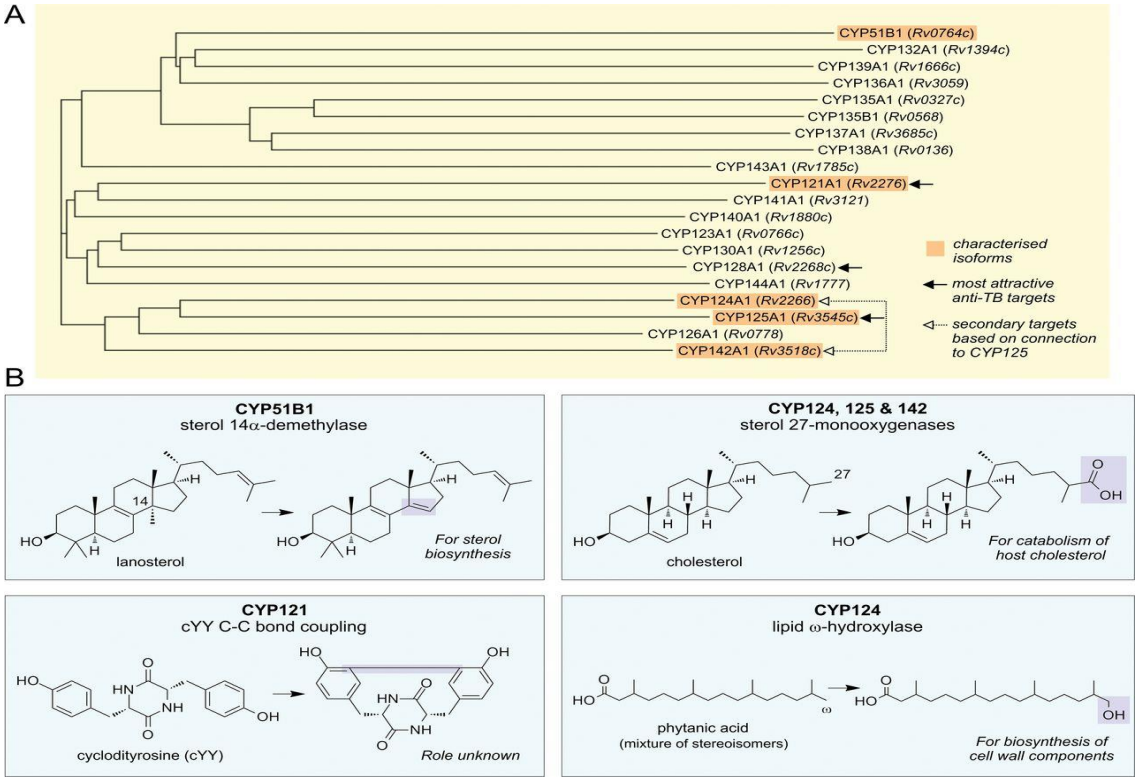


Figure 2.1. Schematic representation of 20 Cytochrome P450s (A) and the ones that are functionally characterized (B) (Ortiz de Montellano, 2015).

Table 2.1. Functional analysis of *M. tuberculosis* P450s.

P450 name	Function	Reference
CYP121A1	Play a key role in the formation of an extra molecular C-C bond formation between 2 tyrosyl carbon atoms of CYY(Cyclic dipeptide Cyclo-L-Try-L-Try)	(Belin et al., 2009)
CYP124A1	Play a key role in the oxidation of branched-chain fatty acids	(Johnston et al., 2009)
CYP128A1	*In vivo* studies demonstrated that it is involved in the oxidation of menaquinone	(Sogi et al., 2016)

CYP125A1-CYP142A1	Play a key role in cholesterol catabolism	(McLean et al., 2009, Capyk et al., 2009, Van der Geize et al., 2007), and
CYP130A1	Catalyzes the N-demethylation of dextromethorphan yielding 3-methoxymorphinan	(Ortega Ugalde et al., 2018)

2.9. Role of mycobacterial P450s in secondary metabolites production

Analysis of secondary metabolite biosynthetic gene clusters (BGCs) in sixty mycobacterial species revealed many P450s are in the BGCs (Table 2.2) (Senate et al., 2019)**.** A total number of 204 P450s were found to be part of 9 types of BGCs in mycobacterial species (Table 2.2). All the 204 P450s belonging to 31 P450 families (Table 2.3). Among these 31 P450 families CYP139 family members were highly present as part of secondary metabolite gene clusters (Table 2.3) (Senate et al., 2019) . Latest study further confirmed that CYP139 family members are indeed involved in synthesis of secondary metabolites per se helping the mycobacterial species in their pathogenesis (Syed et al., 2019). This indicates that mycobacterial P450s are indeed involved in production of different secondary metabolites that are useful to *mycobacterium* to survive in the host (Senate et al., 2019).

Table 2.2. Comparative analysis of biosynthetic gene cluster and their P450s in sixty mycobacterial species (Senate et al., 2019). The abbreviation of biosynthetic gene clusters is the standard abbreviation available at Anti-SMASH (Blin et al., 2021).

Biosynthetic gene cluster type	Number of P450s
Other	73
Nrps	34
T3pks-T1pks	30
T1pks	24
T1pks-Nrps	19

T3pks	14
Arylpolyene	5
T2pks	3
Terpene	2
Total	204

Table 2.3. The list of mycobacterial P450s is part of the secondary metabolite biosynthetic gene cluster (Senate et al., 2019).

P450 family	No. of P450s
CYP139	32
CYP124	24
CYP128	22
CYP121	17
CYP187	14
CYP150	13
CYP144	10
CYP105	9
CYP1128	8
CYP135	8
CYP189	6
CYP279	6
CYP138	4
CYP1110	3
CYP125	3

CYP153	3
CYP102	2
CYP1034	2
CYP140	2
CYP143	2
CYP147	2
CYP185	2
CYP188	2
CYP1121	1
CYP1122	1
CYP126	1
CYP136	1
CYP161	1
CYP164	1
CYP292	1
CYP1123	1
	204

2.10. Role of secondary metabolites in mycobacterial species physiology

In mycobacterial species, polyketide synthases (PKS) are involved in the biosynthesis of complex lipids. Therefore, it is suggested that PKSs, in conjunction with fatty acid synthases, can generate diverse and unusual lipids that constitute this organism's complex cell wall of this organism thus evading the host immune system. For example, MAR/MAP BGC products are found to be part of the cell envelope in *M. marinum*, possibly complicating its access to the host immune system or drug actions (Parvez, 2018). Akaeolide has cytotoxic activity against fibroblasts, suggesting it may play a role in tissue weakening in the host (Zhou et al., 2015);

JBIR-100 exhibits cytotoxic activities and inhibition of proton pumps such as vacuolar-type ATPases (V-ATPases) activities and is thus linked with an increasing number of diseases such as osteopetrosis, male infertility and renal acidosis (Ueda et al., 2010, Huss and Wieczorek, 2009).

Lorneic acid A inhibits *phosphodiesterase* PDE5 blocking the degradation of cGMP (Iwata et al., 2009); thus, it might play a role in pulmonary hypertension. *Meridamycin* has been found to bind FK506-binding proteins (FKBP12) (Salituro et al., 1995). FKBP12 proteins play a key role in regulating fundamental aspects of cell biology and have been found to be critical in mice survival (Aghdasi et al., 2001). *Nigericin* inhibits the Golgi functions in eukaryotic cells and is a well-known activator of the NLRP3 inflammasome (Wawrocki and Druszczynska, 2017, Rao et al., 2008, Katsnelson et al., 2015), indicating a bacterial infection. One secondary metabolite, namely mycolactone, a lipid-like toxin with cytotoxic, immunosuppressive, and tissue necrosis activity, is involved in developing Buruli ulcers by *M. ulcerans* (Sarfo et al., 2016).

CHAPTER 3: RESULTS AND DISCUSSION

3.1. Saprophytic to pathogen lifestyle led to the loss of P450s in mycobacterial Species

Genome-wide data mining and annotation for P450s in 2666 mycobacterial species revealed the presence of 62815 P450s and 90 P450 fragments/pseudo genes in their genomes (Tables 4 and S1). The 2666 mycobacterial species includes, *Mycobacterium tuberculosis* complex (MTBC) (2128 species), *M. chelonae-abscessus* complex (MCAC) (255 species), *M. avium* complex (MAC) (106 species), Mycobacteria causing leprosy (MCL) (4 species), Nontuberculous mycobacteria (NTM) (203 species) and Saprophytes (SAP) (10 species). Comparative analysis with other bacterial species revealed that mycobacterial species have a high average number of P450s (24 P450s) compared to other bacterial species, only exceeded by *Streptomyces* species (27 P450s) (Table 3.1), indicating *Streptomyces* species still have more P450s as mentioned previously (Mnguni et al., 2020, Senate et al., 2019). However, a point to be noted is that only 126 *Streptomyces* species data are available for comparison with 2666 mycobacterial species. Thus, analysis of more *Streptomyces* species probably bridges the gap and shows a similar trend to mycobacterial species. Nonetheless, it is clear that among the bacterial population, actinomycetes have the highest average number of P450s in their genomes (Table 3.1), indicating that P450s play a key role in their primary and secondary metabolism that including adaptation to diverse ecological niches as described elsewhere (Mnguni et al., 2020, Senate et al., 2019, Parvez et al., 2016). Among mycobacterial species, *Mycolicibacterium rhodesiae* JS60 has the highest (95 P450s), and *Mycobacterium leprae* Br492 has the lowest number of P450s (3 P450s) in their genomes (Table S1).

