In silico analysis of cholesterol catabolic genes/proteins in the genus *Mycobacterium*

By

Rochelle van Wyk

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

21 January 2018

Promoter: Prof Khajamohiddin Syed

Co-Promoters: 1. Prof Samson Sitheni Mashele

2. Dr Dedré Olivier

3. Ms Mari van Wyk

DECLARATION

I, **ROCHELLE VAN WYK** (SA ID number:), 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).

ROCHELLE VAN WYK **DATE**

DECLARATION ON EDITING

M.B. BRADLEY

P.O. Box 37326
Faerie Glen
Pretoria 0043

072 369 5149

DECLARATION ON EDITING

Customer:	**Date**
Prof K Syed Department of Health Sciences Faculty of Health and Environmental Sciences Central University of Technology	2016/11/06
Contact: Prof Syed	

Thesis submitted for editing

Titles of the thesis as indicated:

In silico analysis of cholesterol catabolic genes/proteins in the genus *Mycobacterium*

The above entitled thesis was submitted to me for language editing, which was done on 23 October 2017.

M.B. BRADLEY (MA) - Language editor

Cell no 0723695149

ACKNOWLEDGEMENTS

This thesis became a reality with the kind support and help of many individuals. I would like to gratefully acknowledge all of them:

- My promotors, Prof Khajamohiddin Syed (University of Zululand), Prof Samson Sitheni Mashele, Dr Dedré Olivier and Ms Mari van Wyk – my deepest gratitude for their full support, expert guidance, understanding and encouragement throughout my research. Without the incredible patience and timely wisdom and counsel, my thesis work would not have been.
- Research and Innovation Fund, Central University of Technology, Free State – thank you for funding my studies.
- The National Research Foundation – I am eternally grateful for financial support during my full-time research.
- Distinguished members of the review panel for your time and consideration.
- My family – especially my beloved husband, beautiful children, and wonderful parents for the encouragement.
- Fellow research students – for directly or indirectly having lent their helping hand in this venture.
- Above all, I want to offer this endeavour to our God Almighty, Creator, Saviour and Protector, for the wisdom He bestowed upon me, the strength, peace of mind and good health to finish this research. “I can do all things through Christ who strengthens me.”

DEDICATION

To all those who wonder if they should or could…

"Whether you think you can, or you think you can't – you're right."

\- Henry Ford

TABLE OF CONTENTS

CHAPTER 2: ANNOTAION AND CLASSIFICATION OF GENES/PROTEINS INVOLVED IN CHOLESTEROL DEGRADATION IN *MYCOBACTERIUM TUBERCULOSIS* H37RV

CHAPTER 3: *IN SILICO* DETERMINATION OF THE CHOLESTEROL DEGRADATION ABILITY OF MYCOBACTERIAL SPECIES

LIST OF ABBREVIATIONS AND ACRONYMS

3-HSA	3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione
3,4-DHSA	3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione
3β-HSD	3β-hydroxysteroid dehydrogenase
4,9-DSHA	4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-dien-4-oic acid
9OHADD	9-hydroxy-1,4-androstene-3,17-dione
17β-HSD	17β-hydroxysteroid dehydrogenase
α	Alpha
β	Beta
Δ^1KstD	3-ketosteroid-Δ1-dehydrogenase
λ	Lambda
ω	Omega
%	Percent
ABC	ATP-binding cassette
AD	Androst-4-ene-3,17-dione
ADD	Androsta-1,4-diene-3,17-dione
ADH	Alcohol dehydrogenase
AMP	Adenosine monophosphate
AP	Apurinic/apyrimidinic
ATP	Adenosine triphosphate
A-D	A to D
Bcp	Bacterioferritin comigratory protein
BLAST	Basic Local Alignment Search Tool

BCG	Bacillus Calmette-Guérin
C	Carbon
CC	Carbon-carbon
CDC1551	*Mycobacterium tuberculosis* CDC1551
CHP	Conserved hypothetical protein
ChoD	Cholesterol oxidase
ChOx	Cholesterol oxidase
CoA	Co-enzyme A
CYP	Cytochrome P450
DNA	Deoxyribonucleic acid
DOHNAA	9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid
e.g.	Example
et al.	*Et alia* (and others)
H37Rv	*Mycobacterium tuberculosis* H37Rv
HHD	2-hydroxy-hexa-2,4-dienoic acid
HIV	Human immunodeficiency virus
HP	Hypothetical protein
HsaC	3,4-DHSA dioxygenase
HsaD	4,9-DSHA hydrolase
HsaEFG	2-hydroxypentadienoate hydratase (E), 4-hydroxy-2-ketovalerate aldolase (F), aldehyde dehydrogenase (G)
HSD	Hydroxysteroid dehydrogenase
i.e.	*Id est* (that is)
IFN	Interferon
igr	Intracellular growth

IpdAB	Methylhexahydroindanedione propionate (HIP) CoA transferase A and B
KEGG	Kyoto encyclopedia of genes and genomes
Ksh	3-ketosteroid 9α-hydroxylase
KshA	Kerosteroid-9α-hydroxylase, oxygenase
KshB	Ketosteroid-9α-hydroxylase, reductase
KstD	Ketosteroid dehydrogenase
KstR	TetR-type transcriptional repressor
KstR2	TetR-type transcriptional repressor
LldD	L-lactate dehydrogenase
MAC	*Mycobacterium avium* complex
MCAC	*Mycobacterium chelonae-abscessus* complex
MCE	Mammalian cell entry
MCL	Mycobacteria causing leprosy
MDR	Multiple drug-resistant
MgtE	Mg^{2+} transport transmembrane protein
MmpL	*Mycobacterium* membrane protein laboratory
MTBC	*Mycobacterium tuberculosis* complex
NAD^+	Nicotinamide adenine dinucleotide
NCBI CDD	National center for biotechnology information conserved domain database
NTM	Nontuberculous mycobacteria
P450	Cytochrome P450
PDIM	Phthiocerol dimycocerosate
PE	Protein family with highly conserved Proline-

	Glutamate residues near the start of their encoded proteins
PGRS	Polymorphic GC-rich-repetitive sequence
pks	Polyketide synthase
PPE	Protein family with highly conserved Proline-Proline-Glutamate residues near the start of their encoded proteins
PQQ	Pyrrolo-quinoline quinone
PTP/PTPase	Phosphotyrosine protein phosphatase /protein-tyrosine-phosphatase
RHA1	*Rhodococcus jostii* RHA1
RNA	Ribonucleic acid
SAP	Saprophytes
TB	Tuberculosis
TDR	Total drug-resistant
TetR	Tetracycline repressor
WHO	World Health Organization
XDR	Extensively drug-resistant

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

It is well known that *Mycobacterium tuberculosis*, the causative agent of one of the deadliest human diseases, tuberculosis, uses human cholesterol as a carbon source both in the latent and active phases of its lifestyle. The discovery of the ability of *M. tuberculosis* to degrade and use cholesterol as a sole source of carbon and energy has opened up the possibility of using genes/proteins involved in cholesterol degradation as novel drug targets. If one can find the highly conserved genes/proteins across the mycobacteria that are capable of degrading cholesterol, then in future these genes can possibly be used as universal drug targets against mycobacterial infections. However, to date, data on how many mycobacterial species utilise cholesterol has not been reported. Furthermore, performing laboratory experiments is laborious and time- and money-consuming, considering each of the mycobacterial species has different lifestyle and culture conditions. The study is aimed at using the available genomic data to perform comparative genomic studies to unravel the nature of cholesterol catabolic genes/proteins in the genus *Mycobacterium* to determine which mycobacterial species are capable of degrading cholesterol. This study is a first of its kind comprehensive analysis of the genes/proteins involved in cholesterol degradation across 93 mycobacterial species, using bioinformatic tools.

Ninety-three mycobacterial species whose genomes are available for public use at the KEGG database were used in this study. Literature on cholesterol degradation by bacteria was collected and the cholesterol degradation pathway was deduced. The intermediate metabolites and enzymes involved in each of the steps were identified and mapped using ChemDraw software. A software program that extracts homolog data across 93 mycobacterial species was developed. The hit proteins' domains/functions were identified using software programs: NCBI Batch Web CD-search tool and the KEGG functional database. Based on the sequence identity, functional motifs and functional data, if available, the hit proteins were sorted into specific enzymatic reactions of cholesterol degradation.

After thorough literature analysis, 152 genes/proteins were identified as cholesterol catabolic genes/proteins and grouped into four different categories. The four categories are: (i) genes predicted to be specifically required for growth on cholesterol, (ii) cholesterol catabolic genes proven to be or predicted to be essential for the survival of *M. tuberculosis* in macrophage cells and in murine infection, (iii) genes/proteins that are up-regulated during growth on cholesterol, and (iv) genes involved in cholesterol degradation by *M. tuberculosis* H37Rv, but not confirmed or predicted to be essential. *In silico* analysis of 152 genes across 93 mycobacterial species revealed that 51 mycobacterial species are unable to degrade cholesterol. The specifics on mycobacterial species' ability to utilise cholesterol, as per their different categories, are listed in the table below:

In silico analysis of cholesterol-utilising capability of mycobacterial species

Category	No of species	Ability to utilise cholesterol as carbon source	
		Negative	Positive
Mycobacterium tuberculosis complex	39	10	29
Mycobacterium chelonae-abscessus complex	10	10	
Mycobacterium avium complex	15	5	10
Mycobacteria causing leprosy	2	2	
Nontuberculous mycobacteria	8	5	3
Saprophytes	19	19	

Results from this study are based on the *in silico* analysis and need to be experimentally validated.

Keywords: Mycobacteria, Species, Cholesterol, Degradation, Breakdown, Genes, Proteins, Drug target, *Mycobacterium tuberculosis*

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1. Introduction

Mycobacterium tuberculosis infection, tuberculosis (TB), is one of the leading causes of death worldwide, killing an estimated two million people annually (WHO, 2016; Sotgiu *et al*., 2017). It is estimated that one third of the world's population (approximately two billion people) is infected with this highly pathogenic organism (WHO, 2015). Once it has entered the human body and after ingestion by macrophages, this intracellular pathogen can survive in a modified phagosome and cause latent infection for years and sometimes decades without any symptoms (Glickman and Jacobs, 2001). Tubercle bacilli are able to persist in this dormant state, from which they may reactivate and cause TB (Glickman and Jacobs, 2001). The reactivation of latent phase *M. tuberculosis* into the active phase is observed among people whose immune system is weakened by human immunodeficiency virus (HIV) infection, use of immunosuppressive drugs, malnutrition and/or aging (Flynn and Chan, 2001). Over the past decades, the threat of TB has become greater with the development from single-drug-resistant to multiple-drug-resistant (MDR) strains and recently, the surfacing of extensive drug resistance (XDR) that threatens to compromise available drugs severely (Migliori *et al*., 2012). With the documentation of total drug-resistant (TDR) strains (Migliori *et al*., 2012), along with the insufficiency of new drug targets, it is clear that more research is needed to discover novel drug targets.

M. tuberculosis can infect, grow and survive in the harsh environment of the macrophage and other host cells using mechanisms that are not yet well understood (Clark-Curtiss and Haydel, 2003; Koul *et al*., 2004). Host cholesterol levels are thought to play a role in the development of *M. tuberculosis* infection (Kim *et al*., 2010), with high levels of cholesterol in the diet significantly enhancing the bacterial burden in the lung (Schäfer *et al*., 2009) and impairing immunity to *M. tuberculosis* (Martens *et al*., 2008). Specifically, cholesterol is required for the phagocytosis of mycobacteria into macrophages (Gatfield and Pieters, 2000; Peyron *et al*., 2000). In fact, mycobacteria bind and enter phagocytes through cholesterol-enriched membrane microdomains (lipid rafts) (Muñoz *et al*., 2009). In addition, cholesterol plays a crucial role in the mediation of the infected phagosomal association of tryptophan-aspartate-containing coat protein (Pieters and Gatfield, 2002), leading to the inhibition of phagosome–lysosome fusion (Nguyen and Pieters, 2005). This experimental evidence suggests an important role for cholesterol during *M. tuberculosis* infection and persistence.

Research studies demonstrated that *M. tuberculosis* can grow using cholesterol as the sole carbon and energy source (Pandey and Sassetti, 2008). Therefore, cholesterol has recently been identified as an important lipid for mycobacterial infection (De Chastellier and Thilo, 2006; Ouellet *et al*., 2011). The relatively abundant cholesterol distributed in host cells is an important growth substrate for these bacteria in different infection stages (e.g. intracellular growth or intracellular persistence) (García *et al*., 2012). *M. tuberculosis* growing in human cells appears to obtain energy from host lipids rather than other nutrients such as carbohydrates (Dubnau *et al*., 2005).

In light of the above facts and recent momentum on cholesterol catabolism as a therapeutic target in *M. tuberculosis*, Ouellet *et al.* (2011) clearly state that more research needs to be done in order to understand cholesterol degradation pathways in mycobacteria. Especially, comparative genomic studies should be performed to unravel and map the cholesterol degradation pathway in different mycobacteria. If one can find the highly conserved genes across the mycobacteria that are capable of degrading cholesterol, then in future these genes can possibly be used as universal drug targets against mycobacterial infections.

1.2. Literature Review

1.2.1. Cholesterol

Cholesterol is a polycyclic steroid compound that is widely distributed in the biosphere. This 27 carbon-atom compound is an amphiphilic lipid, structurally bearing four alicyclic rings and the aliphatic side chain. The side chain is a flat and rigid carbon skeleton fused to four alicyclic rings (Figure 1.1.) (García *et al.*, 2012).

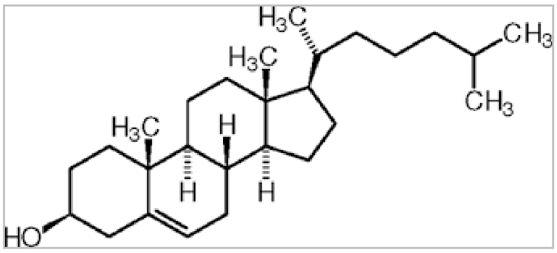

Figure 1.1. The chemical structure of cholesterol.

It is not only abundantly distributed in the biosphere, but also in animals as an essential structural component of animal cell membranes (Slaytor and Bloch, 1965) and is required for viability and cell proliferation (Dahl and Dahl, 1988; Yeagle, 1993) as well. Cholesterol plays a great physiologic role in plasma membranes. The most important function of cholesterol in the lipid bilayer is to modulate the physicochemical properties of cellular membranes (Yeagle, 1985; Yeagle, 1988; Finegold, 1992). This steroid lipid is an abundant structural component of cell membranes (Slaytor and Bloch, 1965). About 20% of membrane lipid is cholesterol (Marieb and Hoehn, 2010). The structural composition of the plasma membrane results in a laterally more condensed membrane with increased density of the phospholipids (Lund-Katz *et al*., 1988; Smaby *et al*., 1994; McIntosh, 1999). This increases the mechanical strength and decreases the permeability of the membrane (Yeagle, 1985; Needham and Nunn, 1990). As a result, the relatively high rates of both lateral and rotational diffusion of the lipid bilayer are maintained (Yeagle, 1988; Vist and Davis, 1990; Davis and Finegold, 1993). On membranes, cholesterol wedges its plate-like hydrocarbon rings between the phospholipid tails, rendering stability to the membrane while increasing the mobility of other lipids (Marieb and Hoehn, 2010) and the fluidity of the membrane (Kusumi *et al*., 1983 and 1986). In addition, it also raises hydrophobic barriers for polar molecules and increases rigidity barriers for non-polar molecules (Subczynski and Wisniewska, 2000). Moreover, cholesterol can play a role in cellular signalling by modulating the physical properties of the lipid bilayer, thereby affecting the activity of receptors and enzymes residing on it, or directly as a regulator of enzymes in the biosynthesis of cholesterol (Dahl and Dahl, 1988; Jackson *et al*., 1997; Edwards and Ericsson, 1998).

