Genome-wide analysis and genome mapping of essential cytochrome P450 monooxygenase CYP125 in mycobacteria

By

Richie Monyaki

Dissertation submitted in fulfilment of the requirements of the degree

Master of Health Sciences in Biomedical Technology

Department of Health Sciences

Central University of Technology, Free State

08 November 2016

Promoter: Prof. Samson Sitheni Mashele

Co-Promoter: Prof. Khajamohiddin Syed

DECLARATION

I, **RICHIE MONYAKI**, (South African ID number: [redacted]) hereby certify that the dissertation submitted by me for the degree MASTER OF HEALTH SCIENCES IN BIOMEDICAL TECHNOLOGY, is my own independent work; and complies with the Code of Academic Integrity, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology (Free State). I hereby declare, that this research project has not been previously submitted before to any university or faculty for the attainment of any qualification. I further waive copyright of the dissertation in favour of the Central University of Technology (Free State).

Furthermore, I declare that some of the contents used in the thesis are my own work published in Nature publishing group journal "Scientific Reports" (as listed in Abstract of the thesis) where I serve as a co-author.

MONYAKI RICHIE **DATE**

Dedication

This thesis is dedicated to my

Father

Tshehla Paulus Monyaki

&

Mother

Dieketseng Elizabeth Monyaki

ACKNOWLEDGEMENTS

I pass my sincere gratitude to the following individuals for the success of my study to:

- My supervisor Prof. Samson Setheni Mashele for guidance and his un-measurable support.
- My Co-supervisor Prof. Khajamohiddin Syed, for guidance and full support in every aspect. I could not have made it without him.
- Research and Innovation Fund, Central University of Technology, Free State for funding my studies.
- National Research Foundation scholarship for funding my studies.
- Mrs Dieketseng Monyaki, My Mother, for the unmeasurable love, support, guidance throughout my life and also through the duration of this study.
- My late Father Mr Tshehla Paulus Monyaki for showing me the potential I have and teaching me on how to tackle the difficulties in life and coming out the as the better person throughout every situation. This thesis is dedicated to you as you couldn't be around to see my success. You are always in my heart.
- My brother Buza Monyaki and my relatives for the support and trust in me.
- My friends, I am so grateful for your love and support.
- My colleagues at the Faculty of Health and Environmental Sciences, Department of Biomedical Technology at the Central University of Technology especially my research team for their overwhelming friendship and motivation.
- To my Creator for my existence, protection, guidance and blessings I am provided with.

"The function of education is to teach one to think intensively and to think critically. Intelligence plus character - that is the goal of true education."

Martin Luther King, Jr

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS

%	Percentage
>	Greater than
=	Equals to
C-H	Carbon–hydrogen
CYPs/P450s	Cytochrome P450 monooxygenases
CYPED	Cytochrome P450 Engineering Database
et al.,	*Et alia* (and others)
H_2O	Water
H^+	Hydrogen ion
HIV	Human Immune Virus
KEGG	Kyoto Encyclopedia of Genes and Genomes
M. avium	*Mycobacterium avium*
MAC	*Mycobacterium avium* complex
MCAC	*Mycobacterium chelonae-abscessus* complex
MDR	Multidrug-resistant
MEGA	Molecular Evolutionary Genetics Analysis
M. tuberculosis	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
NADP	Nicotinamide adenine dinucleotide phosphate
$NADP^+$	Reduced nicotinamide adenine dinucleotide phosphate

NCBI CDD	National Center for Biotechnology Information Conserved Domain Database
NS	New Subfamily
NTM	Nontuberculous mycobacteria
N-terminal	Amino terminal end
nm	Nanometre
O_2	Oxygen
SAP	Saprophytes
TB	Tuberculosis
TCE	Trichloroethylene
TRaSH	Transposon-site-hybridization-mutagenesis
WHO	World Health Organization
XDR	Extensively drug-resistant
USA	United States of America

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Abstract

Tuberculosis, an infectious lung disease, is a leading cause of death worldwide caused by *Mycobacterium tuberculosis*. Genome-wide screening for genes essential for the survival of *M. tuberculosis* has revealed that cytochrome P450 monooxygenase CYP125A1 is critical for *M. tuberculosis* survival. CYP125A1 play key role in oxidation of cholesterol and help *M. tuberculosis* to utilize cholesterol as a carbon source during its inhabitant in host organism. Despite this great importance, to date, genome wide identification, annotation and phylogenetic analysis of CYP125A1 and its genome mapping with respect to gene-cluster analysis across mycobacterial species has not been performed. Also, to date, P450s from prokaryote organisms has not been subjected to evolutionary analysis. This study addresses these two research gaps.

Genome data-mining and annotation of CYP125 P450s across 60 mycobacterial species revealed presence of a total number of 120 CYP125 P450s that can be grouped into five subfamilies (A, D, E, F, NS). Analysis of CYP125 P450s distribution in different mycobacterial categories revealed that *Mycobacterium tuberculosis* complex (MTBC) species showed lowest copies of CYP125 in their genomes compared to other categories. This study revealed that CYP125 P450 is not present in Mycobacteria causing leprosy (MCL) species. *Mycobacterium avium* complex (MAC) and Saprophytes (SAP) species showed highest number of CYP125 subfamilies in their genomes. Analysis of CYP125 P450 subfamily patterns in mycobacterial categories revealed MAC species have highest diversity of CYP125 subfamilies followed by species belong to SAP and Nontuberculous mycobacteria (NTM). Presence of more than one copy of CYP125 in some mycobacterial categories suggests important role of this P450s in their physiology. Analysis of subfamily patterns in mycobacterial categories revealed MAC species have highest diversity of CYP125

subfamilies followed by species belong to SAP and NTM. Among subfamilies, subfamily A is more dominant across mycobacterial species. Subfamily D is present only in species belonging to MAC and NTM. Subfamily E is present only in species belonging to MAC. Subfamily F and NS is present in MAC and SAP.

Analysis of CYP125 gene clusters in the genus *Mycobacterium* revealed presence of 28 CYP125 gene-clusters. Gene clusters 1 to 20 comprised of quite a number of CYP125 P450s ranging from 2 to 23 and gene clusters 21 – 28 named as unique gene clusters considering each of the CYP125 P450 in this cluster have different genes both in the upstream and downstream of CYP125. Overall, SAP species showed highest CYP125 gene cluster diversity (10 clusters including 1 unique cluster) followed by MAC (8 clusters including 3 unique clusters), NTM (5 clusters including 3 unique clusters), MCAC (4 clusters) and MTBC (2 clusters including one unique cluster). This study is first of its kind on analysis of gene-clusters in prokaryote P450s. Some of the CYP125 P450s in different clusters have reverse complement arrangement of genes compared to other CYP125s in the same cluster. These P450s are under investigation for further analysis of possible gene rearrangement events in the chromosome.

Results generated in this study on genome data mining, identification, annotation and phylogenetic analysis of CYP125 is published as part of major article on mycobacterial P450s where I am a co-author. The article details are:

R Monyaki (co-author) (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. Scientific Reports 6, Article number: 33099.

Apart from my Masters study, I supervised two B. Tech student projects and also worked on a few other bioinformatics projects and earned co-authorship in high impact factor journal listed below:

R Monyaki (co-author) (2015) Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. Scientific Reports 5, Article number: 11572.

Article as first author on CYP125 gene-cluster analysis under preparation for submission to Plos ONE.

In addition to the above credits, I was featured on national TV and in newspapers for discovering a novel drug target. I also presented work at both national and international (Canada) conferences.

Chapter 2

Introduction and Literature review

2.1. Introduction to P450s

Cytochrome P450 monooxygenases (CYPs/P450s) are mixed function oxidoreductases ubiquitously distributed across all biological kingdoms (Nelson, 2013). P450s are well known for their role in essential cellular anabolic and catabolic processes. P450s are mainly characterized by their ability to absorb light at the 450nm peak by the heme cofactor and also their potential to oxidize a wide range of metabolic substrates and xenobiotic compounds. Independent studies carried out by Klingenberg (1958) and Garfinkel (1958) resulted in discovery of a carbon monoxide-binding pigment with a unique absorption maximum at 450 nm in rat and pig liver microsomes. P450s participate in a large number of primary, secondary and xenobiotic metabolic reactions. P450s are found to be involved in the production of metabolites critical for specific processes such as pathogenesis, the utilization of specific substrates, and/or the detoxification of xenobiotics (described in Ortiz de Montellano, 2015). This phenomenon has resulted in the interweaving of the evolution of the P450s with organism adaptation to several ecological niches (Syed *et al*., 2014a). Their ability to neutralize toxic chemical substrates and also synthesize metabolites has brought forward the hypothesis that they may have evolved through the chemical warfare waged among plants, animals, insects, and microbes (Gonzalez and Nebert, 1990; Lewis *et al*., 1998).

Upon the breakthrough of the discovery of P450s, studies on these enzymes have sparked interest among chemists, biochemists and biotechnologists worldwide. The work done on these P450's during the last 55 years has shown great advancements not only from the perspective of proceeding essential understanding but also at looking at the industrial

perspectives. Their applications in the synthesis of oxyfunctionalized building blocks closely linked with the retrieval of new important compounds in demand (such as specialty chemicals and pharmaceutical synthons) are of immense importance (described in Syed and Yadav, 2012). Moreover, P450s have a great potential for the development of biosensors, as well as in bioremediation (described in Syed and Yadav, 2012).

The most common reaction catalysed by P450s is a monooxygenase reaction i.e., insertion of one atom of oxygen into the aliphatic position of an organic substrate (RH) while the other oxygen atom is reduced to water as shown below:

$$\mathbf{RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+}$$

In addition to the typical monooxygenase reaction, P450s also perform catalytically diverse reactions as shown in Figure 2.1.

Hydrocarbon hydroxylation
Alkene epoxidation
Alkyne oxygenation
Arene epoxidation
Aromatic hydroxylation
N-Dealkylation
S- Dealkylation
O- Dealkylation
N-Hydroxylation
N-Oxidation
S-Oxidation
Oxidative deamination
Oxidative dehalogenation
Alcohol and aldehyde oxidations
Dehydrogenation
Dehydratations
Reductive dehalogenation
N-Oxide reduction
Epoxide reduction
Reductive β-scission of alkyl peroxides
NO reduction
Isomerizations
Oxidative C-C bond cleavage

Figure 2.1. Reactions catalysed by P450s (taken from Sono *et al*., 1996; Bernhardt, 2006)

2.2. P450 nomenclature

P450s usually represent with code name such as "CYP1A2" where "CY" stands for cytochrome meaning hemoprotein; "P" is an abbreviation for pigment which absorbs light at 450 nm; the number that follow the "P" i.e. 1, represent the family; the letter "A" represent the subfamily and the number "2" represents the position in the subfamily (Nebert *et al*., 1987 and 1991; Nelson, 1999).

The great diversification of P450's amongst living life forms such as insects, bacteria and fungi has brought forward a dire need to provide new gradations of nomenclature which are above the family/subfamily level. A new concept of clans has thus been introduced as a new level above the family rank (Nelson, 1998). The first nomenclature naming system was based on 100 P450 families applying a three digit system but owing to the growing nature of these P450's a new numbering scheme was introduced as a four digit setup (Nelson, 2009). The initial ranges of the three digit system were as follows: Bacteria (CYP101-299); Animals (CYP301-499); lower eukaryotes (CYP501-699) and plants (CYP701-999).The new ranges were then introduced and they arranged as follows: Bacteria (CYP1001-2999); animals (CYP3001-4999); lower eukaryotes (CYP5001-6999); and plants (CYP7001-9999) (Nelson, 2006).

The P450s were named based on the percentage identity where >40% identity and >55% identity were grouped under the same family and subfamily, respectively (Nebert *et al*., 1987; Nebert *et al*., 1991; Nelson, 2009). This criterion for naming of P450s is set by the International P450 Nomenclature Committee. Some of the P450s that have nearly 40% identity to the named P450s were classified based on their phylogenetic position (Nebert *et al*., 1987; Nebert *et al*., 1991; Nelson, 2009). P450s that have below 40% identity will be assigned to a new family with the help of Prof David Nelson, the University of Tennessee Health Science Center, Tennessee, USA for naming the new P450.

2.3. Applications of P450s

P450s perform catalytically diverse enzymatic reactions with stereo- and regio-specific manner (Figure 2.1). Due to these unique characteristics, P450s are exploited for various applications.

Looking at the field of biotransformation, the selective biocatalytic oxyfunctionalization of nonactivated hydrocarbons is considered as potentially the most useful (Myles and Whitesides, 1990). P450s contain heme B as prosthetic group that enables not only the activation of molecular oxygen (which is also possible by using flavin-containing enzymes) but also the oxidation of kinetically inert non-activated C–H bonds. Industrial applications of P450s have so far been restricted to whole-cell systems, which mostly solve the problem of cofactor delivery and regeneration. In such instances, physiological effects such as limited substrate uptake and reduced efflux of products out of cells, substrate or product toxicity, product degradation, as well as elaborate downstream processing are additional limiting factors that must be taken into account and often require optimization (van Beilen, 2003).

2.3.1. Production of drug metabolites

P450s have been identified to play a crucial role in the field of drug transformation (Figure 2.2). They are generally accountable for the primary oxidation of xenobiotics. In humans, P450s are found to be one of the essential enzymes in phase I drug metabolism reactions. Among the 57 P450s isoenzymes that have been expressed in human, special focus was given to certain P450s namely: CYP1A2, 2C9, 2C19, 2D6, and 3A4 as these P450s were identified to mediate close to 75-80% of the drug metabolism (Evans and Relling, 1999; Guengerich, 2003 and 2015). It is very important to carry out a thorough research on the molecular properties of drug metabolites as this plays an essential role in the evaluation of drug toxicity, drug–drug interaction and drug-induced side effects. Since drug metabolite standards are in

most cases not available or difficult to synthesize by chemical means, P450s are the most important enzymes for the biotransformation of drugs and the preparation of metabolites.

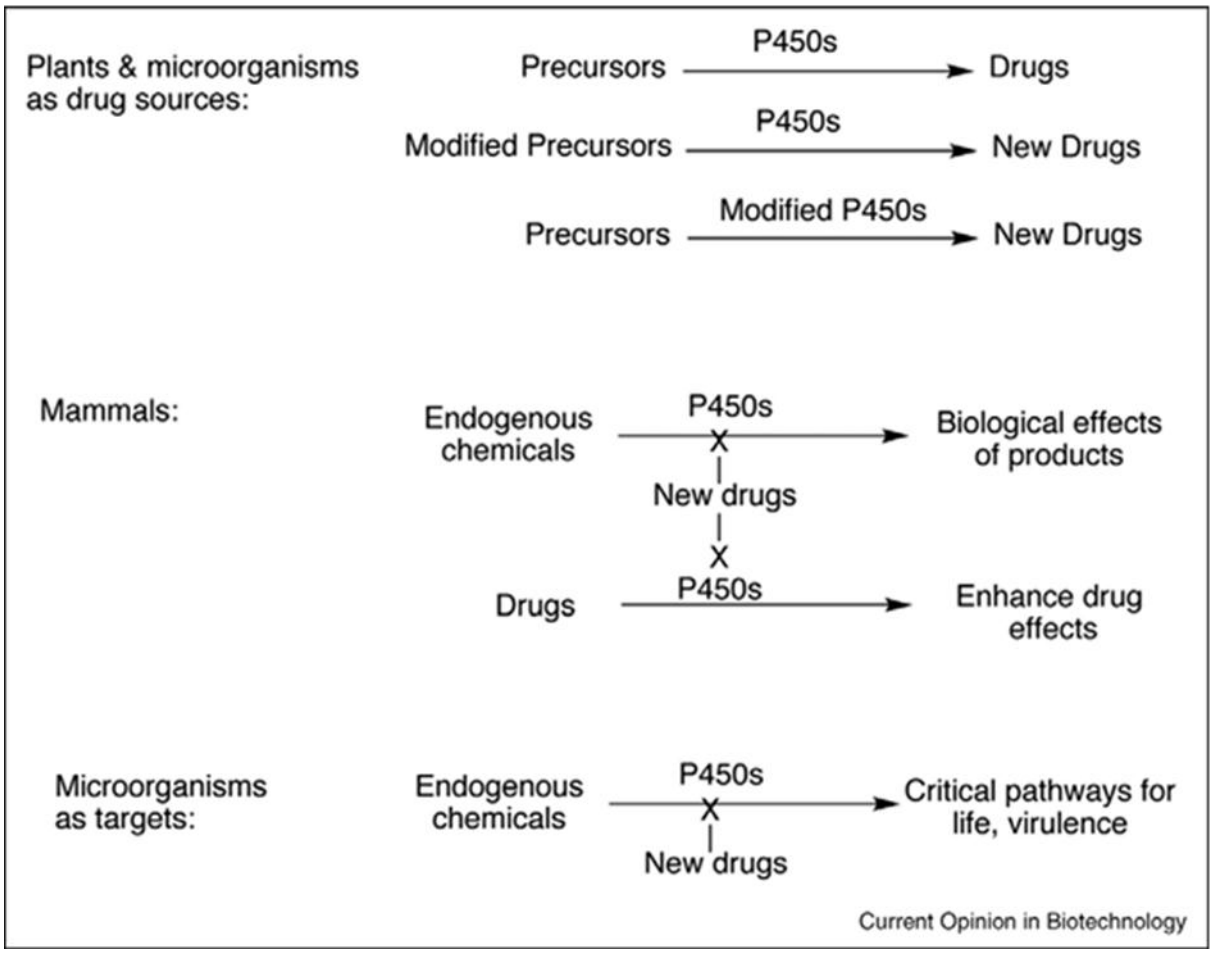


Figure 2.2. Cytochrome P450 monooxygenases role as drug targets and new drug discovery processes (taken from Lamb *et al*., 2007). This is the summary of roles played by P450s in various applications that have been accomplished that include drug synthesis and drug targets.

2.3.2. Phytoremediation

Phytoremediation is the generic term used for the group of technologies that employ the direct use of plants to clean up environmental pollution. This process sometimes encounters problems like slow rate of removal or incomplete metabolism, so in order to counteract these limitations, new enzymatic activities has to be introduced within plants by the use of genetic engineering. During some cases of genetic engineering, some of the mammalian or bacterial P450s has been expressed in certain plants to remediate polluted soil, groundwater and air (Reichenauer and Germida, 2008). An example of this phenomenon was observed when the expression of the human CYP2E1 in hydroponically grown tobacco enhanced the metabolism of the volatile hydrocarbon trichloroethylene (TCE) up to 640-fold. The oxidation product 2,2,2-trichloroacetaldehyde (chloral) generated by CYP2E1 is further metabolized in the plant to the corresponding alcohol (Figure 2.3) (Doty *et al*., 2000).

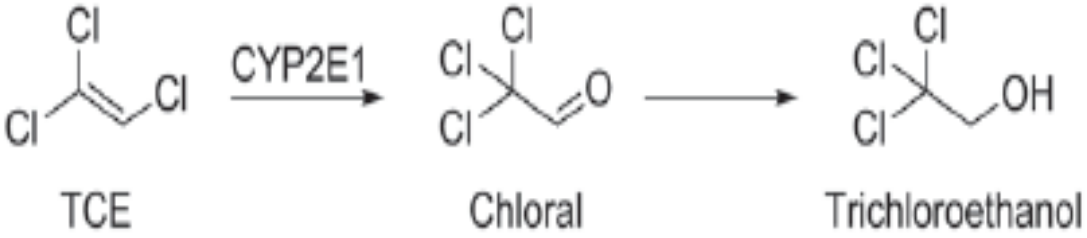


Figure 2.3. CYP2E1 mediated transformation for trichloroethylene (TCE) (taken from Girhard *et al*., 2015).

Phytoremediation has also been found to be used in the extraction of herbicides and in this case it has been found that P450s expressed by transgenic plants enhance the process. Human CYP1A1 expressed in *Oryza sativa* (rice) either by conjunction or separately with CYP2B6 and CYP2C19, has been found to display a high resistance towards a wide range of herbicides which includes metolachlor, norflurazon and mixtures thereof (Kawahigashi *et al*., 2005 and 2006).

2.3.3. Selective oxidations of macrolide antibiotics

The macrolides are a class of natural products that belong to the class of polyketides. These set of macromolecules have been found to have antibiotic or antifungal activity and can be used as pharmaceutical drugs. Their core structure is synthesized by polyketide synthases based on general precursor molecules and then further diversified by other P450-catalyzed hydroxylation and epoxidation activities (Zhan, 2009).

Thirty five percent of all marketed antibiotic formulations contain an active ingredient derived from *Actinomycetes*. Since most antibiotics are semisynthetic derivatives of a few natural products, *Actinomycetes* produce an impressive 76% of all original natural product scaffolds used as anti-infective agents (Gomez-Escribano and Bibb, 2014). Therefore, the "deorphanization" of *Actinomycetes* P450s is considered quite important for pharmacology, with ramifications for the use of clinical therapeutics (Lamb *et al*., 2013; Guengerich, 2015). A well-characterized P450 involved in ring decoration of macrolide antibiotics is PikC from *Streptomyces venezuelae* catalyzing regioselective C12-hydroxylation of narbomycin—the final step in pikromycin biosynthesis (Figure 2.4).

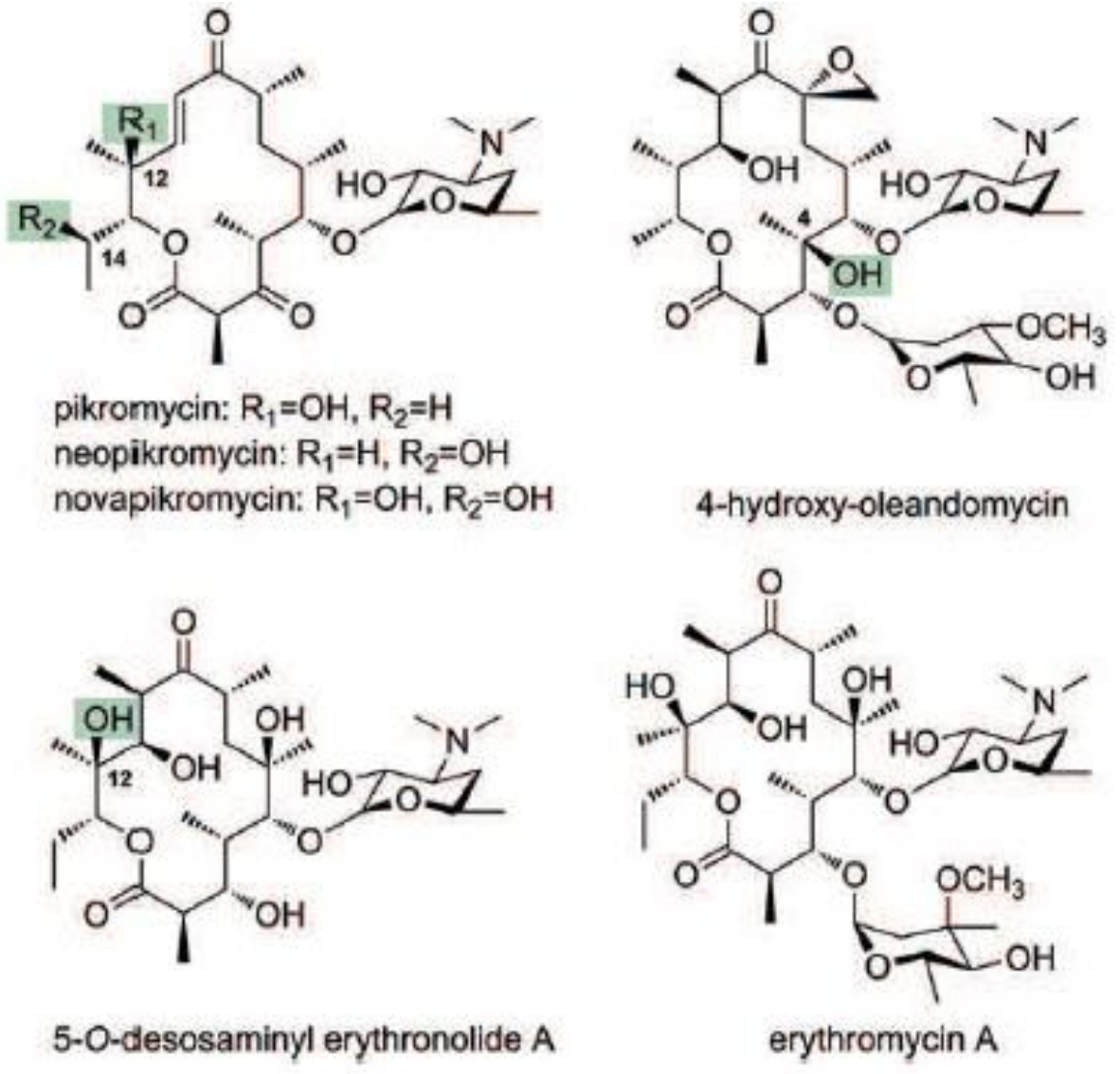


Figure 2.4. Macrolide antibiotics originating from erythromycin A and their hydroxylated derivatives produced by P450 PikC (taken from Girhard *et al*., 2015)

2.4. Genome data-mining for P450s

P450's are ubiquitously distributed in species belong to different biological kingdoms (Nelson, 2013). Genome data mining plays an important role in identification of P450s with novel oxidation activities (Furuya and Kino, 2010). The recent advancements in technologies throughout the years have unravelled new strategies that have led to the advancements of alternative screening strategies that are aimed to overcome the obstacles of modern microbial

screenings. One example is the screening of the metagenomic libraries of non-culturable microorganisms (Schmeisser *et al*., 2007). The *in silico* screening of annotated P450 sequences from different publicly available online databases has been considered as one of the most promising strategies to date. The number of annotated sequences is increasing by a large margin due to the large number of genome sequencing projects. Chronological events throughout the years since P450s were first discovered in 1958 (Klingenberg, 1958; Garfinkel, 1958) up to the 2013, has shown an exponential growth in number of P450s (from 1000 to >21000). The data on P450s annotation has shown less than 1000 P450 sequences between 1958 and 1998 (Nelson, 1999), then the number approached 4000 in 2004 (Nelson, 2006), 18,000 in 2011 (Nelson, 2011), and increased over 21,000 in 2013 (Nelson, 2013).

The P450s sequences and their annotation details are kept at the "official" P450 database also known as "the Cytochrome P450 Homepage" that is maintained by Prof David Nelson, University of Tennessee, Tennessee, USA (http://drnelson.uthsc.edu/CytochromeP450.html) (Nelson, 2009). This database provides information on classification of the currently known P450s. This database contains classification of 18 937 P450 genes, *inter alia* including bacteria with 1254 genes, fungi with 5729 genes, plants with 7446 genes, insects with 3452 genes, and mammals with 1056 genes (status as of August 2013).

Cytochrome P450 Engineering Database (CYPED; http://www.cyped.uni-stuttgart.de ; Universität Stuttgart; 2014/03/20) is another P450 database where structure based information is available (Fischer *et al*., 2007; Sirim *et al*., 2009). CYPED includes more than 16,000 sequences of P450s. In addition, information on 741 structures of P450s is integrated into this database to facilitate protein engineering.

The growing nature in which new P450 sequences are being identified has made it difficult to keep up with the characterization of their biochemical properties. To this point

only a few number of currently annotated P450 sequences have been functionally characterized. Nevertheless, reports on the biotechnological exploitation of naturally occurring and highly selective oxidations by P450 enzymes are accumulating and has been reviewed by Schulz and co-worker's (2012).

The rapid influx of genome sequences calls for robust computational tools that can effectively support large-scale comparative analyses of genomes and specific gene families. The basis of P450 mining in diverse species has been carried out by following the methodology that has been described meticulously in several published literature (Park *et al*., 2008; Syed *et al*., 2013; Syed and Mashele, 2014; Syed *et al*., 2014a; Syed *et al*., 2014b; Jawallapersand *et al*., 2014; Kgosiemang *et al*., 2014; Sello *et al*., 2015; Mthakathi *et al*., 2015; Qhanya *et al*., 2015; Parvez *et al*., 2016). Park and co-workers (2008) developed an automated pipeline system for genome data mining and subsequence naming of P450s (Figure 2.5). Syed and co-workers developed a simple procedure for genome data-mining and subsequence annotation of P450s (Syed *et al*., 2013; Syed and Mashele, 2014;Syed *et al*., 2014a; Syed *et al*., 2014b; Jawallapersand *et al*., 2014; Kgosiemang *et al*., 2014; Sello *et al*., 2015; Mthakathi *et al*., 2015; Qhanya *et al*., 2015; Parvez *et al*., 2016). According to the authors the genome data mining and annotation of P450s follows as: briefly, the whole proteome of species will be subjected to the NCBI Batch Web CD-Search Tool (http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). NCBI Batch Web CD-Search Tool to identify P450s based on conserved motifs that include the oxygen-binding motif (FXXGXXXCXG) and the heme-binding motif (EXXR). Proteins that belong to a P450 superfamily were selected and further subjected to BLAST analysis against bacterial P450s at the Cytochrome P450 Homepage (Nelson, 2009). Based on the International P450 Nomenclature Committee rule, proteins with >40% identity and >55% identity were grouped under the same family and subfamily, respectively. P450s that showed less than 40% identity

to known P450s at the Cytochrome P450 Homepage (Nelson, 2009) were assigned to new P450 families and subfamilies as per International P450 Nomenclature Committee rules.

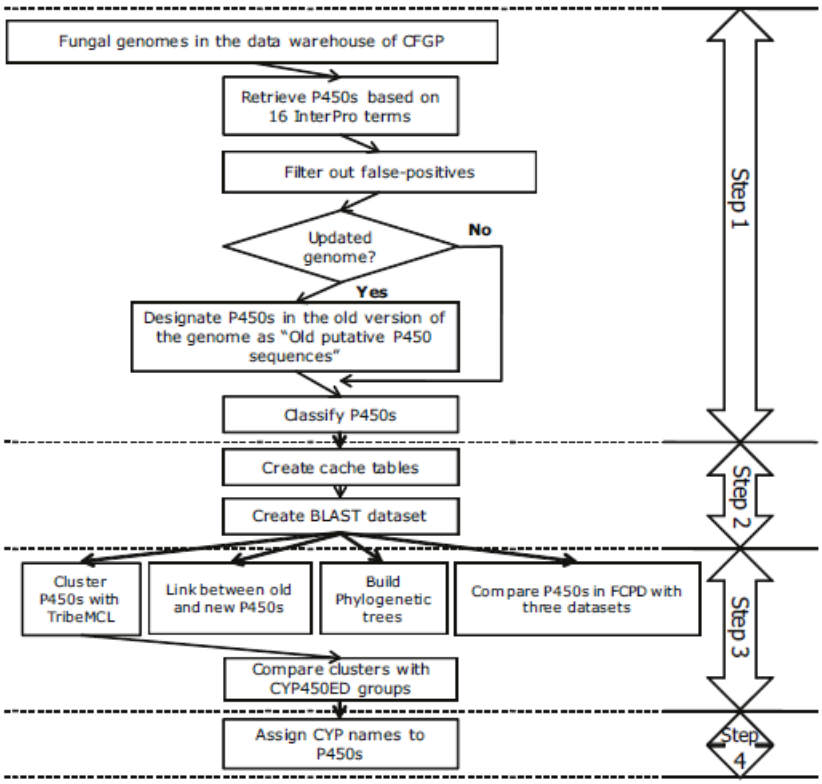


Figure 2.5. Protocol for automated identification and classification of P450s (taken from Park *et al*., 2008).

2.5. Phylogenetic analysis of P450s

Phylogenetic analysis of P450s is very critical to assign P450s to correct families and subfamilies and also to group them into higher order such as clans. Although different methods employed for phylogenetic analysis, overall, analysis includes identifying the P450s position with other named P450s to double check the naming accuracy. Sometimes, P450s that showed close to 40% identity will be named based on their position on phylogenetic tree. Apart from this, the phylogenetic trees were extensively utilized to group P450s into higher order such as clans or clades (Chen *et al*., 2014; Sello *et al*., 2015; Parvez *et al*., 2016).

2.6. Synteny analysis and genome mapping of P450s

The physical localization of P450 genes on the chromosomes were known as Synteny analysis (Figure 2.6). As shown in Figure 2.6, P450s in an organism will be mapped on their respective chromosomes with their position.

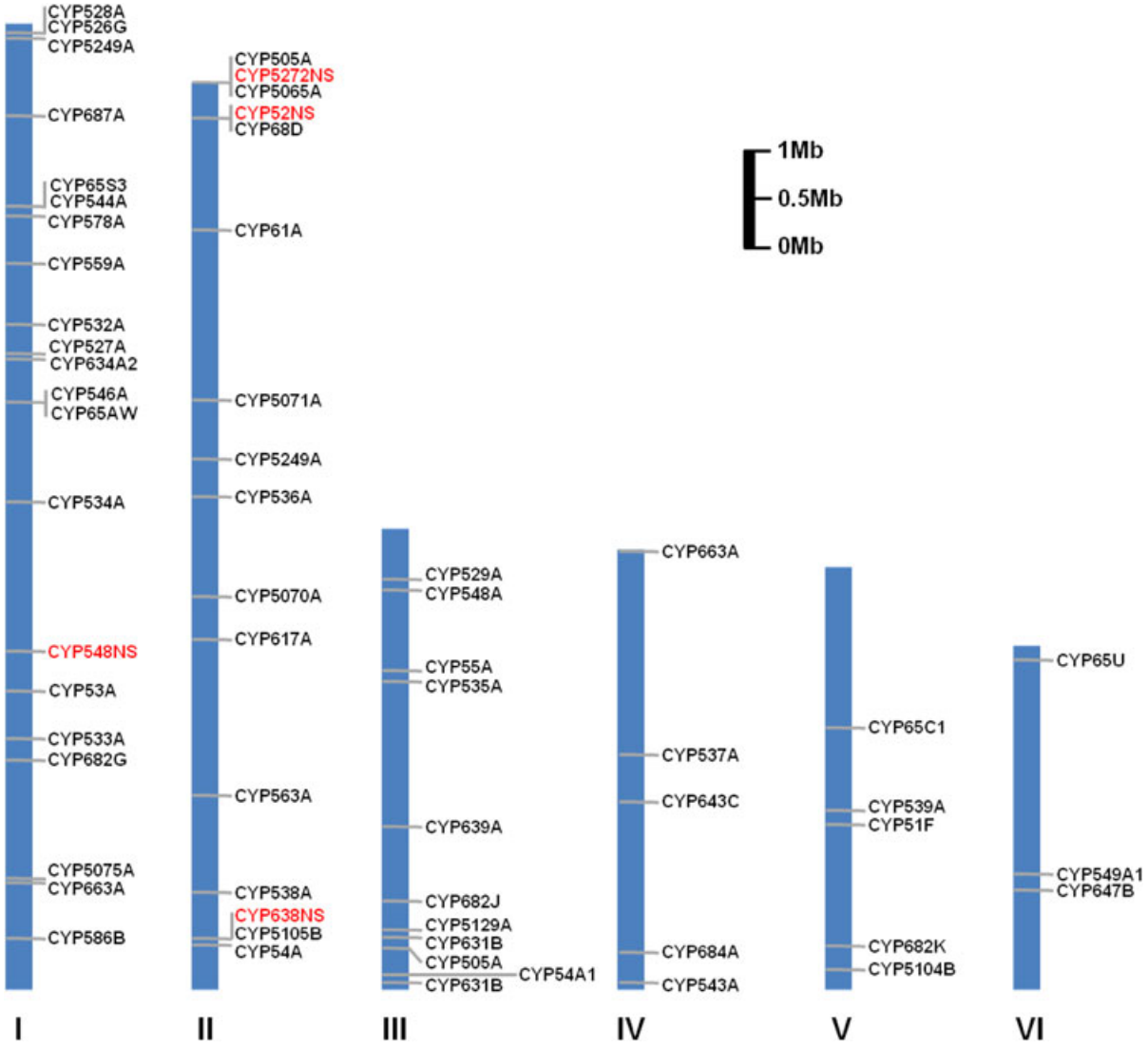


Figure 2.6. Synteny analysis of P450s in *Thielavia terrestris* (taken from Syed *et al*., 2014b).

Synteny analysis provides excellent information on genome-duplication of P450s where localization of P450s belongs to the same family on same chromosome next to each

other is a direct indication that these P450s are possibly duplicated during the evolution (Qhanya *et al*., 2015; Sello *et al*., 2015).

Genome mapping of P450s refer to the identification of neighbouring genes with respect to P450s. Genome mapping reveals information on possible role of P450s in a physiological function by involving in a metabolic pathway. For example, CYP128A1 of *Mycobacterium tuberculosis* function is predicted based on its physical localization with genes involved in the biosynthesis of menaquinone-like molecule (Holsclaw *et al*., 2008). In addition, CYP121A1 of *M. tuberculosis* function identified based on its location in an operon with genes involved in the biosynthesis of cyclodipeptide cyclo(L-Tyr-L-Tyr) (cYY) molecule (Belin *et al*., 2009).