Analysis of the P450 profile of six mycobacterial categories revealed a gradual loss of P450s from SAP to MTBC (Tables 5 and S1). The order is as follows: SAP (35-95:51) > MAC (28-69:48) = NTM (9-89:48) > MCAC (22-52:26) > MTBC (5-74:20) > MCL (3-7:5) where the minimum and maximum number of P450s, and after the semicolon, the average number of P450s, are shown in parenthesis. This suggests that during the progression from saprophytic to pathogenic lifestyle, mycobacterial species lost P450s (Tables 5 and S1). This phenomenon of gradual loss of P450s in mycobacterial species from SAP to MTBC/MCL was previously reported (Parvez et al., 2016). However, in the previous study, only 60 mycobacterial species were analyzed (Parvez et al., 2016). Observation of the same phenomenon in this study, where many species (2666 species) were examined, strongly supports that mycobacterial species lost P450s in their genomes to adapt to diverse ecological niches. As described elsewhere (Parvez

et al., 2016, Mnguni et al., 2020, Senate et al., 2019, Syed et al., 2019, van Wyk et al., 2019), the P450s retained in these species played a crucial role in their adaptation to diverse ecological niches. A detailed analysis of P450 key features in six mycobacterial categories is presented in Table 3.2.

Table 3.1. Comparative analysis of key features of P450s and their association with secondary metabolism in different bacterial species. Abbreviation: No., number of; BGCs: biosynthetic gene clusters.

Category	*Salinispora* species	*Streptomyces* species	Mycobacterial species	Cyanobacterial species	*Bacteroidetes* species	*Firmicutes* species	Alphaproteobacterial species	Betaproteobacterial species	Gammaproteobacterial species	Deltaproteobacterial species	Epsilonproteobacterial species
Species analyzed	126	203	2666	114	334	972	599	513	1261	107	216
Species with P450s	126	203	2666	114	77	229	229	290	169	23	53
Percentage of species with P450s	100	100	100	100	23	24	38	57	13	21	25
No. of P450s	2643	5460	62815	341	98	712	873	603	277	333	53
No. of families	45	253	182	36	21	14	143	79	81	74	2
No. of subfamilies	103	698	345	79	28	53	214	119	102	171	2
Dominant P450 family	CYP105	CYP107	CYP135	CYP110	CYP1103	CYP107	CYP202	CYP116	CYP133 & CYP107	CYP107	CYP172
Average No. of P450s	21	27	24	3	1	3	4	2	2	14	1
No. of P450s part of BGCs	1236	1231	9399	27	8	126	21	107	49	69	0
No. of P450 families part of BGCs	35	135	68	6	5	10	16	18	22	37	0

Percentage of P450s part of BGCs	47	23	15	8	8	18	2	18	18	21	0
Reference(s)	(Malinga et al., 2022)	(Mnguni et al., 2020, Senate et al., 2019)	This work	(Khumalo et al., 2020)	(Nkosi et al., 2022)	(Padayachee et al., 2020)	(Nzuza et al., 2021)	(Msweli et al., 2022)	(Msomi et al., 2021)	(Msweli et al., 2022)	(Msweli et al., 2022)

Table 3.2. Comparative analysis of key features of P450s and their association with secondary metabolism in mycobacterial categories. Abbreviation: MTBC, *Mycobacterium tuberculosis* complex; MCAC, M. *chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, Nontuberculous mycobacteria and SAP, Saprophytes; No., number of; BGCs: biosynthetic gene clusters.

Category	MTBC	MCAC	NTM	MAC	SAP	MCL
Total No. of species analyzed	2128	255	163	106	10	4
Total No. of P450s	42917	6519	7760	5093	505	21
P450 fragments/pseudo	2	10	69	3	6	0
Average No. of P450s	20	26	48	48	51	5
Min No. of P450s	5	22	9	28	35	3
Maximum No. of P450s	74	52	89	69	95	7
No. of P450 families	66	37	145	59	66	9
No of P450 subfamilies	95	48	261	88	101	9
Dominant P450 family	CYP135	CYP125	CYP125	CYP150	CYP189	CYP136 & CYP184
No of the P450s part of BGCs	8153	438	450	328	30	0
No. of P450 families part of BGCs	19	13	56	11	15	0
Percentage of P450s part of BGCs	19,0	6,7	5,8	6,4	5,9	0
Percentage of P450 families part of BGCs	28,8	35,1	38,6	18,6	22,7	0

3.2. P450 family and subfamily blooming/expansion in mycobacterial species

Based on the International P450 Nomenclature Committee rules (Nelson et al., 1993, Nelson, 2006, Nelson, 1998), the percentage identity of > 40% for a family and > 55% for a subfamily, the 62815 P450s from 2666 mycobacterial species can be grouped into 182 P450 families and 345 P450 subfamilies (Figure 3.1 and Table S2). Among mycobacterial species, *M. rhodesiae* JS60 had the highest number of P450s and also had the highest number of P450 families (53) and subfamilies (75) (Table S1). No P450 family was conserved in mycobacterial species (Table S1). Among 182 P450 families, 21 P450 families contributed 85% of total P450s in

mycobacterial species (Figure 3.1 and Table S2), indicating their important role in these species. Among the P450 families, CYP135 is dominant with 4702 members, followed by CYP125 with 4085 members, CYP123 with 3019 members, and CYP136 with 3017 members (Figure 3.1 and Table S2). It is safe to say that these P450 families bloomed (present more than one copy in a species) in mycobacterial species, considering the number of species analyzed in this study. The P450 families include CYP138, CYP140, CYP144, CYP51, CYP130, CYP142, CYP124, CYP143, CYP126, CYP139, CYP128, CYP137, CYP132, CYP141, and CYP121 has members between 2000-3000 indicating their expansion in mycobacterial species (Figure 3.1 and Table S2). Contrary to the P450 families that bloomed/expanded, 43 P450 families had a single member, and 25 P450 families had only two members (Table S2).

Among P450 families, CYP107 had the highest number of subfamilies (15), followed by CYP125 (13 subfamilies) and CYP105 (11 subfamilies) (Table S2). The Blooming/expansion phenomenon is also observed at a P450 subfamily level (Tables 6 and S2). P450 subfamily analysis revealed that specific subfamilies were bloomed or expanded in mycobacterial species indicating a selective preference for distinct subfamilies by the species (Tables 6 and S2).

Comparative analysis with other bacterial species revealed that mycobacterial species have the highest number of P450 families and subfamilies but only next to *Streptomyces* species (Table 3.1). This suggests that *Streptomyces* species have the highest P450 family and subfamily diversity compared to mycobacterial species (Table 3.1).

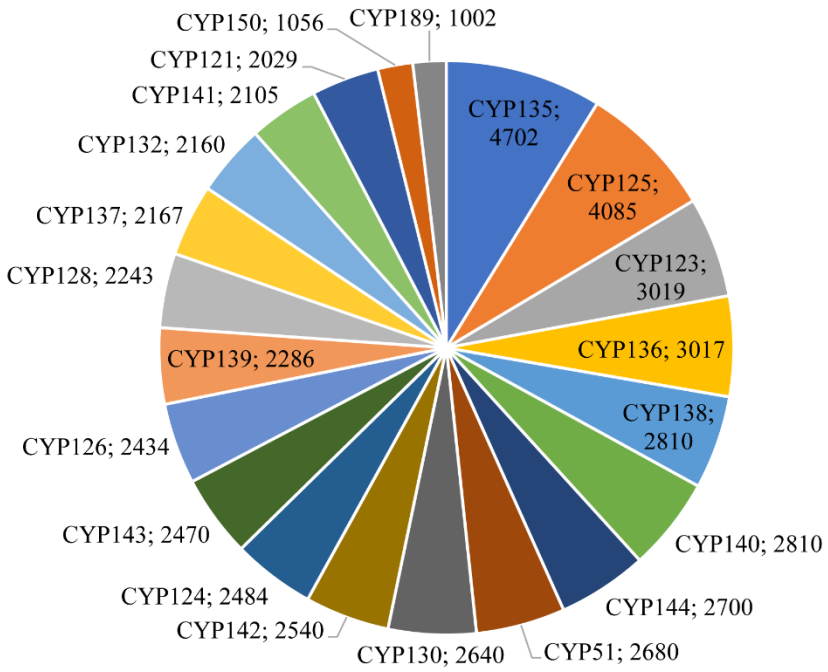


Figure 3.1. Comparative analysis of P450 families in 2666 mycobacterial species. P450 families with >1000 members were presented in the figure. The P450 family name and number of P450s are shown in the figure. Detailed information on P450 families and subfamilies is presented in Table S2.

Table 3.3. Analysis of P450 subfamily blooming/expansion in mycobacterial species. Subfamily information for P450 families with >500 members is presented in the table, along with the nature of the subfamily (blooming/expansion). A detailed analysis of P450 subfamilies is shown in Table S2.