Depending on the physiological demand and biomolecule necessity, cholesterol can be catabolised for the biosynthesis of many physiologically important steroids. Steroids possess important biological functions in eukaryotic organisms (Merino *et al*., 2013). Some are membrane components (e.g., cholesterol and phytosterols), others function as hormones (e.g., testosterone and estradiol) and some are bile salts and detergents used in the solubilisation and intestinal absorption of fats, cholesterol and lipid-soluble vitamins (Wollam and Antebi, 2011). Figure 1.2. illustrates the chemical structure of cholesterol and some biologically important steroids derived from cholesterol.

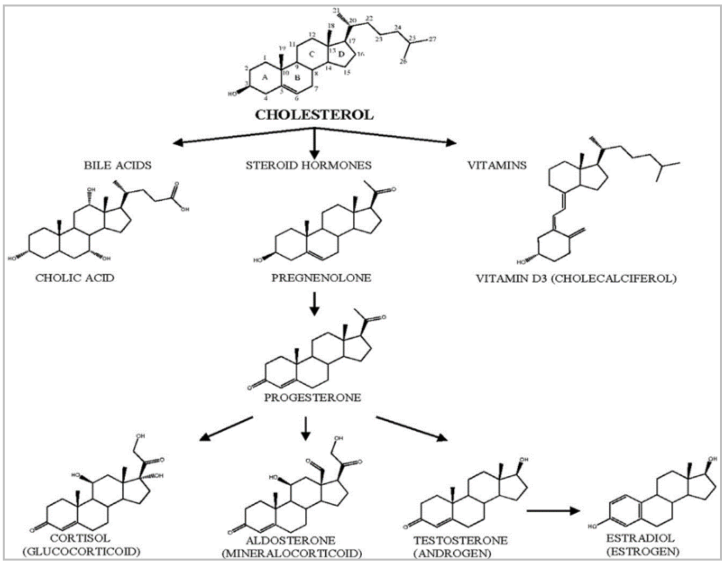

Figure 1.2. Cholesterol and its catabolic products (taken from García *et al.*, 2012).

Cholesterol is naturally abundant in both the hydrosphere and lithosphere. Although ubiquitous in nature, cholesterol is a recalcitrant molecule to biodegradation because of its low number of functional groups (one carbon-carbon double bond and a single hydroxyl group), its low solubility in water (3×10^{-8}M) and its complex spatial conformation constituted by four alicyclic rings and two quaternary carbon atoms (García *et al*., 2012). The high rate of ubiquity and persistence of cholesterol and some derived compounds such as coprastonol, have been used as reference biomarkers for environmental pollution analysis (Veiga *et al*., 2005). Cholesterol and its metabolites have many functions in both prokaryotes and eukaryotes; functions relating to biology, medicine and biotechnology.

1.2.2. *Mycobacterium tuberculosis* and cholesterol

The ubiquity of cholesterol and related sterols in the environment has made them a common carbon source for many different microorganisms, some of them being important pathogens, such as *M. tuberculosis* (García *et al*., 2012). The ability of *M. tuberculosis* to survive in the nutrient-deficient vacuole of the host cell has led to studies identifying its ability to utilise the host's cholesterol as an important source of carbon (Griffin *et al*., 2012).

In order to understand how *M. tuberculosis* uses cholesterol, research has been intensified to understand the cholesterol degradation pathway. However, cholesterol catabolism by bacteria has not yet been fully elucidated in any of the bacterial strains having cholesterol degrading abilities, thus the metabolic pathways are suggested by merging biochemical and genetic studies done on different organisms. Based on the available data, the cholesterol degradation pathway in *M. tuberculosis* can be divided into two major phases: (i) initial degradation of the aliphatic side chain, and (ii) subsequent degradation of the A–D rings (Figure 1.3.) (Ouellet *et al*., 2011). It remains unclear whether there is an obligatory order of the degradation reactions in *M. tuberculosis*. The evidence obtained from the literature on rhodococcal sterol catabolism suggests that intermediates of ring and side chain degradation can be exchanged between the two pathway branches (Rosłoniec *et al*., 2009). However, the situation seems to differ in *M. tuberculosis,* as blockage of the side chain degradation resulted in the accumulation of cholest-4-en-3-one as a major metabolite (Ouellet *et al.*, 2011), suggesting that the ring-degrading enzymes (e.g. KsaAB and HsaA-C) act optimally after the side chain is removed.

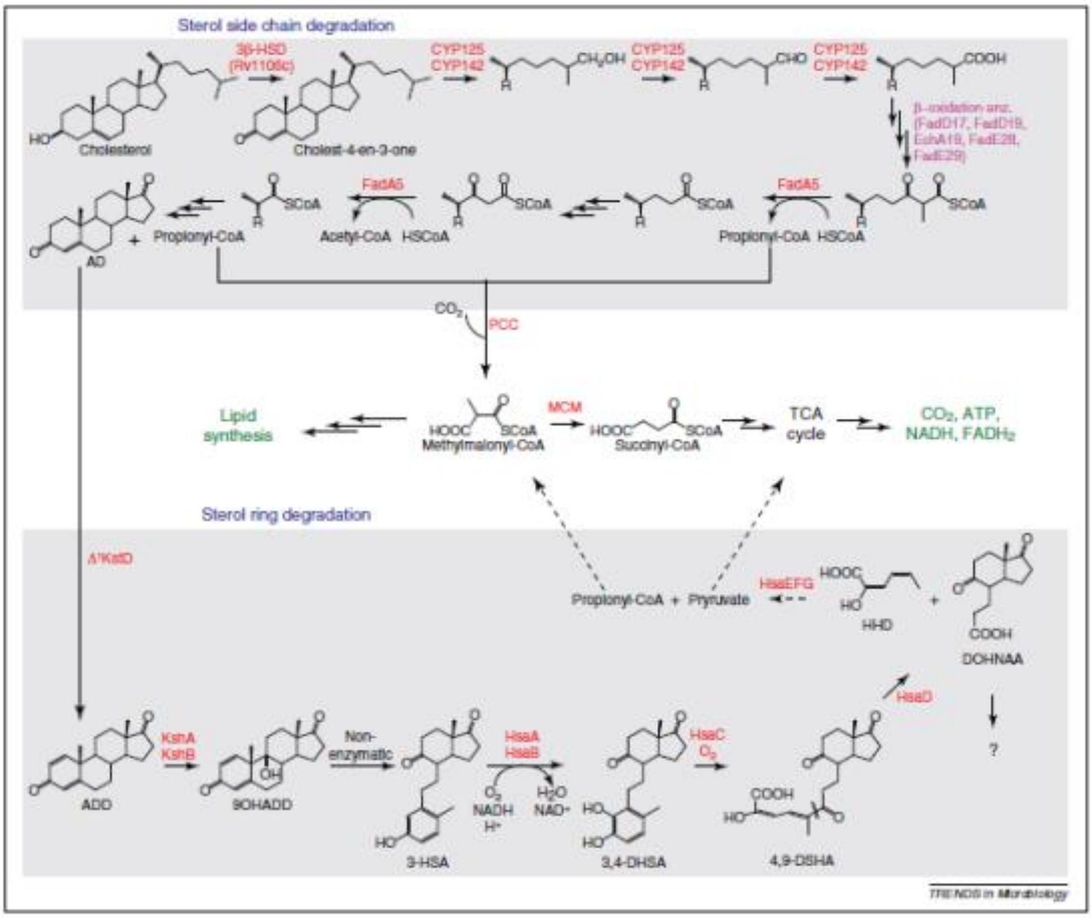

Figure 1.3. Proposed degradation pathway and flux of metabolites derived from catabolism of the cholesterol aliphatic side chain and ring nucleus (taken from Ouellet *et al*., 2011).

1.2.3. Genes/proteins involved in the cholesterol degradation by *M. tuberculosis*

To proliferate within the macrophages, *M. tuberculosis* cells undergo a shift in metabolism from using carbohydrates to primarily using host lipids (Pandey and Sassetti, 2008; Nesbitt *et al*., 2010). With determination of the complete genome of *M. tuberculosis*, revealing at least 250 genes predicted to be involved in lipid metabolism (Cole *et al*., 1998), and recent identification of a large regulon of cholesterol catabolic genes, a lot of time has been devoted to identify and annotate these specific genes in the search for possible novel drug targets against this organism. Van Der Geize *et al*. (2007) identified a complete set of genes required for cholesterol degradation in *Rhodococcus jostii*. Furthermore, they identified these 51 up-regulated proteins in *M. tuberculosis* and *M. bovis* Bacillus Calmette-Guérin (BCG). This suggests that cholesterol catabolic genes are conserved among mycobacterial species. For annotation of the cholesterol catabolic genes they compared the sequence similarity of the gene products of *R. jostii* RHA1 and *M. tuberculosis* H37Rv strains and compiled a list of genes predicted to be involved in cholesterol catabolism (Table 1.1.) (Van Der Geize *et al*., 2007). They further stated that *hsd4A*, *hsd4B*, *fadD19*, *fadE26* and *ltp3* comprise one set of these genes that is highly up-regulated and encodes all of the enzymes necessary to perform one full cycle of β-oxidation (Van Der Geize *et al*., 2007). Up-regulated to a lesser extent, is a second incomplete set of β-oxidation genes, *echA19*, *fadD17*, *fadE27* and *ltp4*, that are probably also involved in side chain degradation (Van Der Geize *et al*., 2007).

Table 1.1. Annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway (taken from Van Der Geize *et al*., 2007).

Gene[a]	RHA1[b]	H37Rv[c]	BCG[d]	Identity[e]	Annotation of gene product	Best hit[f]	Identity[g]
mce4F	Ro04703	Rv3494c	Bcg3558c	37	MCE family protein	NA	NA
mce4E	Ro04702	Rv3495c	Bcg3559c	35	MCE family protein	NA	NA
mce4D	Ro04701	Rv3496c	Bcg3560c	43	MCE family protein	NA	NA
mce4C	Ro04700	Rv3497c	Bcg3561c	41	MCE family protein	NA	NA
mce4B	Ro04699	Rv3498c	Bcg3562c	46	MCE family protein	NA	NA
mce4A	Ro04698	Rv3499c	Bcg3563c	41	MCE family protein	CAA50257	32
supB	Ro04697	Rv3500c	Bcg3564c	66	Sterol uptake permease subunit (ABC transporter)	AAT51760	49
supA	Ro04696	Rv3501c	Bcg3565c	72	Sterol uptake permease subunit (ABC transporter)	AAT51759	55
choD	Ro04305	Rv3409c	Bcg3479c	60	Cholesterol oxidase	P12676	19
hsd4A	Ro04695	Rv3502c	Bcg3566c	59	17β-hydroxysteroid dehydrogenase	BAD66689	36
hsd4B	Ro04531	Rv3538	Bcg3602	62	2-Enoyl acyl-CoA hydratase	CAA55037	30
kshA	Ro04538	Rv3526	Bcg3590	59	Ketosteroid-9α-hydroxylase, oxygenase	AAL96829	57
kshB	Ro05833	Rv3571	Bcg3636	54	Ketosteroid-9α-hydroxylase, reductase	AAL96830	69
kstD	Ro04532	Rv3537	Bcg3601	62	3-Ketosteroid-Δ1-dehydrogenase	AAL82579	40
hsaB	Ro04542	Rv3567c	Bcg3632c	70	3-HSA hydroxylase, reductase	BAC67692	16
hsaC	Ro04541	Rv3568c	Bcg3633c	81	3,4-DHSA dioxygenase	BAB15809	42
hsaD	Ro04540	Rv3569c	Bcg3634c	75	4,9-DSHA hydrolase	BAC67693	31
hsaA	Ro04539	Rv3570c	Bcg3635c	76	3-HSA hydroxylase, oxygenase	BAC67691	37
hsaF	Ro04535	Rv3534c	Bcg3598c	79	4-Hydroxy-2-oxovalerate aldolase	P51017	48
hsaG	Ro04534	Rv3535c	Bcg3599c	85	Acetaldehyde dehydrogenase	BAB97164	61
hsaE	Ro04533	Rv3536c	Bcg3600c	71	2-Hydroxypentadienoate hydratase	BAB97166	59
fadE26	Ro04693	Rv3504	Bcg3568	77	Acyl-CoA dehydrogenase	P71539	25
fadE27	Ro04692	Rv3505	Bcg3569	54	Acyl-CoA dehydrogenase	P16219	24
fadD17	Ro04691	Rv3506	Bcg3570	56	Fatty acid-CoA synthetase	Q4LDG0	20
fadD19	Ro04689	Rv3515c	Bcg3578c	64	Fatty acid-CoA ligase	AAB87139	38
echA19	Ro04688	Rv3516	Bcg3579	73	Fatty acid-CoA hydratase	P31551	33
ltp4	Ro04684	Rv3522	Bcg3586	72	3-Ketoacyl-CoA thiolase	NA	NA
ltp3	Ro04683	Rv3523	Bcg3587	79	SCPx related 3-ketoacyl-CoA thiolase	AAA40098	20
	Ro06698	NA	NA	NA	Cyclohexanone monooxygenase	AAG01290	59
	Ro06693	NA	NA	NA	5-Valerolactone hydrolase	BAC22650	68

Notes:

[a] Name assigned based on current study.

[b] Identification number for the RHA1 gene.

[c] Identification number for the reciprocal best hit in *M. tuberculosis* H37Rv.

[d] Identification number for the reciprocal best hit *M. bovis* BCG.

[e] Percent amino acid sequence identity of the RHA1 and H37Rv and BCG orthologues based on full sequence alignment. Nucleotide sequence identity between H37Rv and BCG genes is >98%.

[f] Accession number of functionally characterised best hit in National Center for Biotechnology Information database.

[g] Percentage amino acid sequence identity of the RHA1 enzyme and its experimentally characterised best hit based on full sequence alignment. NA, not available (either no homologous gene in H37Rv or BCG, or no functionally characterised homolog reported).

Griffin *et al*. (2011) used high-resolution phenotypic profiling with highly parallel Illumina sequencing to characterise transposon libraries in order to identify the genes that are important for the growth of *M. tuberculosis* precisely. This deep sequencing-based mapping approach allowed them to identify 96 genes that are predicted to be important for growth on cholesterol (Table 1.2.).

Table 1.2. Genes predicted to be specifically required for growth on cholesterol (taken from Griffin *et al*., 2011).

Locus	Synonym	Locus	Synonym	Locus	Synonym
Rv0009	ppiA	Rv1919c	-	Rv3526	kshA
Rv0153c	ptbB	Rv1963c	mce3R	Rv3531c	-
Rv0202c	mmpL11	Rv2048c	pks12	Rv3534c	hsaF
Rv0244c	fadE5	Rv2118c	-	Rv3536c	hsaE
Rv0362	mgtE	Rv2206	-	Rv3537	kstD
Rv0391	metZ	Rv2239c	-	Rv3540c	ltp2
Rv0450c	mmpL4	Rv2416c	eis	Rv3542c	-
Rv0485	-	Rv2462c	tig	Rv3543c	fadE29
Rv0495c	-	Rv2506	-	Rv3544c	fadE28
Rv0655	mkl	Rv2668	-	Rv3545c	cyp125

Locus	Synonym	Locus	Synonym	Locus	Synonym
Rv0693	pqqE	Rv2681	-	Rv3546	fadA5
Rv0694	lldD1	Rv2684	arsA	Rv3548c	-
Rv0695	-	Rv2710	sigB	Rv3549c	-
Rv0696	-	Rv2799	-	Rv3551	-
Rv0761c	adhB	Rv2914c	pknI	Rv3553	-
Rv0805	-	Rv2985	mutT1	Rv3559c	-
Rv0876c	-	Rv3050c	-	Rv3560c	fadE30
Rv1071c	echA9	Rv3274c	fadE25	Rv3561	fadD3
Rv1084	-	Rv3419c	gcp	Rv3563	fadE32
Rv1096	-	Rv3421c	-	Rv3564	fadE33
Rv1129c	-	Rv3492c	-	Rv3568c	hsaC
Rv1130	-	Rv3493c	-	Rv3569c	hsaD
Rv1131	gltA1	Rv3494c	mce4F	Rv3570c	hsaA
Rv1183	mmpL10	Rv3495c	lprN	Rv3571	kshB
Rv1193	fadD36	Rv3496c	mce4D	Rv3572	-
Rv1428c	-	Rv3497c	mce4C	Rv3573c	fadE34
Rv1432	-	Rv3498c	mce4B	Rv3575c	-
Rv1608c	bcpB	Rv3499c	mce4A	Rv3779	-
Rv1626	-	Rv3500c	yrbE4B	Rv3820c	papA2
Rv1627c	-	Rv3501c	yrbE4A	Rv3824c	papA1
Rv1798	-	Rv3502c	hsd4A	Rv3825c	pks2
Rv1906c	-	Rv3515c	fadD19	Rv3911	sigM

In *M. tuberculosis*, 41 of the steroid catabolic pathway genes, including those involved in the A and B rings' degradation, are specifically up-regulated during survival in the macrophages (Van Der Geize *et al*., 2007).