2.7. *Mycobacterium Tuberculosis*

The genus *Mycobacterium* contains a vast number of strict and opportunistic pathogens that post a threat towards humans and other animals (Ventura *et al*., 2007). Among various species within this genus, the principal leading pathogens in humans is *M. tuberculosis*, the causative agent of Tuberculosis (TB) (Koch, 1882), and *Mycobacterium leprae*, causative agent of leprosy (Cole, 2002). Mycobacterial species which serve as opportunistic pathogens comprise of *Mycobacterium avium*, *Mycobacterium simiae*, *Mycobacterium kansasii* and *Mycobacterium haemophilum*, which are more common among immunocompromised patients (Bhambri *et al*., 2009). *Mycobacterium ulcerans* is also responsible for other opportunistic mycobacterial infections of which in most instances it produces a destructive, primarily tropical skin disease which, if however not treated rapidly, tends to produce chronic ulcers with necrotic centres (also known as Buruli ulcer) (Stinear *et al*., 2007); *Mycobacterium marinum*, a pathogen of fish and amphibians, (close relative of *M. tuberculosis*), is responsible for fish-tank or swimming-pool granuloma which essentially concerns people exposed to fish or water (Li *et al*., 2005).

Rapidly growing mycobacteria are found to be medically important due to their association with traumatic and surgical wound infections, skin, soft tissue infections, pulmonary disease etc, and they are essentially limited to *Mycobacterium fortuitum*, *Mycobacterium chelonae and Mycobacterium abscessus* (Brown *et al*., 1992). The principal animal mycobacterial pathogens includes *Mycobacterium bovis*, the causative agent of bovine tuberculosis, *Mycobacterium paratuberculosis and M. avium* causes Johne's disease or paratuberculosis in cattle and pigs (Green *et al*., 1989; Li *et al*., 2005). *Mycobacterium vanbaalenii* is a rod-shaped, non-motile, non-sporulating bacterium; it was the first bacterium isolated by virtue of its ability to metabolize the polycyclic aromatic hydrocarbons (Khan *et al*., 2002).

Mycobacterial infections are notoriously difficult to treat. This has brought forward a dire need for a worldwide research in search for a novel drug targets against mycobacterial pathogens to find new therapeutic methods and to eradicate or to minimize the infectivity rate.

TB is a major global epidemic and kills over one million people a year (WHO, 2013). The disease which was once thought to be practically eradicated in the western parts of the world, *M. tuberculosis* has resurfaced in recent years becoming a major worldwide threat to human health. The reappearance of *M. tuberculosis* owes its respect towards its synergy with the HIV virus whereby *M. tuberculosis* thrives in immune-compromised HIV-infected individuals and also because of the development and propagation of *M. tuberculosis* strains that are becoming resistant to existing anti-tubercular drugs (Brosch *et al*., 2001). This has brought forward for it to be regarded as the leading cause of human mortality among infectious diseases.

2.7.1. Essential or important P450s in *M. tuberculosis*

The greatest significant breakthrough came from the research carried out by Cole and co-workers (1998) in which they sequenced the genome of *M. tuberculosis* H37Rv. Genome sequencing of *M. tuberculosis* revealed presence of 20 cytochrome P450 monooxygenases (P450s) in its genome (Cole *et al*., 1998). The preponderance of P450 genes in the *M. tuberculosis* genome is an idiosyncratic feature for a prokaryote and it might underlie a possible importance of the P450 gene-family in the life history of *M. tuberculosis*. As expected, three P450s, CYP121A1 (McLean *et al*., 2008), CYP125A1 (Sassetti and Rubin, 2003) and CYP128A1 (Sassetti *et al*., 2003) were found essential for survival of *M. tuberculosis*. Further, gene knockout studies showed CYP121A1 essential for *in vitro* *M. tuberculosis* growth (McLean *et al*., 2008) and CYP125A1 is essential for infection in mice (Sassetti and Rubin, 2003; McLean *et al*., 2009) and survival in macrophages (Chang *et al*., 2007). Since CYP128A1 mutant could not be initially obtained during *in vitro* growth, it can be only presumed that it plays an important role during infection (Ouellet *et al*., 2010). *In vitro M. tuberculosis* latency model studies including carbon starvation model (Betts *et al*., 2003) and hypoxia model (Rustad *et al*., 2008) showed up-regulation of three *M. tuberculosis* P450s, CYP128A1 and CYP135A1 and CYP123A1, suggesting their potential role during *M. tuberculosis* latency. Based on meta-analysis of expression data CYP123A1 is selected as best drug candidate against dormant phase of *M. tuberculosis* (Murphy & Brown 2007). It is noteworthy that CYP123A1 and CYP135A1 are present only in TB causing bacteria suggesting their essential role (Parvez *et al*., 2015).

2.8. Rationale for the study

Tuberculosis continues to be a leading cause of death globally, despite global efforts in disease-control programs during the previous 20 years (Raviglione *et al*., 2012). In 2013, 9 million people fell ill with TB and 1.5 million died from the disease (WHO, 2013). In 2010,

about 10 million children were orphaned as a result of TB deaths among parents. TB is a leading killer of people living with HIV, causing one quarter of all deaths. The WHO declared TB a "global health emergency" in 1993 (Lawn and Zumla, 2011) and in 2006. Even though urbanized countries seem to be doing considerably great in controlling or preventing TB, current studies shows that there is an existence of multidrug-resistant TB (MDR-TB) and extensively drug-resistant TB (XDR-TB) in urbanized countries as well (Migliori *et al*., 2012). The presently available anti-TB drugs were developed over 40 years ago and they have become ineffective against drug-resistant TB (MDR- and XDR-TB strains). Globally in 2013, an estimated 480 000 people developed multidrug resistant TB (MDR-TB) (WHO, 2013). Long-term TB medication (6-12 months) is causing severe side effects and it smoothed the progress of the development of drug-resistant TB.

Study on *M. tuberculosis* P450s revealed that CYP125A1 can be a good drug target (Ouellet et al., 2010; Hudson *et al*., 2012) and researchers are in progress in targeting this P450 by developing inhibitors. However, humans have 57 P450s and they oxidize cholesterol, a natural substrate for CYP125. If any inhibitor developed based on cholesterol have potential to cross reach with human P450s and may lead side-effects. If one can target promoter binding elements of CYP125 there is a chance of no cross reaction as promoter and its binding partners in eukaryotes and prokaryotes are entirely different. In this direction, this study will be the first of its kind on genome data mining and genome mapping of CYP125 in the genus *Mycobacterium*. Study results will pave the way to explore promoter and its binding elements of CYP125 of *M. tuberculosis* so that in future one can use them as novel drug target. If one can inhibit CYP125 expression means bacteria can be killed successfully and the drugs developed based on promoter and its binding elements may have no cross reaction with human P450s.

2.8.1. Aims and objectives

This study is aimed to perform genome wide data mining, identification, annotation and phylogenetic analysis of CYP125 P450s in 60 mycobacterial species. Furthermore, CYP125 P450s will be subjected to gene-cluster analysis as part of genome-mapping.

2.8. References

- Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C. and Dugave, C., 2009. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *106*(18), pp.7426-7431.
- Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology*, *124*(1), pp.128-145.
- Betts, J.C., McLaren, A., Lennon, M.G., Kelly, F.M., Lukey, P.T., Blakemore, S.J. and Duncan, K., 2003. Signature gene expression profiles discriminate between isoniazid-, thiolactomycin-, and triclosan-treated *Mycobacterium tuberculosis*. *Antimicrobial agents and chemotherapy*, *47*(9), pp.2903-2913.
- Bhambri, S., Bhambri, A. and Del Rosso, J.Q., 2009. Atypical mycobacterial cutaneous infections. *Dermatologic clinics*, 27(1), pp.63-73.
- Brosch, R., Pym, A.S., Gordon, S.V. and Cole, S.T., 2001. The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends in microbiology*, *9*(9), pp.452-458.
- Brown, B.A., Wallace, R.J., Onyi, G.O. and De Rosas, V., 1992. Activities of four macrolides, including clarithromycin, against *Mycobacterium fortuitum*, *Mycobacterium chelonae*, and *M. chelonae*-like organisms. *Antimicrobial Agents and Chemotherapy*, *36*(1), pp.180-184.
- Chang, J.C., Harik, N.S., Liao, R.P. and Sherman, D.R., 2007. Identification of mycobacterial genes that alter growth and pathology in macrophages and in mice. *Journal of Infectious Diseases*, *196*(5), pp.788-795.

- Chen, W., Lee, M.K., Jefcoate, C., Kim, S.C., Chen, F. and Yu, J.H., 2014. Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution*, *6*(7), pp.1620-1634.
- Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.3. and Tekaia, F., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*(6685), pp.537-544.
- Cole, S.T., 2002. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiology*, *148*(10), pp.2919-2928.
- Myles, D.C. and Whitesides, G.M., 1990. Biotransformations in preparative organic chemistry, the use of isolated enzymes and whole cell systems in synthesis: By HG Davies, RH Green, DR Kelly, and Stanley M. Roberts. Academic Press, San Diego, 1989.
- Doty, S.L., Shang, T.Q., Wilson, A.M., Tangen, J., Westergreen, A.D., Newman, L.A., Strand, S.E. and Gordon, M.P., 2000. Enhanced metabolism of halogenated hydrocarbons in transgenic plants containing mammalian cytochrome P450 2E1. *Proceedings of the National Academy of Sciences*, *97*(12), pp.6287-6291.
- Evans, W.E. and Relling, M.V., 1999. Pharmacogenomics: translating functional genomics into rational therapeutics. *science*, *286*(5439), pp.487-491.
- Fischer, M., Knoll, M., Sirim, D., Wagner, F., Funke, S. and Pleiss, J., 2007. The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics*, *23*(15), pp.2015-2017.

- Furuya, T. and Kino, K., 2010. Genome mining approach for the discovery of novel cytochrome P450 biocatalysts. *Applied microbiology and biotechnology*, *86*(4), pp.991-1002.
- Garfinkel, D., 1958. Studies on pig liver microsomes. I. Enzymic and pigment composition of different microsomal fractions. *Archives of biochemistry and biophysics*, *77*(2), pp.493-509.
- Girhard, M., Bakkes, P.J., Mahmoud, O. and Urlacher, V.B., 2015. P450 Biotechnology. In *Cytochrome P450* (pp. 451-520). Springer International Publishing.
- Gomez-Escribano, J.P. and Bibb, M.J., 2014. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *Journal of industrial microbiology & biotechnology*, *41*(2), pp.425-431.
- Gonzalez, F.J. and Nebert, D.W., 1990. Evolution of the P450 gene superfamily: animal-plant 'warfare', molecular drive and human genetic differences in drug oxidation. *Trends in Genetics*, *6*, pp.182-186.
- Green, E.P., Tizard, M.L.V., Moss, M.T., Thompson, J., Winterbourne, D.J., McFadden, J.J. and Hermon-Taylor, J., 1989. Sequence and characteristics or IS900, an insertion element identified in a human Crohn's disease isolate or *Mycobacterium paratuberculosis*. *Nucleic Acids Research*, *17*(22), pp.9063-9073.
- Guengerich, F.P., 2003. Cytochromes P450, drugs, and diseases. *Molecular interventions*, *3*(4), p.194.
- Guengerich, F.P., 2015. Human cytochrome P450 enzymes. In *Cytochrome P450* (pp. 523-785). Springer International Publishing.
- Holsclaw, C.M., Sogi, K.M., Gilmore, S.A., Schelle, M.W., Leavell, M.D., Bertozzi, C.R. and Leary, J.A., 2008. Structural characterization of a novel sulfated

menaquinone produced by stf3 from *Mycobacterium tuberculosis*. *ACS chemical biology*, *3*(10), pp.619-624.

- Hudson, S.A., McLean, K.J., Munro, A.W. and Abell, C., 2012. *Mycobacterium tuberculosis* cytochrome P450 enzymes: a cohort of novel TB drug targets. *Biochemical Society Transactions*, *40*(3), pp.573-579.
- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and 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*(9), p.e107209.
- Kawahigashi, H., Hirose, S., Ohkawa, H. and Ohkawa, Y., 2005. Transgenic rice plants expressing human CYP1A1 remediate the triazine herbicides atrazine and simazine. *Journal of agricultural and food chemistry*, *53*(22), pp.8557-8564.
- Kawahigashi, H., Hirose, S., Ohkawa, H. and Ohkawa, Y., 2006. Phytoremediation of the herbicides atrazine and metolachlor by transgenic rice plants expressing human CYP1A1, CYP2B6, and CYP2C19. *Journal of agricultural and food chemistry*, *54*(8), pp.2985-2991.
- Kgosiemang, I.K.R., Mashele, S.S. and Syed, K., 2014. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum Saccharomycotina. *J. Pure Appl. Microbiol*, *8*, pp.291-302.
- Khan, A.A., Kim, S.J., Paine, D.D. and Cerniglia, C.E., 2002. Classification of a polycyclic aromatic hydrocarbon-metabolizing bacterium, *Mycobacterium* sp. strain PYR-1, as *Mycobacterium vanbaalenii* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *52*(6), pp.1997-2002.
- Klingenberg, M., 1958. Pigments of rat liver microsomes. *Archives of biochemistry and biophysics*, *75*(2), pp.376-386.

- Koch R., 1882. Die Äetiologie der Tuberkulose. *Berliner klinische Wochenschrift*, 15, pp.221–230.
- Lamb, D.C., Waterman, M.R. and Zhao, B., 2013. *Streptomyces* cytochromes P450: applications in drug metabolism. *Expert opinion on drug metabolism & toxicology*, *9*(10), pp.1279-1294.
- Lamb, D.C., Waterman, M.R., Kelly, S.L. and Guengerich, F.P., 2007. Cytochromes P450 and drug discovery. *Current opinion in biotechnology*, *18*(6), pp.504-512.
- Lawn, S.D., Zumla, A.I., 2011. Tuberculosis. *Lancet*, 378(9785), pp.57–72.
- Lewis, D.F.V., Watson, E. and Lake, B.G., 1998. Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics. *Mutation Research/Reviews in Mutation Research*, *410*(3), pp.245-270.
- Li, L., Bannantine, J.P., Zhang, Q., Amonsin, A., May, B.J., Alt, D., Banerji, N., Kanjilal, S. and Kapur, V., 2005. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), pp.12344-12349.
- 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. and Rosenkrands, I., 2008. Characterization of active site structure in CYP121 a cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *Journal of Biological Chemistry*, *283*(48), pp.33406-33416.
- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and 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*(51), pp.35524-35533.

- Migliori, G.B., Sotgiu, G., Gandhi, N.R., Falzon, D., DeRiemer, K., Centis, R., Hollm-Delgado, M.G., Palmero, D., Pérez-Guzmán, C., Vargas, M.H. and D'Ambrosio, L., 2012. Drug resistance beyond XDR-TB: results from a large individual patient data meta-analysis. *European Respiratory Journal*, pp.erj01363-2012.
- Mthakathi, N.T., Kgosiemang, I.K.R., Chen, W., Mohlatsane, M.E., Mojahi, T.J., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Cytochrome P450 monooxygenase analysis in free-living and symbiotic microalgae *Coccomyxa* sp. C-169 and *Chlorella* sp. NC64A. *Algae*, *30*(3), p.233.
- Murphy, D.J. and Brown, J.R., 2007. Identification of gene targets against dormant phase *Mycobacterium tuberculosis* infections. *BMC infectious diseases*, *7*(1), p.1.
- Nebert, D. W.; Nelson, D. R.; Coon, M. J.; Estabrook, R. W.; Feyereisen, R.; Fujii-Kuriyama,Y.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Loper, J. C.; Sato, R.;Waterman, M. R.; Waxman, D. J., 1991. The P450 superfamily: Update on new sequences, gene mapping and recommended nomenclature. *DNA and cell biology*, 10(1): pp. 1–14.
- Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Geungerich, F.P.,Gunsaius., I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., and Sato, R., Waterman, M.R., 1987. The P450 gene superfamily: recommended nomenclature. *DNA*, 6(1), pp.1–11.
- Nelson, D.R., 1998. Metazoan cytochrome P450 evolution. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, *121*(1), pp.15-22.
- Nelson, D.R., 1999. Cytochrome P450 and the individuality of species. *Archives of biochemistry and biophysics*, *369*(1), pp.1-10.

- Nelson, D.R., 2006. Cytochrome P450 Nomenclature, 2004. *Cytochrome P450 Protocols*, pp.1-10.
- Nelson, D.R., 2009. The cytochrome p450 homepage. Human genomics, 4(1), p.1.
- Nelson, D.R., 2011. Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1814*(1), pp.14-18.
- Nelson, D.R., 2013. A world of cytochrome P450s. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1612), p.20120430.
- Ortiz de Montellano, P. R., 2015. Cytochrome P450: Structure, mechanism and biochemistry. Springer International Publishing, Switzerland. ISBN: 978-3-319-12107-9. DOI: 10.1007/978-3-319-12108-6.
- Ouellet, H., Guan, S., Johnston, J.B., Chow, E.D., Kells, P.M., Burlingame, A.L., Cox, J.S., Podust, L.M. and De Montellano, P.R.O., 2010. *Mycobacterium tuberculosis* CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Molecular microbiology*, *77*(3), pp.730-742.
- Park, J., Lee, S., Choi, J., Ahn, K., Park, B., Park, J., Kang, S. and Lee, Y.H., 2008. Fungal cytochrome P450 database. *BMC genomics*, *9*(1), p.1.
- Parvez, M., Qhanya, L.B., Mthakathi, N.T., Kgosiemang, I.K.R., Bamal, H.D., Pagadala, N.S., Xie, T., Yang, H., Chen, H., Theron, C.W. and Monyaki, R., 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Scientific Reports*, *6*.
- Qhanya, L.B., Matowane, G., Chen, W., Sun, Y., Letsimo, E.M., Parvez, M., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete Biotrophic plant pathogens. *PloS one*, *10*(11), p.e0142100.

- Raviglione, M., Marais, B., Floyd, K., Lönnroth, K., Getahun, H., Migliori, G.B., Harries, A.D., Nunn, P., Lienhardt, C., Graham, S. and Chakaya, J., 2012. Scaling up interventions to achieve global tuberculosis control: progress and new developments. *The Lancet*, *379*(9829), pp.1902-1913.
- Reichenauer, T.G. and Germida, J.J., 2008. Phytoremediation of organic contaminants in soil and groundwater. *ChemSusChem*, *1*(8-9), pp.708-717.
- Rustad, T.R., Harrell, M.I., Liao, R. and Sherman, D.R., 2008. The enduring hypoxic response of *Mycobacterium tuberculosis*. *PloS one*, *3*(1), p.e1502.
- Sassetti, C.M. and Rubin, E.J., 2003. Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences*, *100*(22), pp.12989-12994.
- Sassetti, C.M., Boyd, D.H. and Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, *48*(1), pp.77-84.
- Schmeisser, C., Steele, H. and Streit, W.R., 2007. Metagenomics, biotechnology with non-culturable microbes. *Applied microbiology and biotechnology*, *75*(5), pp.955-962.
- Schulz, S., Girhard, M. and Urlacher, V.B., 2012. Biocatalysis: key to selective oxidations. *ChemCatChem*, *4*(12), pp.1889-1895.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific reports*, *5*.
- Sirim, D., Wagner, F., Lisitsa, A. and Pleiss, J., 2009. The cytochrome P450 engineering database: Integration of biochemical properties. *BMC biochemistry*, *10*(1), p.1.

- Sono, M., Roach, M.P., Coulter, E.D. and Dawson, J.H., 1996. Heme-containing oxygenases. *Chemical Reviews*, *96*(7), pp.2841-2888.
- Stinear, T.P., Seemann, T., Pidot, S., Frigui, W., Reysset, G., Garnier, T., Meurice, G., Simon, D., Bouchier, C., Ma, L. and Tichit, M., 2007. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome research*, *17*(2), pp.192-200.
- Syed, K. and 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*(4), p.e95616.
- Syed, K. and Yadav, J.S., 2012. P450 monooxygenases (P450ome) of the model white rot fungus *Phanerochaete chrysosporium*. *Critical reviews in microbiology*, *38*(4), pp.339-363.
- Syed, K., Nelson, D.R., Riley, R. and Yadav, J.S., 2013. Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore species *Bjerkandera adusta*, *Ganoderma* sp. and *Phlebia brevispora*. *Mycologia*, *105*(6), pp.1445-1455.
- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and Pagadala, N.S., 2014b. Genome-wide identification, annotation and characterization of novel thermostable cytochrome P450 monooxygenases from the thermophilic biomass-degrading fungi *Thielavia terrestris* and *Myceliophthora thermophila*. *Genes & Genomics*, *36*(3), pp.321-333.
- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014a. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450

monooxygenase families enriched in model basidiomycete fungi. *PLoS One*, *9*(1), p.e86683.

- van Beilen, J.B., Duetz, W.A., Schmid, A. and Witholt, B., 2003. Practical issues in the application of oxygenases. *Trends in biotechnology*, *21*(4), pp.170-177.
- Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R. and Jacks, T., 2007. Restoration of p53 function leads to tumour regression in vivo. *Nature*, *445*(7128), pp.661-665.
- WHO, 2013. Tuberculosis. http://www.who.int/mediacentre/factsheets/fs104/en/. Accessed on 23 March 2014.
- Zhan, J., 2009. Biosynthesis of bacterial aromatic polyketides. *Current topics in medicinal chemistry*, *9*(17), pp.1598-1610.

Chapter 3:

Annotation and phylogenetic analysis of cytochrome P450 monooxygenase CYP125 family in the genus *Mycobacterium*

3.1. Introduction

Cytochrome P450 monooxygenases (CYPs/P450s) are mixed function oxidoreductases ubiquitously distributed in species belong to all biological kingdoms (Nelson, 2013). P450s are well known for their role in essential cellular anabolic and catabolic processes. Genome sequencing analysis of deadliest human pathogen *Mycobacterium tuberculosis* revealed presence of 20 P450s in its genome (Cole *et al*., 1998). Among *M*. *tuberculosis* P450s CYP125A1 has been found to be essential for *M*. *tuberculosis* survival on cholesterol (Sassetti and Rubin, 2003; McLean *et al*., 2009; Ouellet *et al*., 2010) and studies are in progress to explore this P450 as a novel drug target (Ouellet *et al*., 2010; Hudson *et al*., 2012). Below Table 3.1 shows properties of CYP125A1.

Table 3.1. Properties of CYP125A1 (taken from McLean *et al*., 2015)

Microaaray/genetic analysis	References	Key facts	References
Essential for infection in mice and induced in macrophages	Sassetti and Rubin., 2003; Schnappinger *et al*., 2003	Part of *igr* operon with fadE28, fadE29, IgrD-E, and ltp2 (Rv3544c-3540c)	Chang *et al*., 2009; Thomas *et al*., 2011
In KstR reglon and igr	Kendall *et al*.,	Cholesterol/cholest-	McLean *et al*.,

operon, essential gene for growth and virulence in macrophages and mice	2007; Thomas *et al*., 2011	4-en-3-one 26-oxidase. Structurally characterized	2009; Ouellet *et al*., 2010
Expressed in dormancy model and upregulated during infection of dendritic cells	Murphy and Brown, 2007; Tailleux *et al*., 2008		

CYP125A1 catalyze C26 ω-hydroxylation(s) of the side chain of cholesterol, and of its ketone derivative cholest-4-en-3-one (McLean *et al*., 2009; Capyk *et al*., 2009; Driscoll *et al*., 2010; Ouellet *et al*., 2010; Johnston *et al*., 2010) (Figure 3.1a.). CYP125A1 also found to produce five additional products, resulting from deformylation of the aliphatic cholesterol side-chain aldehyde intermediate (Sivaramakrishnan *et al*., 2012) (Figure 3.1b).

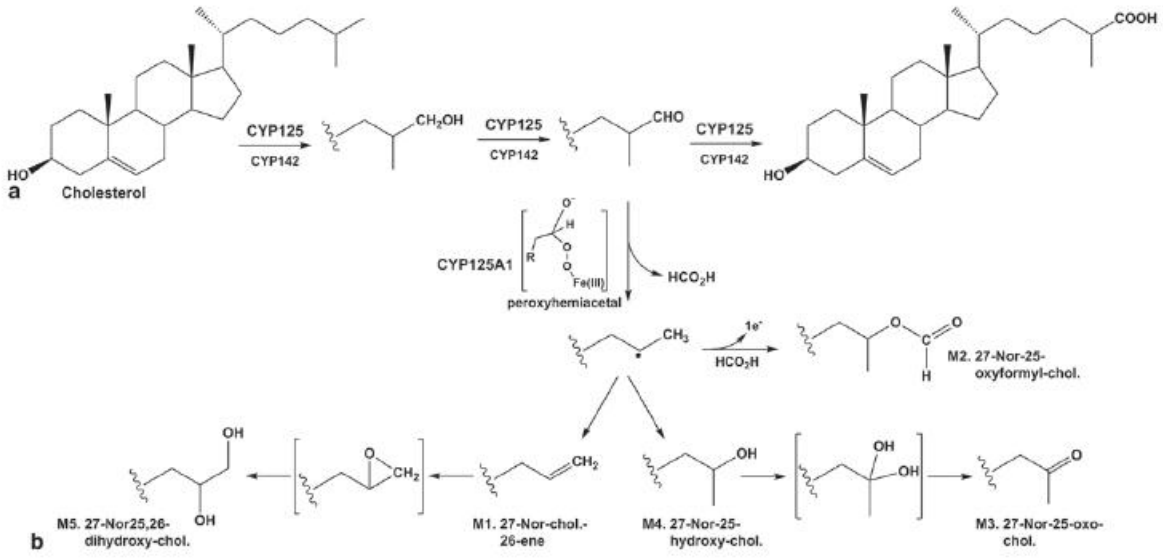


Figure 3.1. The oxidation of cholesterol and cholesten-4-en-3-one (taken from McLean *et al*., 2015). (a) The CYP125A1/CYP142A1 (and CYP124A1)-dependent conversion of cholesterol and cholesten-4-en-3-one through C26-oxidation reactions to the acid via the hydroxyl and aldehyde forms (McLean *et al*., 2009; Capyk *et al*., 2009; Driscoll *et al*., 2010; Ouellet *et al*., 2010; Johnston *et al*., 2010). (b) CYP125A1-catalyzed deformylation of the side chain of cholesterol andcholesten-4-en-3-one (chol.) *via* a peroxyhemiacetal adduct, predicted to be derived from the reaction of the heme iron ferric–peroxo anion ($Fe^{3+}O_2$) species with the aldehyde intermediate, leading to C–C bond cleavage.

In addition to functional analysis, structural analysis of CYP125A1 has been thoroughly studied (McLean *et al*., 2009). However, to date, CYP125A1 distribution across mycobacterial species has not been carried out. Hence, in this study, the aim is to conduct, genome wide CYP125 P450 identification, annotation and phylogenetic analysis in 60 mycobacterial species.

3.2. Methods

3.2.1. Species used in the study

Mycobacterial species genomes that are published and are publicly available were used in this study. A total of 60 mycobacterial species belonging to six different mycobacterial categories (Parvez *et al*., 2016) were selected and analysed for this study (Table 3.2). The six categories include *Mycobacterium tuberculosis* complex (MTBC) (27 species), *M. chelonae-abscessus complex* (MCAC) (6 species), *M. avium* complex (MAC) (8 species), Mycobacteria causing leprosy (MCL) (2 species), Nontuberculous mycobacteria (NTM) (6 species) and Saprophytes (SAP) (11 species).

3.2.2. Genome data mining and annotation of CYP125 in *Mycobacterium*

Mycobacterial genomes that are publicly available at different genome databases as listed in Table 3.2 were mined for CYP125 as described elsewhere (Syed *et al*., 2013; Syed and

Mashele, 2014; Syed *et al*., 2014a; Syed *et al*., 2014b; Jawallapersand *et al*., 2014; Kgosiemang *et al*., 2014; Sello *et al*., 2015; Mthakathi *et al*., 2015; Qhanya *et al*., 2015; Parvez *et al*., 2016). Briefly, CYP125A1 of *M. tuberculosis* (McLean *et al*., 2009) is blasted against other mycobacterial species genomes. The proteins that showed >40% identity were selected and subjected to BLAST analysis against bacterial P450s at the Cytochrome P450 Homepage (Nelson, 2009). Based on the International P450 Nomenclature Committee rule, proteins with >55% identity were grouped under the same subfamily. Some mycobacterial CYP125 P450s were annotated and made available at the Cytochrome P450 Homepage (Nelson, 2009). In this case, the same nomenclature for CYP125 P450s was continued. The CYP125 P450s that showed less than 55% identity to known CYP125 P450s at the Cytochrome P450 Homepage (Nelson, 2009) were assigned to new P450 subfamilies (NS). Annotated CYP125 P450 sequences are listed under Appendix.

3.2.3. Phylogenetic analysis of mycobacterial CYP125 P450s

The phylogenetic analysis of CYP125 P450s was carried out as described elsewhere (Sello *et al*., 2015). Briefly, the evolutionary history the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013).

Table 3.2. Information on mycobacterial species with codes in parenthesis and their respective genome database links used in the study.

Species name	Database	Weblink
Mycobacterium tuberculosis C(TBCG); *Mycobacterium tuberculosis* F11(TBFG); *Mycobacterium tuberculosis* H37Ra(MRA); *Mycobacterium tuberculosis* H37Rv(Rv); *Mycobacterium tuberculosis* Haarlem(TBHG); *Mycobacterium tuberculosis* KZN 1435(TBMG); *Mycobacterium tuberculosis* KZN 605(TBXG); *Mycobacterium tuberculosis* KZN 4207(TBSG); *Mycobacterium tuberculosis* RGTB327(MRGA); *Mycobacterium tuberculosis* strains CCDC5079(CCDC); *Mycobacterium tuberculosis* 7199-99(MT); *Mycobacterium tuberculosis* Beijing/NITR203(J111); *Mycobacterium tuberculosis* CAS/NITR204(J113); *Mycobacterium tuberculosis* EAI5(M943); *Mycobacterium tuberculosis* EAI5/NITR206(J114); *Mycobacterium tuberculosis* Erdman= ATCC 35801(ERDMAN); *Mycobacterium abscessus* ATCC 19977(MAB); *Mycobacterium Avium* 104(MAV); *Mycobacterium Avium* subsp.	TB Database	http://genome.tbdb.org/tbdb_sysbio/GenomesIndex.html

paratuberculosis K10(MAP); *Mycobacterium ulcerans* Agy99(MUL); *Mycobacterium Marinum(MMAR)*; *Mycobacterium* sp. MCS(Mmcs); *Mycobacterium vanbaalenii* PYR-1(Mvan); *Mycobacterium smegmatis* MC2 155(MSMEG); *Mycobacterium chubuense* NBB4 (Mycch)		
Mycobacterium tuberculosis CDC1551(MT); *Mycobacterium bovis* AF 2122/97(Mb); *Mycobacterium* sp. KMS(Mkms); *Mycobacterium gilvum* PYR-GCK(Mflv)	Xbase	http://www.xbase.ac.uk/mycodb/
Mycobacterium africanum GM041182(MAF); *Mycobacterium tuberculosis* UT205(UDA); *Mycobacterium canetii* CIPT 140010059(MCAN); *Mycobacterium canetii* CIPT 140060008(BN44); *Mycobacterium canetii* CIPT 140710010(BN42); *Mycobacterium bovis* BCG Pasteur 1173P2(BCG); *Mycobacterium bovis* BCG Korea 1168P(K60); *Mycobacterium bovis* BCG Mexico(BCGMEX); *Mycobacterium bovis* BCG Tokyo 172(JTY); *Mycobacterium abscessus* subsp. *bolletii* 50594(MASS); *Mycobacterium avium* subsp. paratuberculosis MAP4(MAP4); *Mycobacterium intracellulare* ATCC	KEGG	http://www.kegg.jp/kegg-bin/show_organism?org=maf

13950(OCU); *Mycobacterium intracellulare* MOTT-02(OCO); *Mycobacterium Intracellulare* MOTT-64(OCQ); *Mycobacterium intracellulare* MOTT-36Y(W7S); *Mycobacterium Indicus pranii* MTCC 9506(MIP); *Mycobacterium leprae* Br4923(MLBR); *Mycobacterium Leprae* TN(ML); *Mycobacterium* sp. JDM601(JDM601); *Mycobacterium liflandii* 128FXT(MULP); *Mycobacterium massiliense(MYCMA)*; *Mycobacterium kansassii* ATCC 12478(MKAN); *Mycobacterium* sp. JLS(Mjls); *Mycobacterium gilvum* Spyr1(Mspyr1); *Mycobacterium smegmatis* JS623(Mycrhn); *Mycobacterium rhodesiae* NBB3; *Mycobacterium neoaurum* VKM Ac-1815D(d174)		
Mycobacterium abscessus 47J26(MYCAB); *Mycobacterium abscessus*103(LA61); *Mycobacterium abscessus* subsp. bolletiiMA 1948(LA62); *Mycobacterium abscessus* VO6705(MYCAB)	UniProt	http://www.uniprot.org/taxonomy/36809

Notes on databases

TB database: TBDB contains annotated genome and expression (microarray and RT–PCR) data and a suite of data analysis tools designed to serve as a unique resource for TB research and for the discovery of new drugs, vaccines and biomarkers

UniProt: The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. It contains high-quality manually annotated and non-redundant protein sequence records. Manual annotation consists of analysis, comparison and merging of all available sequences for a given protein, as well as a critical review of associated experimental and predicted data. UniProt curators extract biological information from the literature and perform numerous computational analyses.

KEGG: A database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies

xBASE: A genome database aimed at helping laboratory-based bacteriologists make best use of bacterial genome sequence data, with a particular emphasis on comparative genomics. The latest version, xBASE 2.0 (http://xbase.bham.ac.uk), now provides comprehensive coverage of all bacterial genomes and features an updated modularized backend and an improved user interface, which includes a taxonomy browser and a powerful full-text search facility.

3.3. Results and Discussion

Genome-wide data mining and annotation of CYP125 P450s in 60 mycobacterial species belonging to six different categories revealed presence of 120 P450s belong to five different subfamilies i.e. A, D, E. F and NS (Table 3.3).

Table 3.3. Genome data mining and annotation of CYP125 P450s in 60 mycobacterial species.

Name of the species	**P450 count**	**Subfamilies**				
		A	**D**	**E**	**F**	**NS**
***Mycobacterium tuberculosis* complex (MTBC)**						
Mycobacterium africanum GM041182	1	1	0	0	0	0
Mycobacterium tuberculosis C	1	1	0	0	0	0
Mycobacterium tuberculosis F11	1	1	0	0	0	0
Mycobacterium tuberculosis H37Ra	1	1	0	0	0	0
Mycobacterium tuberculosis H37Rv	1	1	0	0	0	0
Mycobacterium tuberculosis Haarlem	1	1	0	0	0	0
Mycobacterium tuberculosis KZN 1435	1	1	0	0	0	0
Mycobacterium tuberculosis KZN 605	1	1	0	0	0	0
Mycobacterium tuberculosis KZN 4207	1	1	0	0	0	0
Mycobacterium tuberculosis RGTB327	0	0	0	0	0	0
Mycobacterium tuberculosis CDC1551	1	1	0	0	0	0
Mycobacterium tuberculosis strains CCDC5079	1	1	0	0	0	0
Mycobacterium tuberculosis 7199-99	1	1	0	0	0	0
Mycobacterium tuberculosis Beijing/NITR203	1	1	0	0	0	0

Mycobacterium tuberculosis CAS/NITR204	1	1	0	0	0	0
Mycobacterium tuberculosis EAI5	0	0	0	0	0	0
Mycobacterium tuberculosis EAI5/NITR206	1	1	0	0	0	0
Mycobacterium tuberculosis Erdman= ATCC 35801	1	1	0	0	0	0
Mycobacterium tuberculosis UT205	1	1	0	0	0	0
Mycobacterium canetii CIPT 140010059	1	1	0	0	0	0
Mycobacterium canetii CIPT 140060008	1	1	0	0	0	0
Mycobacterium canetii CIPT 140710010	1	1	0	0	0	0
Mycobacterium bovis AF 2122/97	1	1	0	0	0	0
Mycobacterium bovis BCG Pasteur 1173P2	1	1	0	0	0	0
Mycobacterium bovis BCG Korea 1168P	1	1	0	0	0	0
Mycobacterium bovis BCG Mexico	1	1	0	0	0	0
Mycobacterium bovis BCG Tokyo 172	1	1	0	0	0	0
***Mycobacterium chelonae-abscessus* complex (MCAC)**						
Mycobacterium abscessus ATCC 19977	4	4	0	0	0	0
Mycobacterium abscessus subsp. bolletii 50594	3	3	0	0	0	0
Mycobacterium abscessus 47J26	4	4	0	0	0	0
*Mycobacterium abscessus*103	4	4	0	0	0	0
Mycobacterium abscessus subsp. *Bolletii* MA1948	4	4	0	0	0	0
Mycobacterium abscessus VO6705	4	4	0	0	0	0
***Mycobacterium avium* complex (MAC)**						
Mycobacterium Avium 104	3	2	0	0	0	1
Mycobacterium Avium subsp. *paratuberculosis* K10	4	2	0	0	1	1
Mycobacterium avium subsp. *paratuberculosis* MAP4	4	2	0	0	1	1

Mycobacterium intracellulare ATCC 13950	2	1	0	1	0	0
Mycobacterium intracellulare MOTT-02	2	1	1	0	0	0
Mycobacterium Intracellulare MOTT-64	2	1	1	0	0	0
Mycobacterium intracellulare MOTT-36Y	2	1	1	0	0	0
Mycobacterium Indicus pranii MTCC 9506	5	3	1	0	0	1
Mycobacteria causing Leprosy (MCL)						
Mycobacterium leprae Br4923	0	0	0	0	0	0
Mycobacterium Leprae TN	0	0	0	0	0	0
Nontuberculous mycobacteria (NTM)						
Mycobacterium sp. JDM601	5	5	0	0	0	0
Mycobacterium liflandii 128FXT	2	2	0	0	0	0
Mycobacterium ulcerans Agy99	1	1	0	0	0	0
Mycobacterium Marinum	2	2	0	0	0	0
Mycobacterium massiliense	0	0	0	0	0	0
Mycobacterium kansassii ATCC 12478	1	0	1	0	0	0
Saprophytes (SAP)						
Mycobacterium sp. JLS	3	3	0	0	0	0
Mycobacterium sp. KMS	3	3	0	0	0	0
Mycobacterium sp. MCS	3	3	0	0	0	0
Mycobacterium vanbaalenii PYR-1	4	3	0	0	1	0
Mycobacterium smegmatis MC2 155	3	3	0	0	0	0
Mycobacterium chubuense NBB4	4	3	0	0	1	0
Mycobacterium gilvum PYR-GCK	4	3	0	0	1	0
Mycobacterium gilvum Spyr1	4	3	0	0	1	0

Mycobacterium smegmatis JS623	4	3	0	0	1	0
Mycobacterium rhodesiae NBB3	5	4	0	0	0	1
Mycobacterium neoaurum VKM Ac-1815D	0	0	0	0	0	0

Analysis of CYP125 P450s distribution in different categories revealed that MTBC species showed lowest copies of CYP125 in their genomes compared to other categories (Table 3.4). This study revealed that CYP125 P450 is not present in MCL species. MAC and SAP species showed highest number of CYP125 subfamilies in their genomes (Table 3.4).