P450 family	Count	Subfamily	Count	Nature of the subfamily
CYP123	3019	A	2908	Bloomed
CYP125	4085	A	3978	Bloomed
CYP136	3017	A	2687	Bloomed
CYP105	674	Q	377	Expanded
CYP108	714	B	703	Expanded
CYP121	2029	A	2029	Expanded
CYP124	2484	A	2425	Expanded
CYP126	2434	A	2429	Expanded
CYP128	2243	A	2143	Expanded
CYP130	2640	A	2640	Expanded
CYP132	2160	A	2160	Expanded
CYP135	4702	A	2149	Expanded
		B	2543	Expanded
CYP137	2167	A	2167	Expanded
CYP138	2810	A	2733	Expanded
CYP139	2286	A	2286	Expanded
CYP140	2810	A	2401	Expanded
CYP141	2105	A	2105	Expanded
CYP142	2540	A	2507	Expanded

CYP143	2470	A	2387	Expanded
CYP144	2700	A	2421	Expanded
CYP150	1056	A	932	Expanded
CYP164	514	A	505	Expanded
CYP189	1002	A	987	Expanded
CYP51	2680	B	2680	Expanded

3.3. Different mycobacterial categories have distinct P450 profiles

P450s play a key role in organisms' adaptation vis a vis lifestyle impacts P450 content in their genome (Msweli et al., 2022, Padayachee et al., 2020, Senate et al., 2019, Parvez et al., 2016, Mnguni et al., 2020, Nkosi et al., 2022). This phenomenon is evident in mycobacterial species, as mycobacterial categories have distinct P450 profiles (Figure 1.2 and Tables 5 and S3). Among mycobacterial categories, NTM had the highest number of P450 families and subfamilies (145 and 261), followed by SAP (66 and 101), MTBC (66 and 95), MAC (59 and 88), MCAC (37 and 48), and MCL (9 families and subfamilies) (Tables 5 and S3). A point to be noted is that despite the lower number of species analyzed for NTM (163 species) compared to MTBC (2128 species) and MCAC (255 species), NTM has the highest number of P450 families (145) and subfamilies (261) (Tables 5 and S3). The same can be attributed when compared to SAP and MTBC, where only ten species of SAP and 2128 species of MTBC had the same number of P450 families (Tables 5 and S3). This indicates that NTM had a high P450 family and subfamily diversity among mycobacterial categories, indicative of diverse P450s in the species.

P450 family conservation analysis revealed different P450 families conserved in different mycobacterial categories. CYP135 is conserved in MTBC, CYP125 is conserved in MCAC, CYP105 and CYP150 are conserved in MAC, CYP189 is conserved in SAP, and CYP164 is conserved in MCL (Tables S1 and S4). No P450 family was conserved in NTM, possibly due to the high P450 family diversity mentioned above. However, CYP125 was the dominant P450 family in NTM (Table 3.2). Analysis of unique and shared P450 families among mycobacterial categories revealed that NTM has the highest number of unique P450 families (71 families), followed by SAP (23 families), MAC (15 families), MTBC (11

families), and MCAC (6 families) (Tables 3.4 and S4). No P450 family was unique for MCL (Tables 3.4 and S4). As indicated in Table 3.4, many P450 families shared among mycobacterial groups. CYP105, CYP125, CYP150, and CYP189 families are dominantly shared across mycobacterial categories (Tables 3.4 and S4). This indicates that these P450 families play an important role in mycobacterial species; thus, they not only retained but also bloomed/expanded (Table 3.3). MCL shares the CYP136 and CYP164 families with other bacterial groups (Table 3.4).

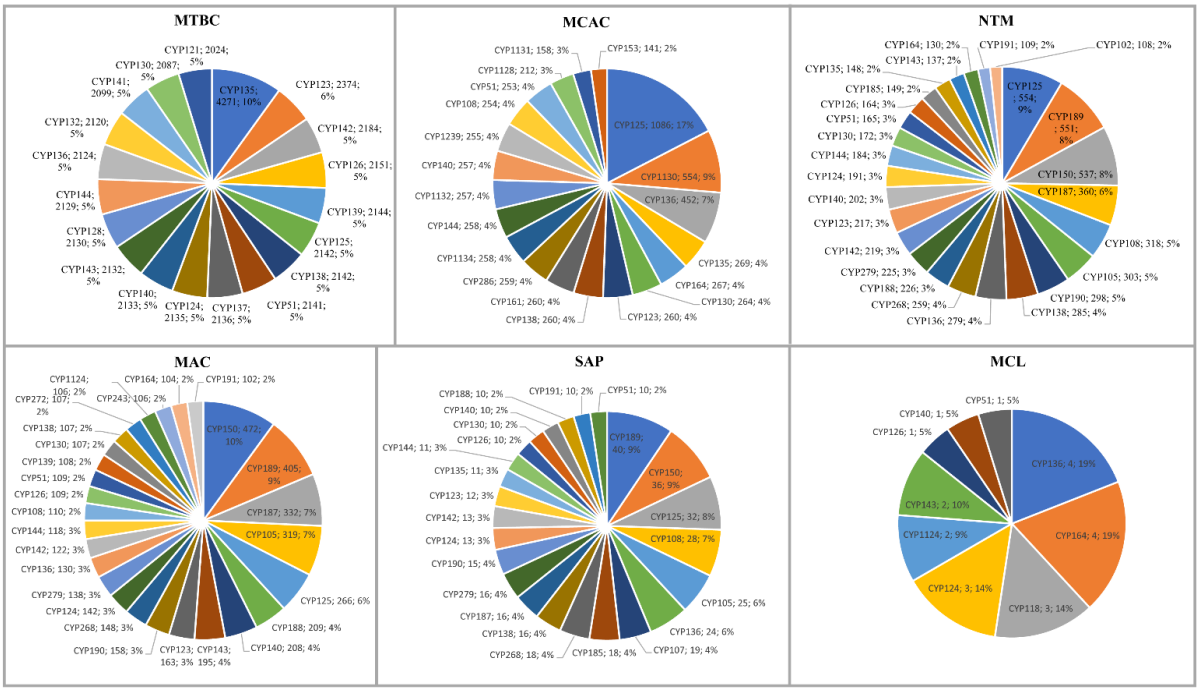


Figure 3.2. Comparative analysis of P450s families in six mycobacterial categories. The P450 family, members and their percentage in the total number of P450s in a category are presented in the figure. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, Nontuberculous mycobacteria and SAP, Saprophytes. Detailed information on P450 families with their specific count for each mycobacterial category is presented in Table S3.

Table 3.4. Analysis of unique and shared P450 families among mycobacterial categories. P450 family dominantly shared among mycobacterial categories is listed in parenthesis. A detailed analysis of P450 families and their member count is presented in Table S4. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, M. *chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, Nontuberculous mycobacteria and SAP, Saprophytes.

Category	MTBC	MCAC	NTM	MAC	SAP	MCL
MTBC	11	22 (CYP125)	49 (CYP189)	36 (CYP150)	34 (CYP189)	8 (CYP136 & CYP164)
MCAC	22 (CYP125)	6	28 (CYP125)	20 (CYP105)	19 (CYP125)	5 (CYP136 & CYP164)
NTM	49 (CYP189)	28 (CYP125)	71	34 (CYP189)	29 (CYP189)	7 (CYP164)
MAC	36 (CYP150)	20 (CYP105)	34 (CYP189)	15	25 (CYP189)	9 (CYP136 & CYP164)
SAP	34 (CYP189)	19 (CYP125)	29 (CYP189)	27 (CYP189)	23	5 (CYP136)
MCL	8 (CYP136 & CYP164)	5 (CYP136 & CYP164)	7 (CYP164)	9 (CYP136 & CYP164)	5 (CYP136)	0

3.4. CYP121, CYP124, and CYP128 P450s are part of the same biosynthetic gene cluster

Analysis of P450s involved in natural metabolite biosynthesis in mycobacterial species revealed a total of 9399 P450s out of 62815 (15%) are part of BGCs (Tables 4 and S4). Mycobacterial species have the highest percentage of P450s that are part of BGCs compared to cyanobacterial species, *Bacteroidetes* species, *Firmicutes* species, and proteobacterial species (Table 3.1). However, *Streptomyces*-, and *Salinispora*-species had the highest percentage of P450s part of BGCs compared to mycobacterial species (Table 3.1). This indicates that actinomycetes have more P450s involved in natural metabolite biosynthesis among bacterial species and particularly *Salinispora* species have more P450s part BGCs (Table 3.1). Among 182 P450 families identified in mycobacterial species, only 68 P450 families (37%) are part of BGCs (Figure 3.3 and Table S5).