Expression of many of these cholesterol catabolic genes of *M. tuberculosis* has been shown to be controlled by two TetR-type transcriptional repressors, *kstR* and *kstR2* (Figure 1.4.) (Kendall *et al*., 2010; Ouellet *et al*., 2011). KstR controls the expression of most of the genes found in two chromosomal segments, R*v3494c* – *Rv3547* and *Rv3566c* – *Rv3574*, but there are many other KstR-regulated genes spread throughout the *M. tuberculosis* genome, including *Rv1106c* that codes for 3β-hydroxysteroid dehydrogenase (3β-HSD) – one of the first enzymes involved in cholesterol degradation by this organism (Figure 1.3.) (Ouellet *et al*., 2011). It was shown that *kstR2* (*Rv3557c*) controls the expression of a subset of 15 genes (*Rv3548* – *Rv3565*) within the *kstR* regulon (Kendall *et al*., 2010). Both KstR and KstR2 negatively autoregulate themselves, but they act independently of each other (Kendall *et al*., 2010).

Also playing a crucial role in the pathogenesis is the *igr* locus (Figure 1.4.), which is required for intracellular growth (therefore named *igr* for *i*ntracellular *gr*owth) of *M. tuberculosis* in macrophages and mice and essential for cholesterol metabolism, since *igr*-deficient bacteria cannot grow using cholesterol as a primary carbon source (Chiang *et al*., 2007; Chang *et al*., 2009; Miner *et al*., 2009). The *igr* operon is made up of six genes – *cyp125* (a cytochrome P450), *fadE28* and *fadE29* (two probable acyl-CoA dehydrogenases, *Rv3544c* and *Rv3543c*), *Rv3541-2c* (two conserved hypothetical proteins, also thought to be putative enoyl-CoA hydratases) and *ltp2* (a putative lipid carrier protein, *Rv3540c*) (Ouellet *et al*., 2011; García-Fernández *et al*., 2013). In research done by Thomas *et al*. (2011) they concluded, based on the structure of the isolated metabolite, enzyme activity and bioinformatic annotations, that the *igr* operon is necessary to completely metabolise the side chain of cholesterol metabolites.

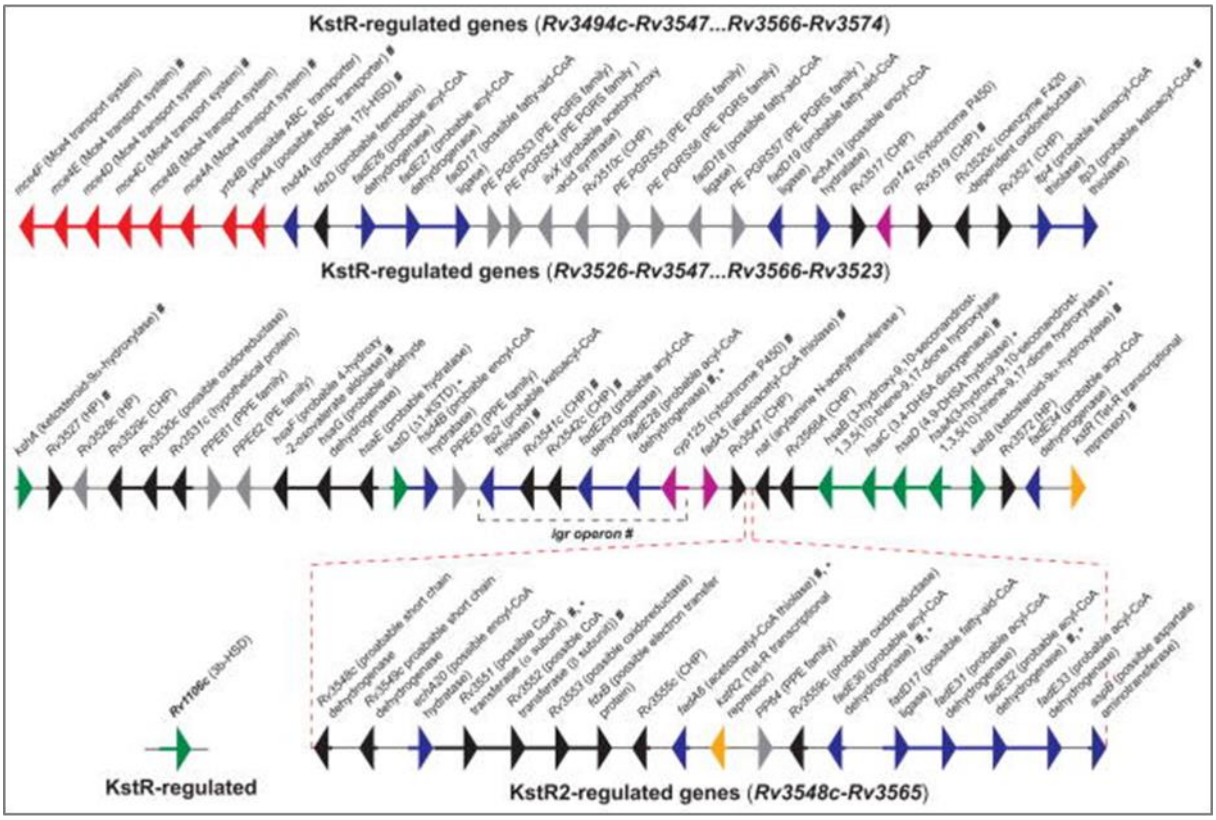

Figure 1.4. The cholesterol catabolic genes of *M. tuberculosis* (taken from Ouellet *et al*., 2011).

Assigned or proposed catalytic activities of genes are colour-coded: red = uptake, magenta = side-chain degradation, green = AB-rings degradation, blue = β-oxidation, orange = transcriptional regulator, black = unassigned function, grey = not regulated by KstR or KstR2. * Genes predicted to be essential for survival in macrophage cells and in murine infection. # Genes proven to be essential for survival in macrophage cells and in murine infection. Abbreviations: ABC-transporter, ATP-binding cassette transporter; PE, protein-containing Pro-Glu motifs; PE-PGRS, protein consisting of the PE domain followed by a C-terminal extension with multiple tandem repetitions of Gly-Gly-Ala or Gly-Gly-Asn; CHP, conserved hypothetical protein; HP, hypothetical protein; PPE, protein containing Pro-Pro-Glu motifs.

Like most bacteria, *M. tuberculosis* does not make its own sterols, but research has shown that cholesterol is required for infection and that the mycobacteria can use cholesterol as a sole source of carbon (Brzostek *et al*., 2009; Ouellet *et al*., 2011). An ABC-like transport system, coded by *mce4* genes, was identified to be involved in cholesterol import into *M. tuberculosis* (Ouellet *et al*., 2011). There are actually four *mce* gene clusters (*mce1-4*) that consist of 12-gene operons (Rengarajan *et al*., 2005). The Mce4 transport system is essential for cholesterol transport into bacterial cells and also plays an important role in pathogenesis, since a mutant lacking this transporter fails to persist in the lungs of chronically infected mice and cannot grow in (IFN-λ)-activated macrophages (Pandey and Sassetti, 2008). This means that the Mce4 transport system transports host cholesterol into the bacterial cell where cholesterol is degraded to produce energy for the bacteria to survive and cause disease. Also, the ability of Mce4 proteins to bind cholesterol could act as a signal allowing the pathogen to interact with the host (Mohn *et al*., 2008).

Cholesterol-degrading enzymes also play a crucial role in the pathogenesis. For instance, the cholesterol oxidase (ChOx) from *Rhodococcus equi*, a primary pathogen of horses and an opportunistic pathogen of humans, has been described as the major membrane-damaging factor (Navas *et al*., 2001), although its role in pathogenesis is controversial (Pei *et al*., 2006). The cholesterol oxidase found in *M. tuberculosis* also appears to be important for the growth of the pathogen in peritoneal macrophages and lungs of mice (Brzostek *et al*., 2005).

Using ^{14}C-labeled cholesterol derivatives in the investigation of cholesterol degradation by *M. tuberculosis*, it was shown that the carbon atoms from the sterol framework go to energy production and the ones from the aliphatic side chain go to lipid synthesis (Pandey and Sassetti, 2008), increasing the lipid virulence factor phthiocerol dimycocerosate (PDIM) in the organism cell wall (Yang *et al*., 2009).

1.2.4. Side chain degradation

It is generally accepted that the cholesterol side chain is shortened by β-oxidation reactions (Ouellet *et al*., 2011). Before the saturated side chain of cholesterol can enter into the *M. tuberculosis* β-oxidation pathway, it must first be chemically functionalised at the ω-position (Ouellet *et al*., 2011). Of the four chemical steps necessary to prepare the side chain for β-oxidation, the first three are oxidation reactions catalysed by cytochrome P450 enzymes CYP125 (*Rv3545c*), CYP142 (*Rv3518c*) and CYP124 (*Rv2266*) (Figure 1.3.) (Rosłoniec *et al*., 2009; Ouellet *et al*., 2011). They are capable of oxidising the side chains of cholesterol and cholest-4-en-3-one to the terminal alcohol (by introducing a hydroxyl group onto the side chain), aldehyde and carboxylic acid metabolites (Ouellet *et al*., 2011). A sterol-CoA ligase catalyses the final ATP-dependent step (Ouellet *et al*., 2011).

Research has demonstrated that CYP125 does not play a key role in cholesterol catabolism in the *M. tuberculosis* H37Rv strain and suggests that this strain possesses compensatory activities (Johnston *et al*., 2010). However, investigation of the *in vitro* enzyme specificities found that CYP125 and CYP142 are the dominant P450 enzymes responsible for initiating sterol side chain degradation in *M. tuberculosis* (Johnston *et al*., 2010), although in the CDC1551 strain, CYP142 is present as a pseudogene (García-Fernández *et al*., 2013). *In vitro* analysis has also demonstrated that CYP142 can support the growth of the H37Rv strain on cholesterol in the absence of *cyp125A1* (Johnston *et al*., 2010). Using western blot analysis, they found that CYP124A1 was not detectably expressed in the H37Rv or CDC1551 strains, but CYP142 was found in H37Rv and not in CDC1551 (Johnston *et al*., 2010). In the absence of CYP125 or CYP142, cholest-4-en-3-one accumulates and inhibits bacterial growth on cholesterol (Ouellet *et al*., 2011).

β-Oxidation is the pathway of the breakdown of fatty acids in the form of acyl-CoA molecules, i.e. the metabolic pathway of fatty acids oxidation (Figure 1.5.) (Świzdor *et al*., 2012). Before the oxidative reactions of the β-oxidation cycle, the fatty acid is activated in a reaction catalysed by an ATP-dependent ligase, to its thioester with coenzyme A (CoA). The thioester then undergoes dehydrogenation catalysed by acyl-CoA dehydrogenase to form the enoyl-CoA, which is then hydrated to the hydroxyacyl-CoA by enoyl-CoA hydratase. 3-Hydroxyacyl-CoA dehydrogenase then catalyses the oxidation of the hydroxyl group. The thiolase described in the next step by Świzdor *et al*., (2012) to catalyse the thiolytic cleavage of β-ketoacyl-CoA into two molecules of acyl-CoA as products, seems to correspond to the FadA5 described below. A single round of the β-oxidation cycle of unbranched chain fatty acids produces acetyl-CoA and a CoA thioester of an acid shorter by two carbon atoms. The shortened fatty acyl-CoA then undergoes a further round of the β-oxidation cycle.

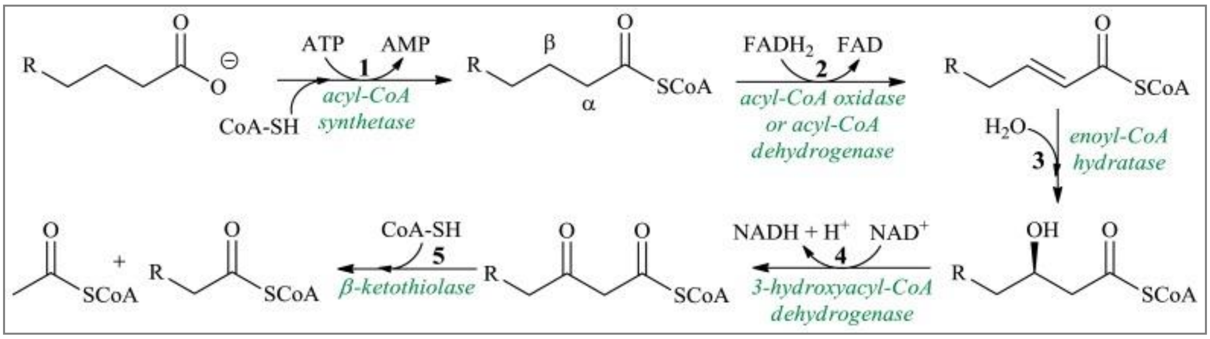

Figure 1.5. Scheme of the β-oxidation cycle (taken from Świzdor *et al*., 2012).

Genes believed to be encoding β-oxidation enzymes have been identified in the cholesterol regulons of *M. tuberculosis* (Ouellet *et al*., 2011). One of these enzymes, a thiolase encoded by *fadA5*, catalyses the thiolysis of acetoacetyl-CoA *in vitro*, which is consistent with removal of the side chain by β-oxidation, producing androstene metabolites, androst-4-ene-3,17-dione (AD) and androsta-1,4-diene-3,17-dione (ADD) (Figure 1.3.) (Nesbitt *et al*., 2010). This activity is required for growth on cholesterol and virulence, especially during the late (chronic) stage of mouse infection, prior to the onset of the immune response (Nesbitt *et al*., 2010; García-Fernández *et al*., 2013). Another set of enzymes, acyl-CoA dehydrogenases, is required to catalyse unsaturation reactions in β-oxidation of steroid-CoA substrates, and the *M. tuberculosis* genome contains six sets of these enzyme genes (*fadE*'s) (Wipperman *et al*., 2013). Regulated by cholesterol, each set of these genes is found adjacent to another within the same operon (Wipperman *et al*., 2013).

According to Schnappinger *et al*. (2003), their research indicated the induction of 18 genes predicted to encode all the enzymes necessary for the biochemical activation and β-oxidation of fatty acids, including fatty acid-CoA synthase (*fadD3*, *fadD9*, *fadD10*, *fadD19*), acyl-CoA dehydrogenase (*fadE5*, *fadE14*, *fadE22-24*, *fadE27-29*, *fadE31*), enoyl-CoA hydratase (*echA19*), hydroxybutyryl-CoA dehydrogenase (*fadB2*, *fadB3*), and acetyl-CoA transferase (*fadA5*, *fadA6*).