Table 3.4. Analysis of CYP125 P450 family in different mycobacterial categories.

Category	No of species used for analysis	Range (min-max)	Number of subfamilies
MTBC	27	0-1	1
MCAC	6	3-4	1
MAC	8	2-5	5
MCL	2	0	0
NTM	6	0-5	2
SAP	11	0-5	5

Analysis of subfamily patterns in mycobacterial categories revealed MAC species have highest diversity of CYP125 subfamilies followed by species belong to SAP and NTM (Tables 3.3 and 3.4). Among subfamilies, subfamily A is more dominant across mycobacterial species. Subfamily D is present only in species belonging to MAC and NTM. Subfamily E is present only in species belonging to MAC. Subfamily F and NS is present in MAC and SAP.

Phylogenetic analysis of CYP125 P450s revealed that CYP125 P450s belonging to the same mycobacterial group were clustered together in the tree showing conservation of CYP125 as per category (Figure 3.2). Alignment of CYP125 P450s belongs to the same subfamily together further authenticate that annotation of P450s in this study are correct. Presence of more than one copy of CYP125 in some mycobacterial categories suggests important role of this P450s in their physiology.

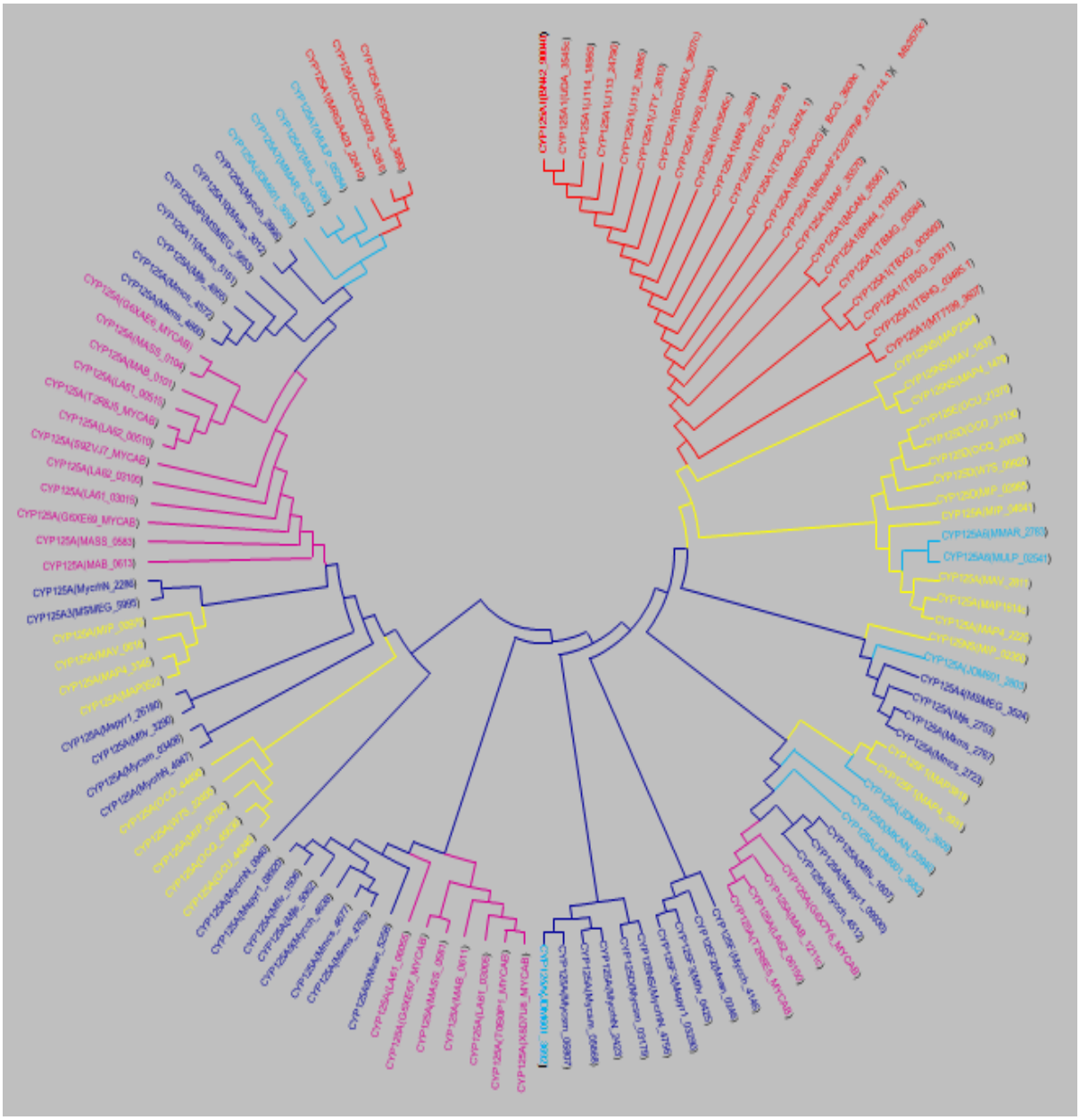


Figure 3.2. Phylogenetic analysis of CYP125 P450s in mycobacterial species. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 28.32425881 is shown. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 119 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 406 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al*., 2013). Colour codes: red-MTBC; pink-MCAC; yellow-MAC; sky blue-NTM; navy blue-SAP.

3.4. Conclusion

Annotation of CYP125 P450 across 60 mycobacterial species has revealed a total number of 120 CYP125 P450s grouped into five subfamilies. Analysis of CYP125 P450s distribution in different categories revealed that MTBC species showed lowest copies of CYP125 in their genomes compared to other categories. This study revealed that CYP125 P450 is not present in MCL species. MAC and SAP species showed highest number of CYP125 subfamilies in their genomes. Analysis of CYP125 P450 subfamily patterns in mycobacterial categories revealed MAC species have highest diversity of CYP125 subfamilies followed by species belong to SAP and NTM. Presence of more than one copy of CYP125 in some mycobacterial categories suggests important role of this P450s in their physiology.

3.5. References

- Capyk, J.K., Kalscheuer, R., Stewart, G.R., Liu, J., Kwon, H., Zhao, R., Okamoto, S., Jacobs, W.R., Eltis, L.D. and Mohn, W.W., 2009. Mycobacterial cytochrome p450 125 (cyp125) catalyzes the terminal hydroxylation of c27 steroids. *Journal of Biological Chemistry*, *284*(51), pp.35534-35542.
- Chang, J.C., Miner, M.D., Pandey, A.K., Gill, W.P., Harik, N.S., Sassetti, C.M. and Sherman, D.R., 2009. *igr* Genes and *Mycobacterium tuberculosis* cholesterol metabolism. *Journal of bacteriology*, *191*(16), pp.5232-5239.
- Cole, S., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.3. and Tekaia, F., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*(6685), pp.537-544.
- Driscoll, M.D., McLean, K.J., Levy, C., Mast, N., Pikuleva, I.A., Lafite, P., Rigby, S.E., Leys, D. and Munro, A.W., 2010. Structural and biochemical characterization of Mycobacterium tuberculosis CYP142 evidence for multiple cholesterol 27-hydroxylase activities in a human pathogen. *Journal of Biological Chemistry*, *285*(49), pp.38270-38282.
- Hlavica, P., 2015. Mechanistic basis of electron transfer to cytochromes p450 by natural redox partners and artificial donor constructs. In *Monooxygenase, Peroxidase and Peroxygenase Properties and Mechanisms of Cytochrome P450* (pp. 247-297). Springer International Publishing.
- Hudson, S.A., McLean, K.J., Munro, A.W. and Abell, C., 2012. Mycobacterium tuberculosis cytochrome P450 enzymes: a cohort of novel TB drug targets. *Biochemical Society Transactions*, *40*(3), pp.573-579.

- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and 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*(9), p.e107209.
- Johnston, J.B., Ouellet, H. and de Montellano, P.R.O., 2010. Functional redundancy of steroid C26-monooxygenase activity in Mycobacterium tuberculosis revealed by biochemical and genetic analyses. *Journal of Biological Chemistry*, *285*(47), pp.36352-36360.
- Kendall, S.L., Withers, M., Soffair, C.N., Moreland, N.J., Gurcha, S., Sidders, B., Frita, R., Ten Bokum, A., Besra, G.S., Lott, J.S. and Stoker, N.G., 2007. A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in Mycobacterium smegmatis and Mycobacterium tuberculosis. *Molecular microbiology*, *65*(3), pp.684-699
- Kgosiemang, I.K.R., Mashele, S.S. and Syed, K., 2014. Comparative genomics and evolutionary analysis of cytochrome P450 monooxygenases in fungal subphylum Saccharomycotina. *J. Pure Appl. Microbiol*, *8*, pp.291-302.
- 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, P. R.). *Springer International Publishing*, Ch. 6, 261–407.
- McLean, K.J., Lafite, P., Levy, C., Cheesman, M.R., Mast, N., Pikuleva, I.A., Leys, D. and 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*(51), pp.35524-35533.
- Mthakathi, N.T., Kgosiemang, I.K.R., Chen, W., Mohlatsane, M.E., Mojahi, T.J., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Cytochrome P450 monooxygenase analysis in

free-living and symbiotic microalgae Coccomyxa sp. C-169 and Chlorella sp. NC64A. *Algae*, *30*(3), p.233.

- Murphy, D.J. and Brown, J.R., 2007. Identification of gene targets against dormant phase Mycobacterium tuberculosis infections. *BMC infectious diseases*, *7*(1), p.1.
- Nelson, D.R., 2009. The cytochrome p450 homepage. *Human genomics*, *4*(1), p.1.
- Nelson, D.R., 2013. A world of cytochrome P450s. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1612), p.20120430.
- Ouellet, H., Guan, S., Johnston, J.B., Chow, E.D., Kells, P.M., Burlingame, A.L., Cox, J.S., Podust, L.M. and De Montellano, P.R.O., 2010. Mycobacterium tuberculosis CYP125A1, a steroid C27 monooxygenase that detoxifies intracellularly generated cholest-4-en-3-one. *Molecular microbiology*, *77*(3), pp.730-742.
- Parvez, M., Qhanya, L.B., Mthakathi, N.T., Kgosiemang, I.K.R., Bamal, H.D., Pagadala, N.S., Xie, T., Yang, H., Chen, H., Theron, C.W. and Monyaki, R., 2016. Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. *Scientific Reports*, *6*.
- Qhanya, L.B., Matowane, G., Chen, W., Sun, Y., Letsimo, E.M., Parvez, M., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Genome-Wide Annotation and Comparative Analysis of Cytochrome P450 Monooxygenases in Basidiomycete Biotrophic Plant Pathogens. *PloS one*, *10*(11), p.e0142100.
- Saitou, N. and Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), pp.406-425.
- Sassetti, C.M. and Rubin, E.J., 2003. Genetic requirements for mycobacterial survival during infection. *Proceedings of the National Academy of Sciences*, *100*(22), pp.12989-12994.

- Sassetti, C.M., Boyd, D.H. and Rubin, E.J., 2003. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, *48*(1), pp.77-84.
- Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C. and Schoolnik, G.K., 2003. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages insights into the phagosomal environment. *The Journal of experimental medicine*, *198*(5), pp.693-704.
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. *Scientific reports*, *5*.
- Sivaramakrishnan, S., Ouellet, H., Matsumura, H., Guan, S., Moënne-Loccoz, P., Burlingame, A.L. and Ortiz De Montellano, P.R., 2012. Proximal ligand electron donation and reactivity of the cytochrome P450 ferric–peroxo anion. *Journal of the American Chemical Society*, *134*(15), pp.6673-6684.
- Syed, K. and 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*(4), p.e95616.
- Syed, K. and Yadav, J.S., 2012. P450 monooxygenases (P450ome) of the model white rot fungus Phanerochaete chrysosporium. *Critical reviews in microbiology*, *38*(4), pp.339-363.
- Syed, K., Nelson, D.R., Riley, R. and Yadav, J.S., 2013. Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore

species Bjerkandera adusta, Ganoderma sp. and Phlebia brevispora. *Mycologia*, *105*(6), pp.1445-1455.

- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and Pagadala, N.S., 2014. Genome-wide identification, annotation and characterization of novel thermostable cytochrome P450 monooxygenases from the thermophilic biomass-degrading fungi Thielavia terrestris and Myceliophthora thermophila. Genes & Genomics, 36(3), pp.321-333.
- Syed, K., Shale, K., Nazir, K.N.H., Krasevec, N., Mashele, S.S. and Pagadala, N.S., 2014b. Genome-wide identification, annotation and characterization of novel thermostable cytochrome P450 monooxygenases from the thermophilic biomass-degrading fungi *Thielavia terrestris* and *Myceliophthora thermophila*. Genes & Genomics, 36(3), pp.321-333.
- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014a. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. PLoS One, 9(1), p.e86683.
- Tailleux, L., Waddell, S.J., Pelizzola, M., Mortellaro, A., Withers, M., Tanne, A., Castagnoli, P.R., Gicquel, B., Stoker, N.G., Butcher, P.D. and Foti, M., 2008. Probing host pathogen cross-talk by transcriptional profiling of both Mycobacterium tuberculosis and infected human dendritic cells and macrophages. *PloS one*, *3*(1), p.e1403.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution*, *30*(12), pp.2725-2729.

- Thomas, S.K., Iravatham, C.C., Moni, B.H., Kumar, A., Archana, B.V., Majid, M., Priyadarshini, Y., Rani, P.S., Valluri, V., Hasnain, S.E. and Ahmed, N., 2011. Modern and ancestral genotypes of Mycobacterium tuberculosis from Andhra Pradesh, India. *PloS one*, *6*(11), p.e27584.
- Zuckerkandl, E. and Pauling, L., 1965. Evolutionary divergence and convergence in proteins. *Evolving genes and proteins*, *97*, pp.97-166.

APPENDIX

>CYP125A(MAB_0611)
MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIARAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP
ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS*

>CYP125A(MAB_0613)
MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHYCIGANLARLTINLMFN
AIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ*

>CYP125A(MAB_1211c)
MTAMKTAAELGLPEGFDFTDPELYGNRMPHEEFATLRREAPVWWNPQPRTVGGFADEGYWVISKHRDVREVSLHTDTFSSGRKGAIPRLEDHIS
PEEFQATLSVLINKDAPEHTQLRGLVSRMFTPRSIAALRITLEERAERIVRAALEGGHGEFVREVASELPMQAIAELIGVPEEDRVKLFEWSNQ
MTGYDEADVEIDPRVGAAQILGYSYQLAEQRRDCPGNDVVSRLLTGTVDGEQLTPEQFGFFVVMLSVAGNETTRNATTMGMMAFLEHPGQWELF
KSARPSTTVDEIVRYTSPLISQQRTALQDTVISDVRIRAGERVVMLYPSANFDEEVFENPHTFDITRDPNPHLGFGGTGAHYCLGANLAKAELE
IIFNKIADRMPDISRIGDAPRFHSGWINGIKKFDTAYCPVTH*

>CYP125A(MAB_0101)
MTTCPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAVNGLHARLVERAHGIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATTATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR
PSTAADEIIRWASPIIAFQRTALQDVELGGVQIRKDQRVGMFYASANFDEDVFDDPFAFNIERDPNPHLAFGGHGIHY
CLGANLARLEIGIMFDALADRLPDLMPTGAPTRFRSGWINGVVALPANYHGSGPRG*

>CYP125A1(MAF_35570)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125NS(MAV_1637)
MNVNAATAACGDDPAERGLAMTTAAVDLSDFSLWCNGFPDELFTELRRTRPLFHHDLTPGVAATVHRDFWVATKHRHAVRLHRDTESFTAADGP
LIQPVAMFSSSPTIITMDPPELNKRRKLISNAFNPRAIAKLEDGIRARAARMIDNLLAHGGGDWIEDVADALPMTVIGDILGIPERDRPRIFDL
FDRILKALAPDAHPRGGVELELFASVFDYAMQLTADKRRNPTGDIWSTLATAVITGEDGEEFRLPANELEFFFFVLAFAGSDTTKNALAIGLQA
FLANPGQVERYRADEALRPTAVEEVLRWASPVAYWTRTAKVDVEMDGQRIAKGERVVSMLRSANRDEEVFDAPFTFDIGRQPNPHVAFGGGGPH
HCLGAMLARAELRAVFDELLLRCDDIEIGPAKAAYPNLITNMSIYDEMPISLRRR*

>CYP125A(MAV_0616)
MPSPNLPPGFDLLDPDVCVKGLPVAELAELRKSAPIYWMDVPGGTGGFGDKGYWAITKHKDVKEISVRSDIFSSQQDCAIPVWPKEMTREQIDL
QRNVMLNMDAPHHTRLRKIISRGFTPRAVGRLRDELDARAQNIAKTAAAAGAGDFVEQVSCELPLQAIAGLLGVPQEDRDKIFRWSNEMTGNED
PEYAHIDPAMSSAELIMYAMKMAEERAKNPGDDIVTQLIQADLDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFADNPDQWELFKKERP
ETAPDEIVRWATPVTAFQRTALEDYELSGVQIKKGQRVVMFYRSANFDEEVFEDPHRFNILRNPNPHVGFGGTGAHYCIGANLARMTISLIFNA
VADHMPDLKPLSAPERLRSGWLNGIKHWQVDYTGKCPVAH*

>CYP125A(MAV_2811)
MATVEPTTKPVPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPIWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSLRSDVFSSLQKTALPRYKDG
TVAEQVERGKFVLLNMDAPQHTRLRKIISRAFTPRAVERLRDDLRERARRIVEAAAAEGSGDFVEQVSCELPLQAIASLMGVPQEDRKKLFHWS
NEMVGDQDPEFASNDAITASVELIMYGMQMAADRAKNPGEDLVTKLVQADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELFKRERPATAADEIVRWATPVTSFQRTALQDYELSGVKIKKGQRVVMFYRSANFDEDVFDDPFTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADAMPDLESIGKPERLRSGWLNGIKHWQVDYHTNGSSKCPVAH*

>CYP125NS(MAP2344)
MNVNAATAACGDDPAERGSAMTTAAVDLSDFSLWCNGFPDELFAELRRTRPLFHHDLTPGVAATVHRDFWVATKHRHAVRLHRDTESFTAADGP
LIQPVAMFSSSPTIITMDPPELNKRRKLISNAFNPRAIAKLEDGIRARAARMIDSLLAHGGGDWIEDVADALPMTVIGDILGIPERDRPRIFDL
FDRILKALAPEAHPRGGVELELFASVFDYAMQLTADKRRNPTGDIWSTLATAVITGEDGEEFRLPANELEFFFFVLAFAGSDTTKNALAIGLQA

FLANPEQVERYCADEALRPTAVEEVLRWASPVAYWTRTAKVDVEMDGQRIAKGERVVSMLRSANRDEEVFDAPFTFDIGRQPNPHVAFGGGGPH
HCLGAMLARAELRAVFDELLLRCDDIEIGPAKAAYPNLITNMSIYDEMPISLRRR*

>CYP125A(MAP1614c)
MATVEPTTKPVPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPIWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSLRSDVFSSLQKTALPRYKDG
TVAEQVERGKFVLLNMDAPQHTRLRKIISRAFTPRAVERLRDDLRERARRIVEAAAAEGSGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFHWS
NEMVGDQDPEFASNDAITASVELIMYGMQMAADRAKNPGEDLVTKLVQADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELFKRERPATAADEIVRWATPVTSFQRTALQDYELSGVKIRKGQRVVMFYRSANFDEDVFDDPFTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADAMPDLESIGKPERLRSGWLNGIKHWQVDYHTNGSSKCPVAH*

>CYP125A(MAP0522)
MPSPNLPPGFDLLDPDVCVKGLPVAELAELRKSAPIYWVDVPGGTGGFGDKGYWAITKHKDVKEISVRSDIFSSQQDCAIPVWPKEMTREQIDL
QRNVMLNMDAPHHTRLRKIISRGFTPRAVGRLRDELDARAQNIAKTAAAAGAGDFVEQVSCELPLQAIAGLLGVPQEDRDKIFRWSNEMTGNED
PEYAHIDPAMSSAELIMYAMKMAEERAKNPGDDIVTQLIQADLDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFADNPDQWELFKKERP
ETAPDEIVRWATPVTAFQRTALEDYELSGVQIKKGQRVVMFYRSANFDEEVFEDPHRFNILRNPNPHVGFGGTGAHYCIGANLARMTISLIFNA
VADHMPDLKPLSAPERLRSGWLNGIKHWQVDYTGKCPVAH*

>CYP125F1(MAP3818)
MRTPVTVGQHRHPFGRDIYVGRSGYVTEDAISIGGVNLADPDTYRAGMPYGAFRKLRERAPVAWHPQKDGSGFWALTGYEEIHAVSRDSATWSS
QINGAMFDAPPPGEVPPVMIFMDPPQHTALRKLINKGFTPRQVTRLNEHIVEMAKQIVDDVIERGECEFADDVAGALPSYVIAEMLGIPLEDGR
RLYQITEILHTGSVGDSDDERQQAMVEMFQYGVELAVRKRAEPGDDIATSLLHAEVDGQSLSDLEFNLFFMLLIDAGGDTTRNLVAAGILALLE
HPQELQRLKADPSLMPTAIEEMLRYTSPVTAFLRTATKDTELRGVPVKAGERVAMFYPSGNRDDSHFADPDRLDVGRAPNPHLAFGGGGTHFCL
GANLARVEASAMVPEVLSRMNDLELAGPVERLRSDLINGIRSMPVRFTPGKRLGTA*

>CYP125A1(MbovAF2122/97,NP_857214.1)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(MbovBCG)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQD
RGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMA
FAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHYC
IGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A(MIP_04041)
MSTVEPSTKSVPNLPPGFDFTDPDIHAERLPVEELAELRRTAPIWWNEQPIGAGGFDDGGFWVVSKHKDVKEVSLRSDVFSSLQKTALPRYKDG
TVGEQIERGKFVLLNMDAPQHTRLRKIISRAFTPRAIERLREDLAERARHIVEQAAAEGRGDFVEQVSCELPLQAIAGLMGVPQEERKKLFHWS
NEMVGDQDPEFANSDAITASVELIMYGMQMAAERTKNPGEDLVTKLVQADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELFKKERPGTAADEIVRWATPVTSFQRTALEDYELSGVRIKKGQRVVMVYRSANFDEDVFDDPFTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADGIPDLESIGKPERLRSGWLNGIKHWQVDYHTDGKAKAPAAH

>CYP125A(MIP_00975)
MPSPNLPPGFDFLDPDVSVKGLPVAELAEVRKSEPIFWVDVPGGTGGFGDKGYWAITKHKDVKEISVRSDIFSSQQDCAIPVWPQEMTREQIDL
QRNVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELDARAQNIAKTAASSGSGDFVEQVSCELPLQAIAGLLGVPQEDRDKIFRWSNEMTGNED
PEYAHIDPAMSSAELIMYAMKMAEERAKNPGDDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFADNPDQWELFKKERP
DTAPDEIVRWATPVTAFQRTALEDYELSGVQIKKGQRVVMFYRSANFDEEVFEDPHSFNIMRNPNPHVGFGGTGAHYCIGANLARMTISLIFNA
VADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGKCPVAH

>CYP125NS(MIP_02350)
MTAAAVDLSDLSLWCNGFPDDLFAELRRSSPLFRHELTPGVAQTVHRDFWMATKHRHAVRLHRDTESFTAADGPLIQPVAMFSSFPTIITLDPP
ELNRRRKLISNAFNPRAIAKLEEGIRARAARMIDDLLAAGGGDWIEDVADALPMSVIGDIIGIPDEDRPRIFGNFDQILKALAPEAHPSGRVEL
DLFASVFRYAMELTAEKRRNPTDDIWSTLATAVITGEDGERFSLPENELEFFFFVLAFAGSDTTKNALAIGLQAFVRNPQQIERYRAQEALRPG
AVEEVLRWASPVAYWTRTAKVDVEMDGQRIAKGERVVSMLRSANRDEEVFDSPFVFDIGRQPNPHVAFGGGGPHHCLGAMLARAELRAVFDELL
LRCDDIEIGPAKAAHPNLTTNMSIYDEMAISLRERR

>CYP125D(MIP_02985)
MSIAKPTLVKSLVPQNLDTAADRDAAAVLDPDTFVTGAPYDAMTRLRAASPVHPVQLPGLPRAWLLTKHADVRLVSRDTDTFTSSKGNTLVEAE
AGPNSAMLPGIDPPRHVHFRKLINQGFTVRNVQRLEPRMRQVARGIVAAITDKREFDAVTDISAEMSLQVIADVLGVPAEDRMDVFRWSNAIGS
LGIEDPDYAPTPEALGQAAAEMFAYCGELVEHRRKHGLTDDILSALLAAEVDGEKLNRDQLNEFFLLLAIAGNETTRNTLSHGILALAERPEQQ

ALLARDPAAIKPAVEELLRWATPVMHFRRTVVRDVEIRGQRIPSGDWVLMHYLSANRDEEVFDRPDQFDVTRPDAGHAAFGGGGVHFCLGAQLA
RLELRVMLEELYANVPGLAVTGPPDRLRSSFFHGIKRLPCTT

>CYP125A(MIP_06760)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRRIVSRAFTPRAVERLRAELSERARCIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDYPVQWELFKARRP
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125A6(MMAR_2783)
MPAAEPTATSVPNLPPGFDFTDPDIYAERLPVAELAEMRRSAPIWWNEQPTGCGGFDDGGFWVVTKHKDVKEISLRSDVFSSLQKTALPRYKDG
TVDEQIERGKFVLLNMDAPQHTRLRKIVSRAFTPRAVERLRDDLRERARRIVEAAAAEGRGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFHWS
NEMVGDQDPEFATNDALTASVELIMYGMQMAADRAKNPGQDLVTKLVEADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELYKRERPVTTADEIVRWATPVTSFQRTALQDYELSGVRIKKGQRVVMFYRSANFDEDVFDDPYTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADVMPDLESISQPERLRSGWLNGIKHWQVDYHSDSSGKCPVAH*

>CYP125A7(MMAR_5032)
MPCPNLPPGFDFTDPDIYAERLPVEEFAELRSSEPIWWDEQLPGQGGGFHDGGFWAITKLKDVKEVSRRSDVFSSYENGVIPRFKNDIAREDID
VQRFVMLNMDAPHHTRLRKIISRGFTPRAIGRLHDELNDRAQNIAKAAAAAGSGDFVEQVSCELPLQAIAGLLGIPQEDRGKLFDWSNEMTGTE
DPEFAHIDAKASSVELIGYAMKMAEEKAKNPGDDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFADNPEQWELYKRER
PETAADEIVRWATPVTSFQRTALEDYELSGVQIKKGQRVLMFYRSANFDEEVFEDPFSFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLTPIAAPERLRSGWLNGIKHWQVDYTGKCPVSH*

>CYP125A5P(MSMEG_5853)
MATMNTCPFGNGYDFTDPDVLFRGIPVEEFAVLRKTAPVWWNQQGESIFDDGGYWVISRHEDIKTISRDGGEVWSTNAKGAVMRLPDGVTAEQL
DLTKALLINHDAPEHTRLRKLVSRLFTPRSVAALEEKLAVAAHDIVAEAKAEGSGNFVEDIAMKLPLLAIADLIGVPEADREKIFAWSNAIINT
DDPDFDSDPTVANAELMGYAYTMAEERRRCPADDIVTRLVQADVSGESLGEVEFAFFVILLAVAGNETTRNAITHGMNAFFENPDQWELFKRER
PITAVDEIVRWATPVHCFQRTAVVDTEIGGVPIKAGQRAGLFYSSANYDEDVFDDPFRFDILRDPNPHLGFGGNGAHYCIGANLARMEIRLMFD
EIADQIPDITKVGEPQRLRSGWINGVKDLQVSYRG*

>CYP125A4(MSMEG_3524)
MVMSDSALHLPAGFDFTDPDIYAERLPVDELAELRRVAPIWWNAQPIGAGGFDDGGFWVVTKHKDVKEISLRSDVFSSLEKTALPRYPEGTVQD
QIEQGRFVLLNMDAPHHTHLRKIISRAFTPRAVERLRDDLAERARAIVRAAAEEGSGDFVEQVACELPLQAIAGLMGVPQEDRRKLFDWSNQMV
GNQDPEFVANDGASAAVELITYGMQLAAQRAAAPGDDLVTKLVQADVEGHKLSDDEFGFFVVLLAVAGNETTRNSITQGMMAFTDHRDQWELFK
RERPVTTADEIVRWATPVTSFQRTALADTEVSGVRIKKGQRVVMFYRSANFDEDVFTDPYRFDILRDPNPHVGFGGTGAHYCIGANLARMTIDL
IFNAIADEMPDLTPISEPVRLRSGWLNGIKHWQVDYRGDAAKHRAAQASSQADR*

>CYP125A3(MSMEG_5995)
MPTPNIPSDFDFLDATLNLERLPVEELAELRKSEPIHWVDVPGGTGGFGDKGYWLVTKHADVKEVSRRSDVFGSSPDGAIPVWPQDMTREAVDL
QRAVLLNMDAPQHTRLRKIISRGFTPRAIGRLEDELRSRAQKIAQTAAAQGAGDFVEQVSCELPLQAIAELLGVPQDDRDKLFRWSNEMTAGED
PEYADVDPAMSSFELISYAMKMAEERAVNPTEDIVTKLIEADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFAQNPDQWELYKKERP
ETAADEIVRWATPVSAFQRTALEDVELGGVQIKKGQRVVMSYRSANFDEEVFEDPHTFNILRSPNPHVGFGGTGAHYCIGANLARMTINLIFNA
IADNMPDLKPIGAPERLKSGWLNGIKHWQVDYTGAGKASVSGAPGTCPVAH*

>CYP125A(Mjls_2753)
MSSDRLRPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPVWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSRRSDVFSSLEKTALPRYRDGTVGE
QIERGKYVLLNQDAPHHTHLRQIVSRAFTPRAVERLRAELDARAQQIARTAREQGSGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFDWSNQMV
GDQDPEFAGNDAIGASVELIMYGMQMAADRVANPGDDLVTKLVQADVEGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQWELFK
RERPATAADEIVRWATPVTSFQRTALCDTELSGVTIKKGQRVVMFYRSANFDEDVFTDPYSFGILRDPNPHVGFGGTGAHYCIGANLARMTIDL
MFNAIADHMPDLTPVGKPERLRSGWLNGIKHWQVDYTGSAAKPPAAQ

>CYP125A(Mkms_4660)
MTTMESRCPFGPGFDFTDPDVLVQGIPVNEFAQLRKTAPVWWNEQQESIFDDGGYWVISRHEDIKSISRNGDLWSTNAKGAVMRLPEGVTAEQL
DLTKALLINHDAPEHTRLRKIVSRLFTPRSVAALEEKLAISARQIVAAAREKGSGDFVTDIAMSLPLQAIADLIGVPEADREKLFHWTNCIMNT
DDPDFDSDPTVANAELMGYAYNMAEERRRCPADDIVTRLIQADIDGESLGDVEFAFFVILLAVAGNETTRNAMTHGMNAFFEHPDQWELFVRER
PETAVDEIVRWATPVHCFQRTALADVELGGVTIREGQRAGLFYSSANYDEDVFQSPFEFDILRDPNPHLGFGGNGAHYCIGANLARMEIKLIFN
ELADQIPDIAKLGEPQRLRSGWINGVKELPVSYRG*

>CYP125A(Mkms_2767)
MSSDRLRPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPIWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSRRSDVFSSLEKTALPRYRDGTVGE
QIERGKYVLLNQDAPHHTHLRQIVSRAFTPRAVERLRAELDARAQQIARTAREQGSGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFDWSNQMV
GDQDPEFAGNDAIGASVELIMYGMQMAADRVANPGDDLVTKLVQADVEGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQWELFK

RERPATAADEIVRWATPVTSFQRTALCDTELSGVTIKKGQRVVMFYRSANFDEDVFTDPYSFDILRDPNPHVGFGGTGAHYCIGANLARMTIDL
MFNAIADHMPDLTPVGKPERLRSGWLNGIKHWQVDYTGSAAKPPAAQ*

>CYP125A(Mkms_4763)
MPGPNSCPIAPDFDFLDANLNLERLPVAELAELRKSEPVHWVDVPGGTGGFGDKGYWLVTKHADVKDVSKRNDVFGSSPDGAIPVWPQDMTRDA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRARAQKIAETAAAEGAGDFVEQVSCELPLQAIAELLGVPQDDRDKLFRWSNEMTA
GEDPEYADVDPAMSSFELITYAMKMAEERAKNPTEDIVTKLIEADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQNPDQWELYKE
RPETAADEIVRWATPVSAFQRTALEDTELGGVQIKKGQRVVMSYRSANFDEEVFENPYQFDILRNPNPHVGFGGTGAHYCIGANLAKMTINLIF
NAIADKMPDLKPIGQPERLKSGWLNGIKHWQVDYTGAGGPAIEQKCPVAH*

>CYP125A(Mmcs_4572)
MTTMESRCPFGPGFDFTDPDVLVQGIPVNEFAQLRKTAPVWWNEQQESIFDDGGYWVISRHEDIKSISRNGDLWSTNAKGAVMRLPEGVTAEQL
DLTKALLINHDAPEHTRLRKIVSRLFTPRSVAALEEKLAISARQIVAAAREKGSGDFVTDIAMSLPLQAIADLIGVPEADREKLFHWTNCIMNT
DDPDFDSDPTVANAELMGYAYNMAEERRRCPADDIVTRLIQADIDGESLGDVEFAFFVILLAVAGNETTRNAMTHGMNAFFEHPDQWELFVRER
PETAVDEIVRWATPVHCFQRTALADVELGGVTIREGQRAGLFYSSANYDEDVFQSPFEFDILRDPNPHLGFGGNGAHYCIGANLARMEIKLIFN
ELADQIPDIAKLGEPQRLRSGWINGVKELPVSYRG*

>CYP125A(Mmcs_2723)
MSSDRLRPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPIWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSRRSDVFSSLEKTALPRYRDGTVGE
QIERGKYVLLNQDAPHHTHLRQIVSRAFTPRAVERLRAELDARAQQIARTAREQGSGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFDWSNQMV
GDQDPEFAGNDAIGASVELIMYGMQMAADRVANPGDDLVTKLVQADVEGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQWELFK
RERPATAADEIVRWATPVTSFQRTALCDTELSGVTIKKGQRVVMFYRSANFDEDVFTDPYSFDILRDPNPHVGFGGTGAHYCIGANLARMTIDL
MFNAIADHMPDLTPVGKPERLRSGWLNGIKHWQVDYTGSAAKPPAAQ*