Among 68 P450 families that are part of BGCs, CYP139 was the most dominant with 2171 members, followed by CYP128 with 1960 members, CYP121 with 1953 members, and CYP124 with 1946 members (Figure 3.3 and Table S5). In total, these four P450 families contributed 85% of P450s part of BGCs in mycobacterial species (Figure 3.3 and Table S5). A point to be noted is that these four P450 families possibly play a key role in natural metabolite

biosynthesis and, thus, are expanded in mycobacterial species (see Section 2.2). One example is CYP139 BGC clusters known to produce metabolites that provide mycobacterial species with advantageous traits in diverse niches competing with other microbial or viral agents, which might help these microbes infect hosts by interfering with the host's metabolism and immune system (Syed et al., 2019). CYP128 was found to play an essential role in producing a metabolite that acts as a negative regulator of virulence (Sogi et al., 2016, Ngcobo et al., 2020). CYP121 is a crucial P450 and drug target against pathogenic mycobacterial species and is involved in synthesizing a metabolite named mycocyclosin (Belin et al., 2009, Nguyen et al., 2019, Ortiz de Montellano, 2018, Rajput et al., 2019). CYP124 is a lipid hydroxylase and secondary drug target in pathogenic mycobacterial species (Johnston et al., 2009, Johnston et al., 2010, Ortiz de Montellano, 2018). A recent study reported that CYP128 and CYP121, and CYP124 are part of the same BGC, indicating their collective role in synthesizing complex natural metabolites that play an important role in mycobacterial species (Ngcobo et al., 2020). In this current study, we also observed that these three P450s are part of the same BGC (Table S4), further supporting the previous observation and conclusion on the collective role of these P450s in synthesizing natural metabolites in mycobacterial species (Ngcobo et al., 2020).

Five P450 families, CYP144 with 216 members, CYP135 with 198 members, CYP1128 with 193 members, CYP150 with 132 members, and CYP187 with 130 members, contributed 9% of P450s part of BGCs in mycobacterial species (Figure 3.3 and Table S5). The remaining 59 P450 families contributed only 6% of P450s part of BGCs, indicating their minor role in natural metabolite biosynthesis in mycobacterial species (Figure 3.3 and Table S5). A point to be noted is that CYP135, despite being the most dominant P450 family and CYP144 and CYP150 families being expanded in mycobacterial species (Figure 3.1), only a fraction of these family members are part of BGCs (Figure 3.3 and Table S5). Furthermore, CYP125, CYP123, CYP136, CYP138, and CYP140 families, despite being bloomed/expanded (see section 2.2), their role in natural metabolite biosynthesis is negligible as only a few members are found to be part of BGCs (Figure 3.3 and Table S5). This suggests that these P450s might be involved in key primary metabolism and thus are bloomed/expanded in these species. This also supports previous observations that the dominant P450 family may not necessarily play a role in natural metabolite biosynthesis (Khumalo et al., 2020, Malinga et al., 2022, Mnguni et al., 2020, Nkosi et al., 2022, Nzuza et al., 2021, Padayachee et al., 2020, Parvez et al., 2016). Detailed information on P450s part of BGCs, their species name, and BGC type, the similar known-

gene cluster is presented in Table S5, and the data on the analysis of P450s that are part of BGCs is shown in Table S6.

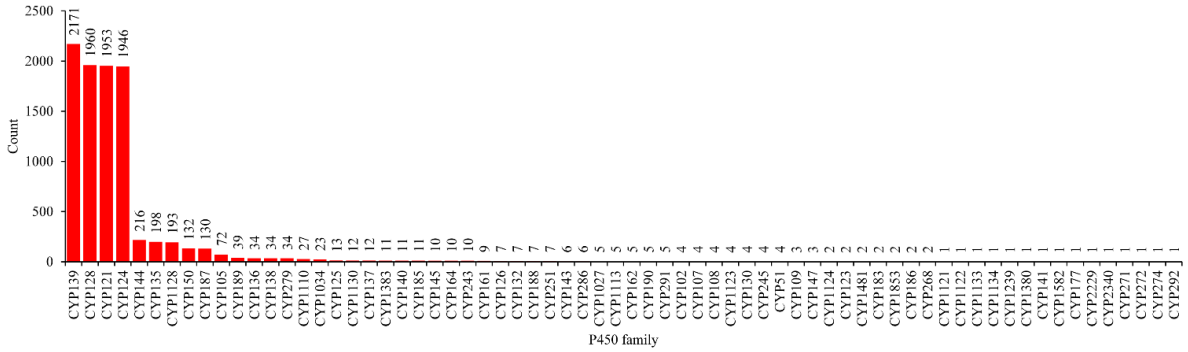


Figure 3.3. Comparative analysis of P450s involved in natural metabolite biosynthesis in mycobacterial species. The number next to the bars represents the number of members in a P450 family. Detailed information on the P450s part of biosynthetic gene clusters (BGCs), their species names, and BGC types are presented in Table S5.

3.5. More P450s are involved in natural metabolite biosynthesis in pathogenic mycobacterial species

Analysis of P450s involved in natural metabolite biosynthesis revealed that more P450s are part of BGCs in pathogenic mycobacteria compared to non-pathogenic (Figure 3.4 and Tables 5 and S6). The number of P450s part of BGCs in MTBC was the highest (8153 P450s), followed by NTM (450 P450s), MCAC (438 P450s), MAC (328 P450s), and SAP (30 P450s) (Figure 3.4 and Tables 5 and S6). As expected, due to reduced genome size and having few P450s in their genomes, MCL has no P450s as part of BGCs, and thus, we did not include MCL for comparative analysis, same as followed elsewhere (Parvez et al., 2016). One can argue that the number of species analyzed for MTBC is the highest compared to other categories, and thus one can see the highest number of P450s part of BGCs (Table 3.2). To clarify and nullify this argument, we have compared the percentage of P450s part of BGCs in different mycobacterial categories (Table 3.2). Analysis of the percentage of P450s part of BGCs revealed that indeed MTBC has the highest percentage of P450s (19%) part of BGCs, followed by MCAC (6.7%), MAC (6.4%), SAP (5.9%), and NTM (5.8%) (Table 3.2). This suggests that pathogenic mycobacterial species such as MTBC indeed have more P450s as part of BGCs and, thus, more P450s in these species involved in the biosynthesis of natural metabolites. The point to be noted

is that MTBC has more P450s part of BGCs (Table 3.2) despite having the lowest number of P450s in their genomes compared to MCAC, MAC, NTM, and SAP (Table 3.1). This indicates that P450s in MTBC may play a vital role in the biosynthesis of natural metabolites, thus helping these organisms survive in the host, as mentioned in section 2.4.

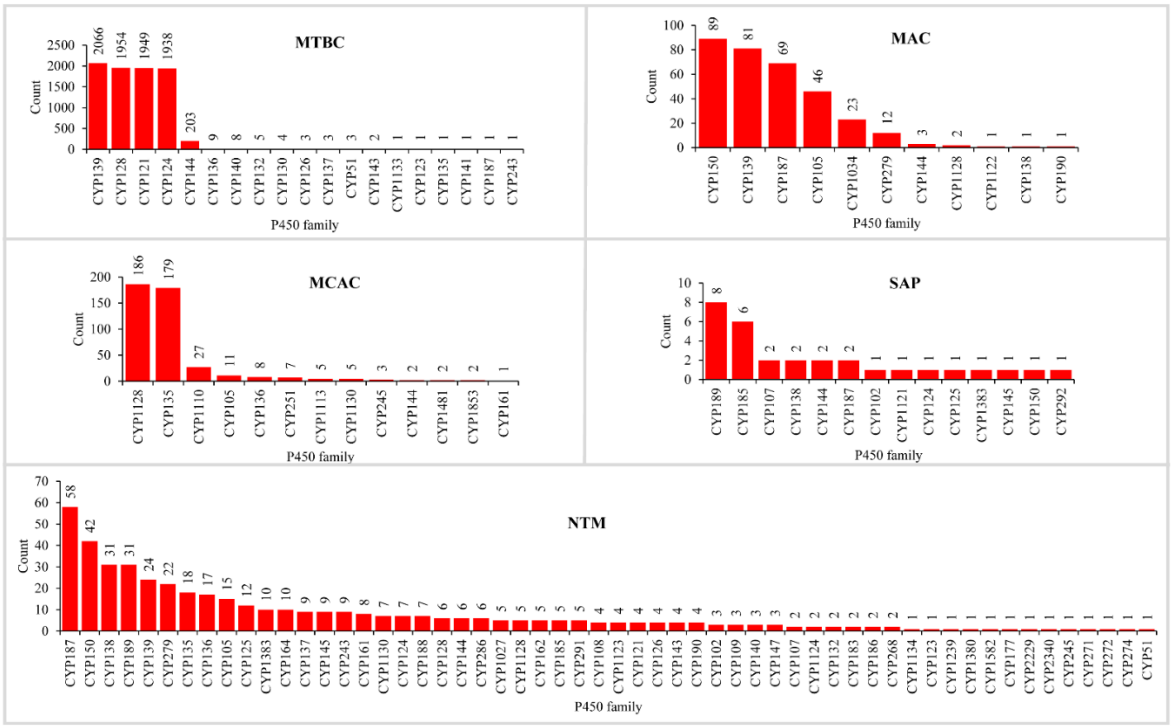


Figure 3.4. Comparative analysis of P450s that are part of biosynthetic gene clusters (BGCs) in mycobacterial categories. Detailed information on the P450s part of BGCs, their species names, and BGC types are presented in Table S5. P450 family analysis in individual categories is presented in Table S6. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, Nontuberculous mycobacteria; and SAP, Saprophytes.