Griffin *et al*. (2011) also found that *hsd4A*, another predicted β-oxidation gene, is required for growth on cholesterol, along with *ltp2*, *fadE29*, *fadE28*, *fadA5*, *fadE30*, *fadE32*, *fadE33*, *fadE34*, *hsd4B* and also *fadE5*, *echA9*, *fadD36* and *fadE25*.

1.2.5. Sterol ring degradation

The first step in the breakdown of the sterol ring is the conversion of cholesterol to cholest-4-en-3-one (Figure 1.3.) (Ouellet *et al*., 2011). This reaction may be catalysed by either a 3β-HSD or a cholesterol oxidase (ChoD) (Ouellet *et al*., 2011). As mentioned earlier, *Rv1106c* encodes a 3β-HSD. This enzyme uses NAD^+ as a cofactor and oxidises cholesterol (among others) to its 3-keto-4-ene product, cholest-4-en-3-one (Ouellet *et al*., 2011). *Rv3409c* encodes ChoD and is required for *M. tuberculosis* virulence (Brzostek *et al*., 2007). However, in a study by Yang *et al*. (2011) they found that *Rv3409c* is not required for growth on cholesterol as a sole carbon source and concluded that 3β-HSD is required for the initial conversion of cholesterol and that a second ChoD activity is not present in *M. tuberculosis*. In addition to this, mice infection experiments confirmed the significance of ChoD in the pathogenesis of *M. tuberculosis*, where it drives the oxidation of 3β-hydroxy-5-ene to 3-keto-4-ene (Brzostek *et al*., 2007).

3-ketosteroid-Δ1-dehydrogenase (Δ^1KstD) is assumed to be coded by the *Rv3537* gene that is part of the cholesterol regulon (Van Der Geize *et al*., 2007; Ouellet *et al*., 2011). This enzyme catalyses the trans-axial elimination of the C1(α) and C2(β) hydrogen atoms (C1-C2 dehydrogenation) of the 3-ketosteroid A ring of 4-androstenedione (AD) to yield 1,4-androstenedione (ADD) (Figure 1.3.) (Ouellet *et al*., 2011) and targeted disruption of this gene inhibited growth on cholesterol (Brzostek *et al*., 2009). In research done by Brzostek *et al*. (2009) they found direct evidence that *M. tuberculosis* degrades cholesterol exclusively via the AD/ADD intermediates and that KstD plays an essential role in this process.

In the next step, 9-hydroxylation of the 3-ketosteroid is catalysed by KshAB (3-ketosteroid 9α-hydroxylase), a two-component Rieske oxygenase (Figure 1.3) (Petrusma *et al*., 2014). KshA (*Rv3526*) is the oxygenase component and KshB (*Rv3571*) is the reductase component (Petrusma *et al*., 2014). Research has shown that Δ*kshA* and Δ*kshB* deletion mutants were unable to utilise cholesterol and are essential in *M. tuberculosis* pathogenicity (Hu *et al*., 2010).

These two steps – the 9-hydroxylation of the 3-ketosteroid together with the C1-C2 dehydrogenation – are key to opening of the B ring and aromatisation of the A ring via 9-hydroxy-1,4-androstene-3,17-dione (9OHADD) (Figure 1.3.) (Ouellet *et al*., 2011). This intermediate is unstable and spontaneously hydrolyses to 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione (3-HSA) (Petrusma *et al*., 2014).

The *hsaACDB* genes in *M. tuberculosis* are part of a single operon (Dresen *et al*., 2010) and transposon mutagenesis studies have indicated that their activity is critical for the survival of *M. tuberculosis* in macrophages (Rengarajan *et al*., 2005; Dresen *et al*., 2010). The *hsaA* and *hsaB* genes encode for a putative oxygenase and reductase, respectively, of a flavin-dependent mono-oxygenase that hydroxylates (C4-hydroxylation) 3-HAS, a phenol, to a catechol, 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione (3,4-DHSA) (Figure 1.3) (Dresen *et al*., 2010). 3,4-DHSA is then oxygenated and cleaved by HsaC, an iron-dependent extradiol dioxygenase, to produce 4,5-9,10-diseco-3-hydroxy-5,9,17-

trioxoandrosta-1(10),2-dien-4-oic acid (4,9-DSHA) (Figure 1.3) (Ouellet *et al*., 2011). When this enzyme that fragments ring A of the sterol is interrupted, the toxic catechol intermediate accumulates (Ouellet *et al*., 2011). HsaD, a member of the *α/β* hydrolase family, is involved in the aerobic degradation of aromatic compounds in microbes and is coded by *hsaD*, one of the genes identified to be required for survival in macrophages (Ouellet *et al*., 2011). It is proposed to catalyse the hydrolysis of a carbon-carbon bond in 4,9-DSHA to yield 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid (DOHNAA) and 2-hydroxy-hexa-2,4-dienoic acid (HHD) (Figure 1.3.). HHD is then metabolised to tricarboxylic acid cycle intermediates (Lack *et al*., 2010), and propionyl-CoA (Ouellet *et al*., 2011), probably by HsaEFG (*hsaEFG*) (Griffin *et al*., 2011), whereas the metabolic fate of DOHNAA (corresponding to the C and D ring fragment) is still uncertain (Figure 1.3.) (Lack *et al*., 2010). Griffin *et al*. (2011) identified genes *fadE28*, *fadE29* and *fadD3* to probably be involved in the degradation of DOHNAA.

1.2.6. Cholesterol catabolism inhibitors

With increasingly more research demonstrating the importance of cholesterol catabolism in *M. tuberculosis* infection and persistence, targeting this pathway could prove effective in the search for novel potential drug targets. Blockage at certain steps in the pathway has been shown to result in cell death or bacteriostasis.

KshAB, HsaC are particularly attractive enzymes for antimycobacterials, since their inhibition causes bacteriotoxicity (Hu *et al*., 2010, Ouellet *et al*., 2011). Research done by Eltis and co-workers also suggests IpdAB (a predicted CoA transferase involved in the degradation of rings CD) as a potential drug target, and they have identified several compounds that strongly inhibit HsaC (https://www.microbiology.ubc.ca/research/labs/eltis/research/cholesterol-catabolism-mycobacterium-tuberculosis). They also indicate that KshAB, HsaC and IpdAB have no close human homologs, so host toxicity and bacterial cross-resistance are less likely, and since they are not targets of current therapies, they might be valuable therapeutic targets.

With *cyp125* predicted as one of the genes specifically required for growth on cholesterol (Griffin *et al*., 2011) and with the inhibition of bacterial growth on cholesterol in the absence of CYP125 (Ouellet *et al*., 2011), it seems to be a promising drug target for eliminating non-replicating latent *M. tuberculosis* (Hudson *et al*., 2012). Since CYP142 counters the growth defect of a CYP125 *M. tuberculosis* mutant strain on cholesterol, this one too may present an additional secondary drug target for latent infections (Johnston *et al*., 2010). Several antifungal azole drugs bind to CYP125 and CYP142 (García-Fernández *et al*., 2013) and research has shown that econazole especially exhibits antimycobacterial activities against both latent *M. tuberculosis* infection and multidrug-resistant strains (Ahmad *et al*., 2006a and b).

Azasteroids are compounds that inhibit enzymes in steroid biosynthetic pathways and inhibition of 3β-HSD for targeting the *M. tuberculosis* cholesterol catabolic pathway using a series of azasteroids was tested, but cross-reactivity with human 3β-HSD necessitates future work to develop more *M. tuberculosis*-specific inhibitors (Thomas *et al*., 2011).

1.3. Rationale, Aims and Objectives of the Study

As indicated above, genes/proteins involved in cholesterol degradation are now considered as novel drug targets in the fight against the deadly human pathogen, *M. tuberculosis*. However, to date, studies analysing the cholesterol degradation capability of different species belonging to the genus *Mycobacterium* have not been reported. Furthermore, performing wet-laboratory experiments on a large scale with different mycobacterial species is quite laborious and time- and money-consuming, in addition to the special laboratory safety requirements for working with certain safety-risk strains.

Considering the above facts and taking the advantage of genome sequencing of quite a number of mycobacterial species into account, this study aims to perform comprehensive *in silico* comparative analysis of the genes/proteins involved in cholesterol degradation in the genus *Mycobacterium*. Study results are aimed at identifying mycobacterial species capable of using cholesterol as a carbon source. To achieve this task, a comprehensive cholesterol degradation pathway will be deduced from the available literature. Genes/proteins involved in the cholesterol degradation pathway will be identified and comprehensive comparative analysis of homolog genes/proteins in different mycobacterial species will be performed using a newly developed software program. Gene/protein sequences will be collected and subjected to protein family assignment, phylogenetic and functional analysis. Finally, based on the presence or the absence of genes/proteins that are critical for cholesterol degradation, mycobacterial species' ability to degrade cholesterol will be determined.

References

- Ahmad, Z., Sharma, S. and Khuller, G.K., 2006a. Azole antifungals as novel chemotherapeutic agents against murine tuberculosis. *FEMS microbiology letters*, *261*(2), pp. 181-186.
- Ahmad, Z., Sharma, S. and Khuller, G.K., 2006b. The potential of azole antifungals against latent/persistent tuberculosis. *FEMS microbiology letters*, *258*(2), pp. 200-203.
- Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J. and Dziadek, J., 2007. Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*. *FEMS microbiology letters*, *275*(1), pp. 106-112.
- Brzostek, A., Pawelczyk, J., Rumijowska-Galewicz, A., Dziadek, B. and Dziadek, J., 2009. *Mycobacterium tuberculosis* is able to accumulate and utilize cholesterol. *Journal of bacteriology*, *191*(21), pp. 6584-6591.
- Brzostek, A., Śliwiński, T., Rumijowska-Galewicz, A., Korycka-Machała, M. and Dziadek, J., 2005. Identification and targeted disruption of the gene encoding the main 3-ketosteroid dehydrogenase in *Mycobacterium smegmatis*. *Microbiology*, *151*(7), pp. 2393-2402.
- 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.
- Chiang, Y.R., Ismail, W., Müller, M. and Fuchs, G., 2007. Initial steps in the anoxic metabolism of cholesterol by the denitrifying *Sterolibacterium denitrificans*. *Journal of biological chemistry*, *282*(18), pp. 13240-13249.
- Clark-Curtiss, J.E. and Haydel, S.E., 2003. Molecular genetics of *Mycobacterium tuberculosis* pathogenesis. *Annual reviews in microbiology*, *57*(1), pp. .517-549.

- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E. and Tekaia, F., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*(6685), p.537-544.
- Dahl, C. and Dahl, J., 1988. Cholesterol and cell function. In *Biology of cholesterol* (pp. 148-153). Boca Raton, FL: CRC Press Inc.
- Davis, J.H. and Finegold, L., 1993. The molecular dynamics, orientational order, and thermodynamic phase equilibria of cholesterol/phosphatidylcholine mixtures: 2H nuclear magnetic resonance. *Cholesterol in membrane models*, pp. 67-136.
- De Chastellier, C. and Thilo, L., 2006. Cholesterol depletion in *Mycobacterium avium*-infected macrophages overcomes the block in phagosome maturation and leads to the reversible sequestration of viable mycobacteria in phagolysosome-derived autophagic vacuoles. *Cellular microbiology*, *8*(2), pp. 242-256.
- Dresen, C., Lin, L.Y.C., D'Angelo, I., Tocheva, E.I., Strynadka, N. and Eltis, L.D., 2010. A flavin-dependent monooxygenase from *Mycobacterium tuberculosis* involved in cholesterol catabolism. *Journal of biological chemistry*, *285*(29), pp. 22264-22275.
- Dubnau, E., Chan, J., Mohan, V.P. and Smith, I., 2005. Responses of *Mycobacterium tuberculosis* to growth in the mouse lung. *Infection and immunity*, *73*(6), pp. 3754-3757.
- Edwards, P.A. and Ericsson, J., 1998. Signaling molecules derived from the cholesterol biosynthetic pathway: mechanisms of action and possible roles in human disease. *Current opinion in lipidology*, *9*(5), pp. 433-440.
- Finegold, L.X., 1992. *Cholesterol in membrane models* (pp. 2-9). Boca Raton, FL: CRC Press Inc.

- Flynn, J.L. and Chan, J., 2001. Tuberculosis: latency and reactivation. *Infection and immunity*, *69*(7), pp. 4195-4201.
- García, J.L., Uhía, I. and Galán, B., 2012. Catabolism and biotechnological applications of cholesterol degrading bacteria. *Microbial biotechnology*, *5*(6), pp. 679-699.
- García-Fernández, E., Frank, D.J., Galán, B., Kells, P.M., Podust, L.M., García, J.L. and Ortiz de Montellano, P.R., 2013. A highly conserved mycobacterial cholesterol catabolic pathway. *Environmental microbiology*, *15*(8), pp. 2342-2359.
- Gatfield, J. and Pieters, J., 2000. Essential role for cholesterol in entry of mycobacteria into macrophages. *Science*, *288*(5471), pp. 1647-1651.
- Glickman, M.S. and Jacobs, W.R., 2001. Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline. *Cell*, *104*(4), pp. 477-485.
- Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog*, *7*(9), p. e1002251.
- Griffin, J.E., Pandey, A.K., Gilmore, S.A., Mizrahi, V., Mckinney, J.D., Bertozzi, C.R. and Sassetti, C.M., 2012. Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. *Chemistry and biology*, *19*(2), pp. 218-227.
- Hu, Y., Van Der Geize, R., Besra, G.S., Gurcha, S.S., Liu, A., Rohde, M., Singh, M. and Coates, A., 2010. 3-ketosteroid 9α-hydroxylase is an essential factor in the pathogenesis of *Mycobacterium tuberculosis*. *Molecular microbiology*, *75*(1), pp. 107-121.
- 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):573-579.

- Jackson, S.M., Ericsson, J., Edwards, P.A. and Bittman, R., 1997. *Cholesterol: Its functions and metabolism in biology and medicine* (pp. 1-22). Plenum Press. New York (USA).
- 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., Burgess, P., Balhana, R., Withers, M., ten Bokum, A., Lott, J.S., Gao, C., Uhia-Castro, I. and Stoker, N.G., 2010. Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: kstR and kstR2. *Microbiology*, *156*(5), pp. 1362-1371.
- Kim, M.J., Wainwright, H.C., Locketz, M., Bekker, L.G., Walther, G.B., Dittrich, C., Visser, A., Wang, W., Hsu, F.F., Wiehart, U. and Tsenova, L., 2010. Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO molecular medicine*, *2*(7), pp. 258-274.
- Koul, A., Herget, T., Klebl, B. and Ullrich, A., 2004. Interplay between mycobacteria and host signalling pathways. *Nature reviews microbiology*, *2*(3), pp. 189-202.
- Kusumi, A., Subczynski, W.K., Pasenkiewicz-Gierula, M., Hyde, J.S. and Merkle, H., 1986. Spin-label studies on phosphatidylcholine-cholesterol membranes: effects of alkyl chain length and unsaturation in the fluid phase. *Biochimica et biophysica acta (BBA)-Biomembranes*, *854*(2), pp. 307-317.
- Kusumi, A., Tsuda, M., Akino, T., Ohnishi, S. and Terayama, Y., 1983. Protein-phospholipid-cholesterol interaction in the photolysis of invertebrate rhodopsin. *Biochemistry*, *22*(5), pp. 1165-1170.