>CYP125A(Mmcs_4677)
MPGPNSCPIAPDFDFLDANLNLERLPVAELAELRKSEPVHWVDVPGGTGGFGDKGYWLVTKHADVKDVSKRNDVFGSSPDGAIPVWPQDMTRDA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRARAQKIAETAAAEGAGDFVEQVSCELPLQAIAELLGVPQDDRDKLFRWSNEMTA
GEDPEYADVDPAMSSFELITYAMKMAEERAKNPTEDIVTKLIEADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQNPDQWELYKK
ERPETAADEIVRWATPVSAFQRTALEDTELGGVQIKKGQRVVMSYRSANFDEEVFENPYQFDILRNPNPHVGFGGTGAHYCIGANLAKMTINLI
FNAIADKMPDLKPIGQPERLKSGWLNGIKHWQVDYTGAGGPAIEQKCPVAH*

>CYP125A1(TBCG_03474.1)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(TBFG_13578.4)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(MRA_3584)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(Rv3545c)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(TBHG_03485.1)
MSGNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM

AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH*

>CYP125A1(TBMG_03584)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNTLRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A7(MUL_4106)
MPCPNLPPGFDFTDPDIYAERLPVEEFAELRSSEPIWWDEQFPGQGGGFHDGGFWAITKLKDVKEVSRRSDVFSSYENGVIPRFKNDIAREDID
VQRFVMLNMDAPHHTRLRKIISRGFTPRAIGRLHDELNDRAQNIAKAAAAAGSGDFVEQVSCELPLQAIAGLLGIPQEDRGKLFDWSNEMTGTE
DPEFAHIDAKASSVELIGYAMKMAEEKAKNPGDDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFADNPEQWELYKRER
PGTAADEIVRWATPVTSFQRTALEDYELSGVQIKKGQRVLMFYRSANFDEEVFEDPFSFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLKPIAAPERLRSGWLNGIKHWQVDYTGKCPVSH*

>CYP125F2(Mvan_0246)
MATDAISIGGVDLADPDTYVGGMPHGAFRELRRHAPVAWHPYGDNPGFWALTGYDEVLAVSRDSRTWSSQTTGVFLDVPAPEDSYQLSLMMLTM
DPPRHTALRALVSRGFTPRHLARLNARTADMARDILDAALQRGECEFVDDVAGALPSYVIAELLGIPLDDGRRLYALTEIMNTRPLHDPELMQT
QVELFGYAGDLAASKRAAPGDDIATALLHAEVDGQRLTDLEFNLFFMLLLNAGGDTTRNLVAAGTLALIEHPEQWARLAADPSLMPTAIEEMLR
WTSPVNVFTRTATRDTEVGGVPLRAGERVAMFYPSANRDEKHFADPDRFDIGRAPNHHLAFGGGGTHFCLGASLARVEATAIFGEILTRTAHIE
LAGPVERVRSVLMNGIRSMPVRLTPASVPA*

>CYP125A10(Mvan_3012)
MATPTLPPGFDFTDPDLNLERLPVEELAELRRCAPIWWNEQTSGGAGPFGDGGYWVVTKHRDVKEISKHSEVFSSQQKTALPRYPEGSTTEQVE
TGSLVLLNMDAPRHTHLRKIISRGFTPRAVERLREDLAQRAHNIAKSAAAAGAGDFVEQVSCELPLQAIAGLLGVPLEDRKKLFDWSNQMVSDD
DPEFAHYDNRNAATELIMYAMQLAALRAEQPGEDIVTKLIEADVDGHKLTDDEFGFFMVLLAVAGNETTRNSITHGMIAFTEHPDQWELFKRER
PATAVDEIVRWATPVTSFQRTALRDYELSGVQIKKGQRVVMSYRSANFDEEVFDDPFTFDIMRDPNPHVGFGGTGAHYCIGANLARMTIDLMFN
AIADHLPDLSSAGTPDRLRSGWLNGIKHWQVDYTGPSGCPVAH*

>CYP125A11(Mvan_5151)
MTATQSCPFLPHGYDFTDPDVLLKGIPVTEFAELRRTAPVWWNEQADSIFDDGGYWVISRHEDVKAISRNSTQWSTNTKGAVMRLPDGVTAEQL
DLTKALLINHDAPEHTRLRKIVSRLFTPRAIAGMEDRLADAAREIVRSAAEKDSGDFVDDVAMMLPLQAIADLIGVPEEDREKLFHWTNAIMNT
DDPEFDADPTMANAELMGYAYSMAEERRRCPADDIVTRLVQADIDGESLGEVEFAFFVILLAVAGNETTRNAMTHGMNAFFDNPAQWELFKRER
PETAIDEIIRWATPVHCFQRTALEDVEVGGVTIAEGQRVGLFYSSANFDEDVFDRPFDFDILRDPNPHLAFGGNGAHFCIGANLARMEIKLMFN
EIADQIPDISKLAEPQRLRSGWINGVKNLQVAYR*

>CYP125A9(Mvan_5258)
MPGPNSCPISPEFDFLDASLNLERLPVEELAELRKSEPVHWVDVPGGTGGFGDRGYWLVTKHADVKEVSKHNEIFGSSPDGAIPVWPQEMTREA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRARAQRIAATAATEGSGDFVEQVSCELPLQAIAELLGVPQEDRDKLFRWSNEMTA
GEDPEYADVDPAMSSFELITYAMKMAEERAKNPTEDIVTKLIEADIEGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSRNPDQWELYKK
ERPETAADEIVRWATPVSAFQRTALEDTELGGVQIKKGQRVVMSYRSANFDEEVFENPHSFDIMRNPNPHVGFGGTGAHYCIGANLAKMTINLM
FNAIADAMPDLKPIGDPERLKSGWLNGIKHWQVDYTGQCPVQH*

>CYP125A1(MRGA423_22410)
MPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYENGVIPRFKNDIAREDIE
VQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQEDRGKLFHWSNEMTGNE
DPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFAEHPDQWELYKKVR
PETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(CCDC5079_3286)
MPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYENGVIPRFKNDIAREDIE
VQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQEDRGKLFHWSNEMTGNE
DPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFAEHPDQWELYKKVR
PETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A(MASS_0104)
MTACPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAINGLHSRLVERAHSIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATAATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR

PATAADEIIRWASPIIAFQRTALQDVELGGVQICKDQRVGMFYASANFDEDVFDDPFTFNIERDPNPHLAFGGHGIHYCLGANLARLEIGIMFD
ALADRLPDLMPTGAPTRFRSGWINGVVALPANYLGAGPRG

>CYP125A(MASS_0581)
MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIAKAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP
ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS

>CYP125A(MASS_0583)
MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHYCIGANLARLTINLMFN
AIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ

>CYP125NS(MAP4_1479)
MNVNAATAACGDDPAERGSAMTTAAVDLSDFSLWCNGFPDELFAELRRTRPLFHHDLTPGVAATVHRDFWVATKHRHAVRLHRDTESFTAADGP
LIQPVAMFSSSPTIITMDPPELNKRRKLISNAFNPRAIAKLEDGIRARAARMIDSLLAHGGGDWIEDVADALPMTVIGDILGIPERDRPRIFDL
FDRILKALAPEAHPRGGVELELFASVFDYAMQLTADKRRNPTGDIWSTLATAVITGEDGEEFRLPANELEFFFFVLAFAGSDTTKNALAIGLQA
FLANPEQVERYCADEALRPTAVEEVLRWASPVAYWTRTAKVDVEMDGQRIAKGERVVSMLRSANRDEEVFDAPFTFDIGRQPNPHVAFGGGGPH
HCLGAMLARAELRAVFDELLLRCDDIEIGPAKAAYPNLITNMSIYDEMPISLRRR

>CYP125A(MAP4_2225)
MATVEPTTKPVPNLPPGFDFTDPDIYAERLPVEELAEMRRVAPIWWNEQPIGAGGFDDGGFWVVTKHKDVKEVSLRSDVFSSLQKTALPRYKDG
TVAEQVERGKFVLLNMDAPQHTRLRKIISRAFTPRAVERLRDDLRERARRIVEAAAAEGSGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFHWS
NEMVGDQDPEFASNDAITASVELIMYGMQMAADRAKNPGEDLVTKLVQADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELFKRERPATAADEIVRWATPVTSFQRTALQDYELSGVKIRKGQRVVMFYRSANFDEDVFDDPFTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADAMPDLESIGKPERLRSGWLNGIKHWQVDYHTNGSSKCPVAH

>CYP125A(MAP4_3345)
MPSPNLPPGFDLLDPDVCVKGLPVAELAELRKSAPIYWVDVPGGTGGFGDKGYWAITKHKDVKEISVRSDIFSSQQDCAIPVWPKEMTREQIDL
QRNVMLNMDAPHHTRLRKIISRGFTPRAVGRLRDELDARAQNIAKTAAAAGAGDFVEQVSCELPLQAIAGLLGVPQEDRDKIFRWSNEMTGNED
PEYAHIDPAMSSAELIMYAMKMAEERAKNPGDDIVTQLIQADLDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFADNPDQWELFKKERP
ETAPDEIVRWATPVTAFQRTALEDYELSGVQIKKGQRVVMFYRSANFDEEVFEDPHRFNILRNPNPHVGFGGTGAHYCIGANLARMTISLIFNA
VADHMPDLKPLSAPERLRSGWLNGIKHWQVDYTGKCPVAH

>CYP125F1(MAP4_3931)
MRTPVTVGQHRHPFGRDIYVGRSGYVTEDAISIGGVNLADPDTYRAGMPYGAFRKLRERAPVAWHPQKDGSGFWALTGYEEIHAVSRDSATWSS
QINGAMFDAPPPGEVPPVMIFMDPPQHTALRKLINKGFTPRQVTRLNEHIVEMAKQIVDDVIERGECEFADDVAGALPSYVIAEMLGIPLEDGR
RLYQITEILHTGSVGDSDDERQQAMVEMFQYGVELAVRKRAEPGDDIATSLLHAEVDGQSLSDLEFNLFFMLLIDAGGDTTRNLVAAGILALLE
HPQELQRLKADPSLMPTAIEEMLRYTSPVTAFLRTATKDTELRGVPVKAGERVAMFYPSGNRDDSHFADPDRLDVGRAPNPHLAFGGGGTHFCL
GANLARVEASAMVPEVLSRMNDLELAGPVERLRSDLINGIRSMPVRFTPGKRLGTA

>CYP125A1(K60_036830)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(BCGMEX_3607c)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(JTY_3610)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM

AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(MCAN_35561)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRSPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(BN44_110037)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRSPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(BN42_90040)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKERPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVLFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISEPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A(Mycch_2866)
MPTPTLPPGFDFTDPDLNRDRLPVEELAELRRTAPIWWNEQTAGGAGPFGDGGYWVVTKHRDVKEISKRSDVFSSLEKTALPRYPEGSTGEQIE
TGKFVLLNMDAPHHTHLRKIISRGFTPRAVERLRDDLRNRAQAIAKQAAAEGSGDFVEQVSCELPLQAIAGLIGVPLEDRKKLFDWSNQMVSDD
DPEYTHYDNRNAATELIMYAMQLAALRAEQPGEDIVTKLIEADVEGHKLSDDEFGFFMVLLAVAGNETTRNSITHGMVAFTEHPDQWELFKRER
PFTAIDEIVRWATPVSSFQRTALTDYELSGTLIKKGQRVVMSYRSANFDEDVFDDPFTFNILRDPNPHVGFGGTGAHYCIGANLARMTIELMFN
AIADHIPDLTALTAPDRLRSGWLNGIKHWQVDYTGRSGCPVAH

>CYP125F(Mycch_4146)
MIEDAISIGGVDLTDPNSYLTGPPLDAFRKLRERAPVAWHPFQEGPGFLALTGYDEVVAVSRDSATWSSETDGVYFEAPGPDSPADMRGVMMLT
MDPPRHTALRKLVNKGFTPRQVARLNERIAEMARDLVDNVIEQGECDFVQQVAGALPSYVISEMLGIPLEDGFRLYELTEITNAGQVQDARIAE
AGMQIFAYAAELAARKRVEPGDDIATSLLNAEINGQGLTDMEFNFFFILLLNAGGDTTRNLVAGGMLALMENPAELAKLEKDTSMMSTAVEEML
RYISPVMWFLRTATRDTEVRGMPVEKGGRVAMFYPSANRDETKFPDPDTFDITRTPNPHVAFGGGGTHFCLGANLARVESSALLSEVLARMKNV
ELAGPVQRMQSMFINGIHSMPVRFTPAHRLGRR

>CYP125A(Mycch_4512)
MTTTDTGPQSCPFLPSGYDFTDPDVLLAGIPVAEFAQLRKTAPVWWNAQAESIFDDGGYVISRHEDIKSISRNSAAWSTNANGAVMRLPDGVTA
EQLDLTKALLINHDAPEHTRLRKIISRLFTPRAIAGMEEKLAVSAREIVRTAAEKDTGNFVQDVAMLLPLQAIADLIGVPEADRGKLFGWTNAI
MNTDDPEFDSDPTTANAELMGYAYTMAEQRRRCPADDIVTRLIQADIDGEALGDVEFAFFVILLAVAGNETTRNAMTHGMNAFFDNPDQWELFR
RERPETAVDEIIRWATPVHCFQRTALEDVELGGVTIAKGQRAGLFYSSANFDEDVFDRPFAFDILRNPNPHLAFGGNGAHFCIGANLARMEIKL
MFDEIADQIPDISKLAEPQRLRSGWINGVKDLQVSYH

>CYP125A9(Mycch_4638)
MPGPNSCPISPDFDFLDASLNLERLPVEELAELRKSEPIHWVDVPGGTGGFGDKGYWLVTKHADVKEVSKRNDIFGSSPDGAIPTWPQDMTRDA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAIGRLEDELRARAQKIAETAKAEGSGDFVEQVSCELPLQAIAELLGVPQDDRDKIFRWSNEMTA
GEDPEYAEVDPAMSSFELIQYAMKMAEERAKNPTEDIVTKLIEADIEGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQNPEQWELFKK
DRPETAADEIVRWATPVSAFQRTALEDTELGGVKIKKGERVVMSYRSANFDEEVFDDPHSFNILRSPNPHVGFGGTGAHYCIGANLAKMTINLI
FNAVADHMPDLKPVGEPERLKSGWLNGIKHWQVDYTGQCPVSH

>CYP125A(Mflv_1508)
MPGTNSCPISPDFDFLDATLNLERLPVEELAELRHSEPVHWVDVPGGTGGFGDKGYWLVTKHADVKEVSKRNDIFGSSPDGAIPVWPQDMTRDA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRARAQRIAETAAAAGSGDFVEQVSCELPLQAIAELLGVPQDDRDKLFRWSNEMTA
GEDPEYADVDPAMSSFELITYAMKMAEERAKNPTEDIVTKLIEADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQNPEQWELYKK
ERPETAADEIVRWATPVSAFQRTALEDTELGGVQIKKGQRVVMSYRSANFDDEVFDNPHSFDITRNPNPHVGFGGTGAHYCIGANLAKMTINLI
FNAIADAMPDLKPIGEPERLKSGWLNGIKHWQVDYSGKGCPVSH
*

>CYP125A(Mflv_1607)
MTAEDARVQSCPFLPDGYDFTDPDVLLEGVPVAEFAQLRRTAPVWWNAQQESIFDDGGYWVVSRHEDIKSISRNSALWSTNAKGAVMRLPDGVT
AEQLDLTKALLINHDAPEHTRLRKIVSRLFTPRAIAGMEAKLADSARQIVGEAADKAGGNFVDDVATLLPLQAIADLIGVPEEDREKLFHWSNA
IMNTDDPDFDSDPTIANAELMGYAYTMAEQRRRCPADDIVTRLVEADLDGDALGEVEFAFFVILLAVAGNETTRNAMTHGMNAFLDHPDQWELF

KRERPETAIDEIIRWATPVHCFQRTALDDVEIGGVTVARGQRVGLFYSSANFDEDVFDNPFNFDILRNPNPHLAFGGNGAHYCIGANLARMEIK
LMFEAIADRLPDISKRAEPQRLRSGWINGVKDLQVAYR*

>CYP125F3(Mflv_0425)
MPADAITLGGADLADPDTYLAGMPYDAFRTLRREAPVAWHPQVDKPGFWALTGYDEIYAVSRDSETWSSQATGVFLDVPAPEDSYQLALMMLTM
DPPRHTALRALVSRGFTPRHVARLGARTADMARAIVDDALAHGQCEFVDEVAGALPSYVIAELLGIPLEDGRRLYTLTDIMNTRPLHDPELVAA
QMQMFEYAAELAARKRSEPGDDIATALLHAEVDGRRLTDLEFNLFFLLLINAGGDTTRNLVAAGTLALIEHPDQWRRLAADPSLMPTAVEEMLR
WTSPVTVFSRTATRDTDVGGIRLREGERVAMFYPSANRDEKHFADPDRFDIGRMPNPHLAFGGGGTHFCLGASLARVEAAAIFRELITRTREIG
LVGPVERVRSVLMNGIRSMPAQFTPAVVPA*

>CYP125A(Mflv_3290)
MATPTLPPGFDFTDPDLNRERLPVEELAELRRCAPIWWNEQPDGIGGFGDGGFWVVTKHHDVKEISKRSDVFSSEVKTALPRYPEGSTGEQIET
GSLVLLNMDAPRHTHLRKIISRGFTPRAVERLRDDLNARAQNIAKTAASAGSGDFVEQVSCELPLQAIAGLLGVPVDDRKKLFDWSNQMVSDDD
PEYAHYDNRNAATELIMYAMQLAAVRAEQPGEDIVTKLIEADVDGHKLSDDEFGFFMVLLAVAGNETTRNSITHGMIAFTEHPDQWELYKRERP
ITAVDEIVRWATPVTSFQRTALTDYELSGVQITKGQRVVMSYRSANFDEEVFDDPFTFDILRDPNPHVGFGGTGAHYCIGANLARMTIDLMFNA
IADHMPDLSAIGSPDRLRSGWLNGIKHWQVDYSGRGCPVAH*

>CYP125F3(Mspyr1_03290)
MPADAITLGGADLADPDTYLAGMPYDAFRTLRREAPVAWHPQVDKPGFWALTGYDEIYAVSRDSETWSSQATGVFLDVPAPEDSYQLALMMLTD
PPRHTALRALVSRGFTPRHVARLGARTADMARAIVDDALAHGQCEFVDEVAGALPSYVIAELLGIPLEDGRRLYTLTDIMNTRPLHDPELVAAQ
MQMFEYAAELAARKRSEPGDDIATALLHAEVDGRRLTDLEFNLFFLLLINAGGDTTRNLVAAGTLALIEHPDQWRRLAADPSLMPTAVEEMLRW
TSPVTVFSRTATRDTDVGGIRLREGERVAMFYPSANRDEKHFADPDRFDIGRMPNPHLAFGGGGTHFCLGASLARVEAAAIFRELITRTREIGL
VGPVERVRSVLMNGIRSMPVQFTPAVVPA

>CYP125A(Mspyr1_08920)
MPGTNSCPISPDFDFLDATLNLERLPVEELAELRHSEPVHWVDVPGGTGGFGDKGYWLVTKHADVKEVSKRNDIFGSSPDGAIPVWPQDMTRDA
IDLQKAVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRARAQKIAETAAAAGSGDFVEQVSCELPLQAIAELLGVPQGDRDKLFRWSNEMTA
GEDPEYADVDPAMSSFELITYAMKMAEERAKNPTEDIVTKLIEADIEGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQNPEQWELYKK
ERPETAADEIVRWATPVSAFQRTALEDTELGGVQIKKGQRVVMSYRSANFDEEVFENPHSFNIMRNPNPHVGFGGTGAHYCIGANLAKMTINLI
FNAIADAMPDMKPIGDPERLKSGWLNGIKHWQVDYTGKGCPVSH

>CYP125A(Mspyr1_09930)
MTAEDARAQSCPFLPDGYDFTDPDVLLEGVPVAEFAQLRRTAPVWWNAQQESIFDDGGYWVVSRHEDIKSISRNSALWSTNAKGAVMRLPDGVT
AEQLELTKALLINHDAPEHTRLRKIVSRLFTPRAIAGMEAKLADSARQIVGEAADKDGGNFVDDVATLLPLQAIADLIGVPEEDREKLFHWSNA
IMNTDDPDFDSDPTIANAELMGYAYTMAEQRRRCPADDIVTRLVEADLDGDLGEVEFAFFVILLAVAGNETTRNAMTHGMNAFLDHPDQWELYK
RERPETAIDEIIRWATPVHCFQRTALDDVEIGGVTVARGQRVGLFYSSANFDEDVFDNPFNFDILRNPNPHLAFGGNGAHYCIGANLARMEIKL
MFEAIADRLPDISKRAEPQRLRSGWINGVKDLQVAYR

>CYP125A(Mspyr1_26180)
MATPTLPPGFDFTDPDLNRERLPVEELAELRRCAPIWWNEQPDGIGGFGDGGYWVVTKHHDVKEISKRSDVFSSEVKTALPRYPEGSTGEQIET
GSLVLLNMDAPRHTHLRKIISRGFTPRAVERLRDDLNARAQNIAKTAASAGSGDFVEQVSCELPLQAIAGLLGVPVDDRKKLFDWSNQMVSDDD
PEYAHYDNRNAATELIMYAMQLAAVRAEQPGEDIVTKLIEADVDGHKLSDDEFGFFMVLLAVAGNETTRNSITHGMIAFTEHPDQWELYKRERP
ITAVDEIVRGATPVTSFQRTALTDYELSGVQITKGQRVVMSYRSANFDEEVFDDPFTFDILRDPNPHVGFGGTGAHYCIGANLARMTIDLMFNA
IADHMPDLSAIGSPDRLRSGWLNGIKHWQVDYSGRGCPVAH

>CYP125E(OCU_21370)
MSIAKPTLVKSLVPQNLDTAADRDAAAVLDPDTFVTGAPYDAMTRLRATSPVHPVQLPGL
PRSWLLTKHADVRLVSRDTDTFTSSKGNTLVEAEAGPNSAMLPGIDPPRHVHFRKLINQG
FTVRNVQRLEPRMRLVTRDIVDTIIDKGEFDAVTDISAEMSLQVIADVLGVPAEDRMNVF
RWSNAIGSLGIEDPDYAPTPEALGQAAAEMFAYCGELVEHRRKHGLTDDILSALLAAEVD
GEKLNRDQLNEFFLLLAIAGNETTRNTLSHGILALAEHPEQQAQLARDPAAIKPAVEELL
RWATPVMHFRRTVVRDVEIRGQRIPCGDWVLMHYLSANRDEEVFDRPDQFDVTRPDAGHA
AFGGGGVHFCLGAQLARLELRVMLEELYANVPGLAVTGPPDRLRSSFFHGIKRLPCTT

>CYP125A(OCU_44240)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRRIVSRAFTPRAVERLRAELSERARCIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDYPVQWELFKARRP
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125D(OCO_21130)
MSIAKPTLVKSLVPQNLDTAADRDAAAVLDPDTFVTGAPYDAMTRLRATSPVHPVQLPGLPRSWLLTKHADVRLVSRDTDTFTSSKGNTLVEAE
AGPNSAMLPGIDPPRHVHFRKLINQGFTVRNVQRLEPRMRLVARDIVDTIIDKGEFDAVTDISAEMSLQVIADVLGVPAEDRMNVFRWSNAIGS
LGIEDPDYAPTPEALGQAAAEMFAYCGELVEHRRKHGLTDDILSALLAAEVDGEKLNRDQLNEFFLLLAIAGNETTRNTLSHGILALAEHPEQQ

AQLARDPAAIKPAVEELLRWATPVMHFRRTVVRDVEIRGQRIPCGDWVLMHYLSANRDEEVFDRPDQFDVTRPDAGHAAFGGGGVHFCLGAQLA
RLELRVMLEELYANVPGLAVTGPPDRLRSSFFHGIKRLPCTT

>CYP125A(OCO_44490)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRKIVSRAFTPRAVERLRAELSERAQRIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDHPQQWELFKAQRL
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125D(OCQ_20030)
MSIAKPTLVKSLVPQNLDTAADRDAAAVLDPDTFVTGAPYDAMTRLRATSPVHPVQLPGLPRAWLLTKHADVRLVSRDTDTFTSSKGNTLVEAE
AGPNSAMLPGIDPPRHVHFRKLINQGFTVRNVQRLEPKMRQVARGIVAAITDKREFDAVTDISAEMSLQVIADVLGVPAEDRMDVFRWSNAIGS
LGIEDPDYAPTPEALGQAAAEMFAYCGELVEHRRKHGLTDDILSALLAAEVDGEKLNRDQLNEFFLLLAIAGNETTRNTLSHGILALAERPEQQ
ALLARDPAAIKPAVEELLRWATPVMHFRRTVVRDVEIRGQRIPSGDWVLMHYLSANRDEEVFDRPDQFDVTRPDAGHAAFGGGGVHFCLGAQLA
RLELRVMLEELYANVPGLAVTGPPDRLRSSFFHGIKRLPCTT

>CYP125A(OCQ_45630)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRRIVSRAFTPRAVERLRAELSERARCIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDYPVQWELFKARRP
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125D(MKAN_03940)
MSTPKLSSLIVENQDSMAGRDAAVVLNPDTYLAGAPFDALARLRAHAPVHPMQLSGLPTTWLLTRHSDVRLVSRDSETFASSTGNTLVKVEAAP
TSAMLPGIDPPRHVHYRKLINQGFTARNVLRLEPRMRQVARDIVANIVDKGEFDAVTDISAEISLQVIADILGVPAEDRMNVFRWSNAIGSLGI
EDPDYAPTPEALGQAAAEMFAYCGELVAHRQKHGLTDDILSALLAAEVDGDRLNRDQLNEFFLLLAIAGNETTRNTLSHGILALSEHPDQQATL
ARDRDAVQPAVEELLRWATPVMHFRRTVTRDVVIRGQHIPAGDWVLMHYLSANRDEDVFERAAEFDISRPDADHVAFGGGGVHFCLGAQLARLE
LRVMLEELYPCVPGLTVTGPPDRLRSSFFHGIKRLPCAVG

>CYP125A6(MULP_02541)
MPAAEPTATSVPNLPPGFDFTDPDIYAERLPVAELAEMRRSAPIWWNEQPTGCGGFDDGGFWVVTKHKDVKEISLRSDVFSSLQKTALPRYKDG
TVDEQIERGKFVLLNMDAPQHTRLRKIVSRAFTPRAVERLRDDLRERARRIVEAAAAEGRGDFVEQVSCELPLQAIAGLMGVPQEDRKKLFHWS
NEMVGDQDPEFATNDALTASVELIMYGMQMAADRAKNPGQDLVTKLVEADIDGHKLSDDEFGFFVILLAVAGNETTRNSITQGMMAFTDFPDQW
ELYKRERPVTTADEIVRWATPVTSFQRTALEDYELSGVRIKKGQRVVMFYRSANFDEDVFDDPYTFNILRDPNPHVGFGGTGAHYCIGANLARM
TIDLMFNAIADVMPDLESISRPERLRSGWLNGIKHWQVDYHSDSSGKCPVAH

>CYP125A7(MULP_05284)
MPCPNLPPGFDFTDPDIYAERLPVEEFAELRSSEPIWWDEQLPGQGGGFHDGGFWAITKLKDVKEVSRRSDVFSSYENGVIPRFKNDIAREDID
VQRFVLLNMDAPHHTRLRKIISRGFTPRAIGRLHDELNDRAQNIAKAAAAAGSGDFVEQVSCELPLQAIAGLLGIPQEDRGKLFDWSNEMTGTE
DPEFAHIDAKASSVELIGYAMKMAEEKAKNPGDDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFADNPEQWELYKRER
PETAADEIVRWATPVTSFQRTALEDYELSGVQIKKGQRVLMFYRSANFDEEVFEDPFSFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLKPIAAPERLRSGWLNGIKHWQVDYTGKCPVSH

>CYP125A(MycrhN_0940)
MAAPNLPVGFDFTDPDIYATRVPTEEFAEVRRAAPIWWNDQAPDVGGYGDGGFWVVSKHRDVREVSLRSDVFSAAAKTVVPHFKPSVDVEGQIQ
ASKLSLLMMDDPEHARLRKIVSRGFTPRAVERLRAELNERAQRIAAEAASHASGDFVLEVSRELPLQAIAGLLGVPLEDREKLFDWSNKMVGGD
DPEFEEHNSLEAVIELIGYAMELAKLKEKEPGEDIVSTLIDSEADGQLTEAEFGMFVVTLAVAGNETSRNSITQGMMAFTDFPDQWELFKRERP
KTAADEIIRWASPITAFQRTALADTELSGVPIKKGQRLVLFYRSANFDEDVFDDPYTFDILRDPNPHLGFGGTGAHYCVGANLARMTIDLMFNA
IADHIPHLKPVSEPQRLRSSFINGIKHWQVAYQPS

>CYP125A(MycrhN_2286)
MAPLKIPADFDFLDATLNLERLPVEELAELRASEPVHWVDVPGGTGGFGDKGYWLVTKHADVKEVSKRSDIFGSSPDGAIPTWPQDMTRDAIDL
QKAVLLNMDAPQHTRLRKIISRGFTPRAIGRLEDELRARAQKIAETAAAEGSGDFVEQVSCELPLQAIAELLGVPQDDRDKLFRWSNEMTAGED
PEYADVDPAISSFELIQYAMKMAEERAKNPTEDIVTKLIEADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSQHPQQWELYKKERP
STAADEIIRWATPVSAFQRTALEDTELAGVKIKKGERVVMSYRSANFDDEVFENPHDFDILRDPNPHVGFGGTGAHYCIGANLARMTINLIFNA
VADKMPDLKPISEPERLMSGWLNGIKHWQVDYKGTSA

>CYP125A(MycrhN_2423)
MTQSTCPFGPAFDFTDPDVLLQGIPVTEFAELRKTAPVWWNDQQESIFDDGGYWVITRHEDIKAISRNGDLWSTNRKGAVMRLPDGVTAEQLDL
TKALLINHDAPEHTRLRKIVSRLFTPRSVAALEEKLAVAAHQIVGAAKERDFGNFVDDVAMPLPLLAIADLIGVPEADREKLFHWTNSIMNTDD
PDFDSDPTTANAELMGYAYTMAEERRRCPADDIVTRLIQADIDGESLGDVEFAFFVILLAVAGNETTRNAMTHGMNAFFENPGQWELFKRERPE

TAVDEIIRWATPVHCFQRTALADNEIGGVTIREGQRVGLFYSSANFDEDVFESPFEFDILRNPNPHLSFGGNGAHFCIGANLARMEIKLIFNEL
ADQIPDIAKLEEPQRLRSGWINGVKALPVSYRG

>CYP125NS(MycrhN_4756)
MPTTTPVDLSDSALWQNGFPDDLFAHWRRELPIFHHELTEGVAQTVKRDFWMTTKHRHAQRIHRDTDAFTAADGPLIQGIGPIGAFPNVITMDP
PVLTKRRRVMSHAFTPKAIGKLEEGIRRRAAAMIDRLLESGGGDWIEDVADVLPMSVIGDIVGIPDEDRPHIFDTLDRILKTNEADDQTKPEEH
YELFGQIFTYATELTASKRRNPTDDIWSTLTTAVVTDETGQELSIPASELEIFFFVLTLAGSDTTKNALAGGLQAFVANPAEMERYRDDESIRA
RAVEEVLRWSSPVAFWTRTTKVDVEMDGVIIPAGDRVVSMLRSANRDEEVFDDPFVFDIGRTDNPHVTFGGGGPHHCLGAMLARAEIRAALDEL
LLRADDIRLGPPKVTHPNLANNMSIFDGMSISLTRS

>CYP125A(MycrhN_4947)
MPTPNLPPGFDFTDPDIYAERLPIEELAHMRKVAPIWWQKQERGNLAFGDDGFWVVTKHKDVKEVSRRSDVFSSNKKTALPRYRDEADPASLEA
GKVVLLNQDAPHHTHLRKIISRAFTPRAIESLREELRLRARDIVKRAAAEGSGDFVEQVSCELPLQAIAGLMGVPQEDRMKLFEWSNQMVGDQD
PEYGRNDPTAASVELIMYGMQMAAERGKNPGDDLVTKLVQADVEGHKLTDDEFGFFVILLAVAGNETTRNSITQGMMAFTEFPDQWELFKRERP
ATAADEIVRWATPVTSFQRTALEDTELSGVKIKKDDRVVIFYRSANFDEDVFDDPYTFNILRDPNPHVGFGGTGAHYCIGANLARMTIDLIFNA
IADEMPDLTPISAPERLRSGWLNGIKHWQVDYTGAGAT

>CYP125D(Mycsm_03179)
MPITTPVDLSDSALWQNGFPDELFAELRRERPIFHHERTDGVAKTVHRDFWMTTKHRHAQRIHRDTDSFTAVDGPLIQTIGVMTDFPTIIHMDP
PELTKRRRVMSHAFTPKAIARLEEGIRTRATSLVDRLLTAGSGDWIEDVADVLPMSVIGDIIGIPDGDRPQIFDTLDRILKSANGQDATATEQE
QLELFGKVFTYALELTAEKRRNPTDDIWSTLATAVVTDENGEQLSIPANELEIFFYVLTFAGSDTTKNALAFGLQAFVANPSEIARYRADEAIR
SSAVEEVLRWATPVAFWTRSTKVDVQMDGVTIPKGERVVSMLRSANRDEEVFDDPFTFDIGRAENQHVTFGGGGPHHCLGAMLARAEIRAVLDE
LLCRADDIELGQPTVSYPNLTNNMSIFDAMPISLRAR

>CYP125A(Mycsm_03408)
MASPDLPPGFDFTDPDIYAERLPVEELAHMRKVAPIWWNAQTKGNAAFGDDGYWVVTKHKDVKEVSLRSDVFSSNKKTALPRYREDADPESLER
GKVVLLNQDAPHHTHLRKIISRAFTPRAVESLRDELRERAHNIAKAAAAEGSGDFVEQVSCELPLQAIAGLMGVPQDDRKKLFDWSNQMVGDQD
PEFGSNDPMAASIELIMYGMQMAAERSKNPGDDLVTKLVQADVEGHKLTDDEFGFFVILLAVAGNETTRNSITQGMMAFTEFPDQWELFKKERP
ATAADEIVRWATPVTSFQRTALEDTELSGVKIKKGERVVIFYRSANFDEDVFDDPYTFNILRDPNPHVGFGGTGAHYCVGANLARMTIDLIFNA
IADEMPNLTPISPPERLRSGWLNGIKHWQVDYTGKSAVAQ

>CYP125A(Mycsm_05668)
MTQTTCPFGQGFDFTDPDVLLKGIPVTEFAELRKTAPVWWNEQSDSIFDDGGCWVISRHEDIKEISRNGDLWSTNRKGVVMRMPEGTDAEQLEL
TKALLINHDAPEHTRLRKLVSRLFTPRAVATLEGKLAVAAREIVAAAKEKDSGNFVDDIAMKLPLLAIADLIGVPEVDRDKLFHWTNAIMNTED
PDFEADYVAANAELMGYAYTMAEERRRCPADDIVTRLVQADIDGESMEDVEFAFFVILLAVAGNETTRNAMTHGMNAFFENPEQWEIFKRDRPE
TTADEIVRWATPVHCFQRTALADNELGGVTIREGQRVGLFYSSANYDDEVFDRPFEFNILRDPNPHLGFGGNGAHFCIGANLARMEIKLIFNEI
AEQIPDISKLSEPQRLRSGWLNGVKRLQVAYRG

>CYP125A(Mycsm_05807)
MPCPISADFDFLDAELNRAGLPVAELAELRKSEPVHWVDVPGGTGGFGDKGYWLVTKHADVKEVSKRNEVFGSSPDGAIPVWPQEMTRDAIDLQ
RSVLLNMDAPQHTRLRKIISRGFTPRAVGRLEDELRIRAQKIAETAAAQESGDFVEQVSCELPLQAIAGLLGVPQDDRDKLFRWSNEMTAGEDP
EYAHIDPAMSSIELIQYAMQMAAERAKNPTEDIVTQLIEADIEGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIGFSQNPDQWELYKRERPE
TAADEIVRWATPVSAFQRTALEDVELGGVQIKKGQRVVMSYRSANFDEDVFEDPHTFNILRSPNPHVGFGGTGAHYCIGANLAKMTINLIFSAI
ADHMPDLKPIGEPERLKSGWLNGIKHWQVDYTGKCPVAH