Analysis of P450 families part of BGCs revealed that NTM has the highest number of P450 families part of BGCs, followed by MTBC, SAP, MCAC, and MAC (Figure 3.4 and Table S6). The number of P450 families part of BGCs followed the same pattern as the number of P450 families in these categories indicating diverse P450s are indeed involved in natural metabolite biosynthesis in NTM (Figure 3.4 and Table 3.2). A clear picture emerged when we compared the percentage of P450 families part of BGCs, where NTM still had the highest P450 families as part of the BGCs, followed by MCAC, MTBC, SAP, and MAC (Table 3.2). The lowest percentage of P450 families part of BGC indicates blooming/expansion of a particular P450 family. This is true as few P450 families are populated in BGCs in different

mycobacterial categories (Figure 3.4 and Table S6). Four P450 families such as CYP139, CYP128, CYP121, and CYP124 have contributed 97% of P450 families as part of BGCs in MTBC; CYP150, CYP139, CYP187, and CYP105 families contributed 87% in MAC, and CYP1128 and CYP135 families contributed 83% in MCAC (Figure 3.4 and Table S6) indicating these P450 families preferred in these species possibly due to their importance in the natural product synthesis. Pathogenic mycobacterial species seem to have recruited more P450 families belonging to the same family for natural metabolite biosynthesis. In contrast, the non-pathogenic mycobacterial species have less and more diverse P450s for natural metabolite biosynthesis. This suggests that natural metabolites produced by MTBC play a role in their survival in the host (as mentioned in section 2.4). Thus, P450s play a crucial role in synthesizing these metabolites.

CHAPTER 4: METHODOLOGY

4.1. Species and their genome database information

Mycobacterial species genomes (permanent and finished draft genomes) available for public use at the Joint Genome Institute Integrated Microbial Genomes and Microbiomes (JGI IMG/M) (Chen et al., 2021) were used in the study (last accessed on April 2022). Information on the species and their genome IDs used in the study is provided in Table S1.

4.2. Grouping of mycobacterial species

Mycobacterial species were grouped into six categories following the criteria described elsewhere (Parvez et al., 2016). The six categories include *Mycobacterium tuberculosis* complex (MTBC), *M. chelonae-abscessus* complex (MCAC), *M. avium* complex (MAC), Mycobacteria causing leprosy (MCL), nontuberculous mycobacteria (NTM) and Saprophytes (SAP). Briefly, mycobacterial species are grouped into six categories based on their characteristic features, including ecological niches and the nature and site of infection, as described elsewhere (Ventura et al., 2007). Also, a taxonomical grouping of mycobacterial species is considered as described elsewhere (Tortoli, 2012). Mycobacterial species and their categories are presented in Table S1.

4.4. Identification of P450s part of BGCs

P450s that are part of BGCs were identified following the method described elsewhere (Mnguni et al., 2020, Syed et al., 2019, Malinga et al., 2022). Briefly, for each mycobacterial species genome available at JGI IMG/M (Chen et al., 2021), the BGCs were searched for the presence of P450s using the P450 gene ID. The cluster type is noted if a P450 is found as part of the cluster. The gene cluster sequence was downloaded and submitted to anti-SMASH (antibiotics & Secondary Metabolite Analysis Shell) (Blin et al., 2019) to find a similar known cluster. Results were recorded on Excel spreadsheets and represented species-wise BGCs, BGC type, percentage similarity to known gene clusters, and P450s part of specific BGCs.

4.5. P450 key features analysis

All calculations were carried out following the procedure reported previously by our laboratory (Msomi et al., 2021, Msweli et al., 2022). The average number of P450s was calculated using the formula: Average number of P450s = Number of P450s/Number of species. The percentage of P450s that formed part of B.G.C.s was calculated using the formula: Percentage of P450s part of B.G.C.s = 100 × Number of P450s part of BGCs/Total number of P450s present in

species. P450 family/subfamily is considered bloomed when a member count exceeds the number of species and expands when the member count exceeds >500.

4.3. Genome data mining and annotation of P450s

Genome data mining and identification of P450s in mycobacterial species were carried out following the protocol described elsewhere (Mnguni et al., 2020, Syed et al., 2019, Malinga et al., 2022). Briefly, each mycobacterial species genome available at JGI IMG/M (Chen et al., 2021) was searched for P450s using the InterPro code "IPR001128". The hit protein sequences were then searched for the presence of P450 characteristic motifs such as EXXR and CXG (Syed and Mashele, 2014, Gotoh, 1992). Proteins with one of these motifs or short amino acid length are considered P450-fragments. P450 fragments were not considered for the final P450 family and subfamily count. Proteins having both motifs were selected for assigning the family and subfamilies. Following the International P450 Nomenclature Committee rule (Nelson et al., 1993, Nelson, 2006, Nelson, 1998), proteins with >40% identity and >55% identity will be grouped under the same family and subfamily, respectively. P450s with less than 40% identity were assigned to a new P450 family. Mycobacterial species P450s identified in this study and their protein sequences, assigned names, and species are presented in Supplementary Dataset 1.

4.6. Comparative analysis of P450s and BGCs data

For comparative analysis of P450s and BGCs, information for bacterial species belonging to different groups such as phyla, *Firmicutes* (Padayachee et al., 2020), *Bacteroidetes* (Nkosi et al., 2022), *Proteobacteria* (Msweli et al., 2022), and *Cyanobacteria* (Khumalo et al., 2020), and the genera, *Streptomyces* (Senate et al., 2019, Mnguni et al., 2020), and *Salinispora* (Malinga et al., 2022), was resourced from published articles.

CHAPTER 5: CONCLUSIONS AND FUTURE PERSPECTIVES

Cytochrome P450 monooxygenases (CYPs/P450s) play a key role in synthesizing natural metabolites in organisms. They attribute diversity to the metabolites by performing unique regio- and stereo-selective oxidation reactions. Mycobacterial species have complex metabolites that help them survive in diverse ecological niches. Previous studies limited to a few P450 families or a few species indicated that P450s play a role in synthesizing natural metabolites in mycobacterial species. The availability of many mycobacterial genomes allowed us to look into the P450s role in the biosynthesis of natural metabolites concerning their lifestyle. This study's results indicated that despite having a low number of P450s, pathogenic mycobacterial species used most of the available P450s to synthesize natural metabolites. In contrast, non-pathogenic mycobacterial species had fewer P450s playing a role in the biosynthesis of natural metabolites. This suggests that the lifestyle of mycobacterial species changed the P450 profiles vis a vis P450s playing a role in these species' adaptation to different niches as observed in other bacterial species. Characterizing P450 biosynthetic gene cluster metabolites will provide insights into their role in mycobacterial physiology.

SUPPLEMENTARY INFORMATION

The following are supplementary tables and the information provided separately in the excel file and word files.

Table S 1. Genome data-mining and annotation of P450s in mycobacterial species. Mycobacterial species were presented as per their category described elsewhere (Parvez et al., 2016), along with their Genome ID from JGI Mycocosm (Chen et al., 2021). The number of P450s in each mycobacterial species, P450 IDs, and their annotated names as per the International P450 nomenclature system are presented in the table.

Table S 2. Comparative analysis of P450 families and subfamilies in mycobacterial species.

Table S 3. P450 analysis in six mycobacterial categories. Abbreviation: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, nontuberculous mycobacteria, and SAP, Saprophytes.

Table S 4. Comparative analysis of P450 families among six different mycobacterial categories. The number in parenthesis indicates the member count for a family. Abbreviation: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, nontuberculous mycobacteria, and SAP, Saprophytes.

Table S 5. Information on the P450s that are part of mycobacterial species biosynthetic gene clusters (BGCs). Detailed information on species names and codes, the number of P450s part of a BGC in a species, P450 ID and its identification number (ID), the biosynthetic cluster type, and the similar known clusters are presented in the table. Abbreviation: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; MCL, Mycobacteria causing leprosy; NTM, nontuberculous mycobacteria and SAP, Saprophytes.

Table S 6. Comparative analysis of P450 families as part of biosynthetic gene clusters. The number represents the member P450s in a family. Abbreviation: MTBC, *Mycobacterium tuberculosis* complex; MCAC, *M. chelonae-abscessus* complex; MAC, *M. avium* complex; NTM, nontuberculous mycobacteria, and SAP, Saprophytes.

Supplementary Dataset 1. P450s identified and annotated in mycobacterial species are presented with their assigned name, followed by species name, species code, and protein sequence.

REFERENCES

AGHDASI, B., YE, K., RESNICK, A., HUANG, A., HA, H. C., GUO, X., DAWSON, T. M., DAWSON, V. L. & SNYDER, S. H. 2001. FKBP12, the 12-kDa FK506-binding protein, is a physiologic regulator of the cell cycle. *Proceedings of the National Academy of Sciences,* 98**,** 2425-2430.

BADEJO, A. A., DAMILARE, A. & OJUADE, T. D. 2014. Processing effects on the antioxidant activities of beverage blends developed from *Cyperus esculentus, Hibiscus sabdariffa,* and *Moringa oleifera* extracts. *Preventive nutrition and food science,* 19**,** 227.