- Lack, N.A., Yam, K.C., Lowe, E.D., Horsman, G.P., Owen, R.L., Sim, E. and Eltis, L.D., 2010. Characterization of a carbon-carbon hydrolase from *Mycobacterium tuberculosis* involved in cholesterol metabolism. *Journal of biological chemistry*, *285*(1), pp. 434-443.
- Lund-Katz, S., Laboda, H.M., McLean, L.R. and Phillips, M.C., 1988. Influence of molecular packing and phospholipid type on rates of cholesterol exchange. *Biochemistry*, *27*(9), pp. 3416-3423.
- Marieb, E.N. and Hoehn, K.H., 2010. *Human anatomy and physiology* (pp 614-617). 8th ed. San Francisco, USA: Pearson Benjamin Cummings.,
- Martens, G.W., Arikan, M.C., Lee, J., Ren, F., Vallerskog, T. and Kornfeld, H., 2008. Hypercholesterolemia impairs immunity to tuberculosis. *Infection and immunity*, *76*(8), pp. 3464-3472.
- McIntosh, T.J., 1999. Structure and physical properties of the lipid membrane. *Current topics in membranes*, *48*, pp. 23-47.
- Merino, E., Barrientos, A., Rodríguez, J., Naharro, G., Luengo, J.M. and Olivera, E.R., 2013. Isolation of cholesterol-and deoxycholate-degrading bacteria from soil samples: evidence of a common pathway. *Applied microbiology and biotechnology*, *97*(2), pp. 891-904.
- Migliori, G.B., Centis, R., D'Ambrosio, L., Spanevello, A., Borroni, E., Cirillo, D.M. and Sotgiu, G., 2012. Totally drug-resistant and extremely drug-resistant tuberculosis: the same disease? *Clinical infectious diseases*, *54*(9), pp. 1379-1380.
- Miner, M.D., Chang, J.C., Pandey, A.K., Sassetti, C.M. and Sherman, D.R., 2009. Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian journal of experimental biology*, *47*: 407-411.

- Mohn, W.W., Van Der Geize, R., Stewart, G.R., Okamoto, S., Liu, J., Dijkhuizen, L. and Eltis, L.D., 2008. The actinobacterial mce4 locus encodes a steroid transporter. *Journal of biological chemistry*, *283*(51), pp. 35368-35374.
- Muñoz, S., Rivas-Santiago, B. and Enciso, J.A., 2009. *Mycobacterium tuberculosis* entry into mast cells through cholesterol-rich membrane microdomains. *Scandinavian journal of immunology*, *70*(3), pp. 256-263.
- Navas, J., González-Zorn, B., Ladrón, N., Garrido, P. and Vázquez-Boland, J.A., 2001. Identification and mutagenesis by allelic exchange of choE, encoding a cholesterol oxidase from the intracellular pathogen *Rhodococcus equi*. *Journal of bacteriology*, *183*(16), pp. 4796-4805.
- Needham, D. and Nunn, R.S., 1990. Elastic deformation and failure of lipid bilayer membranes containing cholesterol. *Biophysical journal*, *58*(4), pp. 997-1009.
- Nesbitt, N.M., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N.S. and Dubnau, E., 2010. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infection and immunity*, *78*(1), pp. 275-282.
- Nguyen, L. and Pieters, J., 2005. The Trojan horse: survival tactics of pathogenic mycobacteria in macrophages. *Trends in cell biology*, *15*(5), pp. 269-276.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2011. Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends in microbiology*, *19*(11), pp. 530-539.
- Pandey, A.K. and Sassetti, C.M., 2008. Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences*, *105*(11), pp. 4376-4380.

- Pei, Y., Dupont, C., Sydor, T., Haas, A. and Prescott, J.F., 2006. Cholesterol oxidase (ChoE) is not important in the virulence of *Rhodococcus equi*. *Veterinary microbiology*, *118*(3), pp. 240-246.
- Petrusma, M., van der Geize, R. and Dijkhuizen, L., 2014. 3-Ketosteroid 9α-hydroxylase enzymes: Rieske non-heme monooxygenases essential for bacterial steroid degradation. *Antonie van Leeuwenhoek*, *106*(1), pp. 157-172.
- Peyron, P., Bordier, C., Elsa-Noah, N. and Maridonneau-Parini, I., 2000. Nonopsonic phagocytosis of *Mycobacterium kansasii* by human neutrophils depends on cholesterol and is mediated by CR3 associated with glycosylphosphatidylinositol-anchored proteins. *The journal of immunology*, *165*(9), pp. 5186-5191.
- Pieters, J., Gatfield, J., 2002. Hijacking the host; survival of pathogenic mycobacteria inside macrophages. *Trends in microbiology*, *10*(3), pp. 142-146.
- Rengarajan, J., Bloom, B.R. and Rubin, E.J., 2005. Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), pp. 8327-8332.
- Rosłoniec, K.Z., Wilbrink, M.H., Capyk, J.K., Mohn, W.W., Ostendorf, M., Van Der Geize, R., Dijkhuizen, L. and Eltis, L.D., 2009. Cytochrome P450 125 (CYP125) catalyses C26-hydroxylation to initiate sterol side-chain degradation in *Rhodococcus jostii* RHA1. *Molecular microbiology*, *74*(5), pp. 1031-1043.
- Schäfer, G., Guler, R., Murray, G., Brombacher, F. and Brown, G.D., 2009. The role of scavenger receptor B1 in infection with *Mycobacterium tuberculosis* in a murine model. *PloS one*, *4*(12), p. e8448.
- 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. *Journal of experimental medicine*, *198*(5), pp. 693-704.

- Slaytor, M. and Bloch, K., 1965. Metabolic transformations of cholestenediols. *Journal of biological chemistry*, *240*(12), pp. 4598-4602.
- Smaby, J.M., Brockman, H.L. and Brown, R.E., 1994. Cholesterol's interfacial interactions with sphingomyelins and-phosphatidylcholines: hydrocarbon chain structure determines the magnitude of condensation. *Biochemistry*, *33*(31), pp. 9135-9142.
- Sotgiu, G., Sulis, G. and Matteelli, A., 2017. Tuberculosis - a World Health Organization Perspective. *Microbiology spectrum*, *5*(1).
- Subczynski, W.K. and Wisniewska, A., 2000. Physical properties of lipid bilayer membranes: relevance to membrane biological functions. *Acta biochimica polonica - English edition*, *47*(3), pp. 613-626.
- Świzdor, A., Panek, A., Milecka-Tronina, N. and Kołek, T., 2012. Biotransformations utilizing β-oxidation cycle reactions in the synthesis of natural compounds and medicines. *International journal of molecular sciences*, *13*(12), pp. 16514-16543.
- Thomas, S.T., VanderVen, B.C., Sherman, D.R., Russell, D.G. and Sampson, N.S., 2011. Pathway profiling in *Mycobacterium tuberculosis* elucidation of cholesterol-derived catabolite and enzymes that catalyze its metabolism. *Journal of biological chemistry*, *286*(51), pp. 43668-43678.
- Thomas, S.T., Yang, X. and Sampson, N.S., 2011. Inhibition of the *M. tuberculosis* 3β-hydroxysteroid dehydrogenase by azasteroids. *Bioorganic and medicinal chemistry letters*, *21*(8), pp. 2216-2219.
- Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W. and Eltis, L.D., 2007. A gene cluster

encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences*, *104*(6), pp. 1947-1952.

- Veiga, P., Juste, C., Lepercq, P., Saunier, K., Béguet, F. and Gérard, P., 2005. Correlation between faecal microbial community structure and cholesterol-to-coprostanol conversion in the human gut. *FEMS microbiology letters*, *242*(1), pp. 81-86.
- Vist, M.R. and Davis, J.H., 1990. Phase equilibria of cholesterol/dipalmitoylphosphatidylcholine mixtures: deuterium nuclear magnetic resonance and differential scanning calorimetry. *Biochemistry*, *29*(2), pp. 451-464.
- WHO, global tuberculosis report 2015. Available from: http://www.who.int/tb/publications/global_report/gtbr15_main_text.pdf.
- WHO, Tuberculosis: Global tuberculosis report, 2016. Available from: http://www.who.int/tb/publications/factsheet_global.pdf.
- Wipperman, M.F., Yang, M., Thomas, S.T. and Sampson, N.S., 2013. Shrinking the FadE proteome of *Mycobacterium tuberculosis*: insights into cholesterol metabolism through identification of an α2β2 heterotetrameric acyl coenzyme A dehydrogenase family. *Journal of bacteriology*, *195*(19), pp. 4331-4341.
- Wollam, J. and Antebi, A., 2011. Sterol regulation of metabolism, homeostasis, and development. *Annual review of biochemistry*, *80*, pp. 885-916.
- Yang, X., Gao, J., Smith, I., Dubnau, E. and Sampson, N.S., 2011. Cholesterol is not an essential source of nutrition for *Mycobacterium tuberculosis* during infection. *Journal of bacteriology*, *193*(6), pp. 1473-1476.

- Yang, X., Nesbitt, N.M., Dubnau, E., Smith, I. and Sampson, N.S., 2009. Cholesterol metabolism increases the metabolic pool of propionate in *Mycobacterium tuberculosis*. *Biochemistry*, *48*(18), pp. 3819-3821.
- Yeagle, P.L., 1985. Cholesterol and the cell membrane. *Biochimica et biophysica acta (BBA) - Reviews on biomembranes*, *822*(3-4), pp. 267-287.
- Yeagle, P.L., 1988. *Biology of cholesterol* (pp. 121-146). Boca Raton, FL: CRC Press Inc.
- Yeagle, P.L., 1993. The biophysics and cell biology of cholesterol: an hypothesis for the essential role of cholesterol in mammalian cells. *Cholesterol in membrane models*, pp. 1-12.

CHAPTER 2

ANNOTATION AND CLASSIFICATION OF GENES/PROTEINS INVOLVED IN CHOLESTEROL DEGRADATION IN *MYCOBACTERIUM TUBERCULOSIS* H37RV

2.1. Introduction

The discovery of the ability of *Mycobacterium tuberculosis* to degrade and use cholesterol as a sole source of carbon and energy has opened up the possibility of using genes/proteins involved in cholesterol degradation as novel drug targets (Ouellet *et al*. 2011). As discussed in Chapter 1, cholesterol degradation by bacteria has not yet been fully elucidated, but from the available biochemical and genetic data, the cholesterol degradation pathway for *M. tuberculosis* can be deduced.

The current chapter is aimed at deducing the cholesterol degradation pathway and classifying the genes/proteins involved in cholesterol degradation by *M. tuberculosis* H37Rv.

2.2. Methodology

Articles were collected from different databases, mainly Google Scholar, PubMed and Scopus. The literature was reviewed and the data was processed in order to create a schematic diagram of the cholesterol degradation pathway by *M. tuberculosis* H37Rv, showing the intermediate metabolites and the enzymes involved in different reactions. According to Ouellet *et al*. (2011), the cholesterol degradation pathway of *M. tuberculosis* can be divided into two major phases – firstly, the initial degradation of the aliphatic side chain and then the subsequent degradation of the A-D rings. For the purpose of this study, the two phases were drawn up separately, using ChemDraw software (Mills, 2006).

2.3. Results and Discussion

2.3.1. Annotation of the cholesterol degradation pathway in *M. tuberculosis* H37Rv

As mentioned earlier, the cholesterol degradation pathway in *M. tuberculosis* can be divided into two major phases – the initial degradation of the aliphatic side chain and then the degradation of the four alicyclic A-D rings. It has not been confirmed whether there is a specific order to the degradation reactions regarding the side chain and the rings, but for *M. tuberculosis* it has been suggested that the ring-degrading enzymes, KsaAB and HsaA-C, act optimally after the side chain has been removed, since blockage of the side chain degradation resulted in accumulation of cholest-4-en-3-one as a major metabolite (Ouellet *et al.*, 2011).

Based on the available literature (Van Der Geize *et al.*, 2007; Nesbitt *et al.*, 2010; Griffin *et al.*, 2011; Ouellet *et al.*, 2011; Thomas *et al.*, 2011; Świzdor *et al.*, 2012; https://www.microbiology.ubc.ca/research/labs/eltis/research/cholesterol-catabolism-mycobacterium-tuberculosis), the following schematics (Figures 2.1. and 2.2.) have been drawn up using ChemDraw software.

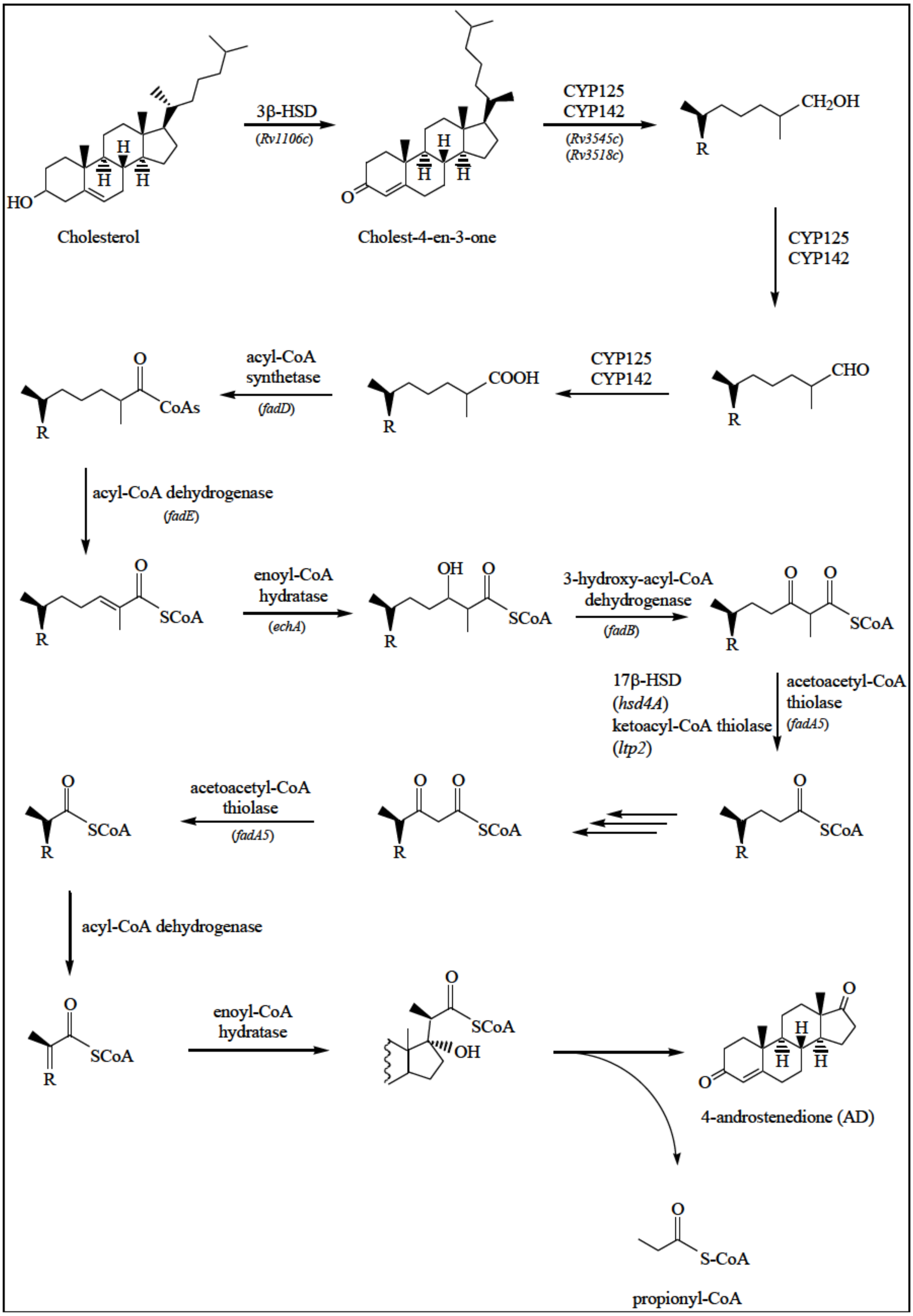

Figure 2.1. Cholesterol side chain degradation.