>CYP125A(JDM601_2803)
MTETPHLLPEGFDFTDPALIGERIPHEEFALLRRTEPIWWNAQPFGVSGYPDEGYWVVTKHADVRAVSLQDDVFSSHENTSLIRTNTTSNQDLH
EASRDNIMLFLDGPKHAKLRRIVSRGFTPRVVAGMRDSLDRQAREIVAAAAEHDTGDFVTEVASRLPLATICELIGVPAAERQQVFDWSNRLVG
GGNGDPQAAADGMQASAELLGYAYQMAEDRKARPRDDIATALVTATIDGEALTALEFGYYVMMLMVAGNETTRNATSQGMVAFFDHPDQWRLFV
AERPATMVDEVVRWATPVISFQRTALRDVELGGVHIAKGQRVGMFYGSANYDEEVFDEPFTFDIRRSPNPHLGFGAPGAHYCIGANLARMQINL
IFGALADIMPDIRRLDRASRSVLPWINGIDAMPVEFGGAAMSTGSR

>CYP125A(JDM601_3609)
MFVLTSTQPACPFGPGFDFTDPDVLLHGMPIAQFAELRKTAPVWWNEQPAHSNIFDDGGYWVVSKHQHIKEISRDNEVWSTNAKGAVMRLPDGI
TADQLELTKALLINHDPPEHTRLRKLVSRLFTPRSVGALEEKLAQSARDIVAAAAEKDSGNFVDDIAMKLPLLAIADLIGVPEADRERLFHWTN
SIMNTDDPDFDSDPAMANAELMGYAYTMAEQRRRCPADDIVTRLVQADMDGESLGETEFAFFVILLAVAGNETTRNAMTHGINAFAENPDQWEL
YKRERPETAVDEIVRWATPVHCFQRTAKIDTELGGVAISKGQRVGLFYSSANYDEEVFERPFAFDVLRDPNPHLAFGGQGTHYCIGANLARMEI
RLMFDEIANQLPDITKLAEPQRLRSGWINGVKDLQVAYHG

>CYP125A(JDM601_3682)
MASPTARSASSIPAGFDPTDPEIWAERIPNAELAALRENEPIKWIEQPDGVGGFNDGGYWAVTRHADVKEISRLDDVFSSEINTAIPRFNDDIQ
REQIDQQRLIMLNQDAPRHTRLRRIVSRGFTPRHILPLHDELQERAQAIAKEALAKGSGDFVVEVASELPLQAIAGLMGVPQSDRGKLFNWTNQ
MTGYDDPEYTEKYDPATSAMEIIAYGLQLAEMKRQNPGSDIVTTLIEADIDGEKLNDDELGFFIILLAVAGSETTRNSITQGMMAFVDHPDQWE

LFKKERPETAADEIVRWASPVSSFQRTATRDYNLNGTQIKEGQRVVMFYRAANFDPEVFDNPQQFNILRDPNPHVGFGGTGAHYCIGTHLARMT
IGLMFNAIADHIPDLKPLDAPSRLQSGWLNAIKRWPVDYTGKA

>CYP125A(JDM601_3692)
MSTLPAGFDFLNPDLIVEGIPEKEFAQLRKTAPVCWIEQAPGKGGGFNDGGYWAVTKLADVKEVSLRSDVFSSYENCVIPRFSDDMQRENIEVQ
RFVMLNMDAPHHTRLRRIISRGFTPRAIGRLRDELHERAQAIVKAAAEAGSGDFVEQVSCELPLQAIAGLLGVPQEDRDKLFQWSNEMTGSEDP
EYADIDPQASSFELITYAMQLAAAKAENPGEDIVTTLINADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITHGMIAFSEHPEQWELFKSERPA
TTADEIVRWASPVICFQRTALEDYELSGAQIKKGQRVVMFYRSANFDEDAFDEPNKFNILRDPNPHVGFGGTGAHYCIGTHLARLTIDLIFNAV
ADHVPDLAPLAKPERLRSGWLNGIKHWQMDYTGKCPVAH

>CYP125A(JDM601_3693)
MSSVKVPPGFDFTDPEIYAERLPDAEFARVRAAAPVTWIDQPDDKSGGFKDGGYWAITSHHDVKEVSRLDEVFSSEINGAIPRYNDDIERENID
VGRLLMLNQDAPRHTRLRRIVSRGFTPRHILPLHDDLERRAQNIAKEALARGTGDFVVEVASELPLQAIAGLMGVPLEDRGKLFNWTNQMTSYD
DPEYAHYDPKTSSMEIISYGLQLAEMKRHNPGNDIVTTLIEADIDGEKLGDELGFFIILLAVAGSETTRNSITQGMMAFTEFPEQWELFRRERP
ETTADEVVRWATPVTSFQRTATRDYELSGVQIKKGQRVVMFYRAANFDPEVFDNPQQFDILRDPNPHVGFGGTGAHYCIGTHLARMSVNLMFNA
IADHIPDLKPLDKPDRLKSGWLNGIKHWRVDYTGKSA

>CYP125D(W7S_09920)
MSIAKPTLVKSLVPQNLDTAADRDAAAVLDPDTFVTGAPYDAMTRLRATSPVHAVQLPGLPRAWLLTKHADVRLVSRGTDTFTSSKGNTLVEAE
AGPNSAMLPGIDPPRHVHFRKLINQGFTVRNVQRLEPKMRQVARGIVAAITDKREFDAVTDISAEMSLQVIADVLGVPAEDRMDVFRWSNAIGS
LGIEDPDYAPTPEALGQAAAEMFAYCGELVDHRRKHGLTDDILSALLAAEVDGEKLNRDQLNEFFLLLAIAGNETTRNTLSHGILALAEHPEQQ
ALLARDPAAIKPAVEELLRWATPVMHFRRTVTRDVEIRGQRIPSGDWVLMHYLSANRDAEVFDRPDQFDVTRPDAGHAAFGGGGVHFCLGAQLA
RLELCVMLEELYANVPGLAVTGPPDRLRSSFFHGIKRLPCTT

>CYP125A(W7S_22405)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRKIVSRAFTPRAVERLRAELSERARCIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDYPDQWELFRARRP
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125A1(MT7199_3607)
MSGNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(J112_19085)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTNSITQGMMA
FAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHYC
IGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(J113_24790)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(J114_18960)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(ERDMAN_3890)
MPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYENGVIPRFKNDIAREDIE
VQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQEDRGKLFHWSNEMTGNE
DPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMMAFAEHPDQWELYKKVR

PETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHYCIGANLARMTINLIFN
AVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(TBXG_003560)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNTLRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(TBSG_03611)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNTLRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A1(UDA_3545c)
MSWNHQSVEIAVRRTTVPSPNLPPGFDFTDPAIYAERLPVAEFAELRSAAPIWWNGQDPGKGGGFHDGGFWAITKLNDVKEISRHSDVFSSYEN
GVIPRFKNDIAREDIEVQRFVMLNMDAPHHTRLRKIISRGFTPRAVGRLHDELQERAQKIAAEAAAAGSGDFVEQVSCELPLQAIAGLLGVPQE
DRGKLFHWSNEMTGNEDPEYAHIDPKASSAELIGYAMKMAEEKAKNPADDIVTQLIQADIDGEKLSDDEFGFFVVMLAVAGNETTRNSITQGMM
AFAEHPDQWELYKKVRPETAADEIVRWATPVTAFQRTALRDYELSGVQIKKGQRVVMFYRSANFDEEVFQDPFTFNILRNPNPHVGFGGTGAHY
CIGANLARMTINLIFNAVADHMPDLKPISAPERLRSGWLNGIKHWQVDYTGRCPVAH

>CYP125A(OEM_44690)
MATPNLPPGFDFTDPDIYAHRLPVREFAELRATEPVWWNEQAPDKGGFGDGGYWAVTKHRDIRDVSLRSDVFSSAAKSIVPRYREDLAAGQIEA
GRASMIMMDDPEHSRLRKIVSRAFTPRAVERLRAELSERARCIVTEAAAAGSGDFVRQVACELPLQAISALLGVPHEDYDKLFDWTNNMIGSDD
PEFAGNDALTSAGELMWYAMQLAARKAEEPGDDIVTTLIQADADGQRLSEAEFGMFVVTLAVAGNETTRNSITQGMMAFTDYPDQWELFRARRP
KTAADEIIRWATPITAFQRTAREDTELGGVAIREGQRVVLFYRSANFDEEVFDDPFTFDILRSPNPHLGFGGTGAHYCIGANLARMTIDVMFNA
LADRLPDLAPLGNPERLRSSFINGIKHWPVDYRGGHPVAS

>CYP125A(G6XE67_MYCAB)

MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIAKAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP
ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS

>CYP125A(G6XE69_MYCAB)

MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHYCIGANLARLTINLMFN
AIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ

>CYP125A(G6XAE6_MYCAB)

MTACPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAVNGLHSRLVERAHSIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATAATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR
PATAADEIIRWASPIIAFQRTALQDVELGGVQICKDQRVGMFYASANFDEDVFDDPFTFNIERDPNPHLAFGGHGIHYCLGANLARLEIGIMFD
ALADRLPDLMPTGAPTRFRSGWINGVVALPANYLGAGPRG

>CYP125A(G6X7Y6_MYCAB)
MTAMKTAAELGLPEGFDFTDPELYGNRMPHEEFATLRREAPVWWNPQPRTVGGFADEGYWVISKHRDVREVSLHTDIFSSGRKGAIPRLEDHIS
PEEFQATLSVLINKDAPEHTQLRGLVSRMFTPRSIAALRITLEERAERIVRAALEGGHGEFVREVASELPMQAIAELIGVPEEDRVKLFEWSNQ
MTGYDEADVEIDPRIGAAQILGYSYQLAEQRRDCPGNDVVSRLLTGTVDGEQLTPEQFGFFVVMLSVAGNETTRNATTMGMMAFLEHPDQWELF

KSARPSTTVDEIVRYTSPLISQQRTALQDTVIGDVRIRAGERVVMLYPSANFDEEVFENPHAFDITRDPNPHLGFGGTGAHYCLGANLAKAELE
IIFNKIADRMPDISRIGDAPRFHSGWINGIKKFDTAYCPVTH

>CYP125A(X8ER31_MYCAB)
MKTAAELGLPEGFDFTDPELYGNRMPHEEFATLRREAPVWWNPQPRTVGGFADEGYWVISKHRDVREVSLHTDTFSSGRKGAIPRLEDHISPEE
FQATLSVLINKDAPEHTQLRGLVSRMFTPRSIAALRITLEERAERIVRAALEGGHGEFVREVASELPMQAIAELIGVPEEDRVKLFEWSNQMTG
YDEADVEIDPRVGAAQILGYSYQLAEQRRDCPGNDVVSRLLTGTVDGEQLTPEQFGFFVVMLSVAGNETTRNATTMGMMAFLEHPGQWELFKSA
RPSTTVDEIVRYTSPLISQQRTALQDTVISDVRIRAGERVVMLYPSANFDEEVFENPHTFDITRDPNPHLGFGGTGAHYCLGANLAKAELEIIF
NKIADRMPDISRIGDAPRFHSGWINGIKKFDTAYCPVTH

>CYP125A(X8EQ35_MYCAB)
MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIARAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP
ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS

>CYP125A(X8EN49_MYCAB)
MTTCPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAVNGLHARLVERAHGIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATTATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR
PSTAADEIIRWASPIIAFQRTALQDVELGGVQIRKDQRVGMFYASANFDEDVFDDPFAFNIERDPNPHLAFGGHGIHYCLGANLARLEIGIMFD
ALADRLPDLMPTGAPTRFRSGWINGVVALPANYHGSGPRG

>CYP125A(X8EM53_MYCAB)
MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHYCIGANLARLTINLMFN
AIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ

>CYP125A(X8D963_MYCAB)
MTAMKTAAELGLPEGFDFTDPELYGNRMPHEEFATLRREAPVWWNPQPRTVGGFADEGYWVISKHRDVREVSLHTDTFSSGRKGAIPRLEDHIS
PEEFQATLSVLINKDAPEHTQLRGLVSRMFTPRSIAALRITLEERAERIVRAALEGGHGEFVREVASELPMQAIAELIGVPEEDRVKLFEWSNQ
MTGYDEADVEIDPRVGAAQILGYSYQLAEQRRDCPGNDVVSRLLTGTVDGEQLTPEQFGFFVVMLSVAGNETTRNATTMGMMAFLEHPGQWELK
SARPSTTVDEIVRYTSPLISQQRTALQDTVISDVRIRAGERVVMLYPSANFDEEVFENPHTFDITRDPNPHLGFGGTGAHYCLGANLAKAELEI
IFNKIADRMPDISRIGDAPRFHSGWINGIKKFDTAYCPVTH

>CYP125A(X8D810_MYCAB)
MTTCPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAVNGLHARLVERAHGIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATTATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR
PSTAADEIIRWASPIIAFQRTALQDVELGGVQIRKDQRVGMFYASANFDEDVFDDPFAFNIERDPNPHLAFGGHGIHYCLGANLARLEIGIMFD
ALADRLPDLMPTGAPTRFRSGWINGVVALPANYHGSGPRG

>CYP125A(X8D8Q9_MYCAB)
MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHYCIGANLARLTINLMFN
AIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ

>CYP125A(X8D7U8_MYCAB)
MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIARAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP
ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS

>CYP125A(T0B0P1_MYCAB)
MVQAQHPHLPDGIDFTDPELFVHGIPERELAELRHTEPIWWNHTERGVAGFDDDGFWVVSKHKDVKEVSLRCEVFSSEQNTAIPRYLPTTPRER
IDATRLIMLNMDPPRHSRLRHIISRGFTPRAISRLRDDLNARAQGIARAAAQLRHGDFVEQVACELPLQAIAGLMGTPLDEREQLFDWSNRLVG
SSDGEDDSAVASAELLMYAMGVAARKTAEPGADICTDLVNADIDGQKLSDDEFGFFVMLLAVAGNETTRNSITHGMHAFTQFPEQWELYKKTRP

ETAADEIVRWATPVTSFQRTALEDTELGGVRIKKGQRVVMMYRSANFDEEVFENPFTFDIMRDPNPHVGFGGNGEHHCVGANLARMTINLMFNA
IADHMPDLASAGEPDRLRSGWLNGVKHWEVDFCPAGYGRAS

>CYP125A(T2R8E5_MYCAB)
MTAMKTAAELGLPEGFDFTDPELYGNRMPHEEFATLRREAPVWWNPQPRTVGGFADEGYWVISKHRDVREVSLHTDTFSSGRKGAIPRLEDHIS
PEEFQATLSVLINKDAPEHTQLRGLVSRMFTPRSIAALRITLEERAERIVRAALEGGHGEFVREVASELPMQAIAELIGVPEEDRVKLFEWSNQ
MTGYDEADVEIDPRVGAAQILGYSYQLAEQRRDCPGNDVVSRLLTGTVDGEQLTPEQFGFFVVMLSVAGNETTRNATTMGMMAFLEHPGQWELF
KSARPSTTVDEIVRYTSPLISQQRTALQDTVISDVRIRAGERVVMLYPSANFDEEVFENPHTFDITRDPNPHLGFGGTGAHYCLGANLAKAELE
IIFNKIADRMPDISRIGDAPRFHSGWINGIKKFDTAYCPVTH

>CYP125A(T2R8J5_MYCAB)
MTTCPFTPGFDFTDPDLIQHRIPAEEFAYLRKTEPIWWNAQPRGVAGFDDDGYWVVTKHADVKEVSRLNEVFSNSVNTTVVRYNEDITAEQLEI
QRENLLIDMDEPKHRILRRIVSPLFTPKAVNGLHARLVERAHGIVEEAAEKSSGNFVSDIASVLPMHAIADLVGIPESDRQQVLDWTNQMFAYD
DPAIGRDTATTATVSMLGYAYAMAEERQLNPQDDILTGLVRGAYDDRPLTPLEFAYFVIQLMVAGNETSRNAITHGVLAFADNPAQWRLYRERR
PSTAADEIIRWASPIIAFQRTALQDVELGGVQIRKDQRVGMFYASANFDEDVFDDPFAFNIERDPNPHLAFGGHGIHYCLGANLARLEIGIMFD
ALADRLPDLMPTGAPTRFRSGWINGVVALPANYHGSGPRG

>CYP125A(S9ZVJ7_MYCAB)

MVHPSLPAGFDFTDPEIYAERLPVEELKELRKTAPIWWQEQPDGVGGFNDGGYWVVTKHKDVKEVSLRSDVFSSWENTAIPRFQDDITREAIEL
QRYVMLNMDAPHHTRLRKIISRGFTPRAIGRLRDELNERAQEIAKAAAASGTGDFVEQVSCELPLQAIAGLLGVPIEDRGKLFNWSNEMTSYDD
PEYADIDPAASSMEILAYSMEMAKQKAENPGEDIVTTLINAEVEGEGKLSDDEFGFFVIMLAVAGNETSRNSITQGMMAFTQFPEQWELYKKER
PETAADEIVRWATPVTSFQRTALEDTELDGVKIKKGQRVVMMYRSANFDEDVFEDPFSFNIMRNPNPHMGFGGSGAHY
CIGANLARLTINLMFNAIADHMPNLAPAGDPKRLQSGWLNGIKHWQVDFTGASGCPVLQ

Chapter 4

CYP125 P450 gene mapping and gene-cluster analysis in the genus *Mycobacterium*

4.1. Introduction

Since their identification five decades ago, quite a large number of cytochrome P450 monooxygenases (CYPs/P450s) have been identified in species across biological kingdoms, especially due to the current genome sequencing rush (Nelson, 2009). Studies on P450 enzymes have been reported from animals owing to their role in drug metabolism (particularly from mammals) (Guengerich, 2015) or analysis of diversity (Nelson, 2013; Sezutsu *et al*., 2013); from fungi owing to their role as drug targets (Kelly and Kelly, 2013; Jawallapersand *et al*., 2014) and for evolutionary analysis (Moktali *et al*., 2012; Chen *et al*., 2014; Syed *et al*., 2014); from bacteria owing to P450 structure-functional analysis (Poulos and Johnson., 2015) and generation of products valuable to humans (McLean *et al*., 2015) and from plants owing to their roles in key cellular processes and defence mechanisms (Mizutani and Ohta, 2010; Hamberger and Bak, 2013; Schuler, 2015). Irrespective of their origins, P450s from all organisms have been exploited for their biotechnological potential (Girhard *et al*., 2015). Although quite a large number of P450s have been identified to date (Nelson, 2009), genome annotation of P450s in recently elucidated organisms genomes has led to the discovery of a novel P450 family containing a novel P450 fusion protein (CYP5619 family) with an N-terminal P450 domain fused to a heme peroxidase/dioxygenase domain (Sello *et al*., 2015), suggesting that much remains to be explored and understood about the evolution of these enzymes.

The above facts suggest that, to date, evolutionary analysis of bacterial P450s has not been reported. This chapter is dedicated to perform genome mapping and gene-cluster analysis of CYP125 P450 in the genus *Mycobacterium*. As part of this chapter, the physical localization (Synteny) of CYP125 P450 genes on the chromosome of different mycobacterial

species was mapped along with up-stream and down-stream genes with respect to CYP125 P450 to understand the gene-cluster analysis and evolutionary pattern of CYP125 in the genus *Mycobacterium*.

Synteny analysis provides excellent information on genome-duplication of P450s where localization of P450s belongs to the same family on same chromosome next to each other is a direct indication that these P450s are possibly duplicated during the evolution (Qhanya *et al*., 2015; Sello *et al*., 2015). Genome mapping of P450s refer to the identification of neighbouring genes with respect to P450s. Genome mapping reveals information on possible role of P450s in a physiological function by involving in a metabolic pathway. For example, CYP128A1 of *Mycobacterium tuberculosis* function is predicted based on its physical localization with genes involved in biosynthesis of menaquinone-like molecule (Holsclaw *et al*., 2008). In addition, CYP121A1 of *M. tuberculosis* function identified based on its operonic arrangement with genes involved in biosynthesis of cyclodipeptide cyclo(L-Tyr-L-Tyr) (cYY) molecule (Belin *et al*., 2009).

4.2. Methodology

4.2.1. P450s

CYP125 genes that are annotated as part of Chapter 3 were used in this study.

4.2.2. Genome mapping and gene-cluster analysis of CYP125

Genome localization of CYP125 P450s was carried out using protein IDs at respective mycobacterial species database located at the KEGG website (http://www.genome.jp/kegg-bin/show_organism?category=Mycobacterium). Three genes upstream of the CYP125 P450 were selected and another three genes downstream were selected and put in a table along with their KEGG gene codes. The genome map was also downloaded and all results were

presented in a table. The direction of the genes was also given in the table. Based on the direction and function of up-stream and down-stream genes, CYP125 P450s were classified into different clusters. The word "order of genes" refer such that they are the same genes/proteins (having the same function) and also the same gene transcription direction.

4.2.3. Visualization of gene-cluster maps

pDRAW32 (http://www.acaclone.com/) programme was used to generate gene-cluster maps. The DNA sequence of entire CYP125 gene along with up-stream and down-stream parts were downloaded from KEGG and used to deduce gene-cluster maps. The genes were then annotated according to their size by analysing the whole downloaded sequence. The genes were assigned to different colours according to KEGG colour codes expect CYP125 gene.

4.3 Results and discussion

Genome mapping and gene cluster analysis of CYP125 is carried out in order to identify same gene-clusters in different mycobacterial species to understand its evolution and also in future when promoter identification is carried out only one gene cluster can be used. Genome mapping and gene-cluster analysis revealed presence of 29 CYP125 gene clusters in 56 mycobacterial species (Figures 4.1 and 4.2). Information on genes, their direction and function along with KEGG gene map is presented in the Appendix. Characteristics feature of gene clusters were presented below:

Cluster 1:

Cluster 1 comprised of 23 CYP125 P450s belong to *Mycobacterium tuberculosis* complex (MTBC species). The order of genes downstream of CYP125 is: acyl-CoA dehydrogenase (fadE28), acyl-CoA dehydrogenase (fadE29) and uncharacterized protein and the order of

genes upstream of CYP125 is: acetyl-CoA acetyltransferase, deazaflavin-dependent nitroreductase and short-chain type dehydrogenase/reductase (Figure 4.1 and Table 4.1).

Cluster 2

Cluster 2 comprised of 5 CYP125 P450s that are exclusively found in *Mycobacterium avium* complex (MAC) (Figure 4.1 and Table 4.1). These CYP125 were from the 5 different species belong to the MAC. The order genes upstream of CYP125 is: TetR family transcriptional regulator, hypothetical protein and hypothetical protein and the order of genes downstream is: thiosulfate/3-mercaptopyruvate sulfurtransferase, alpha-1,6-mannosyltransferase and hydroxylase.

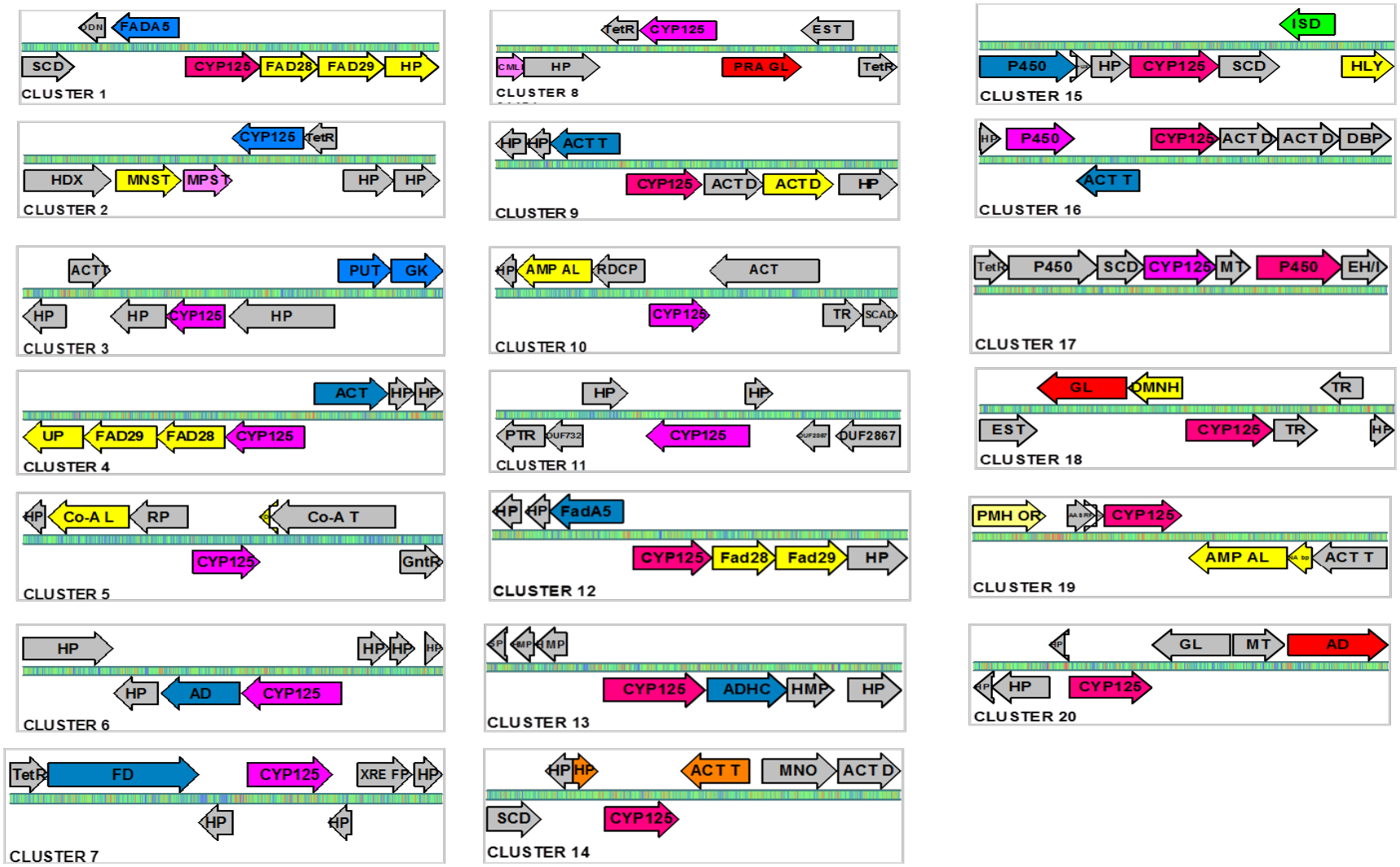


Figure 4.1. CYP125 gene-clusters (clusters 1 to 20) analysis in the genus *Mycobacterium*.

Table 4.1. Analysis of CYP125 gene clusters in the genus *Mycobacterium*. Abbreviations: MTBC: *Mycobacterium tuberculosis* complex; MAC: *Mycobacterium avium* complex; SAP: Saprophytes; NTM: Nontuberculous mycobacteria; MCAC: *Mycobacterium chelonae-abscessus* complex.

Cluster	Category	Number of P450s	Candidate CYP125 P450s
1	MTBC	22	CYP125A1(MAF_35570), CYP125A1(J114_18960), CYP125A1(TBHG_03485.1), CYP125A1(ERDMAN_3890), CYP125A1(MT7199_3607), CYP125A1(UDA_3545c), CYP125A1(TBFG_13578.4), CYP125A1(BN44_110037), CYP125A1(MRA_3584), CYP125A1(BN42_90040), CYP125A1(Rv3545c), CYP125A1(MbovBCG)(BCG_3609c, CYP125A1(TBMG_03584), CYP125A1(MbovAF2122/97,NP_857214.1)(Mb3575c), CYP125A1(TBXG_003560), CYP125A1(K60_036830), CYP125A1(TBSG_03611), CYP125A1(BCGMEX_3607c), CYP125A1(CCDC5079_3286), CYP125A1(JTY_3610), CYP125A1(J112_19085), CYP125A1(J113_24790).
2	MAC	5	CYP125E(OCU_21370), CYP125D(OCQ_20030), CYP125D(OCQ_20030), CYP125D(W7S_09920), CYP125D(OCO_21130)
3	MAC	5	CYP125A(W7S_22405), CYP125A(OCO_44490), CYP125A(OCQ_45630), CYP125A(MIP_06760), CYP125A(OCU_44240).
4	MAC	4	CYP125A(MAP_0522), CYP125A(MAV_0616), CYP125A(MIP_00975), CYP125A(MAP4_3345).
5	MAC	4	CYP125NS(MAP4_1479), CYP125NS(MAV_1637), CYP125NS(MIP_02350), CYP125NS(MAP2344)
6	MAC	4	CYP125A(MAP_1614c), CYP125A(MIP_04041), CYP125A(MAV_2811), CYP125A(MAP4_2225)
7	SAP	8	CYP125A4(MSMEG_3524), CYP125A(Mjls_2753), CYP125A(Mkms_2767), CYP125A(Mmcs_2723), CYP125A(Mflv_3290), CYP125A(Mspyr1_26180), CYP125A10(Mvan_3012), CYP125A(Mycch_2866)
8	SAP	8	CYP125A(Mkms_4660), CYP125A11(Mvan_5151), CYP125A(Mjls_4955), CYP125A(Mmcs_4572), CYP125A5P(MSMEG_5853), CYP125A(Mspyr1_09930), CYP125A(Mflv_1607), CYP125A(Mycch_4512)
9	SAP	7	CYP125A(Mycsm_05807), CYP125A9(Mvan_5258), CYP125A9(Mycch_4638), CYP125A(MycrhN_2286), CYP125A(Mmcs_4677), CYP125A(Mkms_4763),

			CYP125A(Mjls_5062)
10	SAP	2	CYP125D(Mycsm_03179), CYP125NS(MycrhN_4756)
11	SAP	2	CYP125A(Mycsm_03408), CYP125A(MycrhN_4947)
12	SAP	3	CYP125A(Mspyr1_08920), CYP125A(Mflv_1508), CYP125A3(MSMEG_5995)
	NTM	5	CYP125A7(MULP_05284), CYP125A7(MMAR_5032), CYP125A7(MUL_4106), CYP125A(JDM601_3693), CYP125A(JDM601_3692)
13	NTM	2	CYP125A6(MMAR_2783), CYP125A6(MULP_02541)
14	MCAC	4	CYP125A(MASS_0581), CYP125A(MAB_0611), CYP125A(LA61_03005), CYP125A(LA62_03090)
15	MCAC	3	CYP125A(MAB_1211c), CYP125A(LA62_06150), CYP125A(LA61_06055)
16	MCAC	4	CYP125A(LA62_03100), CYP125A(LA61_03015), CYP125A(MAB_0613), CYP125A(MASS_0583)
17	MCAC	4	CYP125A(MAB_0101), CYP125A(LA61_00515), CYP125A(LA62_00510), CYP125A(MASS_0104)
18	SAP	2	CYP125A(Mycsm_05668), CYP125A(MycrhN_2423)
19	SAP	2	CYP125A(MycrhN_0940), CYP125F(Mycch_4146)
20	SAP	2	CYP125F3(Mspyr1_03290),CYP125F3(Mflv_0425)
Unique clusters	MAC	3	CYP125F1(MAP3818), CYP125F1(MAP4_3931), CYP125D(MKAN_03940)
	NTM	3	CYP125A(JDM601_3682), CYP125A(JDM601_3609), CYP125A(JDM601_2803)
	SAP	1	CYP125F2(Mvan_0246).
	MTBC	1	CYP125A1(MCAN_35561)

Cluster 3

Cluster 3 comprised of 5 CYP125 P450s exclusively found in MAC species. The order of genes upstream of CYP125 is : hypothetical protein, UDPglucose--hexose-1-phosphate uridylyltransferase and galactokinase and the order of genes downstream is : hypothetical protein, dihydrolipoyllysine-residue acetyltransferase component of acetoincleaving system and hypothetical protein (Figure 4.1 and Table 4.1).

Cluster 4

Cluster 4 comprised of 4 CYP125 P450 belong to MAC species. The order of genes upstream of CYP125 is: acetyl-CoA C-acetyltransferase, hypothetical protein and hypothetical protein and the order of genes downstream is: FadE28 and FadE29 (involved in geraniol degradation) and uncharacterized protein.

Cluster 5

Cluster 5 comprised of 4 CYP125 P450s belong to MAC species (Figure 4.1 and Table 4.1). The order of genes downstream of CYP125 is: ferredoxin; putative CoA-transferase family protein, Transcriptional regulator, GntR family and the order of genes upstream is: hypothetical protein (oxidoreductase Rieske 2Fe-2S domain protein), putative AMP-binding enzyme (crotonobetaine/carnitine-CoA ligase) and hypothetical protein (Figure 4.1 and Table 4.1).

Cluster 6

Cluster 6 comprised of 4 CYP125 P450s belong to MAC species. The order of the genes downstream of CYP125 is: alcohol dehydrogenase, hypothetical protein, and hypothetical protein and the order of genes upstream of CYP125 is: hypothetical protein, hypothetical protein and hypothetical protein. Two P450s, CYP125A (MAP_1614c) and CYP125A(MIP_04041) has the above mentioned gene arrangements whereas the other two P450s namely, CYP125A(MAV_2811) and CYP125A(MAP4_2225) has the reverse complement arrangement of genes (Figure 4.1 and Table 4.1).

Cluster 7

Cluster 7 was found to be comprised of 8 CYP125 P450s belong to SAP species. The order of genes upstream of CYP125 is: hypothetical protein, formate dehydrogenase and TetR family

transcriptional regulator and the order of the genes downstream of CYP125 P450 is: hypothetical protein, XRE family transcriptional regulator and hypothetical protein (Figure 4.1 and Table 4.1).

Cluster 8

Cluster comprised of 8 CYP125 P450s belong to SAP species. The order of genes downstream of CYP125 is: transcriptional regulator TetR family; hypothetical protein; hypothetical protein and the order of genes upstream is: phosphoribosylamine-glycine ligase; putative esterase; putative transcriptional regulator TetR family (Figure 4.1 and Table 4.1).

Cluster 9

Cluster 9 comprised of 7 CYP125 belong to the SAP species. The order of genes downstream of CYP125 is: acyl-CoA dehydrogenase, acyl-CoA dehydrogenase and putative nucleic-acid-binding protein and the order of genes upstream is : acetyl-CoA C-acetyltransferase, hypothetical protein and deazaflavin-dependent nitroreductase family protein (Figure 4.1 and Table 4.1).

Cluster 10

Cluster 10 comprised of 4 CYP125 P450s belong to SAP species. The order of genes downstream is: putative acyl-CoA transferase; transcriptional regulator; short-chain alcohol dehydrogenase and the order of genes upstream is: Rieske (2Fe-2S) domain-containing protein, acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II, hypothetical protein.

Cluster 11

Cluster 11 comprised of 2 CYP125 P450s belong to the SAP . The order of genes downstream of CYP125 is: hypothetical protein; protein of unknown function (DUF732) and

putative transcriptional regulator and the order of genes upstream is: hypothetical protein, protein of unknown function (DUF2867) and protein of unknown function (DUF2867).

Cluster 12

Cluster 12 comprised of 8 CYP125 belong to the SAP (3 P450s) and Nontuberculous mycobacteria (NTM) (5 P450s). The order of genes upstream is: acetyl-CoA C-acetyltransferase, hypothetical protein and hypothetical protein and the order of genes downstream is: acyl-CoA dehydrogenase domain protein, acyl-CoA dehydrogenase domain protein and predicted nucleic-acid-binding protein containing a Zn-ribbon.

Cluster 13

Cluster 14 comprised of 2 CYP125 P450s belong to the NTM. In both CYP125 P450s the order of genes downstream of P450s is considered. The order of genes downstream is: NADP-dependent alcohol dehydrogenase AdhC; conserved hypothetical membrane protein; conserved hypothetical protein]. The order of genes upstream is different in both P450s.

Cluster 14

Cluster 14 comprised of 4 CYP125 P450s belong to *Mycobacterium chelonae-abscessus* complex (MCAC). The order of genes downstream is: hypothetical protein, hypothetical protein and probable short-chain dehydrogenase/reductase and the order of genes upstream is: probable short-chain dehydrogenase/reductase, putative cytochrome P450 (cholest-4-en-3-one 26-monooxygenase) and probable acyl-CoA dehydrogenase.

Cluster 15

Cluster 15 comprised of 3 CYP125 P450s belong to the MCAC species. The order of genes downstream is: putative short chain dehydrogenase/reductase, probable short-chain Z-

isoprenyl diphosphate synthetase and conserved hypothetical protein and the order of genes upstream is: hypothetical protein, putative ferredoxin, cytochrome P450 (sterol 14-demethylase).

Cluster 16

Cluster 16 comprised of 4 CYP125 P450s belong to the MCAC species. The order of genes upstream is: probable acetyl-CoA acetyltransferase, putative cytochrome P450 (cholest-4-en-3-one 26-monooxygenase) and hypothetical protein and the order of genes downstream is: probable acyl-CoA dehydrogenase, putative acyl-CoA dehydrogenase and hypothetical protein.

Cluster 17

Cluster 17 comprised of 4 CYP125 P450s belong to the MCAC species. The order of genes upstream is: putative short chain dehydrogenase/reductase, probable monooxygenase and putative TetR-family transcriptional regulator and the order of genes downstream is: methyltransferase, probable monooxygenase and probable enoyl-CoA hydratase/isomerase.