BANNANTINE, J. P., LI, L., MWANGI, M., COTE, R., GARAY, J. A. R. & KAPUR, V. 2014. Complete genome sequence of *Mycobacterium avium* subsp. *paratuberculosis,* isolated from human breast milk. *Genome Announc.,* 2**,** e01252-13.

BANSAL-MUTALIK, R. & NIKAIDO, H. 2014. Mycobacterial outer membrane is a lipid bilayer and the inner membrane is unusually rich in diacyl phosphatidylinositol dimannosides. *Proceedings of the National Academy of Sciences,* 111**,** 4958-4963.

BELIN, P., LE DU, M. H., FIELDING, A., LEQUIN, O., JACQUET, M., CHARBONNIER, J. B., LECOQ, A., THAI, R., COURCON, M., MASSON, C., DUGAVE, C., GENET, R., PERNODET, J. L. & GONDRY, M. 2009. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A,* 106**,** 7426-31.

BENTLEY, S. D., COMAS, I., BRYANT, J. M., WALKER, D., SMITH, N. H., HARRIS, S. R., THURSTON, S., GAGNEUX, S., WOOD, J. & ANTONIO, M. 2012. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS neglected tropical diseases,* 6.

BERNHARDT, R. 2006. Cytochromes P450 as versatile biocatalysts. *J Biotechnol,* 124**,** 128-45.

BLIN, K., PASCAL ANDREU, V., DE LOS SANTOS, E. L. C., DEL CARRATORE, F., LEE, S. Y., MEDEMA, M. H. & WEBER, T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res,* 47**,** D625-d630.

BLIN, K., SHAW, S., KLOOSTERMAN, A. M., CHARLOP-POWERS, Z., VAN WEZEL, G. P., MEDEMA, M. H. & WEBER, T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic acids research,* 49**,** W29-W35.

CAMUS, J.-C., PRYOR, M. J., MÉDIGUE, C. & COLE, S. T. 2002. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology,* 148**,** 2967-2973.

CAPYK, J. K., KALSCHEUER, R., STEWART, G. R., LIU, J., KWON, H., ZHAO, R., OKAMOTO, S., JACOBS, W. R., ELTIS, L. D. & MOHN, W. W. 2009. Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. *Journal of Biological Chemistry,* 284**,** 35534-35542.

CHEN, I.-M. A., CHU, K., PALANIAPPAN, K., RATNER, A., HUANG, J., HUNTEMANN, M., HAJEK, P., RITTER, S., VARGHESE, N. & SESHADRI, R. 2021. The

IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. *Nucleic acids research,* 49**,** D751-D763.

CHOO, S. W., WONG, Y. L., YUSOFF, A. M., LEONG, M. L., WONG, G. J., ONG, C. S., NG, K. P. & NGEOW, Y. F. 2012. Genome sequence of the *Mycobacterium abscessus* strain M93. Am Soc Microbiol.

COLE, S., BROSCH, R., PARKHILL, J., GARNIER, T., CHURCHER, C., HARRIS, D., GORDON, S., EIGLMEIER, K., GAS, S. & BARRY III, C. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature,* 393**,** 537.

COLE, S., EIGLMEIER, K., PARKHILL, J., JAMES, K., THOMSON, N., WHEELER, P., HONORE, N., GARNIER, T., CHURCHER, C. & HARRIS, D. 2001. Massive gene decay in the leprosy bacillus. *Nature,* 409**,** 1007-1011.

GALAGAN, J. E., SISK, P., STOLTE, C., WEINER, B., KOEHRSEN, M., WYMORE, F., REDDY, T., ZUCKER, J. D., ENGELS, R. & GELLESCH, M. 2010. TB database 2010: overview and update. *Tuberculosis,* 90**,** 225-235.

GARFINKEL, D. 1958. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Archives of biochemistry and biophysics,* 77**,** 493-509.

GHAZAEI, C. 2018. *Mycobacterium tuberculosis* and lipids: Insights into molecular mechanisms from persistence to virulence. *J Res Med Sci,* 23**,** 63.

GILLESPIE, S. & BAMFORD, K. 2012. *Medical microbiology and infection at a glance,* John Wiley & Sons.

GIRVAN, H. M. & MUNRO, A. W. 2016. Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Current opinion in chemical biology,* 31**,** 136-145.

GOTOH, O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry,* 267**,** 83-90.

GREULE, A., STOK, J. E., DE VOSS, J. J. & CRYLE, M. J. 2018. Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat Prod Rep,* 35**,** 757-791.

HORAN, K. L., FREEMAN, R., WEIGEL, K., SEMRET, M., PFALLER, S., COVERT, T. C., VAN SOOLINGEN, D., LEAO, S. C., BEHR, M. A. & CANGELOSI, G. A. 2006. Isolation of the genome sequence strain *Mycobacterium avium* 104 from multiple patients over a 17-year period. *Journal of clinical microbiology,* 44**,** 783-789.

HUSS, M. & WIECZOREK, H. 2009. Inhibitors of V-ATPases: old and new players. *Journal of Experimental Biology,* 212**,** 341-346.

INGENBLEEK, Y. 2017. Lean Body Mass Harbors Sensing Mechanisms that Allow Safeguarding of Methionine Homeostasis. *Nutrients,* 9.

IWATA, F., SATO, S., MUKAI, T., YAMADA, S., TAKEO, J., ABE, A., OKITA, T. & KAWAHARA, H. 2009. Lorneic acids, trialkyl-substituted aromatic acids from a marine-derived actinomycete. *Journal of natural products,* 72**,** 2046-2048.

JAWALLAPERSAND, P., MASHELE, S. S., KOVACIC, L., STOJAN, J., KOMEL, R., PAKALA, S. B., KRASEVEC, N. & SYED, K. 2014. Cytochrome P450 monooxygenase CYP53 family in fungi: comparative structural and evolutionary analysis and its role as a common alternative anti-fungal drug target. *PLoS One,* 9**,** e107209.

JIN, Y. O., CHEUNG, S., COLEMAN, N. V. & MATTES, T. E. 2010. Association of missense mutations in epoxyalkane coenzyme M transferase with adaptation of *Mycobacterium* sp. strain JS623 to growth on vinyl chloride. *Appl. Environ. Microbiol.,* 76**,** 3413-3419.

JOHNSTON, J. B., KELLS, P. M., PODUST, L. M. & ORTIZ DE MONTELLANO, P. R. 2009. Biochemical and structural characterization of CYP124: a methyl-branched lipid omega-hydroxylase from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A,* 106**,** 20687-92.

JOHNSTON, J. B., OUELLET, H. & ORTIZ DE MONTELLANO, P. R. 2010. Functional redundancy of steroid C26-monooxygenase activity in *Mycobacterium tuberculosis* revealed by biochemical and genetic analyses. *J Biol Chem,* 285**,** 36352-60.

KATSNELSON, M. A., RUCKER, L. G., RUSSO, H. M. & DUBYAK, G. R. 2015. K+ efflux agonists induce NLRP3 inflammasome activation independently of Ca2+ signaling. *The Journal of Immunology,* 194**,** 3937-3952.

KELLY, S. L. & KELLY, D. E. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci,* 368**,** 20120476.

KHUMALO, M. J., NZUZA, N., PADAYACHEE, T., CHEN, W., YU, J. H., NELSON, D. R. & SYED, K. 2020. Comprehensive Analyses of Cytochrome P450 Monooxygenases and Secondary Metabolite Biosynthetic Gene Clusters in *Cyanobacteria*. *Int J Mol Sci,* 21.

KIM, S. S., LEE, S. Y., PARK, K. J., PARK, S. M., AN, H. J., HYUN, J. M. & CHOI, Y. H. 2017. *Gluconacetobacter* sp. gel_SEA623-2, bacterial cellulose producing bacterium isolated from citrus fruit juice. *Saudi J Biol Sci,* 24**,** 314-319.

KLINGENBERG, M. 1958. Pigments of rat liver microsomes. *Arch Biochem Biophys,* 75**,** 376-86.

LAMB, D. C., FOLLMER, A. H., GOLDSTONE, J. V., NELSON, D. R., WARRILOW, A. G., PRICE, C. L., TRUE, M. Y., KELLY, S. L., POULOS, T. L. & STEGEMAN, J. J. 2019. On the occurrence of cytochrome P450 in viruses. *Proceedings of the National Academy of Sciences,* 116**,** 12343-12352.

LEPESHEVA, G. I., FRIGGERI, L. & WATERMAN, M. R. 2018. CYP51 as drug targets for fungi and protozoan parasites: past, present and future. *Parasitology,* 145**,** 1820-1836.

LI, Z., JIANG, Y., GUENGERICH, F. P., MA, L., LI, S. & ZHANG, W. 2020. Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *Journal of Biological Chemistry,* 295**,** 833-849.

MALINGA, N. A., NZUZA, N., PADAYACHEE, T., SYED, P. R., KARPOORMATH, R., GRONT, D., NELSON, D. R. & SYED, K. 2022. An Unprecedented Number of Cytochrome P450s Are Involved in Secondary Metabolism in Salinispora Species. *Microorganisms,* 10**,** 871.