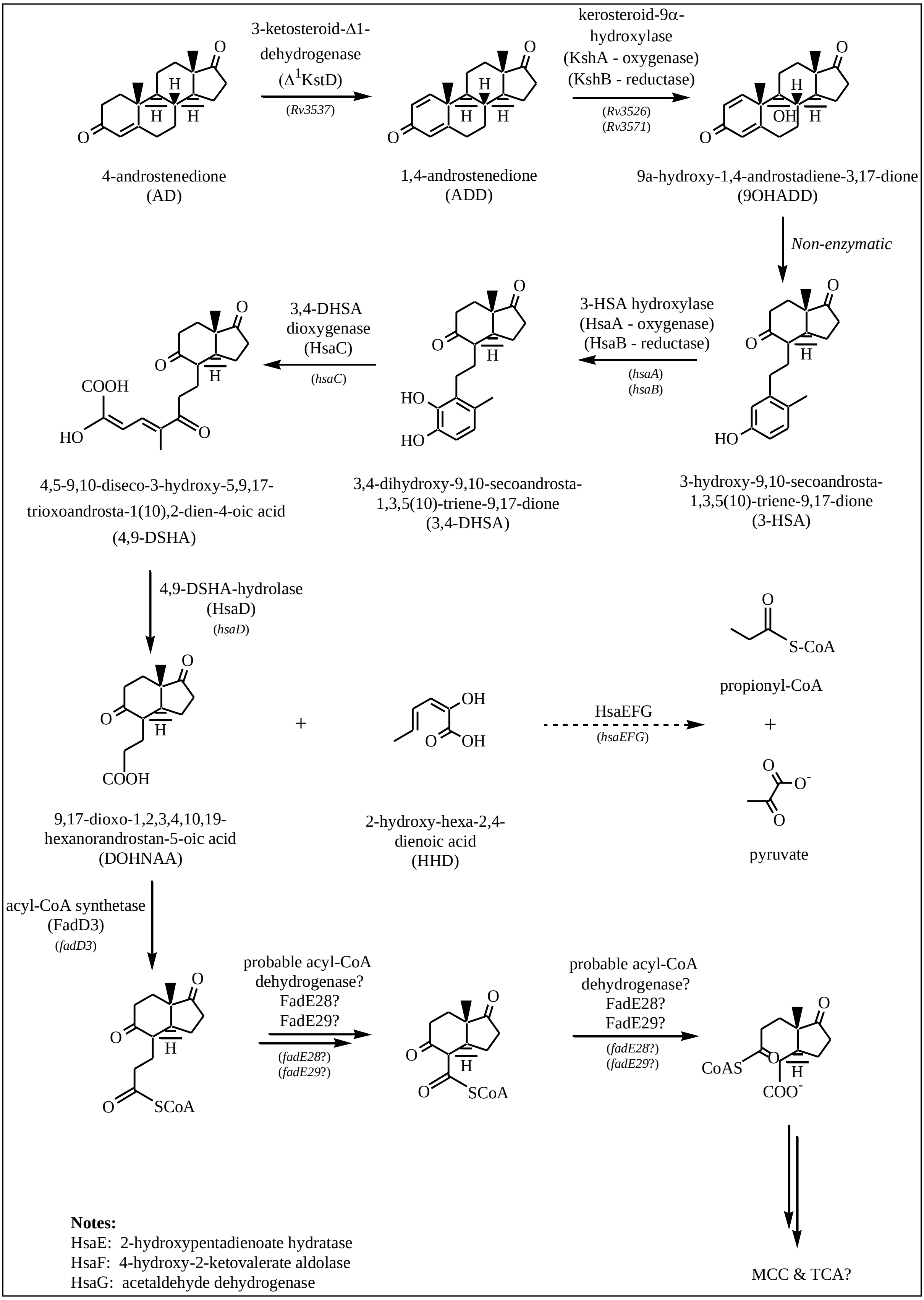

Figure 2.2. Cholesterol ring degradation.

2.3.2. Classification of genes/proteins involved in cholesterol degradation by *M. tuberculosis* H37Rv

Based on literature, all genes known to be involved in cholesterol breakdown in *M. tuberculosis* H37Rv have been classed into four different categories. The following tables elaborate on this classification.

2.3.2.1. Genes predicted to be specifically required for growth on cholesterol

Griffin *et al*. (2011) identified genes that are important for the growth of *M. tuberculosis* on cholesterol. Through a deep sequencing-based mapping approach, they identified 96 genes predicted to be important for *M. tuberculosis* growth on cholesterol (Table 2.1). Independent studies also confirmed that the genes identified by Griffin et al. (2011) to be important for *M. tuberculosis*, grow on cholesterol (Van Der Geize *et al*., 2007; Johnston *et al*., 2010; Nesbitt *et al*., 2010; Ouellet *et al*., 2011; Mukhopadhyay *et al*., 2012; García-Fernández *et al*., 2013).

Table 2.1. Genes predicted to be specifically required for growth on cholesterol (taken from Griffin *et al*., 2011).

Gene	Location	Enzyme
ppiA	*Rv0009*	iron-regulated peptidyl-prolyl cis-trans isomerase A
ptbB	*Rv0153c*	phosphotyrosine protein phosphatase PTPB (protein-tyrosine-phosphatase) (PTPase)
mmpL11	*Rv0202c*	transmembrane transport protein MmpL11
fadE5	*Rv0244c*	acyl-CoA dehydrogenase
mgtE	*Rv0362*	Mg2+ transport transmembrane protein MgtE
metZ	*Rv0391*	O-succinylhomoserine sulfhydrylase
mmpL4	*Rv0450c*	transmembrane transport protein MmpL4
	Rv0485	transcriptional regulatory protein
	Rv0495c	HP
mkl	*Rv0655*	ribonucleotide ABC transporter ATP-binding protein
pqqE	*Rv0693*	coenzyme PQQ synthesis protein E
lldD1	*Rv0694*	L-lactate dehydrogenase (cytochrome) LldD1
	Rv0695	HP
	Rv0696	membrane sugar transferase
adhB	*Rv0761c*	zinc-containing alcohol dehydrogenase NAD dependent ADHB
	Rv0805	HP

Gene	Location	Enzyme
	Rv0876c	transmembrane protein
echA9	*Rv1071c*	3-hydroxyisobutyryl-CoA hydrolase
	Rv1084	HP
	Rv1096	glycosyl hydrolase
	Rv1129c	transcriptional regulator protein
	Rv1130	HP
gltA1	*Rv1131*	citrate synthase
mmpL10	*Rv1183*	transmembrane transport protein MmpL10
fadD36	*Rv1193*	acyl-CoA synthetase
	Rv1428c	HP
	Rv1432	dehydrogenase
bcpB	*Rv1608c*	peroxidoxin BcpB
	Rv1626	two-component system transcriptional regulator
	Rv1627c	lipid-transfer protein
	Rv1798	HP
	Rv1906c	HP
	Rv1919c	HP
mce3R	*Rv1963c*	transcriptional repressor (probably TETR-family) MCE3R
pks12	*Rv2048c*	polyketide synthase pks12
	Rv2118c	RNA methyltransferase
	Rv2206	transmembrane protein
	Rv2239c	HP
eis	*Rv2416c*	HP
tig	*Rv2462c*	trigger factor
	Rv2506	TetR family transcriptional regulator
	Rv2668	HP
	Rv2681	HP
arsA	*Rv2684*	arsenic-transport integral membrane protein ArsA
sigB	*Rv2710*	RNA polymerase sigma factor SigB
	Rv2799	HP
pknI	*Rv2914c*	transmembrane serine/threonine-protein kinase I
mutT1	*Rv2985*	hydrolase MutT1
	Rv3050c	AsnC family transcriptional regulator
fadE25	*Rv3274c*	acyl-CoA dehydrogenase FADE25
gcp	*Rv3419c*	putative DNA-binding/iron metalloprotein/AP endonuclease
	Rv3421c	HP
	Rv3492c	CHP MCE associated protein
	Rv3493c	CHP MCE associated protein
mce4F	*Rv3494c*	Mce4 transport system
mce4E / lprN	*Rv3495c*	Mce4 transport system
mce4D	*Rv3496c*	Mce4 transport system
mce4C	*Rv3497c*	Mce4 transport system
mce4B	*Rv3498c*	Mce4 transport system
mce4A	*Rv3499c*	Mce4 transport system
yrb4B / YrbE4B / supB	*Rv3500c*	possible ABC transporter (Sterol uptake permease subunit)
yrb4A / YrbE4A / supA	*Rv3501c*	possible ABC transporter (Sterol uptake permease subunit)
hsd4A	*Rv3502c*	17β-hydroxysteroid dehydrogenase (17β-HSD)

Gene	Location	Enzyme
fadD19	*Rv3515c*	probable fatty-acid-CoA ligase
kshA	*Rv3526*	kerosteroid-9α-hydroxylase, oxygenase
	Rv3531c	hypothetical protein
hsaF	*Rv3534c*	probable 4-hydroxy-2-oxovalerate aldolase/4-hydroxy-2-ketovalerate aldolase
hsaE	*Rv3536c*	probable hydratase/2-hydroxypentadienoate hydratase
kstD	*Rv3537*	3-ketosteroid-Δ1-dehydrogenase (Δ1-KSTD)
ltp2	*Rv3540c*	probable ketoacyl-CoA thiolase
	Rv3542c	CHP/putative enoyl-CoA hydratase
fadE29	*Rv3543c*	probable acyl-CoA dehydrogenase
fadE28	*Rv3544c*	probable acyl-CoA dehydrogenase
cyp125	*Rv3545c*	cytochrome P450
fadA5	*Rv3546*	acetoacetyl-CoA thiolase
	Rv3548c	probable short chain dehydrogenase/reductase
	Rv3549c	probable short chain dehydrogenase/reductase
ipdA	*Rv3551*	ATP-dependent CoA transferase α subunit
	Rv3553	possible oxidoreductase/2-nitropropane dioxygenase
	Rv3559c	probable oxidoreductase
fadE30	*Rv3560c*	probable acyl-CoA dehydrogenase
fadD3	*Rv3561*	acyl-CoA synthetase (AMP forming)
fadE32	*Rv3563*	probable acyl-CoA dehydrogenase
fadE33	*Rv3564*	probable acyl-CoA dehydrogenase
hsaC	*Rv3568c*	3,4-DHSA dioxygenase
hsaD	*Rv3569c*	4,9-DHSA hydrolase
hsaA	*Rv3570c*	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)
kshB	*Rv3571*	ketosteroid-9α-hydroxylase, reductase
	Rv3572	HP
fadE34	*Rv3573c*	probable acyl-CoA dehydrogenase
	Rv3575c	transcriptional regulatory protein LacI-family
	Rv3779	transmembrane protein alanine and leucine rich
papA2	*Rv3820c*	polyketide synthase associated protein PapA2
papA1	*Rv3824c*	polyketide synthase associated protein
pks2	*Rv3825c*	polyketide synthase PKS2
sigM	*Rv3911*	RNA polymerase sigma factor SigM

Abbreviations: 3-HAS = 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 3,4-DHSA = 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 4,9-DHSA hydrolase = 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-dien-4-oic acid; 17β-HSD = 17β-hydroxysteroid dehydrogenase; Δ1-KSTD = 3-ketosteroid-Δ1-dehydrogenase; ABC = ATP-binding cassette; ADH = alcohol dehydrogenase; AMP = adenosine monophosphate; AP = apurinic/apyrimidinic; ATP = adenosine triphosphate; Bcp = bacterioferritin comigratory protein; CHP = conserved hypothetical protein; CoA = co-enzyme A; DNA = deoxyribonucleic acid; HP =

hypothetical protein; LldD = L-lactate dehydrogenase; MCE = mammalian cell entry; MgtE = Mg2+ transport transmembrane protein; MmpL = *Mycobacterium* membrane protein laboratory; NAD = nicotinamide adenine dinucleotide; pks = polyketide synthase; PQQ = pyrrolo-quinoline quinone; PTP/PTPase = phosphotyrosine protein phosphatase/protein-tyrosine-phosphatase; RNA = ribonucleic acid; TetR = tetracycline repressor

2.3.2.2. Cholesterol catabolic genes proven to be or predicted to be essential for survival of *Mycobacterium tuberculosis* in macrophage cells and in murine infection

In the article by Ouellet *et al*. (2011), some of the cholesterol catabolic genes of *M. tuberculosis* were specified as genes proven to be essential for survival in macrophage cells and in murine infection (Table 2.2), or genes predicted to be essential for survival in macrophage cells and in murine infection (Table 2.3). Of the 24 genes listed in Table 2.2 as proven to be essential for survival in macrophage cells and in murine infection, 17 genes were predicted to be specifically required for growth on cholesterol by Griffin *et al*. (2011) and in other studies (Van Der Geize *et al*., 2007; Johnston *et al*., 2010; Nesbitt *et al*., 2010; Griffin *et al*., 2011; Van Der Geize *et al*., 2011; García-Fernández, *et al*., 2013).

Table 2.2. Genes proven to be essential for survival of *M. tuberculosis* in macrophage cells and in murine infection (as listed by Ouellet *et al*., 2011).

Gene	Location	Enzyme
mce4E / lprN	*Rv3495c*	Mce4 transport system
mce4C	*Rv3497c*	Mce4 transport system
mce4A	*Rv3499c*	Mce4 transport system
yrb4A / YrbE4A / supA	*Rv3501c*	possible ABC transporter (Sterol uptake permease subunit)
hsd4A	*Rv3502c*	17β-hydroxysteroid dehydrogenase (17β-HSD)
	Rv3519	CHP
ltp3	*Rv3523*	probable ketoacyl-CoA thiolase
kshA	*Rv3526*	kerosteroid-9α-hydroxylase, oxygenase
	Rv3527	hypothetical protein (HP)
hsaF	*Rv3534c*	probable 4-hydroxy-2-oxovalerate aldolase/4-hydroxy-2-ketovalerate aldolase
ltp2	*Rv3540c*	probable ketoacyl-CoA thiolase
	Rv3541c	CHP / putative enoyl-CoA hydratase

Gene	Location	Enzyme
	Rv3542c	CHP / putative enoyl-CoA hydratase
fadE28	*Rv3544c*	probable acyl-CoA dehydrogenase
cyp125	*Rv3545c*	cytochrome P450
fadA5	*Rv3546*	acetoacetyl-CoA thiolase
ipdA	*Rv3551*	ATP-dependent CoA transferase α subunit
ipdB	*Rv3552*	ATP-dependent CoA transferase β subunit
fadA6	*Rv3556c*	acetoacetyl-CoA thiolase
fadE30	*Rv3560c*	probable acyl-CoA dehydrogenase
fadE32	*Rv3563*	probable acyl-CoA dehydrogenase
hsaC	*Rv3568c*	3,4-DHSA dioxygenase
kshB	*Rv3571*	ketosteroid-9α-hydroxylase, reductase
kstR	*Rv3574*	Tet-R transcriptional regulator (repressor)

Abbreviations: 3,4-DHSA = 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 17β-HSD = 17β-hydroxysteroid dehydrogenase; ABC = ATP-binding cassette; ATP = adenosine triphosphate; CHP = conserved hypothetical protein; CoA = co-enzyme A; HP = hypothetical protein; MCE = mammalian cell entry; Tet-R = tetracycline repressor

Table 2.3. Genes predicted to be essential for survival of *M. tuberculosis* in macrophage cells and in murine infection (as listed by Ouellet *et al*., 2011).