Cluster 18

Cluster 18 comprised of 2 CYP125 P450s belong to the SAP species. These two P450s has same order of upstream genes i.e. : N-dimethylarginine,dimethylaminohydrolase, phosphoribosylamine-glycine ligase and putative esterase. The order of genes downstream is different.

Cluster 19

Cluster 19 comprised of 2 CYP125 P450s belong to the SAP species. In both P450s only the order of genes downstream is conserved i.e.: pseudogene, anti-anti-sigma regulatory factor, 2-polyprenyl-6-methoxyphenol hydroxylase-like oxidoreductase.

Cluster 20

Cluster 20 comprised of 2 CYP125 P450s belong to the SAP species. In both P450s the order of genes downstream is conserved i.e.: glycosyltransferase, methyltransferase and aldehyde dehydrogenase.

Unique clusters (21-28)

Eight CYP125 P450s have unique order of upstream and downstream genes hence they were grouped under unique clusters (clusters 21-28) (Figure 4.2). CYP125 in the unique clusters belong to MAC and NTM (3 P450s each) and SAP and MTBC (1 P450 each) (Table 4.1.).

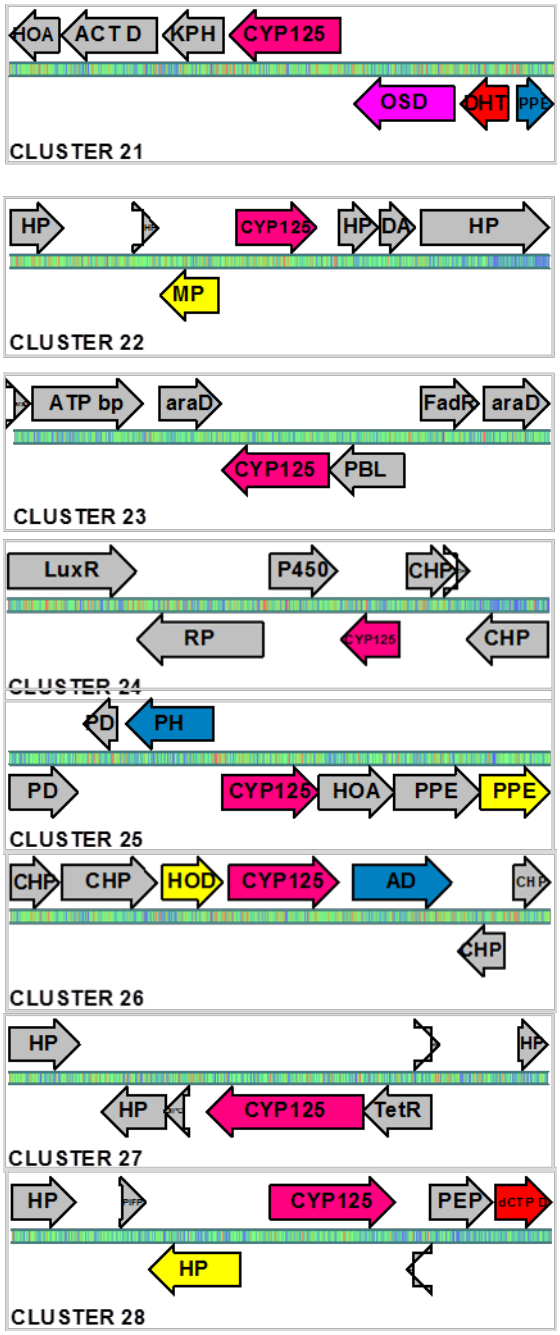


Figure 4.2. CYP125 unique gene-clusters (clusters 21 to 28) analysis in the genus *Mycobacterium*.

4.4. Conclusion

Analysis of CYP125 gene clusters in the genus *Mycobacterium* revealed presence of 28 CYP125 gene-clusters. Gene clusters 1 to 20 comprised of quite a number of P450s ranging from 2 to 23 and gene clusters 21 – 28 named a unique gene clusters considering each of the P450 in this cluster have different genes both in the upstream and downstream of CYP125. Overall, SAP species showed highest CYP125 gene cluster diversity (10 clusters including 1 unique cluster) followed by MAC (8 clusters including 3 unique clusters), NTM (5 clusters including 3 unique clusters), MCAC (4 clusters) and MTBC (2 clusters including 1 unique cluster). This study is first of its kind on analysis of gene-clusters in prokaryote P450s. Some of the CYP125 P450s in different clusters have reverse complement arrangement of genes compared to other CYP125s in the same cluster. These P450s are under investigation for further analysis of possible gene rearrangement events in the chromosome.

4.5. References

- Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C. and Dugave, C., 2009. Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *106*(18), pp.7426-7431.

- Chen, W., Lee, M.K., Jefcoate, C., Kim, S.C., Chen, F. and Yu, J.H., 2014. Fungal cytochrome p450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome biology and evolution*, *6*(7), pp.1620-1634.

- Girhard, M., Bakkes, P.J., Mahmoud, O. and Urlacher, V.B., 2015. P450 Biotechnology. In *Cytochrome P450* (pp. 451-520). Springer International Publishing.

- Guengerich, F.P., 2015. Human cytochrome P450 enzymes. In *Cytochrome P450* (pp. 523-785). Springer International Publishing.

- Hamberger, B. and Bak, S., 2013. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1612), p.20120426.

- Holsclaw, C.M., Sogi, K.M., Gilmore, S.A., Schelle, M.W., Leavell, M.D., Bertozzi, C.R. and Leary, J.A., 2008. Structural characterization of a novel sulfated menaquinone produced by stf3 from *Mycobacterium tuberculosis*. *ACS chemical biology*, *3*(10), pp.619-624.

- Jawallapersand, P., Mashele, S.S., Kovačič, L., Stojan, J., Komel, R., Pakala, S.B., Kraševec, N. and 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*(9), p.e107209.

- Kelly, S.L. and 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?. *Phil. Trans. R. Soc. B*, *368*(1612), p.20120476.

- 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, P. R.). *Springer International Publishing*, Ch. 6, 261–407.

- Mizutani, M. and Ohta, D., 2010. Diversification of P450 genes during land plant evolution. *Annual review of plant biology*, *61*, pp.291-315.

- Moktali, V., Park, J., Fedorova-Abrams, N.D., Park, B., Choi, J., Lee, Y.H. and Kang, S., 2012. Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. *BMC genomics*, *13*(1), p.525.

- Nelson, D.R., 2009. The cytochrome p450 homepage. *Human genomics*, *4*(1), p.1.

- Nelson, D.R., 2013. A world of cytochrome P450s. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1612), p.20120430.

- Poulos, T. L. & Johnson, E. F., 2015. Structures of cytochrome P450 enzymes In *Cytochrome P450* (Ch. 1, 3–32). Springer International Publishing.

- Qhanya, L.B., Matowane, G., Chen, W., Sun, Y., Letsimo, E.M., Parvez, M., Yu, J.H., Mashele, S.S. and Syed, K., 2015. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in Basidiomycete Biotrophic plant pathogens. *PloS one*, *10*(11), p.e0142100.

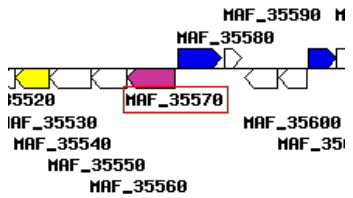
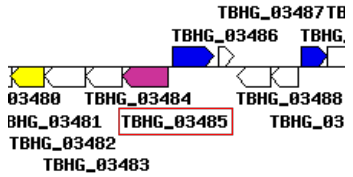
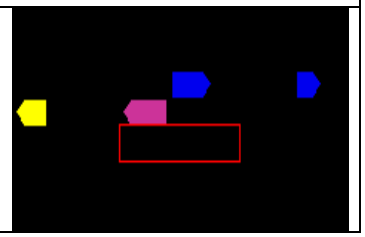
- Schuler, M. A., 2015 P450s in plants, insects, and their fungal pathogens. In *Cytochrome P450* (Ch. 7, 409–449). Springer International Publishing.

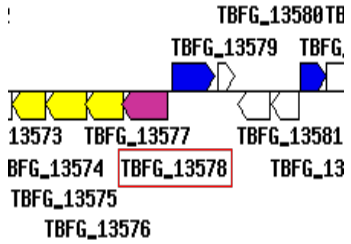
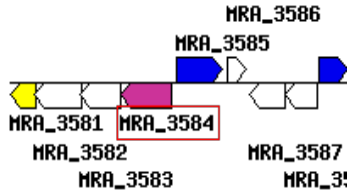
- Sello, M.M., Jafta, N., Nelson, D.R., Chen, W., Yu, J.H., Parvez, M., Kgosiemang, I.K.R., Monyaki, R., Raselemane, S.C., Qhanya, L.B. and Mthakathi, N.T., 2015. Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. Scientific reports, 5.

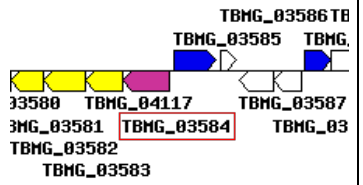
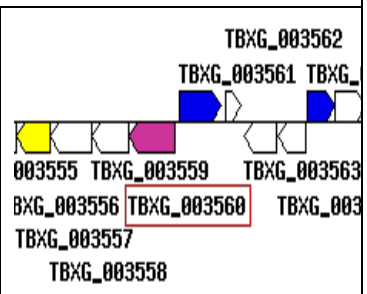
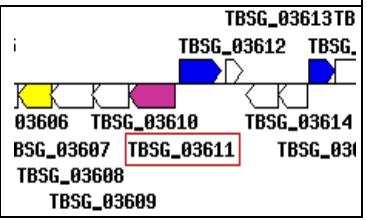
- Sezutsu, H., Le Goff, G. and Feyereisen, R., 2013. Origins of P450 diversity. *Phil. Trans. R. Soc. B*, *368*(1612), p.20120428.

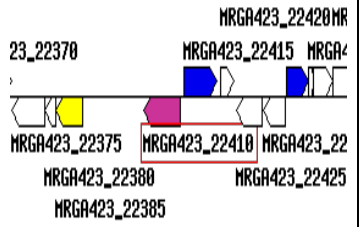
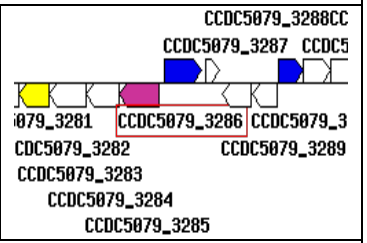
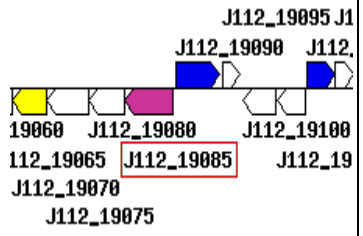
- Syed, K., Shale, K., Pagadala, N.S. and Tuszynski, J., 2014. Systematic identification and evolutionary analysis of catalytically versatile cytochrome P450 monooxygenase families enriched in model basidiomycete fungi. *PLoS One*, *9*(1), p.e86683.

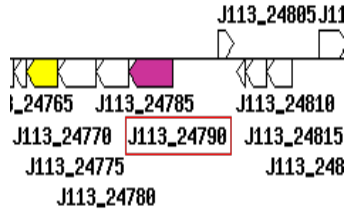
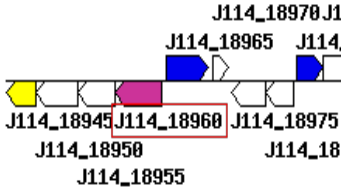
APPENDIX

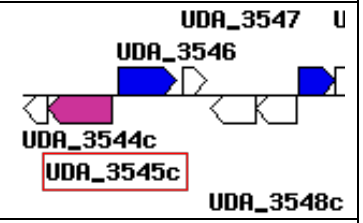
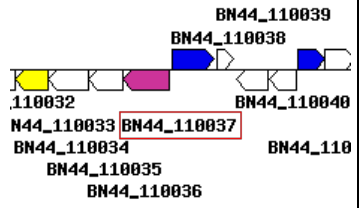
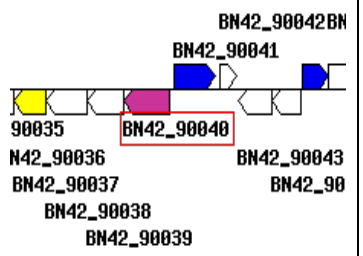
Gene	Gene code	Upstream gene	Direction	Gene code	Downstream gene	Direction	Genetic Map
Mycobacterium africanum GM041182							
cyp125; cytochrome P450	MAF_35570	N/A	N/A	N/A	N/A	N/A	
	MAF_35560	fadE28; acyl-CoA dehydrogenase	←	MAF_35580	fadA5; acetyl-CoA acetyltransferase K00626 acetyl-CoA C-acetyltransferase	→	
	MAF_35550	fadE29; acyl-CoA dehydrogenase	←	MAF_35590	hypothetical protein	→	
	MAF_35540	hypothetical protein; uncharacterized protein	←	MAF_35600	short-chain dehydrogenase/reductase	←	
Mycobacterium tuberculosis Haarlem							
cyp125; cytochrome p450	TBHG_03485.1	N/A	N/A	N/A	N/A	N/A	
	TBHG_03484	acyl-CoA dehydrogenase FadE28	←	TBHG_03486	acetyl-CoA acetyltransferase FadA5	→	
	TBHG_03483	acyl-CoA dehydrogenase FadE29	←	TBHG_03487	deazaflavin-dependent nitroreductase Ddn	→	
	TBHG_03482	hypothetical protein; K07068 uncharacterized protein	←	TBHG_03487	Oxidoreductase	←	
Mycobacterium tuberculosis 7199-99							
cyp125; cytochrome p450	MT7199_3607	N/A	N/A	N/A	N/A	N/A	
	MT7199_3606	putative ACYL-CoA DEHYDROGENASE FADE28	←	MT7199_3608	putative ACETYL-CoA ACETYLTRANSFERASE FADA5 (ACETOACETYL-CoA THIOLASE)	→	
	MT7199_3605	putative ACYL-CoA DEHYDROGENASE FADE29	←	MT7199_3609	ddn; DEAZAFLAVIN-DEPENDENT NITROREDUCTASE DDN	→	

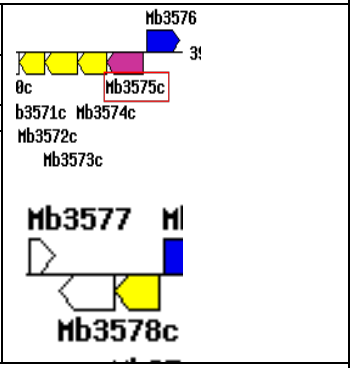
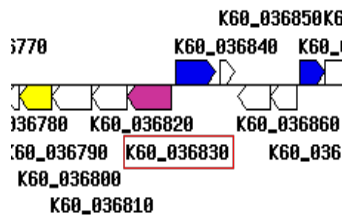
	MT7199_3604	hypothetical protein; K07068 uncharacterized protein	←	MT7199_3610	putative SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE	←	
Mycobacterium tuberculosis F11							
cyp125;cytochrome p450	TBFG_13578	N/A	N/A	N/A	N/A	N/A	
	TBFG_13577	acyl-CoA dehydrogenase fadE28	←	TBFG_13579	K00626 acetyl-CoA C-acetyltransferase	→	
	TBFG_13576	acyl-CoA dehydrogenase fadE29	←	TBFG_13580	conserved hypothetical protein	→	
	TBFG_13575	conserved hypothetical protein; K07068 uncharacterized protein	←	TBFG_13581	hypothetical short-chain type dehydrogenase/reductase	←	
Mycobacterium tuberculosis H37Ra							
cyp125;cytochrome p450	MRA_3584	N/A	N/A	N/A	N/A	N/A	
	MRA_3583	fadE28; acyl-CoA dehydrogenase	←	MRA_3585	fadA5; acetyl-CoA acetyltransferase K00626; acetyl-CoA C-acetyltransferase	→	
	MRA_3582	fadE29; acyl-CoA dehydrogenase	←	MRA_3586	hypothetical protein	←	
	MRA_3581	hypothetical protein; uncharacterized protein	←	MRA_3587	short chain dehydrogenase	←	
Mycobacterium tuberculosis H37Rv							
cyp125;cytochrome p450	Rv3545c	N/A	N/A	N/A	N/A	N/A	
	Rv3544c	fadE28; acyl-CoA dehydrogenase	←	Rv3546	fadA5; acetyl-CoA acetyltransferase	→	

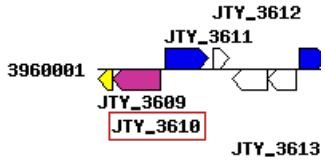
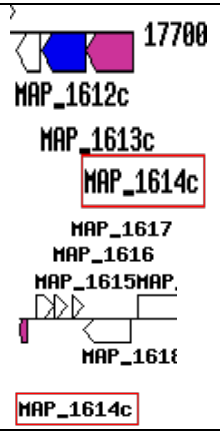
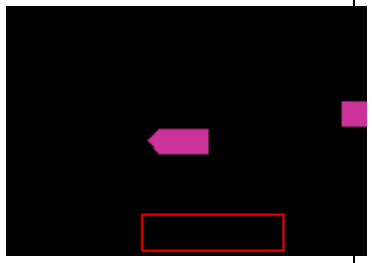
	Rv3543c	fadE29; acyl-CoA dehydrogenase	←	Rv3547	ddn; deazaflavin-dependent nitroreductase	→	
	RV3542c	hypothetical protein; K07068 uncharacterized protein	←	Rv3548c	short-chain type dehydrogenase/reductase	←	
Mycobacterium tuberculosis KZN 1435							
cyp125;cytochrome p450	TBMG_03584	N/A	N/A	N/A	N/A	N/A	
	TBMG_04117	acyl-CoA dehydrogenase	←	TBMG_03585	acetyl-CoA acetyltransferase	→	
	TBMG_03583	acyl-CoA dehydrogenase	←	TBMG_03586	hypothetical protein	→	
	TBMG_03582	hypothetical protein; K07068 uncharacterized protein	←	TBMG_03587	short chain dehydrogenase	←	
Mycobacterium tuberculosis KZN 605							
cyp125; cytochrome p450	TBXG_003560	N/A	N/A	N/A	N/A	N/A	
	TBXG_003559	acyl-CoA dehydrogenase fadE28	←	TBXG_003561	acetyl-CoA acetyltransferase fadA5	→	
	TBXG_003558	acyl-CoA dehydrogenase fadE29	←	TBXG_003562	hypothetical protein	→	
	TBXG_003557	hypothetical protein; K07068 uncharacterized protein	←	TBXG_003563	short-chain type dehydrogenase/reductase	←	
Mycobacterium tuberculosis KZN 4207							
cyp125; cytochrome p450	TBSG_03611	N/A	N/A	N/A	N/A	N/A	
	TBSG_03610	acyl-CoA dehydrogenase fadE28	←	TBSG_03612	acetyl-CoA acetyltransferase fadA5	→	
	TBSG_03609	acyl-CoA dehydrogenase fadE29	←	TBSG_03613	conserved hypothetical protein	→	

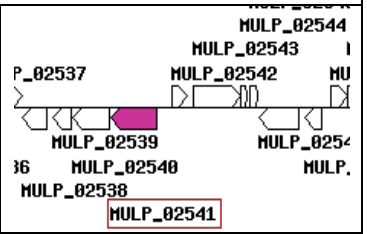
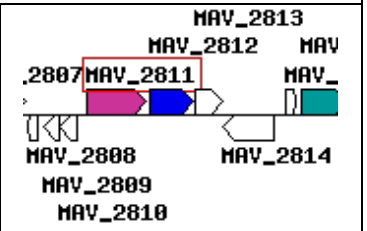
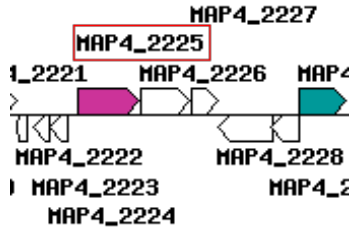
	TBSG_03608	conserved hypothetical protein; K07068 uncharacterized protein	←	TBSG_03614	short-chain type dehydrogenase/reductase	←	
Mycobacterium tuberculosis RGTB327							
cyp125; cytochrome p450	MRGA423_22410	N/A	N/A	N/A	N/A	N/A	
	MRGA423_22485	hypothetical protein; K07068 uncharacterized protein	←	MRGA423_22415	K00626 acetyl-CoA C-acetyltransferase	→	
	MRGA423_22480	hypothetical protein	←	MRGA423_22420	hypothetical protein	→	
	MRGA423_22475	lipid-transfer protein	←	MRGA423_22425	short chain dehydrogenase	←	
Mycobacterium tuberculosis strains CCDC5079							
cyp125; cytochrome p450	CCDC5079_3286	N/A	N/A	N/A	N/A	N/A	
	CCDC5079_3285	acyl-CoA dehydrogenase	←	CCDC5079_3287	K00626 acetyl-CoA C-acetyltransferase	→	
	CCDC5079_3284	acyl-CoA dehydrogenase FADE29	←	CCDC5079_3288	hypothetical protein	→	
	CCDC5079_3283	hypothetical protein; K07068 uncharacterized protein	←	CCDC5079_3289	short chain dehydrogenase	←	
Mycobacterium tuberculosis Beijing/NITR203							
cyp125; cytochrome p450	J112_19085	N/A	N/A	N/A	N/A	N/A	
	J112_19080	acyl-CoA dehydrogenase	←	J112_19090	K00626 acetyl-CoA C-acetyltransferase	→	
	J112_19075	acyl-CoA dehydrogenase	←	J112_19095	hypothetical protein	→	

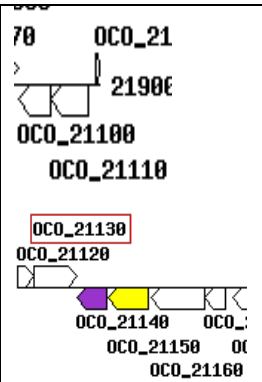
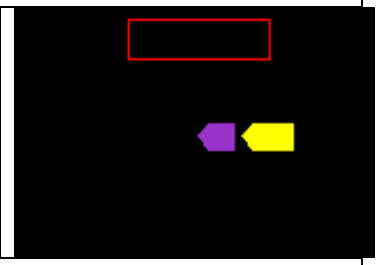
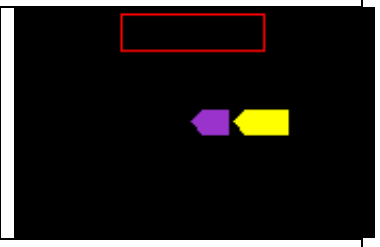
	J112_19070	hypothetical protein; K07068 uncharacterized protein	←	J112_19100	short chain dehydrogenase	←	
Mycobacterium tuberculosis CAS/NITR204							
cyp125; cytochrome p450	J113_24790	N/A	N/A	N/A	N/A	N/A	
	J113_24785	acyl-CoA dehydrogenase	←	J113_24805	hypothetical protein	→	
	J113_24780	acyl-CoA dehydrogenase	←	J113_24810	short chain dehydrogenase	→	
	J113_24775	hypothetical protein; K07068 uncharacterized protein	←	J113_24815	short chain dehydrogenase	←	
Mycobacterium tuberculosis EAI5/NITR206							
cyp125; cytochrome p450	J114_18960	N/A	N/A	N/A	N/A	N/A	
	J114_18955	acyl-CoA dehydrogenase	←	J114_18965	acetyl-CoA C-acetyltransferase	→	
	J114_18950	acyl-CoA dehydrogenase	←	J114_18970	hypothetical protein	→	
	J114_18945	K00626 acetyl-CoA C-acetyltransferase	←	J114_18975	short chain dehydrogenase	←	
Mycobacterium tuberculosis Erdman= ATCC 35801							
cyp125; cytochrome p450	ERDMAN_3890	N/A	N/A	N/A	N/A	N/A	

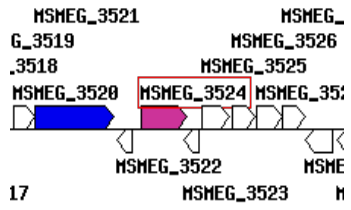
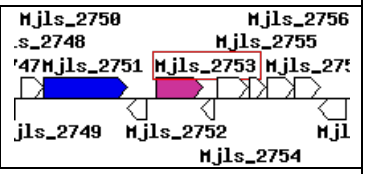
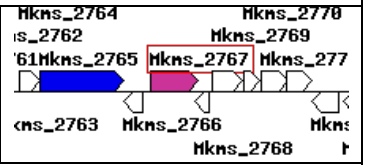
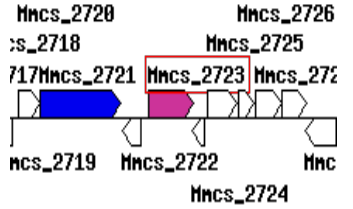
	ERDMAN_3888	fadE28; acyl-CoA dehydrogenase	←	ERDMAN_3891	fadA5; acetyl-CoA acetyltransferase	→	
	ERDMAN_3887	fadE29; acyl-CoA dehydrogenase	←	ERDMAN_3892	hypothetical protein	→	
	ERDMAN_3886	K07068 uncharacterized protein	←	ERDMAN_3893	short chain dehydrogenase	←	
Mycobacterium tuberculosis UT205							
cyp125; cytochrome p450	UDA_3545c	N/A	N/A	N/A	N/A	N/A	
	UDA_3544	fadE28; hypothetical protein	←	UDA_3546	K00626 acetyl-CoA C-acetyltransferase	→	
	UDA_3543c	fadE29; hypothetical protein	←	UDA_3547	hypothetical protein	→	
	UDA_3542c	hypothetical protein; K07068 uncharacterized protein	←	UDA_3548c	hypothetical protein	←	
Mycobacterium canetii CIPT 140060008							
cyp125; cytochrome p450	BN44_110037	N/A	N/A	N/A	N/A	N/A	
	BN44_110036	fadE; Putative FadE28-like Acyl-CoA dehydrogenase	←	BN44_110038	fadA; Putative acetyl-CoA acetyltransferase FadA5 (acetoacetyl-CoA thiolase)	→	
	BN44_110035	fadE; Putative FadE29-like Acyl-CoA dehydrogenase	←	BN44_110039	hypothetical protein	→	
	BN44_110034	hypothetical protein	←	BN44_110040	Putative short-chain type Dehydrogenase/Reductase	←	
Mycobacterium canetii CIPT 140710010							
cyp125; cytochrome p450	BN42_90040	N/A	N/A	N/A	N/A	N/A	
	BN42_90039	Putative FadE28-like Acyl-CoA dehydrogenase	←	BN42_90041	fadA; Putative acetyl-CoA acetyltransferase FadA5	→	
	BN42_90038	Putative FadE29-like Acyl-CoA dehydrogenase	←	BN42_90042	Conserved protein of unknown function	→	
	BN42_90037	K07068 uncharacterized protein	←	BN42_90043	Putative short-chain type Dehydrogenase/Reductase	←	
Mycobacterium bovis BCG Pasteur 1173P2							
cyp125; cytochrome p450	BCG_3609c	N/A	N/A	N/A	N/A	N/A	
	BCG_3608c	acyl-CoA dehydrogenase FADE28	←	BCG_3610	K00626 acetyl-CoA C-	→	

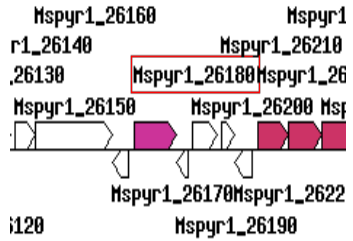
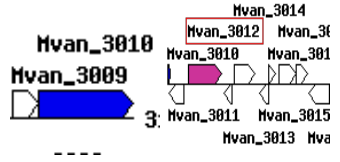
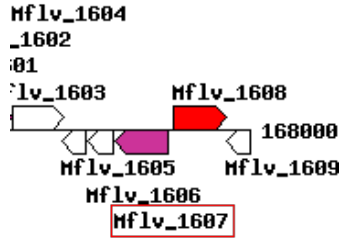
					acetyltransferase		
	BCG_3607c	acyl-CoA dehydrogenase FADE29	←	BCG_3611	hypothetical protein	→	
	BCG_3606c	hypothetical protein; K07068 uncharacterized protein	←	BCG_3612c	short chain dehydrogenase	←	
Mycobacterium bovis AF 2122/97							
cyp125; cytochrome p450	Mb3575c	N/A	N/A	N/A	N/A	N/A	
	Mb3574c	acyl-CoA dehydrogenase FADE28	←	Mb3576	K00626 acetyl-CoA C-acetyltransferase	→	
	Mb3573c	acyl-CoA dehydrogenase FADE29	←	Mb3577	hypothetical protein	→	
	Mb3572c	hypothetical protein; K07068 uncharacterized protein	←	Mb3578c	short chain dehydrogenase	←	
Mycobacterium bovis BCG Korea 1168P							
cyp125; cytochrome p450	K60_036830	N/A	N/A	N/A	N/A	N/A	
	K60_036820	acyl-CoA dehydrogenase FADE28	←	K60_036840	K00626 acetyl-CoA C-acetyltransferase	→	
	K60_036810	acyl-CoA dehydrogenase FADE29	←	K60_036850	hypothetical protein	→	
	K60_036800	hypothetical protein; K07068 uncharacterized protein	←	K60_036860	short chain dehydrogenase	←	
Mycobacterium bovis BCG Mexico							
cyp125; cytochrome p450	BCGMEX_3607c	N/A	N/A	N/A	N/A	N/A	
	BCGMEX_3606c	fadE28; putative acyl-CoA dehydrogenase	←	BCGMEX_3608	fadA5; acetyl-CoA acetyltransferase	→	
	BCGMEX_3605	fadE29; putative acyl-CoA dehydrogenase	←	BCGMEX_3609	hypothetical protein	→	

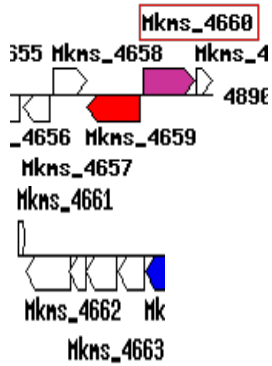
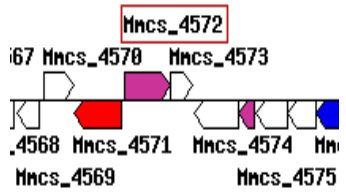
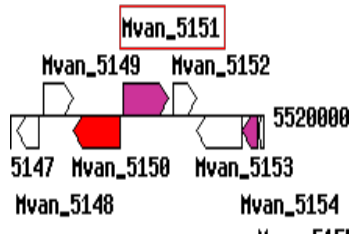
	c						
	BCGMEX_3604 c	hypothetical protein; K07068 uncharacterized protein	←	BCGMEX_3610 c	putative short-chain type dehydrogenase	←	
Mycobacterium bovis BCG Tokyo 172							
cyp125;cytochrome p450	JTY_3610	N/A	N/A	N/A	N/A	N/A	
	JTY_3609	fadE28; putative acyl-CoA dehydrogenase	←	JTY_3611	fadA5; acetyl-CoA acetyltransferase	→	
	JTY_3608	fadE29; putative acyl-CoA dehydrogenase; K00257	←	JTY_3612	hypothetical protein	→	
	JTY_3607	hypothetical protein; K07068 uncharacterized protein	←	JTY_3613	short chain dehydrogenase	←	
Mycobacterium Avium subsp. paratuberculosis K10							
cyp125; cytochrome p450	MAP_1614c	N/A	N/A	N/A	N/A	N/A	
	MAP_1613c	K00001 alcohol dehydrogenase	←	MAP_1615	hypothetical protein	→	
	MAP_1612c	hypothetical protein	←	MAP_1616	hypothetical protein	→	
	MAP_1611	hypothetical protein	→	MAP_1617	hypothetical protein	→	
Mycobacterium Indicus pranii MTCC 9506							
cyp125; cytochrome p450	MIP_04041	N/A	N/A	N/A	N/A	N/A	
	MIP_04040	Alcohol dehydrogenase	←	MIP_04042	hypothetical protein	→	
	MIP_04037	hypothetical protein: Pfam: YceI	←	MIP_04043	hypothetical protein: Pfam: Phage_holin_3_6	→	
	MIP_04036	hypothetical protein	→	MIP_04045	hypothetical protein: Pfam: EspB	→	
Mycobacterium Marinum							
cyp125; cytochrome p450	MMAR_2783	N/A	N/A	N/A	N/A	N/A	

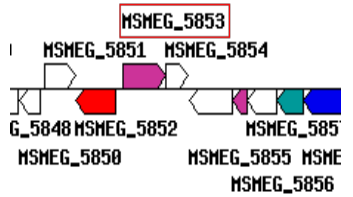
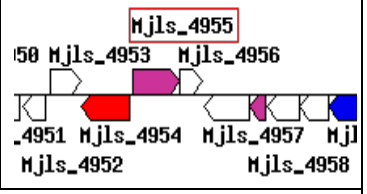
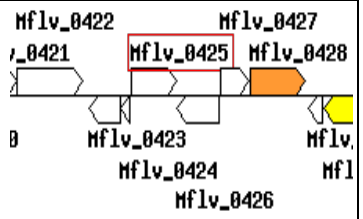
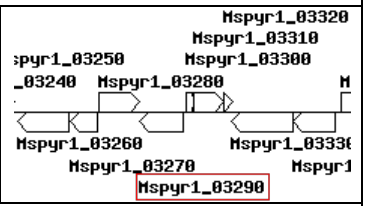
	MMAR_2782	NADP-dependent alcohol dehydrogenase AdhC	←	MMAR_2784	conserved hypothetical membrane protein	→	
	MMAR_2781	conserved hypothetical membrane protein	←	MMAR_2785	conserved hypothetical membrane protein	→	
	MMAR_2780	conserved hypothetical protein	←	MMAR_2786	conserved secreted protein	→	
Mycobacterium liflandii 128FXT							
cyp125; cytochrome p450	MULP_02541	N/A	N/A	N/A	N/A	N/A	
	MULP_02540	NADP-dependent alcohol dehydrogenase AdhC	←	MULP_02542	putative membrane protein	→	
	MULP_02539	hypothetical protein	←	MULP_02543	transposase for IS2404	→	
	MULP_02538	putative membrane protein	←	MULP_02544	pseudogene	→	
Mycobacterium Avium 104							
cyp125;cytochrome p450	MAV_2811	N/A	N/A	N/A	N/A	N/A	
CYP 124	MAV_2810	hypothetical protein	←	MAV_2812	aldehyde dehydrogenase; K00001 alcohol dehydrogenase	→	
	MAV_2809	hypothetical protein	←	MAV_2813	hypothetical protein	→	
	MAV_2808	hypothetical protein	←	MAV_2814	hypothetical protein	←	
Mycobacterium avium subsp. paratuberculosis MAP4							
cyp125; cytochrome p450	MAP4_2225	N/A	N/A	N/A	N/A	N/A	
	MAP4_2224	hypothetical protein:	←	MAP4_2226	Alcohol dehydrogenase	→	
	MAP4_2223	putative membrane protein	←	MAP4_2227	hypothetical protein: Pfam: YceI	→	
	MAP4_2222	hypothetical protein: Pfam: EspB	←	MAP4_2228	hypothetical protein:	←	
Mycobacterium intracellulare ATCC 13950							
cyp125; cytochrome p450	OCU_21370	N/A	N/A	N/A	N/A	N/A	
	OCU_21360	TetR family transcriptional regulator	→	OCU_21380	K01011 thiosulfate/3-mercaptopyruvate sulfurtransferase	←	

	OCU_21350	hypothetical protein:	←	OCU_21390	K14335 alpha-1,6-mannosyltransferase	←	
	OCU_21340	hypothetical protein:	←	OCU_21400	hydroxylase	←	
Mycobacterium intracellulare MOTT-02							
cyp125; cytochrome p450	OCO_21130	N/A	N/A	N/A	N/A	N/A	
	OCO_21120	TetR family transcriptional regulator	→	OCO_21140	K01011 thiosulfate/3-mercaptopyruvate sulfurtr`1ansferase	→	
	OCO_21110	TetR family transcriptional regulator		OCO_21150	K14335 alpha-1,6-mannosyltransferase	←	
	OCO_21000	hypothetical protein:		OCO_21160	hydroxylase	←	
Mycobacterium Intracellulare MOTT-64							
cyp125;cytochrome p450	OCQ_20030	N/A	N/A	N/A	N/A	N/A	
	OCQ_20020	TetR family transcriptional regulator	→	OCQ_20040	K01011 thiosulfate/3-mercaptopyruvate sulfurtransferase	←	
	OCQ_20010	hypothetical protein	←	OCQ_20050	K14335 alpha-1,6-mannosyltransferase	←	
	OCQ_20000	hypothetical protein	←	OCQ_20060	hydroxylase	←	
Mycobacterium intracellulare MOTT-36Y							
cyp125;cytochrome p450	W7S_09920	N/A	N/A	N/A	N/A	N/A	
	W7S_09915	TetR family transcriptional regulator	→	W7S_09925	K01011 thiosulfate/3-mercaptopyruvate sulfurtransferase	→	
	W7S_09910	hypothetical protein	←	W7S_09930	K14335 alpha-1,6-mannosyltransferase	←	
	W7S_09905	hypothetical protein	←	W7S_09935	hydroxylase	←	
Mycobacterium Indicus pranii MTCC 9506							
cyp125;cytochrome p450	MIP_02985	N/A	N/A	N/A	N/A	N/A	
CYP124	MIP_02984	Transcriptional regulator, TetR family	→	MIP_02988	K01011 thiosulfate/3-mercaptopyruvate sulfurtransferase	←	