MCLEAN, K. J., CARROLL, P., LEWIS, D. G., DUNFORD, A. J., SEWARD, H. E., NEELI, R., CHEESMAN, M. R., MARSOLLIER, L., DOUGLAS, P. & SMITH, W. E. 2008. Characterization of active site structure in CYP121: A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *Journal of Biological Chemistry,* 283**,** 33406-33416.

MCLEAN, K. J., LAFITE, P., LEVY, C., CHEESMAN, M. R., MAST, N., PIKULEVA, I. A., LEYS, D. & MUNRO, A. W. 2009. The structure of *Mycobacterium tuberculosis* CYP125 molecular basis for cholesterol binding in a P450 needed for host infection. *Journal of Biological Chemistry,* 284**,** 35524-35533.

MCLEAN, K. J., LEYS, D. & MUNRO, A. W. 2015. Microbial cytochrome P450s In Cytochrome P450: Structure, mechanism, and biochemistry 4th edn.(eds Oritz de Montellano, PR) Ch. 6, 261–407. Springer International Publishing.

MEDEMA, M. H., KOTTMANN, R., YILMAZ, P., CUMMINGS, M., BIGGINS, J. B., BLIN, K., DE BRUIJN, I., CHOOI, Y. H., CLAESEN, J., COATES, R. C., CRUZ-MORALES, P., DUDDELA, S., DUSTERHUS, S., EDWARDS, D. J., FEWER, D. P., GARG, N., GEIGER, C., GOMEZ-ESCRIBANO, J. P., GREULE, A., HADJITHOMAS, M., HAINES, A. S., HELFRICH, E. J., HILLWIG, M. L., ISHIDA, K., JONES, A. C., JONES, C. S., JUNGMANN, K., KEGLER, C., KIM, H. U., KOTTER, P., KRUG, D., MASSCHELEIN, J., MELNIK, A. V., MANTOVANI, S. M., MONROE, E. A., MOORE, M., MOSS, N., NUTZMANN, H. W., PAN, G., PATI, A., PETRAS, D., REEN, F. J., ROSCONI, F., RUI, Z., TIAN, Z., TOBIAS, N. J., TSUNEMATSU, Y., WIEMANN, P., WYCKOFF, E., YAN, X., YIM, G., YU, F., XIE, Y., AIGLE, B., APEL, A. K., BALIBAR, C. J., BALSKUS, E. P., BARONA-GOMEZ, F., BECHTHOLD, A., BODE, H. B., BORRISS, R., BRADY, S. F., BRAKHAGE, A. A., CAFFREY, P., CHENG, Y. Q., CLARDY, J., COX, R. J., DE MOT, R., DONADIO, S., DONIA, M. S., VAN DER DONK, W. A., DORRESTEIN, P. C., DOYLE, S., DRIESSEN, A. J., EHLING-SCHULZ, M., ENTIAN, K. D., FISCHBACH, M. A., GERWICK, L., GERWICK, W. H., GROSS, H., GUST, B., HERTWECK, C., HOFTE, M., JENSEN, S. E., JU, J., KATZ, L., KAYSSER, L., KLASSEN, J. L., KELLER, N. P., KORMANEC, J., KUIPERS, O. P., KUZUYAMA, T., KYRPIDES, N. C., KWON, H. J., LAUTRU, S., LAVIGNE, R., LEE, C. Y., LINQUAN, B., LIU, X., LIU, W., et al. 2015. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol,* 11**,** 625-31.

MENZIES, D., PAI, M. & COMSTOCK, G. 2007. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. *Annals of internal medicine,* 146**,** 340-354.

MNGUNI, F. C., PADAYACHEE, T., CHEN, W., GRONT, D., YU, J.-H., NELSON, D. R. & SYED, K. 2020. More P450s are involved in secondary metabolite biosynthesis in *Streptomyces* compared to *Bacillus*, *Cyanobacteria* and *Mycobacterium*. *International Journal of Molecular Sciences,* 21**,** 4814.

MONOT, M., HONORÉ, N., GARNIER, T., ZIDANE, N., SHERAFI, D., PANIZ-MONDOLFI, A., MATSUOKA, M., TAYLOR, G. M., DONOGHUE, H. D. & BOUWMAN, A. 2009. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nature genetics,* 41**,** 1282.

MOSTOWY, S., ONIPEDE, A., GAGNEUX, S., NIEMANN, S., KREMER, K., DESMOND, E. P., KATO-MAEDA, M. & BEHR, M. 2004. Genomic analysis distinguishes *Mycobacterium africanum*. *Journal of clinical microbiology,* 42**,** 3594-3599.

MSOMI, N. N., PADAYACHEE, T., NZUZA, N., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021. In silico analysis of P450s and their role in secondary metabolism in the bacterial class *Gammaproteobacteria*. *Molecules,* 26**,** 1538.

MSWELI, S., CHONCO, A., MSWELI, L., SYED, P. R., KARPOORMATH, R., CHEN, W., GRONT, D., NKOSI, B. V. Z., NELSON, D. R. & SYED, K. 2022. Lifestyles Shape the Cytochrome P450 Repertoire of the Bacterial Phylum *Proteobacteria*. *International Journal of Molecular Sciences,* 23**,** 5821.

MTHETHWA, B. C., CHEN, W., NGWENYA, M. L., KAPPO, A. P., SYED, P. R., KARPOORMATH, R., YU, J. H., NELSON, D. R. & SYED, K. 2018. Comparative analyses of cytochrome P450s and those associated with secondary metabolism in *Bacillus* species. *Int J Mol Sci,* 19.

NELSON, D. R. 1998. Cytochrome P450 nomenclature. *Methods Mol Biol,* 107**,** 15-24.

NELSON, D. R. 2006. Cytochrome P450 nomenclature, 2004. *Methods Mol Biol,* 320**,** 1-10.

NELSON, D. R. 2009. The cytochrome p450 homepage. *Hum Genomics,* 4**,** 59-65.

NELSON, D. R. 2018. Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins Proteom,* 1866**,** 141-154.

NELSON, D. R., KAMATAKI, T., WAXMAN, D. J., GUENGERICH, F. P., ESTABROOK, R. W., FEYEREISEN, R., GONZALEZ, F. J., COON, M. J., GUNSALUS, I. C., GOTOH, O. & ET AL. 1993. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol,* 12**,** 1-51.

NGCOBO, N. S., CHILIZA, Z. E., CHEN, W., YU, J.-H., NELSON, D. R., TUSZYNSKI, J. A., PRETO, J. & SYED, K. 2020. Comparative Analysis, Structural Insights, and Substrate/Drug Interaction of CYP128A1 in *Mycobacterium tuberculosis*. *International Journal of Molecular Sciences,* 21.

NGUYEN, R. C., YANG, Y., WANG, Y., DAVIS, I. & LIU, A. 2019. Substrate-assisted hydroxylation and O-demethylation in the peroxidase-like cytochrome P450 enzyme CYP121. *ACS catalysis,* 10**,** 1628-1639.

NKOSI, B. V. Z., PADAYACHEE, T., GRONT, D., NELSON, D. R. & SYED, K. 2022. Contrasting Health Effects of *Bacteroidetes* and *Firmicutes* Lies in Their Genomes: Analysis of P450s, Ferredoxins, and Secondary Metabolite Clusters. *International journal of molecular sciences,* 23**,** 5057.

NZUZA, N., PADAYACHEE, T., SYED, P. R., KRYŚ, J. D., CHEN, W., GRONT, D., NELSON, D. R. & SYED, K. 2021. Ancient Bacterial Class *Alphaproteobacteria* Cytochrome P450 Monooxygenases Can Be Found in Other Bacterial Species. *International journal of molecular sciences,* 22**,** 5542.

OMURA, T. 2011. Recollection of the early years of the research on cytochrome P450. *Proc Jpn Acad Ser B Phys Biol Sci,* 87**,** 617-40.

OMURA, T. & SATO, R. 1962. A new cytochrome in liver microsomes. *J Biol Chem,* 237**,** 1375-6.

ORTEGA UGALDE, S., DE KONING, C. P., WALLRAVEN, K., BRUYNEEL, B., VERMEULEN, N. P. E., GROSSMANN, T. N., BITTER, W., COMMANDEUR, J. N. M. &

VOS, J. C. 2018. Linking cytochrome P450 enzymes from *Mycobacterium tuberculosis* to their cognate ferredoxin partners. *Appl Microbiol Biotechnol,* 102**,** 9231-9242.

ORTIZ DE MONTELLANO, P. R. 2018. Potential drug targets in the *Mycobacterium tuberculosis* cytochrome P450 system. *J Inorg Biochem,* 180**,** 235-245.

PADAYACHEE, T., NZUZA, N., CHEN, W., NELSON, D. R. & SYED, K. 2020. impact of lifestyle on cytochrome P450 monooxygenase repertoire is clearly evident in the bacterial phylum *Firmicutes*. *Scientific reports,* 10**,** 1-12.

PARVEZ, M. 2018. *Cytochrome P450 monooxygenase analysis in the genus mycobacterium: special focus on CYP123/by Mohammad Parvez.* Bloemfontein: Central University of Technology, Free State.