Gene	Location	Enzyme
kstD	*Rv3537*	3-ketosteroid-Δ1-dehydrogenase (Δ1-KSTD)
fadE28	*Rv3544c*	probable acyl-CoA dehydrogenase
ipdA	*Rv3551*	ATP-dependent CoA transferase α subunit
fadA6	*Rv3556c*	acetoacetyl-CoA thiolase
fadE30	*Rv3560c*	probable acyl-CoA dehydrogenase
fadE32	*Rv3563*	probable acyl-CoA dehydrogenase
hsaD	*Rv3569c*	4,9-DHSA hydrolase
hsaA	*Rv3570c*	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)

Abbreviations: 3-HSA = 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 4,9-DHSA = 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-dien-4-oic acid; Δ1-KSTD = 3-ketosteroid-Δ1-dehydrogenase; ATP = adenosine triphosphate; CoA = co-enzyme A

Seven of the eight genes listed in Table 2.3 as predicted to be essential for survival in macrophage cells and in murine infection are also listed in Table 2.1 and five in Table 2.2 These eight genes are also noted in other literature (Van Der Geize *et al*., 2007; Nesbitt *et al*., 2010; Griffin *et al*., 2011; Van Der Geize *et al*., 2011; García-Fernández, *et al*., 2013).

2.3.2.3. Genes/proteins that are up-regulated during growth on cholesterol

Table 2.4 lists the genes for *M. tuberculosis* H37Rv that were predicted to be involved in cholesterol metabolism according to research conducted by Van Der Geize *et al*. (2007). In their study, they found that the complete set of 51 genes specifically expressed during growth on cholesterol in *Rhodococcus jostii* are also found in an 82-gene cluster in the *M. tuberculosis* and *M. bovis* BCG genomes. To annotate the cholesterol catabolic genes, they compared the sequence similarity of the gene products of *R. jostii* RHA1 and *M. tuberculosis* H37Rv strains and compiled a table with 28 genes annotated for *M. tuberculosis* H37Rv (Table 1.1). For the purpose of this chapter, the table has been summarised to show only the genes listed for *M. tuberculosis* H37Rv. Independent studies confirmed the importance of these genes in cholesterol degradation by *M. tuberculosis* (Nesbitt *et al*., 2010; Griffin *et al*., 2011; Ouellet *et al*., 2011 and García-Fernández *et al*., 2013). Eighteen of the 28 genes in Table 2.4 are also noted in Table 2.1 and ten of these relate to the list in Table 2.2, and three to the list in Table 2.3.

Table 2.4. Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway (Table 1.1) (taken from Van Der Geize *et al*., 2007).

Gene	Location	Enzyme
choD	*Rv3409c*	cholesterol oxidase
mce4F	*Rv3494c*	Mce4 transport system
mce4E / lprN	*Rv3495c*	Mce4 transport system
mce4D	*Rv3496c*	Mce4 transport system
mce4C	*Rv3497c*	Mce4 transport system
mce4B	*Rv3498c*	Mce4 transport system
mce4A	*Rv3499c*	Mce4 transport system
yrb4B / YrbE4B / supB	*Rv3500c*	possible ABC transporter (Sterol uptake permease subunit)
yrb4A / YrbE4A / supA	*Rv3501c*	possible ABC transporter (Sterol uptake permease subunit)
hsd4A	*Rv3502c*	17β-hydroxysteroid dehydrogenase (17β-HSD)
fadE26	*Rv3504*	probable acyl-CoA dehydrogenase
fadE27	*Rv3505*	probable acyl-CoA dehydrogenase
fadD17	*Rv3506*	possible fatty-acid-CoA ligase
fadD19	*Rv3515c*	probable fatty-acid-CoA ligase
echA19	*Rv3516*	possible enoyl-CoA hydratase
ltp4	*Rv3522*	probable ketoacyl-CoA thiolase
ltp3	*Rv3523*	probable ketoacyl-CoA thiolase
kshA	*Rv3526*	kerosteroid-9α-hydroxylase, oxygenase
hsaF	*Rv3534c*	probable 4-hydroxy-2-oxovalerate aldolase/4-hydroxy-2-ketovalerate aldolase
hsaG	*Rv3535c*	probable aldehyde dehydrogenase
hsaE	*Rv3536c*	probable hydratase / 2-hydroxypentadienoate hydratase
kstD	*Rv3537*	3-ketosteroid-Δ1-dehydrogenase (Δ1-KSTD)
hsd4B	*Rv3538*	probable enoyl-CoA hydratase
hsaB	*Rv3567c*	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)
hsaC	*Rv3568c*	3,4-DHSA dioxygenase
hsaD	*Rv3569c*	4,9-DHSA hydrolase
hsaA	*Rv3570c*	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)
kshB	*Rv3571*	ketosteroid-9α-hydroxylase, reductase

Abbreviations: 3-HSA = 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 3,4-DHSA = 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 4,9-DHSA hydrolase = 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-dien-4-oic acid; 17β-HSD = 17β-hydroxysteroid dehydrogenase; Δ1-KSTD = 3-ketosteroid-Δ1-dehydrogenase; ABC = ATP-binding cassette; CoA = co-enzyme A; MCE = mammalian cell entry

2.3.2.4. Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv, but not confirmed or predicted to be essential

Table 2.5 lists genes that have been noted as being involved in cholesterol catabolism by *M. tuberculosis* H37Rv, but not confirmed or predicted to be essential as per the published data (Van Der Geize *et al*., 2007; Nesbitt *et al*., 2010; Ouellet *et al*., 2011; Yang *et al*., 2011; Mukhopadhyay *et al*., 2012; Driscoll *et al*., 2010; García-Fernández *et al*., 2013).

Table 2.5. Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data (Van Der Geize *et al*., 2007; Nesbitt *et al*., 2010; Ouellet *et al*., 2011; Yang *et al*., 2011; Mukhopadhyay *et al*., 2012; Driscoll *et al*., 2010; García-Fernández *et al*., 2013).

Gene	Location	Enzyme
fadD10	*Rv0099*	fatty acid-CoA synthase
fadB2	*Rv0468*	hydroxybutyryl-CoA dehydrogenase
	Rv1106c	3β-hydroxysteroid dehydrogenase (3β-HSD)
mbtN (fadE14)	*Rv1346*	acyl-CoA dehydrogenase
fadB3	*Rv1715*	hydroxybutyryl-CoA dehydrogenase
fadD9	*Rv2590*	fatty acid-CoA synthase
fadE22	*Rv3061c*	acyl-CoA dehydrogenase
fadE24	*Rv3139*	acyl-CoA dehydrogenase
fadE23	*Rv3140*	acyl-CoA dehydrogenase
fdxD	*Rv3503c*	probable ferredoxin
PE PGRS53	*Rv3507*	PE PGRS family
PE PGRS54	*Rv3508*	PE PGRS family
ilvX	*Rv3509c*	probable acetohydroxy-acid synthase
	Rv3510c	CHP
PE PGRS55	*Rv3511*	PE PGRS family
PE PGRS56	*Rv3512*	PE PGRS family
fadD18	*Rv3513c*	possible fatty-acid-CoA ligase
PE PGRS57	*Rv3514*	PE PGRS family
whiB3	*Rv3517*	conserved hypothetical protein (CHP) / transcription factor
cyp142	*Rv3518c*	cytochrome P450
	Rv3520c	coenzyme F420-dependent oxidoreductase
	Rv3521	CHP
	Rv3524	probable conserved membrane protein
	Rv3525c	possible siderophore binding protein
	Rv3528c	HP
	Rv3529c	CHP
	Rv3530c	possible oxidoreductase

Gene	Location	Enzyme
PPE61	*Rv3532*	PPE family
PPE62	*Rv3533c*	PPE family
PPE63	*Rv3539*	PE
	Rv3547	CHP
echA20	*Rv3550*	possible enoyl-CoA hydratase
fdxB	*Rv3554*	possible electron transfer protein / ferredoxin
	Rv3555c	CHP
kstR2	*Rv3557c*	Tet-R transcriptional regulator (repressor)
PPE64	*Rv3558*	PPE
fadE31	*Rv3562*	probable acyl-CoA dehydrogenase
aspB	*Rv3565*	possible aspartate aminotransferase
	Rv3566A	CHP
nhoA / nat	*Rv3566c*	arylamine N-acetyltransferase

Abbreviations: 3β-HSD = 3β-hydroxysteroid dehydrogenase; CHP = conserved hypothetical protein; CoA = co-enzyme A; HP = hypothetical protein; PE = protein family with highly conserved Proline-Glutamate residues near the start of their encoded proteins; PGRS = polymorphic GC-rich-repetitive sequence; PPE = protein family with highly conserved Proline-Proline-Glutamate; Tet-R = tetracycline repressor

2.3.3. Genes/proteins selected for the study

Based on the above data, 152 genes were selected for determining mycobacterial species' ability to degrade/utilise cholesterol (Table 2.6).

Table 2.6. List of genes/proteins selected for determining mycobacterial species' ability to degrade/utilise cholesterol.

Gene	Location	Enzyme
mce4E / lprN	*Rv3495c*[a,c,d]	Mce4 transport system
mce4C	*Rv3497c*[a,c,d]	Mce4 transport system
mce4A	*Rv3499c*[a,c,d]	Mce4 transport system
yrb4A / YrbE4A / supA	*Rv3501c*[a,c,d]	possible ABC transporter (Sterol uptake permease subunit)
hsd4A	*Rv3502c*[a,c,d]	17β-hydroxysteroid dehydrogenase (17β-HSD)
kshA	*Rv3526*[a,c,d]	kerosteroid-9α-hydroxylase, oxygenase
hsaF	*Rv3534c*[a,c,d]	probable 4-hydroxy-2-oxovalerate aldolase / 4-hydroxy-2-ketovalerate aldolase
kstD	*Rv3537*[b,c,d]	3-ketosteroid-Δ1-dehydrogenase (Δ1-KSTD)
fadE28	*Rv3544c*[a,b,c]	probable acyl-CoA dehydrogenase
ipdA	*Rv3551*[a,b,c]	ATP-dependent CoA transferase α subunit

Gene	Location	Enzyme
fadE30	*Rv3560c*[a,b,c]	probable acyl-CoA dehydrogenase
fadE32	*Rv3563*[a,b,c]	probable acyl-CoA dehydrogenase
hsaC	*Rv3568c*[a,c,d]	3,4-DHSA dioxygenase
hsaD	*Rv3569c*[b,c,d]	4,9-DHSA hydrolase
hsaA	*Rv3570c*[b,c,d]	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)
kshB	*Rv3571*[a,c,d]	ketosteroid-9α-hydroxylase, reductase
mce4F	*Rv3494c*[c,d]	Mce4 transport system
mce4D	*Rv3496c*[c,d]	Mce4 transport system
mce4B	*Rv3498c*[c,d]	Mce4 transport system
yrb4B / YrbE4B / supB	*Rv3500c*[c,d]	possible ABC transporter (Sterol uptake permease subunit)
fadD19	*Rv3515c*[c,d]	probable fatty-acid-CoA ligase
ltp3	*Rv3523*[a,d]	probable ketoacyl-CoA thiolase
hsaE	*Rv3536c*[c,d]	probable hydratase/2-hydroxypentadienoate hydratase
ltp2	*Rv3540c*[a,c]	probable ketoacyl-CoA thiolase
	Rv3542c[a,c]	CHP / putative enoyl-CoA hydratase
cyp125	*Rv3545c*[a,c]	cytochrome P450
fadA5	*Rv3546*[a,c]	acetoacetyl-CoA thiolase
fadA6	*Rv3556c*[a,b]	acetoacetyl-CoA thiolase
ppiA	*Rv0009*[c]	iron-regulated peptidyl-prolyl cis-trans isomerase A
fadD10	*Rv0099*[e]	fatty acid-CoA synthase
ptbB	*Rv0153c*[c]	phosphotyrosine protein phosphatase PTPB (protein-tyrosine-phosphatase) (PTPase)
mmpL11	*Rv0202c*[c]	transmembrane transport protein MmpL11
fadE5	*Rv0244c*[c]	acyl-CoA dehydrogenase
mgtE	*Rv0362*[c]	Mg2+ transport transmembrane protein MgtE
metZ	*Rv0391*[c]	O-succinylhomoserine sulfhydrylase
mmpL4	*Rv0450c*[c]	transmembrane transport protein MmpL4
fadB2	*Rv0468*[e]	hydroxybutyryl-CoA dehydrogenase
	Rv0485[c]	transcriptional regulatory protein
	Rv0495c[c]	HP
mkl	*Rv0655*[c]	ribonucleotide ABC transporter ATP-binding protein
pqqE	*Rv0693*[c]	coenzyme PQQ synthesis protein E
lldD1	*Rv0694*[c]	L-lactate dehydrogenase (cytochrome) LldD1
	Rv0695[c]	HP
	Rv0696[c]	membrane sugar transferase
adhB	*Rv0761c*[c]	zinc-containing alcohol dehydrogenase NAD dependent ADHB
	Rv0805[c]	HP
	Rv0876c[c]	transmembrane protein
echA9	*Rv1071c*[c]	3-hydroxyisobutyryl-CoA hydrolase
	Rv1084[c]	HP
	Rv1096[c]	glycosyl hydrolase
	Rv1106c[e]	3β-HSD
	Rv1129c[c]	transcriptional regulator protein
	Rv1130[c]	HP
gltA1	*Rv1131*[c]	citrate synthase

Gene	Location	Enzyme
mmpL10	*Rv1183*[c]	transmembrane transport protein MmpL10
fadD36	*Rv1193*[c]	acyl-CoA synthetase
mbtN (fadE14)	*Rv1346*[e]	acyl-CoA dehydrogenase
	Rv1428c[c]	HP
	Rv1432[c]	dehydrogenase
bcpB	*Rv1608c*[c]	peroxidoxin BcpB
	Rv1626[c]	two-component system transcriptional regulator
	Rv1627c[c]	lipid-transfer protein
fadB3	*Rv1715*[e]	hydroxybutyryl-CoA dehydrogenase
	Rv1798[c]	HP
	Rv1906c[c]	HP
	Rv1919c[c]	HP
mce3R	*Rv1963c*[c]	transcriptional repressor (probably TETR-family) MCE3R
pks12	*Rv2048c*[c]	polyketide synthase pks12
	Rv2118c[c]	RNA methyltransferase
	Rv2206[c]	transmembrane protein
	Rv2239c[c]	HP
eis	*Rv2416c*[c]	HP
tig	*Rv2462c*[c]	trigger factor
	Rv2506[c]	TetR family transcriptional regulator
fadD9	*Rv2590*[e]	fatty acid-CoA synthase
	Rv2668[c]	HP
	Rv2681[c]	HP
arsA	*Rv2684*[c]	arsenic-transport integral membrane protein ArsA
sigB	*Rv2710*[c]	RNA polymerase sigma factor SigB
	Rv2799[c]	HP
pknI	*Rv2914c*[c]	transmembrane serine/threonine-protein kinase I
mutT1	*Rv2985*[c]	hydrolase MutT1
	Rv3050c[c]	AsnC family transcriptional regulator
fadE22	*Rv3061c*[e]	acyl-CoA dehydrogenase
fadE24	*Rv3139*[e]	acyl-CoA dehydrogenase
fadE23	*Rv3140*[e]	acyl-CoA dehydrogenase
fadE25	*Rv3274c*[c]	acyl-CoA dehydrogenase FADE25
choD	*Rv3409c*[d]	cholesterol oxidase
gcp	*Rv3419c*[c]	putative DNA-binding/iron metalloprotein/AP endonuclease
	Rv3421c[c]	HP
	Rv3492c[c]	CHP MCE associated protein
	Rv3493c[c]	CHP MCE associated protein
fdxD	*Rv3503c*[e]	probable ferredoxin
fadE26	*Rv3504*[d]	probable acyl-CoA dehydrogenase
fadE27	*Rv3505*[d]	probable acyl-CoA dehydrogenase
fadD17	*Rv3506*[d]	possible fatty-acid-CoA ligase
PE PGRS53	*Rv3507*[e]	PE PGRS family
PE PGRS54	*Rv3508*[e]	PE PGRS family
ilvX	*Rv3509c*[e]	probable acetohydroxy-acid synthase