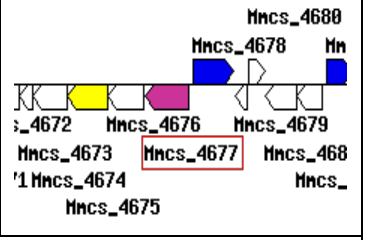
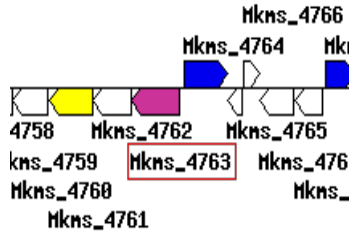
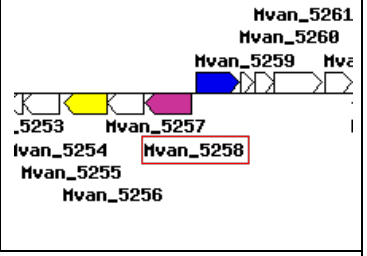
	MIP_02983	Hypothetical protein	←	MIP_02990	K14335 alpha-1,6-mannosyltransferase	←	
	MIP_02982	Hypothetical protein	←	MIP_02991	Acihydroxylase	←	
Mycobacterium smegmatis MC2 155							
cyp125;cytochrome p450	MSMEG_3524	N/A	N/A	N/A	N/A	N/A	
	MSMEG_3522	dopamine receptor D4	←	MSMEG_3523	hypothetical protein	←	
	MSMEG_3521	K00122 formate dehydrogenase	→	MSMEG_3525	XRE family transcriptional regulator	→	
	MSMEG_3520	TetR family transcriptional regulator	→	MSMEG_3526	hypothetical protein	→	
Mycobacterium sp. JLS							
cyp125;cytochrome p450	Mjls_2753	N/A	N/A	N/A	N/A	N/A	
	Mjls_2752	hypothetical protein	←	Mjls_2754	conserved hypothetical protein	←	
	Mjls_2751	K00122 formate dehydrogenase	→	Mjls_2755	transcriptional regulator, XRE family	→	
	Mjls_2750	transcriptional regulator, TetR family	→	Mjls_2756	conserved hypothetical protein	→	
Mycobacterium sp. KMS							
cyp125;cytochrome p450	Mkms_2767	N/A	N/A	N/A	N/A	N/A	
	Mkms_2766	hypothetical protein	←	Mkms_278	hypothetical protein	←	
	Mkms_2765	K00122 formate dehydrogenase	→	Mkms_2769	XRE family transcriptional regulator	→	
	Mkms_2774	TetR family transcriptional regulator	→	Mkms_2770	hypothetical protein	→	
Mycobacterium sp. MCS							
cyp125; cytochrome p450	Mmcs_2723	N/A	N/A	N/A	N/A	N/A	
	Mmcs_2722	conserved hypothetical protein	←	Mmcs_2724	hypothetical protein	←	
	Mmcs_2721	K00122 formate dehydrogenase	→	Mmcs_2725	transcriptional regulator, XRE family	→	
	Mmcs_2720	transcriptional regulator, TetR family	→	Mmcs_2726	conserved hypothetical protein	→	
Mycobacterium gilvum PYR-GCK							
cyp125; cytochrome p450	Mflv_3290	N/A	N/A	N/A	N/A	N/A	
	Mflv_3289	hypothetical protein	←	Mflv_3291	conserved hypothetical protein	←	

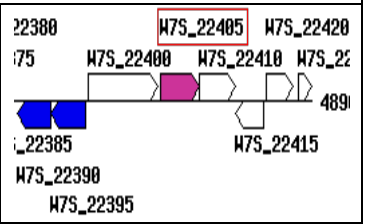
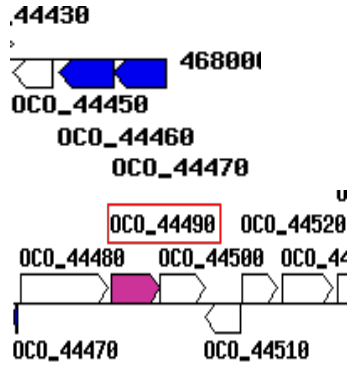
	Mflv_3288	K00122 formate dehydrogenase	→	Mflv_3292	transcriptional regulator, XRE family	→	
	Mflv_3287	transcriptional regulator, TetR family	→	Mflv_3293	conserved hypothetical protein	→	
Mycobacterium gilvum Spyr1							
cyp125;cytochrome p450	Mspyr1_26180	N/A	N/A	N/A	N/A	N/A	
	Mspyr1_26170	hypothetical protein	←	Mspyr1_26190	hypothetical protein	←	
	Mspyr1_26160	anaerobic dehydrogenase, typically selenocysteine-containing	→	Mspyr1_26200	transcriptional regulator, XRE family	→	
	Mspyr1_26150	transcriptional regulator, TetR family	→	Mspyr1_26210	hypothetical protein	→	
Mycobacterium vanbaalenii PYR-1							
cyp125; cytochrome p450	Mvan_3012	N/A	N/A	N/A	N/A	N/A	
	Mvan_3011	conserved hypothetical protein	←	Mvan_3013	conserved hypothetical protein	←	
	Mvan_3010	K00122 formate dehydrogenase	→	Mvan_3014	transcriptional regulator, XRE family	→	
	Mvan_3009	transcriptional regulator, TetR family	→	Mvan_3015	pseudogene	←	
Mycobacterium gilvum PYR-GCK							
cyp125;cytochrome p450	Mflv_1607	N/A	N/A	N/A	N/A	N/A	
	Mflv_1606	transcriptional regulator, TetR family	←	Mflv_1608	K01945 phosphoribosylamine--glycine ligase	→	
	Mflv_1605	hypothetical protein	←	Mflv_1609	putative esterase	←	
	Mflv_1604	hypothetical protein	→	Mflv_1610	putative transcriptional regulator, TetR family	→	
Mycobacterium gilvum Spyr1							
cyp125;cytochrome p450	Mspyr1_09930	N/A	N/A	N/A	N/A	N/A	

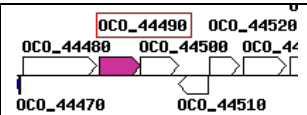
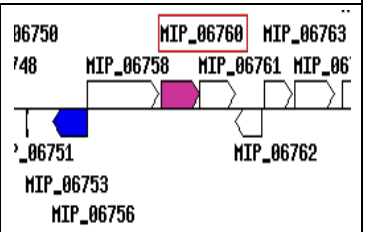
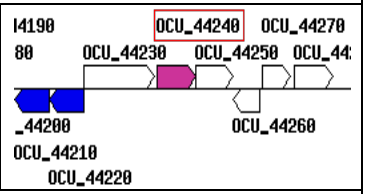
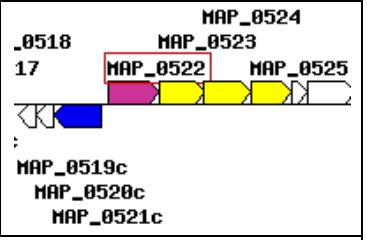
	Mspyr1_09920	transcriptional regulator, TetR family	←	Mspyr1_09940	K01945 phosphoribosylamine--glycine ligase	→	
	Mspyr1_09910	hypothetical protein	←	Mspyr1_09950	Putative esterase	←	
	Mspyr1_09900	hypothetical protein	→	Mspyr1_09960	uncharacterized conserved protein	→	
Mycobacterium sp. KMS							
cyp125; cytochrome p450	Mkms_4660	N/A	N/A	N/A	N/A	N/A	
	Mkms_4659	K01945 phosphoribosylamine--glycine ligase	←	Mkms_4661	TetR family transcriptional regulator	→	
	Mkms_4658	putative esterase	→	Mkms_4662	hypothetical protein	←	
	Mkms_4657	TetR family transcriptional regulator	←	Mkms_4663	carboxymuconolactone decarboxylase	←	
Mycobacterium sp. MCS							
cyp125;cytochrome p450	Mmcs_4572	N/A	N/A	N/A	N/A	N/A	
	Mmcs_4571	K01945 phosphoribosylamine--glycine ligase	←	Mmcs_4573	transcriptional regulator, TetR family	→	
	Mmcs_4570	putative esterase	→	Mmcs_4574	conserved hypothetical protein	←	
	Mmcs_4569	putative transcriptional regulator, TetR family	←	Mmcs_4575	K01607 4-carboxymuconolactone decarboxylase	←	
Mycobacterium vanbaalenii PYR-1							
cyp125;cytochrome p450	Mvan_5151	N/A	N/A	N/A	N/A	N/A	
	Mvan_5150	K01945 phosphoribosylamine--glycine ligase	←	Mvan_5152	transcriptional regulator, TetR family	→	
	Mvan_5149	putative esterase	→	Mvan_5153	conserved hypothetical protein	←	

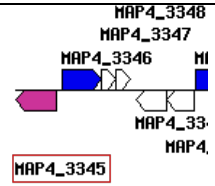
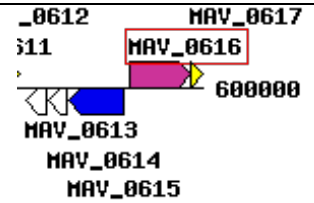
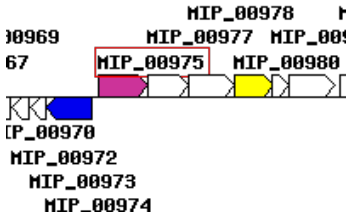
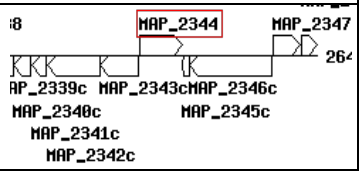
	Mvan_5148	putative transcriptional regulator, TetR family	←	Mvan_5154	K01607 4-carboxymuconolactone decarboxylase	←	
Mycobacterium smegmatis MC2 155							
cyp125;cytochrome p450	MSMEG_5853	N/A	N/A	N/A	N/A	N/A	
	MSMEG_5852	K01945 phosphoribosylamine--glycine ligase	←	MSMEG_5854	TetR family transcriptional regulator	←	
	MSMEG_5851	esterase	←	MSMEG_5855	hypothetical protein	→	
	MSMEG_5850	esterase	←	MSMEG_5856	K01607 4-carboxymuconolactone decarboxylase	→	
Mycobacterium sp. JLS							
cyp125;cytochrome p450	Mjls_4955	N/A	N/A	N/A	N/A	N/A	
	Mjls_4954	phosphoribosylamine--glycine ligase	←	Mjls_4956	transcriptional regulator, TetR family	→	
	Mjls_4953	putative esterase	→	Mjls_4957	conserved hypothetical protein	←	
	Mjls_4952	putative transcriptional regulator, TetR family	←	Mjls_4958	K01607 4-carboxymuconolactone decarboxylase	←	
Mycobacterium gilvum PYR-GCK							
cyp125;cytochrome p450	Mflv_0425	N/A	N/A	N/A	N/A	N/A	
	Mflv_0424	conserved hypothetical protein	←	Mflv_0426	glycosyl transferase, group 1	←	
	Mflv_0423	conserved hypothetical protein	←	Mflv_0427	Methyltransferase type 11	→	
	Mflv_0422	conserved hypothetical protein	→	Mflv_0428	K00130 betaine-aldehyde dehydrogenase	→	
Mycobacterium gilvum Spyr1							
cyp125;cytochrome p450	Mspyr1_03290	N/A	N/A	N/A	N/A	N/A	
	Mspyr1_03280	glycosyltransferase	→	Mspyr1_03300	hypothetical protein	←	
	Mspyr1_03270	methylase involved in ubiquinone/menaquinone biosynthesis	←	Mspyr1_03310	uncharacterized conserved protein	→	
	Mspyr1_03260	NAD-dependent aldehyde dehydrogenase	←	Mspyr1_03320	hypothetical protein	→	
Mycobacterium rhodesiae NBB3							
cyp125;cytochrome	MycrhN_4756	N/A	N/A	N/A	N/A	N/A	

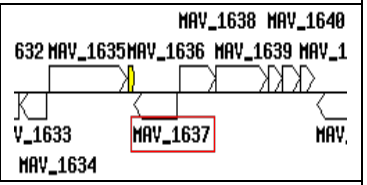
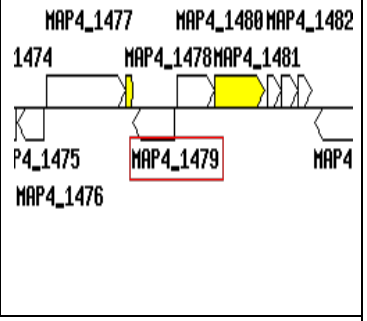
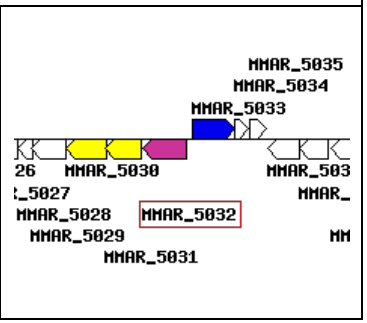
p450							
	MycrhN_4755	putative acyl-CoA transferase	→	MycrhN_4757	Rieske (2Fe-2S) domain-containing protein	→	
	MycrhN_4754	transcriptional regulator	←	MycrhN_4758	acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II; K02182	→	
	MycrhN_4753	short-chain alcohol dehydrogenase	←	MycrhN_4759	hypothetical protein	→	
Mycobacterium smegmatis JS623							
cyp125; cytochrome p450	Mycsm_03179	N/A	N/A	N/A	N/A	N/A	
	Mycsm_03178	putative acyl-CoA transferase	→	Mycsm_03180	Rieske (2Fe-2S) domain-containing protein	→	
	Mycsm_03177	transcriptional regulator	←	Mycsm_03181	acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II; K02182	→	
	Mycsm_03176	short-chain alcohol dehydrogenase	←	Mycsm_03182	hypothetical protein	→	
Mycobacterium rhodesiae NBB3							
cyp125;cytochrome p450	MycrhN_2423	N/A	N/A	N/A	N/A	N/A	
	MycrhN_2422	transcriptional regulator	←	MycrhN_2424	N-dimethylarginine dimethylaminohydrolase; K01482	→	
	MycrhN_2421	transcriptional regulator	→	MycrhN_2425	K01945 phosphoribosylamine--glycine ligase	→	
	MycrhN_2420	hypothetical protein	←	MycrhN_2426	enterochelin esterase-like enzyme	←	
Mycobacterium smegmatis JS623							
cyp125;cytochrome p450	Mycsm_05807	N/A	N/A	N/A	N/A	N/A	
	Mycsm_05806	acyl-CoA dehydrogenase	←	Mycsm_05808	K00626 acetyl-CoA C-acetyltransferase	→	
	Mycsm_05805	acyl-CoA dehydrogenase	←	Mycsm_05809	hypothetical protein	→	
	Mycsm_05804	putative nucleic-acid-binding protein containing a Zn-ribbon	←	Mycsm_05810	deazaflavin-dependent nitroreductase family protein	→	
Mycobacterium chubuense NBB4							
cyp125; cytochrome p450	Mycch_4638	N/A	N/A	N/A	N/A	N/A	

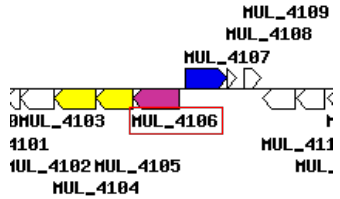
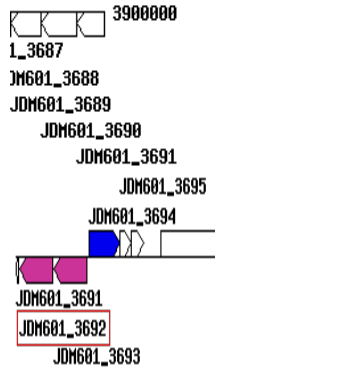
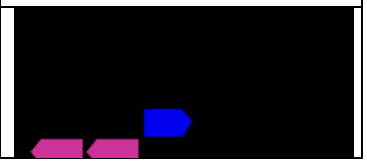
	Mycch_4637	acyl-CoA dehydrogenase	←	Mycch_4639	K00626 acetyl-CoA C-acetyltransferase	→	
	Mycch_4636	acyl-CoA dehydrogenase	←	Mycch_4640	hypothetical protein	→	
	Mycch_4635	putative nucleic-acid-binding protein containing a Zn-ribbon	←	Mycch_4641	deazaflavin-dependent nitroreductase family protein	→	
Mycobacterium sp. MCS							
cyp125;cytochrome p450	Mmcs_4677	N/A	N/A	N/A	N/A	N/A	
	Mmcs_4676	acyl-CoA dehydrogenase-like protein	←	Mmcs_4679	K00626 acetyl-CoA C-acetyltransferase	←	
	Mmcs_4675	acyl-CoA dehydrogenase-like protein	←	Mmcs_4680	conserved hypothetical protein	→	
	Mmcs_4674	protein of unknown function DUF35	←	Mmcs_4681	hypothetical protein	←	
Mycobacterium sp. KMS							
cyp125;cytochrome p450	Mkms_4763	N/A	N/A	N/A	N/A	N/A	
	Mkms_4762	acyl-CoA dehydrogenase domain-containing protein	←	Mkms_4764	K00626 acetyl-CoA C-acetyltransferase	→	
	Mkms_4761	acyl-CoA dehydrogenase domain-containing protein	←	Mkms_4765	hypothetical protein	←	
	Mkms_4760	hypothetical protein	←	Mkms_4766	hypothetical protein	→	
Mycobacterium vanbaalenii PYR-1							
cyp125;cytochrome p450	Mvan_5258	N/A	N/A	N/A	N/A	N/A	
	Mvan_5257	acyl-CoA dehydrogenase domain protein	←	Mvan_5259	K00626 acetyl-CoA C-acetyltransferase	→	
	Mvan_5256	acyl-CoA dehydrogenase domain protein	←	Mvan_5260	conserved hypothetical protein	→	
	Mvan_5255	protein of unknown function DUF35	←	Mvan_5261	hypothetical protein	→	
Mycobacterium rhodesiae NBB3							
cyp125;cytochrome p450	MycrhN_2286	N/A	N/A	N/A	N/A	N/A	

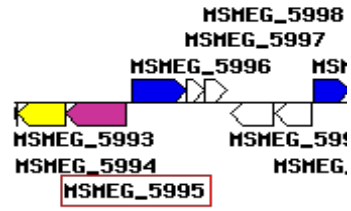
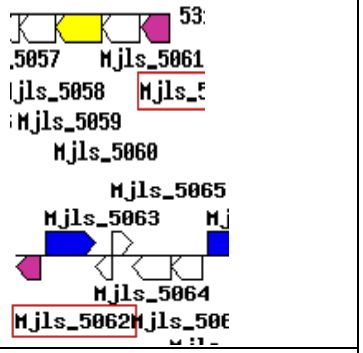
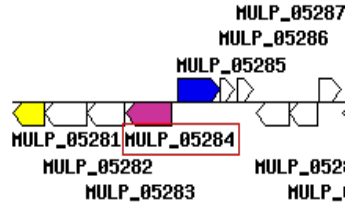
	MycrhN_2285	K00626 acetyl-CoA C-acetyltransferase	→	MycrhN_2287	acyl-CoA dehydrogenase	→	
	MycrhN_2284	hypothetical protein	→	MycrhN_2288	acyl-CoA dehydrogenase	→	
	MycrhN_2283	deazaflavin-dependent nitroreductase family protein	→	MycrhN_2289	putative nucleic-acid-binding protein containing a Zn-ribbon	→	
Mycobacterium intracellulare MOTT-36Y							
cyp125;cytochrome p450	W7S_22405	N/A	N/A	N/A	N/A	N/A	
	W7S_22400	hypothetical protein	→	W7S_22410	hypothetical protein	→	
	W7S_22395	K00965 UDPglucose--hexose-1-phosphate uridylyltransferase	←	W7S_22415	acetoin cleaving system dihydrolipoyllysine-residue acetyltransferase	←	
	W7S_22380	K00849 galactokinase	←	W7S_22420	hypothetical protein	→	
Mycobacterium intracellulare MOTT-02							
cyp125;cytochrome p450	OCO_44490	N/A	N/A	N/A	N/A	N/A	
CYP124	OCO_44480	hypothetical protein	→	OCO_44500	hypothetical protein	→	
	OCO_44470	K00965 UDPglucose--hexose-1-phosphate uridylyltransferase	←	OCO_44510	dihydrolipoyllysine-residue acetyltransferase component of acetoincleaving system	←	
	OCO_44460	K00849 galactokinase	←	OCO_44520	hypothetical protein	→	
Mycobacterium Intracellulare MOTT-64							
cyp125; cytochrome p450	OCQ_45630	N/A	N/A	N/A	N/A	N/A	
CYP124	OCQ_45620	hypothetical protein	→	OCQ_45640	hypothetical protein	→	

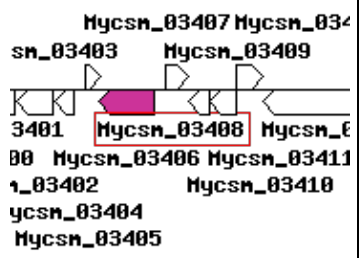
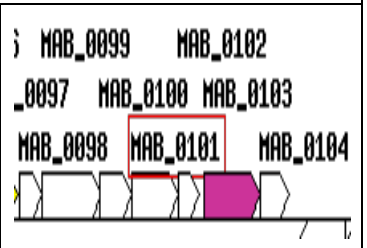
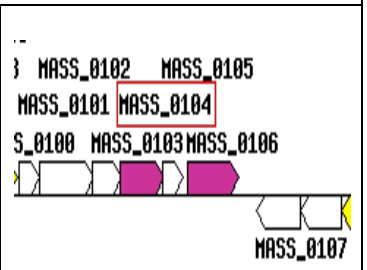
	OCQ_45610	K00965 UDPglucose--hexose-1-phosphate uridylyltransferase	←	OCQ_45650	dihydrolipoyllysine-residue acetyltransferase component of acetoincleaving system	←	
	OCQ_45600	K00849 galactokinase	←	OCQ_45660	hypothetical protein	→	
Mycobacterium Indicus pranii MTCC 9506							
cyp125; cytochrome p450	MIP_06760	N/A	N/A	N/A	N/A	N/A	
	MIP_06758	pseudogene	→	MIP_06761	Hypothetical protein	→	
	MIP_06756	K00965 UDPglucose--hexose-1-phosphate uridylyltransferase	←	MIP_06762	Dihydrolipoyllysine-residue acetyltransferase component of acetoin cleaving system	←	
	MIP_06753	Hypothetical protein	←	MIP_06763	Hypothetical protein	→	
Mycobacterium intracellulare ATCC 13950							
cyp125;cytochrome p450	OCU_44240	N/A	N/A	N/A	N/A	N/A	
CYP124	OCU_44230	hypothetical protein:	→	OCU_44250	hypothetical protein:	→	
	OCU_44220	galactose-1-phosphate uridylyltransferase : K00965 UDPglucose--hexose-1-phosphate uridylyltransferase	←	OCU_44260	dihydrolipoyllysine-residue acetyltransferase component of acetoincleaving system	←	
	OCU_44210	K00849 galactokinase	←	OCU_44270	hypothetical protein:	→	
Mycobacterium Avium subsp. paratuberculosis K10							
cyp125;cytochrome p450	MAP_0522	N/A	N/A	N/A	N/A	N/A	
	MAP_0521c	K00626 acetyl-CoA C-acetyltransferase	←	MAP_0523	FadE28; Geraniol degradation	→	
	MAP_0520c	hypothetical protein	←	MAP_0524	FadE29; Geraniol degradation	→	
	MAP_0519c	hypothetical protein	←	MAP_0525	K07068 uncharacterized protein	→	
Mycobacterium avium subsp. paratuberculosis MAP4							
cyp125;cytochrome p450	MAP4_3345	N/A	N/A	N/A	N/A	N/A	
	MAP4_3344	putative acyl-CoA dehydrogenase FadE28	←	MAP4_3346	acetyl-CoA acetyltransferase FadA5	→	

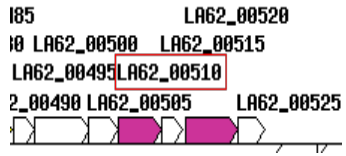
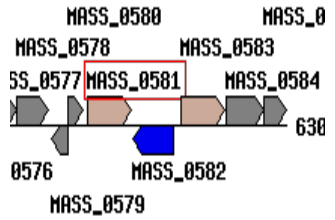
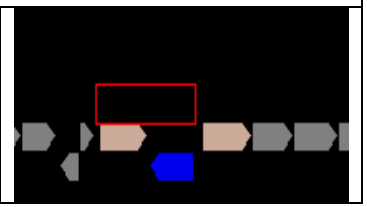
	MAP4_3343	acyl-CoA dehydrogenase FadE29	←	MAP4_3347	hypothetical protein:	→	
	MAP4_3342	hypothetical protein; K07068 uncharacterized protein	←	MAP4_3348	hypothetical protein: Pfam: DUF385 Pyridox_oxidase PGC7_Stella	→	
Mycobacterium Avium 104							
cyp125; cytochrome p450	MAV_0616	N/A	N/A	N/A	N/A	N/A	
	MAV_0615	K00626 acetyl-CoA C-acetyltransferase	←	MAV_0617	acyl-CoA dehydrogenase	→	
	MAV_0614	hypothetical protein	←	MAV_0618	acyl-CoA dehydrogenase	→	
	MAV_0613	AclJ protein	←	MAV_0619	hypothetical protein; K07068 uncharacterized protein	→	
Mycobacterium Indicus pranii MTCC 9506							
cyp125; cytochrome p450	MIP_00975	N/A	N/A	N/A	N/A	N/A	
	MIP_00974	K00626 acetyl-CoA C-acetyltransferase	←	MIP_00977	Crotonobetainyl-CoA dehydrogenase	→	
	MIP_00973	hypothetical protein	←	MIP_00978	Isovaleryl-CoA dehydrogenase	→	
	MIP_00972	AclJ protein	←	MIP_00980	K07068 uncharacterized protein	→	
Mycobacterium Avium subsp. paratuberculosis K10							
cyp125; cytochrome p450	MAP_2344	N/A	N/A	N/A	N/A	N/A	
	MAP_2343c	hypothetical protein Pfam: Rieske Ring hydroxyl_A	←	MAP_2345c	hypothetical protein: Pfam: Fer4_13, Fer4_15,Fer4_19	←	

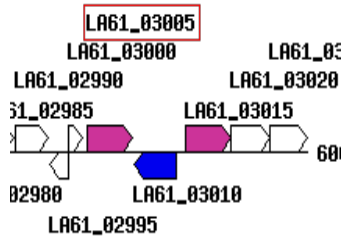
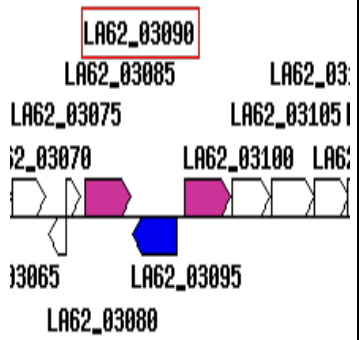
	MAP_2342c	hypothetical protein: Pfam: AMP-binding AMP-binding_C	←	MAP_2346c	hypothetical protein: Pfam: CoA_transf_3	←	
	MAP_2341c	hypothetical protein: Pfam: SnoaL_4	←	MAP_2347	hypothetical protein: Pfam: FCD GntR Rrf2	→	
Mycobacterium Avium 104							
cyp125;cytochrome p450	MAV_1637	N/A	N/A	N/A	N/A	N/A	
	MAV_1636	K05337 ferredoxin	→	MAV_1638	Rieske (2Fe-2S) domain-containing protein	←	
	MAV_1635	caib/baif family protein	→	MAV_1639	AMP-binding enzyme	←	
	MAV_1634	GntR family transcriptional regulator	←	MAV_1640	hypothetical protein	←	
Mycobacterium avium subsp. paratuberculosis MAP4							
cyp125; cytochrome p450	MAP4_1479	N/A	N/A	N/A	N/A	N/A	
	MAP4_1478	K05337 ferredoxin	→	MAP4_1480	oxidoreductase, Rieske 2Fe-2S domain protein	→	
	MAP4_1477	putative CoA-transferase family protein	→	MAP4_1481	putative AMP-binding enzyme; K02182 crotonobetaine/carnitine-CoA ligase	→	
	MAP4_1476	Transcriptional regulator, GntR family	←	MAP4_1482	hypothetical protein:	→	
Mycobacterium Marinum							
cyp125; cytochrome p450	MMAR_5032	N/A	N/A	N/A	N/A	N/A	
	MMAR_5031	acyl-CoA dehydrogenase FadE28	←	MMAR_5033	acetyl-CoA acetyltransferase FadA5	→	
	MMAR_5030	acyl-CoA dehydrogenase FadE29	←	MMAR_5034	conserved hypothetical protein	→	

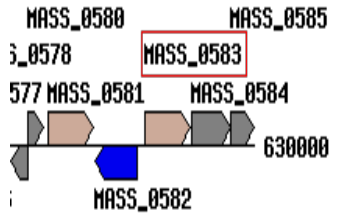
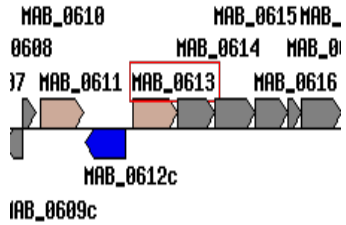
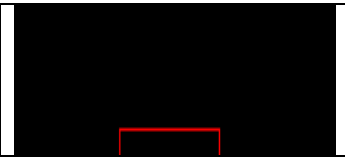
	MMAR_5029	conserved hypothetical protein	←	MMAR_5035	conserved hypothetical protein	→	
Mycobacterium ulcerans Agy99							
cyp125; cytochrome p450	MUL_4106	N/A	N/A	N/A	N/A	N/A	
	MUL_4105	acyl-CoA dehydrogenase FadE28	←	MUL_4107	acetyl-CoA acetyltransferase	→	
	MUL_4104	acyl-CoA dehydrogenase FadE29	←	MUL_4108	pseudogene	→	
	MUL_4103	hypothetical protein	←	MUL_4109	hypothetical protein	→	
Mycobacterium sp. JDM601							
cyp125;cytochrome p450	JDM601_3692	N/A	N/A	N/A	N/A	N/A	
	JDM601_3691	fadE28; acyl-CoA dehydrogenase FadE28	←	JDM601_3693	conserved hypothetical protein	←	
	JDM601_3690	fadE29; acyl-CoA dehydrogenase FadE29	←	JDM601_3694	fadA5; acetyl-CoA acetyltransferase	→	
	JDM601_3689	conserved hypothetical protein	←	JDM601_3695	conserved hypothetical protein	→	
cyp125; cytochrome p450	JDM601_3693	N/A	N/A	N/A	N/A	N/A	
	JDM601_3692	cyp125A7; cytochrome P450 125A7 Cyp125A7; K15981 cholest-4-en-3-one 26-monooxygenase	←	JDM601_3694	fadA5; acetyl-CoA acetyltransferase	→	
	JDM601_3691	fadE28; acyl-CoA dehydrogenase FadE28	←	JDM601_3695	conserved hypothetical protein	→	

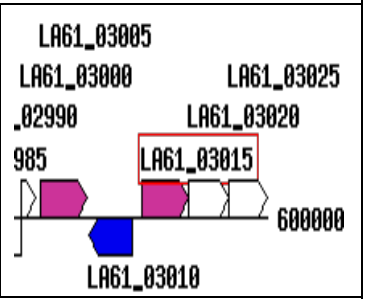
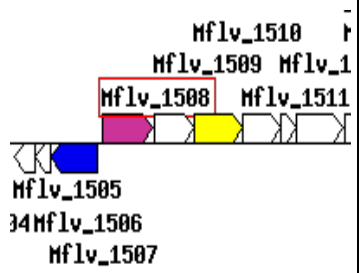
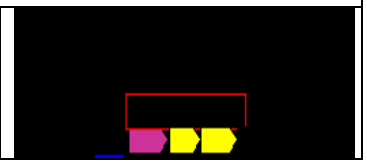
	JDM601_3690	fadE29; acyl-CoA dehydrogenase FadE29	←	JDM601_3696	conserved hypothetical protein	→	
Mycobacterium smegmatis MC2 155							
cyp125;cytochrome p450	MSMEG_5995	N/A	N/A	N/A	N/A	N/A	
	MSMEG_5994	acyl-CoA dehydrogenase; K00257	←	MSMEG_5996	K00626 acetyl-CoA C-acetyltransferase	→	
	MSMEG_5993	acyl-CoA dehydrogenase; K00257	←	MSMEG_5997	hypothetical protein	→	
	MSMEG_5992	hypothetical protein	←	MSMEG_5998	hypothetical protein	→	
Mycobacterium sp. JLS							
cyp125; cytochrome p450	Mjls_5062	N/A	N/A	N/A	N/A	N/A	
	Mjls_5061	acyl-CoA dehydrogenase domain protein	←	Mjls_5063	K00626 acetyl-CoA C-acetyltransferase	→	
	Mjls_5060	acyl-CoA dehydrogenase domain protein	←	Mjls_5064	conserved hypothetical protein	←	
	Mjls_5059	protein of unknown function DUF35	←	Mjls_5065	conserved hypothetical protein	→	
Mycobacterium liflandii 128FXT							
cyp125; cytochrome p450	MULP_05284	N/A	N/A	N/A	N/A	N/A	
	MULP_05283	acyl-CoA dehydrogenase FadE28	←	MULP_05285	fadA5; acetyl-CoA acetyltransferase FadA5	→	
	MULP_05282	acyl-CoA dehydrogenase FadE29	←	MULP_05286	hypothetical protein	→	
	MULP_05281	putative nucleic-acid-binding protein	←	MULP_05287	hypothetical protein	→	
Mycobacterium rhodesiae NBB3							
cyp125;cytochrome p450	MycrhN_4947	N/A	N/A	N/A	N/A	N/A	
	MycrhN_4946	hypothetical protein	→	MycrhN_4948	hypothetical protein	→	

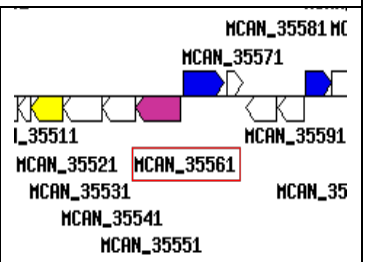
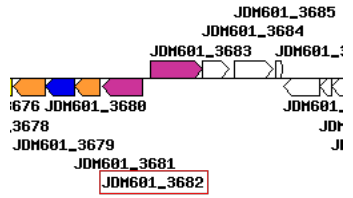
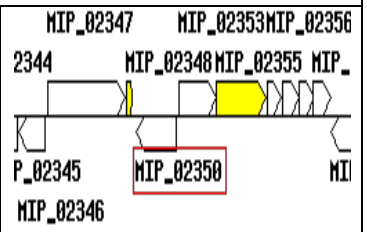
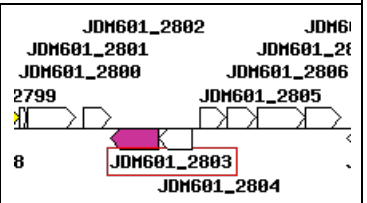
	MycrhN_4945	putative transcriptional regulator	←	MycrhN_4949	Protein of unknown function (DUF2867)	←	
	MycrhN_4944	conserved lipoprotein/antigen	←	MycrhN_4950	Protein of unknown function (DUF2867)	←	
Mycobacterium smegmatis JS623							
cyp125; cytochrome p450	Mycsm_03408	N/A	N/A	N/A	N/A	N/A	
	Mycsm_03407	hypothetical protein	→	Mycsm_03409	hypothetical protein	→	
	Mycsm_03406	Protein of unknown function (DUF732)	←	Mycsm_03410	Protein of unknown function (DUF2867)	←	
	Mycsm_03405	putative transcriptional regulator	←	Mycsm_03411	Protein of unknown function (DUF2867)	←	
Mycobacterium abscessus ATCC 19977							
cyp125; cytochrome p450	MAB_0101	N/A	N/A	N/A	N/A	N/A	
	MAB_0100	Putative short chain dehydrogenase/reductase	→	MAB_0102	Putative methyltransferase	→	
	MAB_0099	Probable monooxygenase	→	MAB_0103	Probable monooxygenase EthA; K10215 monooxygenase	→	
	MAB_0098	Putative TetR-family transcriptional regulator	→	MAB_0104	Probable enoyl-CoA hydratase/isomerase	→	
Mycobacterium abscessus subsp. bolletii 50594							
cyp125;cytochrome p450	MASS_0104	N/A	N/A	N/A	N/A	N/A	
	MASS_0103	putative short chain dehydrogenase/reductase	→	MASS_0105	putative methyltransferase	→	
	MASS_0102	Monooxygenase	→	MASS_0106	monooxygenase EthA; K10215 monooxygenase	→	
	MASS_0101	TetR family transcriptional regulator	→	MASS_0107	xylitol oxidase	←	
Mycobacterium abscessus103							
cyp125;cytochrome p450	LA61_00515	N/A	N/A	N/A	N/A	N/A	