PARVEZ, M., QHANYA, L. B., MTHAKATHI, N. T., KGOSIEMANG, I. K., BAMAL, H. D., PAGADALA, N. S., XIE, T., YANG, H., CHEN, H., THERON, C. W., MONYAKI, R., RASELEMANE, S. C., SALEWE, V., MONGALE, B. L., MATOWANE, R. G., ABDALLA, S. M., BOOI, W. I., VAN WYK, M., OLIVIER, D., BOUCHER, C. E., NELSON, D. R., TUSZYNSKI, J. A., BLACKBURN, J. M., YU, J. H., MASHELE, S. S., CHEN, W. & SYED, K. 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Sci Rep,* 6**,** 33099.

PODUST, L. M. & SHERMAN, D. H. 2012. Diversity of P450 enzymes in the biosynthesis of natural products. *Nat Prod Rep,* 29**,** 1251-66.

QUADRI, L. E. 2014. Biosynthesis of mycobacterial lipids by polyketide synthases and beyond. *Crit Rev Biochem Mol Biol,* 49**,** 179-211.

RAJPUT, S., MCLEAN, K. J., PODDAR, H., SELVAM, I. R., NAGALINGAM, G., TRICCAS, J. A., LEVY, C. W., MUNRO, A. W. & HUTTON, C. A. 2019. Structure–activity relationships of cyclo (L-tyrosyl-L-tyrosine) derivatives binding to *Mycobacterium tuberculosis* CYP121: iodinated analogues promote shift to high-spin adduct. *Journal of medicinal chemistry,* 62**,** 9792-9805.

RAO, S. P., ALONSO, S., RAND, L., DICK, T. & PETHE, K. 2008. The protonmotive force is required for maintaining ATP homeostasis and viability of hypoxic, nonreplicating *Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences,* 105**,** 11945-11950.

RIPOLL, F., PASEK, S., SCHENOWITZ, C., DOSSAT, C., BARBE, V., ROTTMAN, M., MACHERAS, E., HEYM, B., HERRMANN, J.-L. & DAFFÉ, M. 2009. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. *PloS one,* 4.

RUDOLF, J. D., CHANG, C. Y., MA, M. & SHEN, B. 2017. Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function. *Nat Prod Rep,* 34**,** 1141-1172.

SALITURO, G. M., ZINK, D. L., DAHL, A., NIELSEN, J., WU, E., HUANG, L., KASTNER, C. & DUMONT, F. J. 1995. Meridamycin: a novel nonimmunosuppressive FKBP12 ligand from *Streptomyces hygroscopicus*. *Tetrahedron letters,* 36**,** 997-1000.

SARFO, F. S., PHILLIPS, R., WANSBROUGH-JONES, M. & SIMMONDS, R. E. 2016. Recent advances: role of mycolactone in the pathogenesis and monitoring of *Mycobacterium ulcerans* infection/Buruli ulcer disease. *Cellular microbiology,* 18**,** 17-29.

SENATE, L. M., TJATJI, M. P., PILLAY, K., CHEN, W., ZONDO, N. M., SYED, P. R., MNGUNI, F. C., CHILIZA, Z. E., BAMAL, H. D., KARPOORMATH, R., KHOZA, T., MASHELE, S. S., BLACKBURN, J. M., YU, J. H., NELSON, D. R. & SYED, K. 2019. Similarities, variations, and evolution of cytochrome P450s in *Streptomyces* versus *Mycobacterium*. *Sci Rep,* 9**,** 3962.

SOGI, K. M., HOLSCLAW, C. M., FRAGIADAKIS, G. K., NOMURA, D. K., LEARY, J. A. & BERTOZZI, C. R. 2016. Biosynthesis and Regulation of Sulfomenaquinone, a Metabolite Associated with Virulence in *Mycobacterium tuberculosis*. *ACS Infect Dis,* 2**,** 800-806.

SONO, M., ROACH, M. P., COULTER, E. D. & DAWSON, J. H. 1996. Heme-containing oxygenases. *Chemical reviews,* 96**,** 2841-2888.

STINEAR, T. P., SEEMANN, T., HARRISON, P. F., JENKIN, G. A., DAVIES, J. K., JOHNSON, P. D., ABDELLAH, Z., ARROWSMITH, C., CHILLINGWORTH, T. & CHURCHER, C. 2008. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome research,* 18**,** 729-741.

STINEAR, T. P., SEEMANN, T., PIDOT, S., FRIGUI, W., REYSSET, G., GARNIER, T., MEURICE, G., SIMON, D., BOUCHIER, C. & MA, L. 2007. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome research,* 17**,** 192-200.

SYED, K. & MASHELE, S. S. 2014. Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. *PLoS One,* 9**,** e95616.

SYED, P. R., CHEN, W., NELSON, D. R., KAPPO, A. P., YU, J. H., KARPOORMATH, R. & SYED, K. 2019. Cytochrome P450 Monooxygenase CYP139 Family Involved in the Synthesis of Secondary Metabolites in 824 Mycobacterial Species. *Int J Mol Sci,* 20.

TORTOLI, E. 2012. Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infection, Genetics and Evolution,* 12**,** 827-831.

UEDA, J.-Y., HASHIMOTO, J., YAMAMURA, H., HAYAKAWA, M., TAKAGI, M. & SHIN-YA, K. 2010. A new 16-membered tetraene macrolide JBIR-100 from a newly identified *Streptomyces* species. *The Journal of antibiotics,* 63**,** 627-629.

URLACHER, V. B. & GIRHARD, M. 2019. Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology. *Trends Biotechnol,* 37**,** 882-897.

VAN BEILEN, J. B., HOLTACKERS, R., LÜSCHER, D., BAUER, U., WITHOLT, B. & DUETZ, W. A. 2005. Biocatalytic production of perillyl alcohol from limonene by using a novel *Mycobacterium* sp. cytochrome P450 alkane hydroxylase expressed in *Pseudomonas putida*. *Applied and environmental microbiology,* 71, 1737-1744.

VAN DER GEIZE, R., YAM, K., HEUSER, T., WILBRINK, M. H., HARA, H., ANDERTON, M. C., SIM, E., DIJKHUIZEN, L., DAVIES, J. E. & MOHN, W. W. 2007. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. Proceedings of the National Academy of Sciences, 104, 1947-1952.

VAN WYK, R., VAN WYK, M., MASHELE, S. S., NELSON, D. R. & SYED, K. 2019. Comprehensive comparative analysis of cholesterol catabolic genes/proteins in mycobacterial species. International Journal of Molecular Sciences, 20, 1032.

VASCONCELLOS, S. E. G., HUARD, R. C., NIEMANN, S., KREMER, K., SANTOS, A. R., SUFFYS, P. N. & HO, J. L. 2010. Distinct genotypic profiles of the two major clades of *Mycobacterium africanum*. BMC infectious diseases, 10, 80.

VENTURA, M., CANCHAYA, C., TAUCH, A., CHANDRA, G., FITZGERALD, G. F., CHATER, K. F. & VAN SINDEREN, D. 2007. Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev, 71, 495-548.

VISSA, V. D. & BRENNAN, P. J. 2001. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. Genome biology, 2, reviews1023. 1.

WAWROCKI, S. & DRUSZCZYNSKA, M. 2017. Inflammasomes in *Mycobacterium tuberculosis*-driven immunity. Canadian Journal of Infectious Diseases and Medical Microbiology, 2017.

WHITE, R. E. & COON, M. J. 1980. Oxygen activation by cytochrome P-450. Annu Rev Biochem, 49, 315-56.

WU, C.-W., GLASNER, J., COLLINS, M., NASER, S. & TALAAT, A. M. 2006. Whole-genome plasticity among *Mycobacterium avium* subspecies: insights from comparative genomic hybridizations. Journal of bacteriology, 188, 711-723.

WYNNE, J. W., SEEMANN, T., BULACH, D. M., COUTTS, S. A., TALAAT, A. M. & MICHALSKI, W. P. 2010. Resequencing the *Mycobacterium avium* subsp. *paratuberculosis* K10 genome: improved annotation and revised genome sequence. Journal of bacteriology, 192, 6319-6320.

YAMAZAKI, H. 2014. Fifty Years of Cytochrome P450 Research, Japan, Springer Japan.

YAN, Y., WU, J., HU, G., GAO, C., GUO, L., CHEN, X., LIU, L. & SONG, W. 2022. Current state and future perspectives of cytochrome P450 enzymes for C–H and C= C oxygenation. Synthetic and Systems Biotechnology.

YANO, E. M., GOLDZWEIG, C., CANELO, I. & WASHINGTON, D. L. 2006. Diffusion of innovation in women's health care delivery: The Department of Veterans Affairs' adoption of women's health clinics. Women's Health Issues, 16, 226-235.

YOU, Q., VERSCHOOR, C. P., PANT, S. D., MACRI, J., KIRBY, G. M. & KARROW, N. A. 2012. Proteomic analysis of plasma from Holstein cows testing positive for *Mycobacterium avium* subsp. *paratuberculosis* (MAP). Veterinary immunology and immunopathology, 148, 243-251.

ZINK, A. R., SOLA, C., REISCHL, U., GRABNER, W., RASTOGI, N., WOLF, H. & NERLICH, A. G. 2003. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. Journal of clinical microbiology, 41, 359-367.