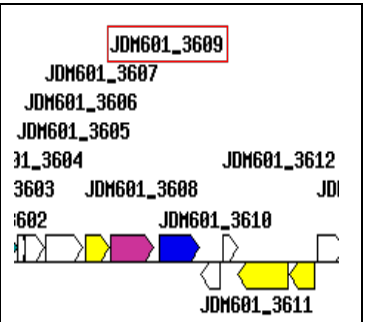
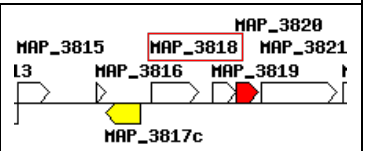
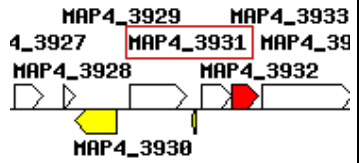
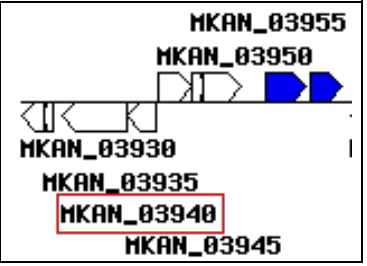
	LA61_00510	NAD-dependent oxidoreductase	→	LA61_00520	methyltransferase	→	
	LA61_00505	monooxygenase	→	LA61_00525	FAD-containing monooxygenase EthA; K10215 monooxygenase	→	
	LA61_00500	TetR family transcriptional regulator	→	LA61_00530	crotonase	→	
Mycobacterium abscessus subsp. bolletii MA 1948							
cyp125;cytochrome p450	LA62_00510	N/A	N/A	N/A	N/A	N/A	
	LA62_00505	NAD-dependent oxidoreductase	→	LA62_00515	methyltransferase	→	
	LA62_00500	monooxygenase	→	LA62_00520	FAD-containing monooxygenase EthA; K10215 monooxygenase	→	
	LA62_00495	TetR family transcriptional regulator	→	LA62_00525	crotonase	→	
Mycobacterium abscessus subsp. bolletii 50594							
cyp125;cytochrome p450	MASS_0581	N/A	N/A	N/A	N/A	N/A	
	MASS_0580	hypothetical protein	→	MASS_0582	acetyl-CoA acetyltransferase	←	
	MASS_0579	hypothetical protein	←	MASS_0583	putative cytochrome P450; K15981 cholest-4-en-3-one 26-monooxygenase	→	
	MASS_0578	short chain dehydrogenase	→	MASS_0584	acyl-CoA dehydrogenase FadE	→	
Mycobacterium abscessus ATCC 19977							
cyp125; cytochrome p450	MAB_0611	N/A	N/A	N/A	N/A	N/A	
	MAB_0610	hypothetical protein	→	MAB_0612	Probable acetyl-CoA acetyltransferase	←	

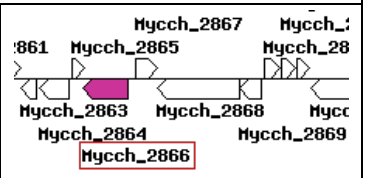
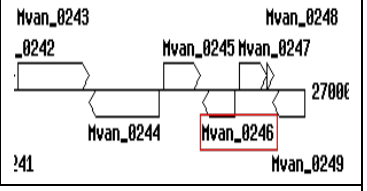
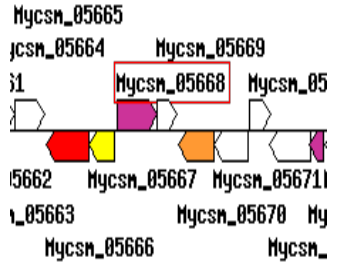
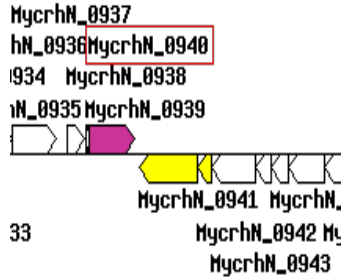
	MAB_0609	hypothetical protein	←	MAB_0613	Putative cytochrome P450; K15981 cholest-4-en-3-one 26-monooxygenase	→	
	MAB_0608	Probable short-chain dehydrogenase/reductase	→	MAB_0614	Probable acyl-CoA dehydrogenase FadE	→	
Mycobacterium abscessus103							
cyp125; cytochrome p450	LA61_03005	N/A	N/A	N/A	N/A	N/A	
	LA61_03000	hypothetical protein	→	LA61_03010	K00626 acetyl-CoA C-acetyltransferase	←	
	LA61_02995	nitroreductase	←	LA61_03015	steroid C27-monooxygenase; K15981 cholest-4-en-3-one 26-monooxygenase	→	
	LA61_02990	short-chain dehydrogenase	→	LA61_03020	acyl-CoA dehydrogenase	→	
Mycobacterium abscessus subsp. bolletii MA 1948							
cyp125; cytochrome p450	LA62_03090	N/A	N/A	N/A	N/A	N/A	
	LA62_03085	hypothetical protein	→	LA62_03095	acetyl-CoA acetyltransferase; K00626 acetyl-CoA C-acetyltransferase	←	
	LA62_03080	nitroreductase	←	LA62_03100	steroid C27-monooxygenase; K15981 cholest-4-en-3-one 26-monooxygenase	→	

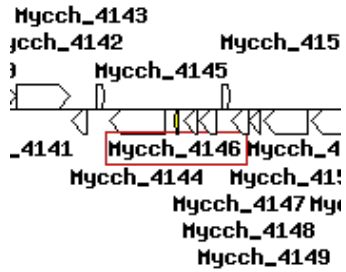
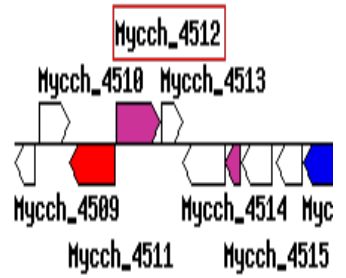
	LA62_03075	short-chain dehydrogenase	→	LA62_03105	acyl-CoA dehydrogenase	→	
Mycobacterium abscessus subsp. bolletii 50594							
cyp125; cytochrome p450	MASS_0583	N/A	N/A	N/A	N/A	N/A	
	MASS_0582	acetyl-CoA acetyltransferase	←	MASS_0584	acyl-CoA dehydrogenase FadE	→	
	MASS_0581	putative cytochrome P450; K15981 cholest-4-en-3-one 26-monooxygenase	→	MASS_0585	putative acyl-CoA dehydrogenase	→	
	MASS_0580	hypothetical protein	→	MASS_0586	hypothetical protein	→	
Mycobacterium abscessus ATCC 19977							
cyp125; cytochrome p450	MAB_0613	N/A	N/A	N/A	N/A	N/A	
	MAB_0612c	Probable acetyl-CoA acetyltransferase	←	MAB_0614	Probable acyl-CoA dehydrogenase FadE	→	
	MAB_0611	Putative cytochrome P450; K15981 cholest-4-en-3-one 26-monooxygenase	→	MAB_0615	Putative acyl-CoA dehydrogenase	→	
	MAB_0610	hypothetical protein	→	MAB_0616	hypothetical protein	→	
Mycobacterium abscessus subsp. bolletii MA 1948							
cyp125; cytochrome p450	LA62_03100	N/A	N/A	N/A	N/A	N/A	
	LA62_03095	acetyl-CoA acetyltransferase; K00626 acetyl-CoA C-acetyltransferase	←	LA62_03105	acyl-CoA dehydrogenase	→	
	LA62_03090	steroid C27-monooxygenase; K15981 cholest-4-en-3-one 26-monooxygenase	→	LA62_03110	acyl-CoA dehydrogenase	→	

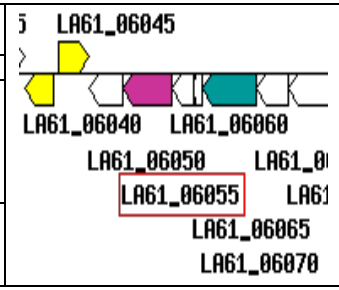
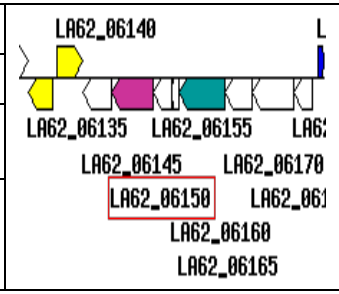
	LA62_03085	hypothetical protein	→	LA62_03115	DNA-binding protein	→	
Mycobacterium abscessus103							
cyp125; cytochrome p450	LA61_03015	N/A	N/A	N/A	N/A	N/A	
	LA61_03010	K00626 acetyl-CoA C-acetyltransferase	←	LA61_03020	acyl-CoA dehydrogenase	→	
	LA61_03005	steroid C27-monooxygenase; K15981 cholest-4-en-3-one 26-monooxygenase	→	LA61_03025	acyl-CoA dehydrogenase	→	
	LA61_03000	hypothetical protein	→	LA61_03030	DNA-binding protein	→	
Mycobacterium gilvum PYR-GCK							
cyp125; cytochrome p450	Mflv_1508	N/A	N/A	N/A	N/A	N/A	
	Mflv_1507	K00626 acetyl-CoA C-acetyltransferase	←	Mflv_1509	acyl-CoA dehydrogenase domain protein	→	
	Mflv_1506	hypothetical protein	←	Mflv_1510	acyl-CoA dehydrogenase domain protein	→	
	Mflv_1505	hypothetical protein	←	Mflv_1511	protein of unknown function DUF35	→	
Mycobacterium gilvum Spyr1							
cyp125; cytochrome p450	Mspyr1_08920	N/A	N/A	N/A	N/A	N/A	
	Mspyr1_08910	K00626 acetyl-CoA C-acetyltransferase	←	Mspyr1_08930	acyl-CoA dehydrogenase; K00257	→	
	Mspyr1_08900	site-specific recombinase XerC	←	Mspyr1_08940	acyl-CoA dehydrogenase; K00257	→	

	Mspyr1_08890	hypothetical protein	←	Mspyr1_08950	predicted nucleic-acid-binding protein containing a Zn-ribbon	→	
Mycobacterium canetii CIPT 140010059							
cyp125; cytochrome p450	MCAN_35561	N/A	N/A	N/A	N/A	N/A	
	MCAN_35551	putative 4-hydroxy-2-oxovalerate aldolase (HOA)	←	MCAN_35571	putative hydratase; K02554 2-keto-4-pentenoate hydratase	→	
	MCAN_35541	PPE62; PPE family protein	←	MCAN_35581	putative dehydrogenase; K05898 3-oxosteroid 1-dehydrogenase	→	
	MCAN_35531	PPE61; PPE family protein	←	MCAN_35591	putative dehydrogenase	←	
Mycobacterium sp. JDM601							
cyp125; cytochrome p450	JDM601_3682	N/A	N/A	N/A	N/A	N/A	
	JDM601_3681	K02554 2-keto-4-pentenoate hydratase	→	JDM601_3683	K05898 3-oxosteroid 1-dehydrogenase	←	
	JDM601_3680	K04073 acetaldehyde dehydrogenase	→	JDM601_3684	dehydratase (MaoC-like)	←	
	JDM601_3679	K01666 4-hydroxy 2-oxovalerate aldolase	→	JDM601_3685	PPE family protein	←	
Mycobacterium Indicus pranii MTCC 9506							
cyp125; cytochrome p450	MIP_02350	N/A	N/A	N/A	N/A	N/A	
CYP124	MIP_02348	K05337 ferredoxin	→	MIP_02353	Chlorophyllide a oxygenase	→	
	MIP_02347	Formyl-coenzyme A transferase	←	MIP_02355	K02182 crotonobetaine/carnitine-CoA ligase	→	
	MIP_02346	Transcriptional regulator, GntR family protein	←	MIP_02356	Hypothetical protein	→	
Mycobacterium sp. JDM601							
cyp125; cytochrome p450	JDM601_2803	N/A	N/A	N/A	N/A	N/A	
	JDM601_2802	araD; ribulose-5-phosphate 4-epimerase AraD	→	JDM601_2804	putative beta-lactamase	←	
	JDM601_2801	ABC transporter ATP-binding protein	→	JDM601_2805	FadR family transcriptional regulator	→	
	JDM601_2800	araD; ribulose-5-phosphate 4-epimerase AraD	→	JDM601_2806	araD; ribulose-5-phosphate 4-epimerase AraD	→	

cyp125; cytochrome p450	JDM601_3609	N/A	N/A	N/A	N/A	N/A	
	JDM601_3608	K00038 3alpha(or 20beta)-hydroxysteroid dehydrogenase	→	JDM601_3610	zinc-containing alcohol dehydrogenase NAD-dependent AdhB; K00121 S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase	→	
	JDM601_3607	conserved hypothetical protein	→	JDM601_3611	conserved hypothetical protein	←	
	JDM601_3606	conserved hypothetical protein	→	JDM601_3612	conserved hypothetical protein	→	
Mycobacterium Avium subsp. paratuberculosis K10							
cyp125; cytochrome p450	MAP_3818	N/A	N/A	N/A	N/A	N/A	
	MAP_3817c	K08977 putative membrane protein	←	MAP_3819	hypothetical protein	→	
	MAP_3816	hypothetical protein	→	MAP_3820	dcd; Dcd; K01494 dCTP deaminase	→	
	MAP_3815	hypothetical protein	→	MAP_3821	hypothetical protein	→	
Mycobacterium avium subsp. paratuberculosis MAP4							
cyp125; cytochrome p450	MAP4_3931	N/A	N/A	N/A	N/A	N/A	
	MAP4_3930	hypothetical protein; K08977 putative membrane protein	←	MAP4_t42	trnG; tRNA-Gly; K14225 tRNA Gly	←	
	MAP4_3929	putative phage integrase family protein	→	MAP4_3932	putative exported protein	→	
	MAP4_3928	hypothetical protein:	→	MAP4_3933	deoxycytidine triphosphate deaminase Dcd; K01494 dCTP deaminase	→	
Mycobacterium kansassii ATCC 12478							
cyp125; cytochrome p450	MKAN_03940	N/A	N/A	N/A	N/A	N/A	
	MKAN_03935	hypothetical protein	→	MKAN_03945	TetR family transcriptional regulator	←	
	MKAN_03930	hypothetical protein	←	MKAN_03950	hypothetical protein	→	
	MKAN_03925	osmotically inducible protein C	←	MKAN_03955	hypothetical protein	→	

Mycobacterium chubuense NBB4							
cyp125;cytochrome p450	Mycch_2866	N/A	N/A	N/A	N/A	N/A	
	Mycch_2865	hypothetical protein	→	Mycch_2867	hypothetical protein	→	
	Mycch_2864	putative transcriptional regulator	←	Mycch_2868	anaerobic dehydrogenase, typically selenocysteine-containing	←	
	Mycch_2863	hypothetical protein	←	Mycch_2869	transcriptional regulator	←	
Mycobacterium vanbaalenii PYR-1							
cyp125;cytochrome p450	Mvan_0246	N/A	N/A	N/A	N/A	N/A	
	Mvan_0245	cytochrome P450	→	Mvan_0247	conserved hypothetical protein	→	
	Mvan_0244	response regulator receiver protein	←	Mvan_0248	conserved hypothetical protein	→	
	Mvan_0243	regulatory protein, LuxR	→	Mvan_0249	conserved hypothetical protein	←	
Mycobacterium smegmatis JS623							
cyp125; cytochrome p450	Mycsm_05668	N/A	N/A	N/A	N/A	N/A	
	Mycsm_05667	N-dimethylarginine,dimethylaminohydrolase; K01482	←	Mycsm_05669	transcriptional regulator	→	
	Mycsm_05666	K01945 phosphoribosylamine--glycine ligase	←	Mycsm_05670	K00215 4-hydroxy-tetrahydrodipicolinate reductase	←	
	Mycsm_05665	putative esterase	→	Mycsm_05671	ring-hydroxylating dioxygenase, large terminal subunit	←	
Mycobacterium rhodesiae NBB3							
cyp125; cytochrome p450	MycrhN_0940	N/A	N/A	N/A	N/A	N/A	
	MycrhN_0939	pseudogene	→	MycrhN_0941	acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II; K00666	←	
	MycrhN_0938	anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	→	MycrhN_0942	putative nucleic-acid-binding protein containing a Zn-ribbon; K07068	←	

	MycrhN_0937	2-polyprenyl-6-methoxyphenol hydroxylase-like oxidoreductase	→	MycrhN_0943	acetyl-CoA acetyltransferase	←	
Mycobacterium chubuense NBB4							
cyp125; cytochrome p450	Mycch_4146	N/A	N/A	N/A	N/A	N/A	
	Mycch_4145	pseudogene	→	Mycch_4147	tRNA-Leu; K14228 tRNA Leu	←	
	Mycch_4144	hypothetical protein	←	Mycch_4148	pseudogene	←	
	Mycch_4145	-polyprenyl-6-methoxyphenol hydroxylase-like oxidoreductase	→	Mycch_4149	hypothetical protein	←	
Mycobacterium chubuense NBB4							
cyp125;cytochrome p450	Mycch_4512	N/A	N/A	N/A	N/A	N/A	
	Mycch_4511	K01945 phosphoribosylamine--glycine ligase	←	Mycch_4513	transcriptional regulator	→	
	Mycch_4510	hypothetical protein	→	Mycch_4514	hypothetical protein	←	
	Mycch_4509	hypothetical protein	←	Mycch_4515	gamma-carboxymuconolactone decarboxylase subunit like protein; K01607	←	
Mycobacterium abscessus ATCC 19977							
cyp125;cytochrome p450	MAB_1211c	N/A	N/A	N/A	N/A	N/A	
	MAB_1210	Putative short chain dehydrogenase/reductase	→	MAB_1212c	hypothetical protein	→	
	MAB_1209	Probable short-chain Z-isoprenyl diphosphate synthetase	←	MAB_1213c	Putative ferredoxin	→	

	MAB_1208	Conserved hypothetical protein	→	MAB_1214c	Probable cytochrome P450; K05917 sterol 14-demethylase	→	
Mycobacterium abscessus103							
cyp125;cytochrome p450	LA61_06055	N/A	N/A	N/A	N/A	N/A	
	LA61_06050	short-chain dehydrogenase	←	LA61_06060	hypothetical protein	←	
	LA61_06045	farnesyl-diphosphate synthase; K12503 short-chain Z-isoprenyl diphosphate synthase	→	LA61_06065	ferredoxin	←	
	LA61_06040	membrane protein; K11068 hemolysin III	←	LA61_06070	cytochrome P450; K05917 sterol 14-demethylase	←	
Mycobacterium abscessus subsp. bolletii MA 1948							
cyp125; cytochrome p450	LA62_06150	N/A	N/A	N/A	N/A	N/A	
	LA62_06145	short-chain dehydrogenase	←	LA62_06155	hypothetical protein	←	
	LA62_061140	farnesyl-diphosphate synthase; K12503 short-chain Z-isoprenyl diphosphate synthase	→	LA62_06160	ferredoxin	←	
	LA62_061135	membrane protein; K11068 hemolysin III	←	LA62_06165	cytochrome P450; K05917 sterol 14-demethylase	←	

Chapter 5

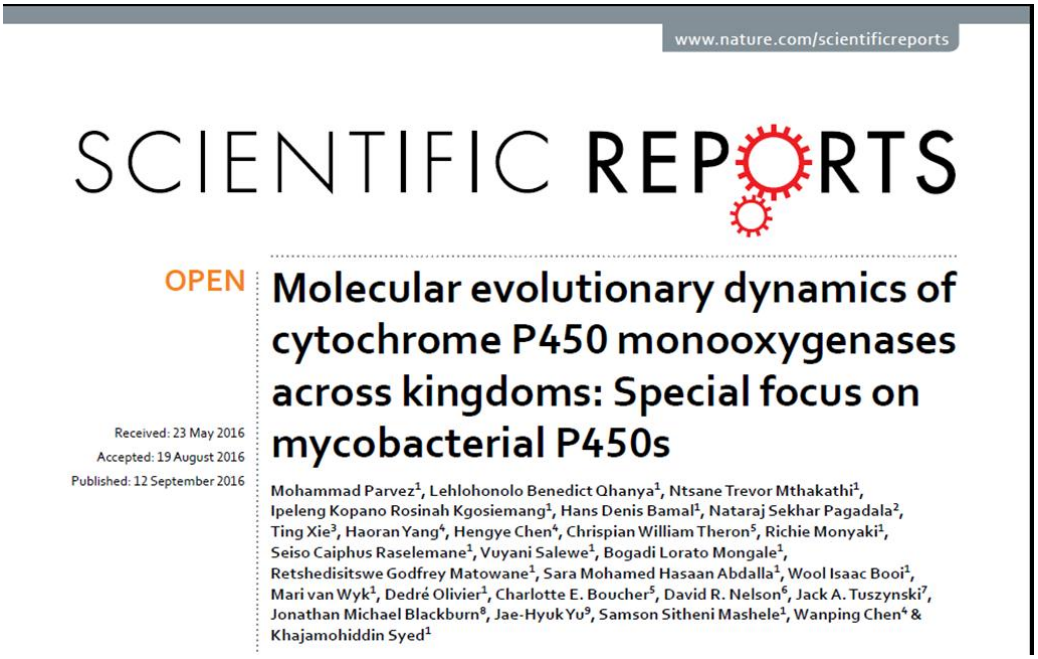
Conclusions and future perspectives

Cytochrome P450 monooxygenases (P450s/CYPs) are heme-thiolate proteins that are distributed in species belong to different biological kingdoms. Since their identification, study on prokaryote P450s especially genome wide annotation and evolutionary analysis has not been carried out. This study address these two research gaps *per se* performing genome wide identification, annotation and phylogenetic analysis of CYP125A1 P450, essential P450 needed for *Mycobacterium tuberculosis* survival. Furthermore, also, genome mapping and gene-cluster analysis of CYP125A1 is completed as part of this study.

Study results will pave the way to select one of CYP125 gene-clusters promoter and its binding elements so that in future one can use them as novel drug target. If we can inhibit CYP125 expression means we can successfully kill the bacteria and the drugs developed based on promoter and its binding elements may have no cross reaction with human P450s considering eukaryote promoter and binding elements different compared to prokaryotes.

Research outputs

Research articles



www.nature.com/scientificreports

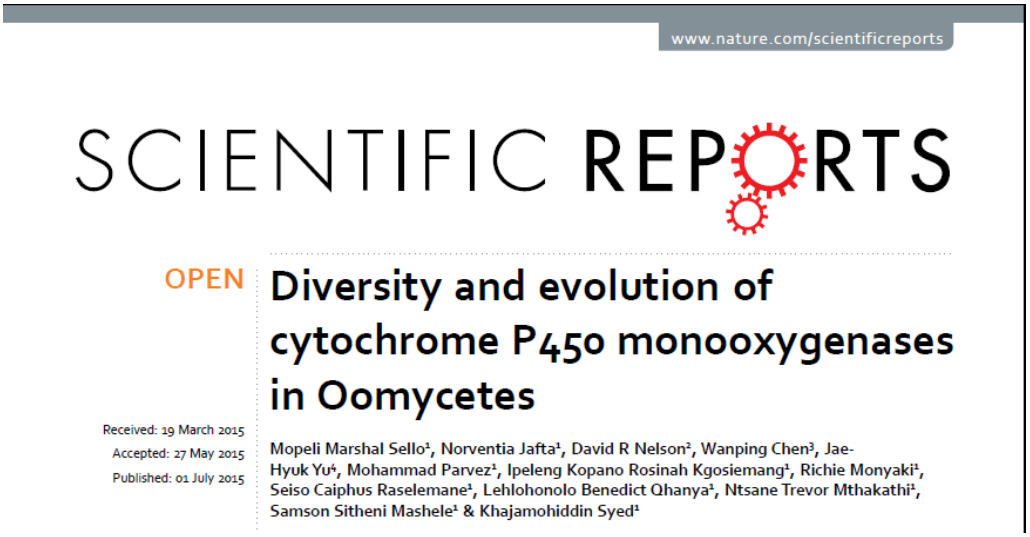
SCIENTIFIC REPORTS

OPEN

Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s

Received: 23 May 2016
Accepted: 19 August 2016
Published: 12 September 2016

Mohammad Parvez[1], Lehlohonolo Benedict Qhanya[1], Ntsane Trevor Mthakathi[1], Ipeleng Kopano Rosinah Kgosiemang[1], Hans Denis Bamal[1], Nataraj Sekhar Pagadala[2], Ting Xie[3], Haoran Yang[4], Hengye Chen[4], Chrispian William Theron[5], Richie Monyaki[1], Seiso Caiphus Raselemane[1], Vuyani Salewe[1], Bogadi Lorato Mongale[1], Retshedisitswe Godfrey Matowane[1], Sara Mohamed Hasaan Abdalla[1], Wool Isaac Booi[1], Mari van Wyk[1], Dedré Olivier[1], Charlotte E. Boucher[5], David R. Nelson[6], Jack A. Tuszynski[7], Jonathan Michael Blackburn[8], Jae-Hyuk Yu[9], Samson Sitheni Mashele[1], Wanping Chen[4] & Khajamohiddin Syed[1]



www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes

Received: 19 March 2015
Accepted: 27 May 2015
Published: 01 July 2015

Mopeli Marshal Sello[1], Norventia Jafta[1], David R Nelson[2], Wanping Chen[3], Jae-Hyuk Yu[4], Mohammad Parvez[1], Ipeleng Kopano Rosinah Kgosiemang[1], Richie Monyaki[1], Seiso Caiphus Raselemane[1], Lehlohonolo Benedict Qhanya[1], Ntsane Trevor Mthakathi[1], Samson Sitheni Mashele[1] & Khajamohiddin Syed[1]

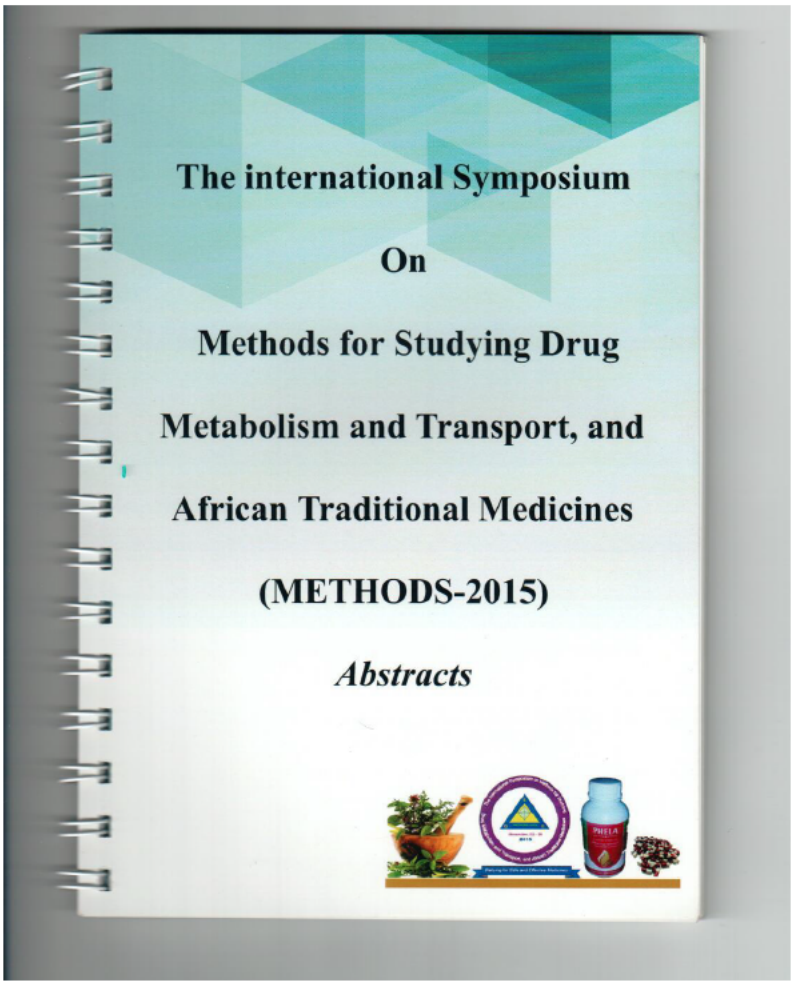
Conference Attendance

P 79-S2

Genome mapping of essential P450 CYP125 in mycobacteria

Monyaki, Richie; Mashele, Samson Sitheni; Syed, Khajamohiddin
Unit for Drug Discovery Research, Department of Health Sciences, Faculty of Health and Environmental Sciences, Central University of Technology, Bloemfontein 9301, South Africa.
rmonyaki@yahoo.com

Tuberculosis (TB) is a leading cause of death worldwide caused by *Mycobacterium tuberculosis*. Genome-wide screening for genes essential for the survival of *M. tuberculosis* has revealed that CYP125A1 is critical for *M. tuberculosis* survival. CYP125A1 play key role in oxidation of cholesterol and help *M. tuberculosis* to utilize cholesterol as carbon source during its inhabitant in host organism. Structural analysis of CYP125A has been elucidated. In this study, genome-mapping of CYP125A1 has been carried out to assess its conservation across sixty mycobacterial species belong to six different categories includes, *M. tuberculosis* complex (MTBC) (27 species); *M. chelonae-abscessus* complex (MCAC) (6 species); *M. avium* complex (MAC) (8 species); Mycobacteria causing Leprosy (MCL) (2 species); Nontuberculous mycobacteria (NTM) (6 species) and Saprophytes (SAP) (11 species). Analysis of genome-mapping revealed conservation of CYP125A1 as per the categories suggesting after speciation CYP125A1 has been conserved both at gene- and protein-level. Furthermore, study also revealed presence of three different ortholog CYP125 P450s that originated before speciation.



DMP 17

Structural analysis of essential P450 CYP125 in ecologically diverse mycobacterial species

R. Monyaki, S.S. Mashele and K. Syed

Unit for Drug Discovery Research, Department of Health Sciences, Faculty of Health and Environmental Sciences, Central University of Technology, Bloemfontein 9300, Free State, South Africa

Tuberculosis (TB) is a leading cause of death worldwide. Genome-wide screening for genes essential for the survival of *Mycobacterium tuberculosis* (Mtb), has revealed that CYP125A1 is critical for *Mtb* survival. In the previous study we performed comparative structural analysis of CYP125 from mycobacterial strains living in diverse environments to assess if this P450 can serve as a common drug target against all mycobacterial strains. 3D models of three CYP125A P450s from *M. avium*, *M. marinum* and *M. vanbaalenii* were generated using CYP125A of *Mtb* H37Rv as a template. 3D models of CYP125A showed all the characteristic P450 helices and sheets and overall they had a similar structure. 3D models were validated using different software programs and validation programs favoured the deduced 3D models, suggesting the models are in good quality. Numerous amino acids were found to be in contact with heme and these residues pretty much conserved across the CYP125A P450s analysed in this study. A large number of amino acid residues were found to be part of the active site in CYP125A P450s. Secondary structure alignment of CYP125A P450s and mapping of the residues suggested that the residues lining the active site are highly conserved. It was determined that CYP125A1 is conserved across different mycobacterial species and can serve as a universal drug target.

Media coverage

pressreader

CUT's groundbreaking discovery

New discovery may help fight aquatic animal infection

The New Age (Free State) · 17 Jun 2015 · News

leading: Research team leader Dr Khajamohiddin Syed and faculty of environmental science acting dean professor Sam Mashele.

Acting dean of the faculty Prof Sam Mashele said some of the deadliest pathogens in the world were found in the fish farming industry.

He said the discovery was made by researchers from the unit for drug discovery in the institution.

"For many years researchers across the world have been trying to understand these micro- organisms in order to control the disease and develop novel drugs against these pathogens, and CUT researchers are leading the way in finding solutions that will bring an end to this socio-economic challenge facing aquatic farming," Mashele said.

pioneers: The CUT science students who helped discover the drug that may help fight aquatic animal infection.

He said microorganisms were widely known in the scientific world and they continued to wreak havoc on the aqua farming sector world-wide.

"They are considered the deadliest of pathogens, causing diminished production of aquatic food," Mashele said.

The team of researchers who discovered the drug were led by Dr Khajamohiddin Syed.

"We collaborated with highly acclaimed international scientists such as professors David Nelson from the University of Tennessee in the US, ae-Hyuk Yu from the University of Wisconsin-Madison and Dr Wanping Chen from Huazhong Agricultural University in China," Mashele said.

"The university researchers are investigating solutions that would sustain the aquatic resources while helping to increase high production levels of aqua farming for commercial purposes, food security and poverty alleviation."

He said their work highlighted the important role aqua farming plays in promoting healthy living and in fighting poverty and hunger.

Syed said aqua farming was a big industry across the world and it involved the farming of fish, shrimps, prawns, squid and octopus and is considered by the UN as an important sector that can provide a livelihood for more than 60 million people in Africa and Asia.

"Consumption of these animals remains a vital source of protein and essential nutrients, especially for developing countries, where they constitute almost half of the total value of their traded communities," Syed said.

"The results of this study have been accepted for publication in the Nature Publication Group journal Scientific Reports, a prestigious multidisciplinary scientific international journal with an impact factor of 5.1."

dag 16 Junie 2015 **Volksblad**

Navorsers by SUT maak deurbraak

Kaydene Davids

'n SPAN navorsers aan die Sentrale Universiteit vir Tegnologie (SUT) het 'n wegspringpunt vir 'n middel ontdek wat groot potensiaal toon om siektes, wat deur parasiete veroorsaak word, in seediere te bestry.

Navorsers van oor die wêreld heen het jare lank probeer om die swamagtige mikro-organismes te verstaan ten einde die siekte, wat onder meer in visse, garnale, inkvisse en seekatte voorkom, te beheer en 'n middel daarteen te ontwikkel.

Maar meer as 50 jaar lank was dit tevergeefs – totdat die navorsers aan die SUT en wêreldbekende wetenskaplikes koppe bymekaargesit het.

Dan Maritz, woordvoerder, sê hul span jong navorsers onder leiding van prof. Samson Sitheni Mashele en dr. Khajemohiddin Syed loop nou voor om oplossings te vind wat die maatskaplik-ekonomiese uitdagings wat die siektes vir veral die visserybedryf bied, te beëindig.

Syed sê miljoene seediere vrek jaarliks weens siektes wat deur parasiete veroorsaak word en dit het 'n regstreekse invloed op die visserybedryf.

Volgens hom is dit alombekend in die wetenskaplike wêreld dat die parasiete (*oomycete*) chaos in die visserybedryf wêreldwyd saai en die lewensduur van die diere aansienlik verkort.

Die navorsers het die *P450 monooxygenase*-proteïen in 13 patogeniese genome bestudeer en bevind dit kan as wegspringpunt gebruik word vir 'n middel om die swamagtige mikro-organismes te beveg.

Die navorsers ondersoek dit nou om oplossings te vind om seehulpbronne te beskerm.

Maritz sê hul (die navorsers se) werk beklemtoon die belangrikheid van die visserybedryf in die bevordering van gesonde lewe en die stryd teen armoede en honger.

Die uitkoms van die navorsing is reeds vir plasing in die *Nature Publication Group Journal* aanvaar.

■ Die wêreldbekende wetenskaplikes wat ook deel van die navorsingspan is, sluit in prof. David R. Nelson van die Universiteit van Tennessee, prof. Jae Hyuk Yu van die Universiteit van Wisconsin-Madison, en dr. Wanping Chen Huazhong van die Landbou-universiteit in China.

Die span navorsers aan die Sentrale Universiteit vir Tegnologie (SUT) wat die wegspringpunt vir 'n middel ontdek het om siektes wat deur parasiete veroorsaak word, in seediere te bestry. Hier is van links, voor: dr. Khajamohiddin Syed; middel: Lorato Mongale, dr. Ntsoaki Malebo, Poojah Jawallapersanel en Mohammad Parvez; agter: prof. Sitheni Mashele, Mopeli Sello, Seiso Raselemane, Lehlohonolo Qwanya, Mthakathi Ntsane, Thapelo Makhele, Euginiah Rampana en Richie Monyaki.
Foto: MLUNGISI LOUW