Gene	Location	Enzyme
	Rv3510c[e]	CHP
PE PGRS55	*Rv3511*[e]	PE PGRS family
PE PGRS56	*Rv3512*[e]	PE PGRS family
fadD18	*Rv3513c*[e]	possible fatty-acid-CoA ligase
PE PGRS57	*Rv3514*[e]	PE PGRS family
echA19	*Rv3516*[d]	possible enoyl-CoA hydratase
whiB3	*Rv3517*[e]	conserved hypothetical protein (CHP)/transcription factor
cyp142	*Rv3518c*[e]	cytochrome P450
	Rv3519[a]	CHP
	Rv3520c[e]	coenzyme F420-dependent oxidoreductase
	Rv3521[e]	CHP
ltp4	*Rv3522*[d]	probable ketoacyl-CoA thiolase
	Rv3524[e]	probable conserved membrane protein
	Rv3525c[e]	possible siderophore binding protein
	Rv3527[a]	hypothetical protein (HP)
	Rv3528c[e]	HP
	Rv3529c[e]	CHP
	Rv3530c[e]	possible oxidoreductase
	Rv3531c[c]	hypothetical protein
PPE61	*Rv3532*[e]	PPE family
PPE62	*Rv3533c*[e]	PPE family
hsaG	*Rv3535c*[d]	probable aldehyde dehydrogenase
hsd4B	*Rv3538*[d]	probable enoyl-CoA hydratase
PPE63	*Rv3539*[e]	PE
	Rv3541c[a]	CHP/putative enoyl-CoA hydratase
fadE29	*Rv3543c*[c]	probable acyl-CoA dehydrogenase
	Rv3547[e]	CHP
	Rv3548c[c]	probable short chain dehydrogenase/reductase
	Rv3549c[c]	probable short chain dehydrogenase/reductase
echA20	*Rv3550*[e]	possible enoyl-CoA hydratase
ipdB	*Rv3552*[a]	ATP-dependent CoA transferase β subunit
	Rv3553[c]	possible oxidoreductase/2-nitropropane dioxygenase
fdxB	*Rv3554*[e]	possible electron transfer protein/ferredoxin
	Rv3555c[e]	CHP
kstR2	*Rv3557c*[e]	Tet-R transcriptional regulator (repressor)
PPE64	*Rv3558*[e]	PPE
	Rv3559c[c]	probable oxidoreductase
fadD3	*Rv3561*[c]	acyl-CoA synthetase (AMP forming)
fadE31	*Rv3562*[e]	probable acyl-CoA dehydrogenase
fadE33	*Rv3564*[c]	probable acyl-CoA dehydrogenase
aspB	*Rv3565*[e]	possible aspartate aminotransferase
	Rv3566A[e]	CHP
nhoA / nat	*Rv3566c*[e]	arylamine N-acetyltransferase
hsaB	*Rv3567c*[d]	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione hydroxylase (3-HSA hydroxylase, reductase)
	Rv3572[c]	HP

Gene	Location	Enzyme
fadE34	*Rv3573c*[c]	probable acyl-CoA dehydrogenase
kstR	*Rv3574*[a]	Tet-R transcriptional regulator (repressor)
	Rv3575c[c]	transcriptional regulatory protein LacI-family
	Rv3779[c]	transmembrane protein alanine and leucine rich
papA2	*Rv3820c*[c]	polyketide synthase associated protein PapA2
papA1	*Rv3824c*[c]	polyketide synthase associated protein
pks2	*Rv3825c*[c]	polyketide synthase PKS2
sigM	*Rv3911*[c]	RNA polymerase sigma factor SigM

Notes:

[a] Genes proven to be essential for survival in macrophage cells and in murine infection.

[b] Genes predicted to be essential for survival in macrophage cells and in murine infection.

[c] Genes predicted to be specifically required for growth on cholesterol.

[d] Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway.

[e] Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data.

Abbreviations: 3-HSA = 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 3,4-DHSA = 3,4-dihydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione; 3β-HSD = 3β-hydroxysteroid dehydrogenase; 4,9-DHSA hydrolase = 4,5-9,10-diseco-3-hydroxy-5,9,17-trioxoandrosta-1(10),2-dien-4-oic acid; 17β-HSD = 17β-hydroxysteroid dehydrogenase; Δ1-KSTD = 3-ketosteroid-Δ1-dehydrogenase; ABC = ATP-binding cassette; ADH = alcohol dehydrogenase; AMP = adenosine monophosphate; AP = apurinic/apyrimidinic; ATP = adenosine triphosphate; Bcp = bacterioferritin comigratory protein; CHP = conserved hypothetical protein; CoA = co-enzyme A; DNA = deoxyribonucleic acid; HP = hypothetical protein; LldD = L-lactate dehydrogenase; MCE = mammalian cell entry; MgtE = Mg2+ transport transmembrane protein; MmpL = *Mycobacterium* membrane protein laboratory; NAD = nicotinamide adenine dinucleotide; PE = protein family with highly conserved Proline-Glutamate residues near the start of their encoded proteins; PGRS = polymorphic GC-rich-repetitive sequence; pks = polyketide synthase; PPE = protein family with highly conserved Proline-Proline-Glutamate; PQQ = pyrrolo-quinoline quinone; PTP/PTPase = phosphotyrosine protein phosphatase/protein-tyrosine-phosphatase; RNA = ribonucleic acid; TetR/TETR = tetracycline repressor.

2.4. Conclusion

In this chapter the cholesterol degradation pathway was annotated, and includes different enzymatic reactions and genes/proteins involved at each of the steps.. The genes/proteins identified in literature to be possibly, probably or definitely involved in the breakdown of cholesterol in *M. tuberculosis* H37Rv were then classified. One hundred and fifty-two genes/proteins that are possibly, probably or definitely involved in the cholesterol degradation pathway were selected for determining the mycobacterial species' ability to degrade/utilise cholesterol.

References

- 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.
- García-Fernández, E., Frank, D.J., Galán, B., Kells, P.M., Podust, L.M., García, J.L. and Ortiz de Montellano, P.R., 2013. A highly conserved mycobacterial cholesterol catabolic pathway. *Environmental microbiology*, *15*(8), pp. 2342-2359.
- Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS pathog*, *7*(9), p. e1002251.
- 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.
- Mills, N., 2006. ChemDraw Ultra 10.0 CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. www. cambridgesoft. com. Commercial Price: 1910fordownload, 2150 for CD-ROM; Academic Price: 710fordownload, 800 for CD-ROM.
- Mukhopadhyay, S., Nair, S. and Ghosh, S., 2012. Pathogenesis in tuberculosis: transcriptomic approaches to unraveling virulence mechanisms and finding new drug targets. *FEMS microbiology reviews*, *36*(2), pp. 463-485.

- Nesbitt, N.M., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N.S. and Dubnau, E., 2010. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infection and immunity*, *78*(1), pp. 275-282.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2011. Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends in microbiology*, *19*(11), pp. 530-539.
- Świzdor, A., Panek, A., Milecka-Tronina, N. and Kołek, T., 2012. Biotransformations utilizing β-oxidation cycle reactions in the synthesis of natural compounds and medicines. *International journal of molecular sciences*, *13*(12), pp. 16514-16543.
- Thomas, S.T., VanderVen, B.C., Sherman, D.R., Russell, D.G. and Sampson, N.S., 2011. Pathway profiling in *Mycobacterium tuberculosis* elucidation of cholesterol-dercived catabolite and enzymes that catalyze its metabolism. *Journal of biological hemistry*, *286*(51), pp. 43668-43678.
- Van der Geize, R., Grommen, A.W.F., Hessels, G.I., Jacobs, A.A.C. and Dijkhuizen, L., 2011. The steroid catabolic pathway of the intracellular pathogen Rhodococcus equi is important for pathogenesis and a target for vaccine development. *PLoS Pathog*, *7*(8), p. e1002181.
- Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W. and Eltis, L.D., 2007. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences*, *104*(6), pp. 1947-1952.

- Yang, X., Gao, J., Smith, I., Dubnau, E. and Sampson, N.S., 2011. Cholesterol is not an essential source of nutrition for *Mycobacterium tuberculosis* during infection. *Journal of bacteriology*, *193*(6), pp. 1473-1476.

CHAPTER 3

IN SILICO DETERMATION OF THE CHOLESTEROL DEGRADATION ABILITY OF MYCOBACTERIAL SPECIES

3.1. Introduction

Genes/proteins involved in cholesterol degradation are now considered as novel drug targets in the fight against the deadliest human pathogen, *M. tuberculosis* (Ouellet et al., 2011). However, to date, a study analysing the cholesterol degradation capability of different species belonging to the genus *Mycobacterium* has not been reported. Furthermore, performing wet-laboratory experiments on a large scale with different mycobacterial species are quite laborious and time- and money-consuming.

Considering the above facts and taking advantage of the genome sequencing of quite a number of mycobacterial species, this study is aimed at performing comprehensive *in silico* comparative analysis of the genes/proteins involved in cholesterol degradation of the genus *Mycobacterium*. Study results are aimed at identifying mycobacterial species capable of using cholesterol as a carbon source.

3.2. Methodology

3.2.1. Species selected for the study

Ninety-three mycobacterial species belonging to six different categories were used in this study (Table 3.1.). The six categories include *Mycobacterium tuberculosis* complex (MTBC) (39 species), *Mycobacterium chelonae-abscessus* complex (MCAC) (10 species), *Mycobacterium avium* complex (MAC) (15 species), Mycobacteria causing leprosy (MCL) (2 species), nontuberculous mycobacteria (NTM) (8 species) and Saprophytes (SAP) (19 species). The criteria for separation of the mycobacterial species into six different groups are based on their characteristic features, including ecological niches, as well as the nature and site of infection as described elsewhere (Ventura et al., 2007; Parvez et al., 2016). Furthermore, taxonomical grouping of mycobacterial species is taken into consideration as described elsewhere (Tortoli, 2012). Detailed information on species, their categories, and genome database links are listed in Table 3.1.

Table 3.1. List of mycobacterial species used in the study. As listed in the table, 93 mycobacterial species were grouped under six different categories based on their characteristic features, including ecological niches, nature of infection and site of infection as described elsewhere (Ventura et al., 2007; Parvez et al., 2016). Taxonomical grouping of the *Mycobacteria* is taken also into consideration, as described elsewhere (Tortoli, 2012). Furthermore, KEGG genome database (Kanehisa et al., 2017) links for each of the species are listed in the table. For some species references were not available despite the genome database being available for public use at KEGG and thus references were not cited.

Species Name	Organism Code	Database Link	Reference
***Mycobacterium tuberculosis* Complex (MTBC)**			
Mycobacterium tuberculosis H37Rv	mtu	http://www.genome.jp/kegg-bin/show_organism?org=mtu	Cole *et al*., 1998
Mycobacterium tuberculosis H37Rv	mtv	http://www.genome.jp/kegg-bin/show_organism?org=mtv	
Mycobacterium tuberculosis CDC1551	mtc	http://www.genome.jp/kegg-bin/show_organism?org=mtc	Fleischmann *et al*., 2002
Mycobacterium tuberculosis H37Ra	mra	http://www.genome.jp/kegg-bin/show_organism?org=mra	Zheng *et al*., 2008
Mycobacterium tuberculosis F11	mtf	http://www.genome.jp/kegg-bin/show_organism?org=mtf	
Mycobacterium tuberculosis KZN 1435	mtb	http://www.genome.jp/kegg-bin/show_organism?org=mtb	
Mycobacterium tuberculosis KZN 4207	mtk	http://www.genome.jp/kegg-bin/show_organism?org=mtk	
Mycobacterium tuberculosis KZN 605	mtz	http://www.genome.jp/kegg-bin/show_organism?org=mtz	
Mycobacterium tuberculosis RGTB327	mtg	http://www.genome.jp/kegg-bin/show_organism?org=mtg	Madhavilatha *et al*., 2012
Mycobacterium tuberculosis RGTB423	mti	http://www.genome.jp/kegg-bin/show_organism?org=mti	Madhavilatha *et al*., 2012
Mycobacterium tuberculosis CCDC5079	mte	http://www.genome.jp/kegg-bin/show_organism?org=mte	Zhang Y *et al*., 2011
Mycobacterium tuberculosis CCDC5079	mtur	http://www.genome.jp/kegg-bin/show_organism?org=mtur	Tang *et al*., 2013
Mycobacterium tuberculosis CCDC5180	mtl	http://www.genome.jp/kegg-bin/show_organism?org=mtl	Zhang Y *et al*., 2011
Mycobacterium tuberculosis CTRI-2	mto	http://www.genome.jp/kegg-bin/show_organism?org=mto	Ilina *et al*., 2013
Mycobacterium tuberculosis UT205	mtd	http://www.genome.jp/kegg-bin/show_organism?org=mtd	Isaza *et al*., 2012
Mycobacterium tuberculosis Erdman = ATCC 35801	mtn	http://www.genome.jp/kegg-bin/show_organism?org=mtn	Miyoshi-Akiyama *et al*., 2012
Mycobacterium tuberculosis Beijing/NITR203	mtj	http://www.genome.jp/kegg-bin/show_organism?org=mtj	Narayanan *et al*., 2013
Mycobacterium tuberculosis 7199-99	mtub	http://www.genome.jp/kegg-bin/show_organism?org=mtub	Roetzer *et al*., 2013

Species Name	Organism Code	Database Link	Reference
Mycobacterium tuberculosis CAS/NITR204	mtuc	http://www.genome.jp/kegg-bin/show_organism?org=mtuc	Narayanan *et al*., 2013
Mycobacterium tuberculosis EAI5/NITR206	mtue	http://www.genome.jp/kegg-bin/show_organism?org=mtue	Narayanan *et al*., 2013
Mycobacterium tuberculosis EAI5	mtx	http://www.genome.jp/kegg-bin/show_organism?org=mtx	Al Rashdi *et al*., 2014
Mycobacterium tuberculosis Haarlem/NITR202	mtuh	http://www.genome.jp/kegg-bin/show_organism?org=mtuh	Narayanan *et al*., 2013
Mycobacterium tuberculosis Haarlem	mtul	http://www.genome.jp/kegg-bin/show_organism?org=mtul	
Mycobacterium tuberculosis BT1	mtut	http://www.genome.jp/kegg-bin/show_organism?org=mtut	
Mycobacterium tuberculosis BT2	mtuu	http://www.genome.jp/kegg-bin/show_organism?org=mtuu	
Mycobacterium tuberculosis HKBS1	mtq	http://www.genome.jp/kegg-bin/show_organism?org=mtq	
Mycobacterium bovis AF2122/97	mbo	http://www.genome.jp/kegg-bin/show_organism?org=mbo	Garnier *et al*., 2003
Mycobacterium bovis BCG Pasteur 1173P2	mbb	http://www.genome.jp/kegg-bin/show_organism?org=mbb	Brosch *et al*., 2007
Mycobacterium bovis BCG Tokyo 172	mbt	http://www.genome.jp/kegg-bin/show_organism?org=mbt	Seki *et al*., 2009
Mycobacterium bovis BCG Mexico	mbm	http://www.genome.jp/kegg-bin/show_organism?org=mbm	Orduna *et al*., 2011
Mycobacterium bovis BCG Korea 1168P	mbk	http://www.genome.jp/kegg-bin/show_organism?org=mbk	Joung *et al*., 2013
Mycobacterium bovis BCG ATCC 35743	mbx	http://www.genome.jp/kegg-bin/show_organism?org=mbx	Pan *et al*., 2011
Mycobacterium bovis ATCC BAA-935	mbz	http://www.genome.jp/kegg-bin/show_organism?org=mbz	
Mycobacterium africanum	maf	http://www.genome.jp/kegg-bin/show_organism?org=maf	Bentley *et al*., 2012
Mycobacterium canettii CIPT 140010059	mce	http://www.genome.jp/kegg-bin/show_organism?org=mce	Bentley *et al*., 2012
Mycobacterium canettii CIPT 140060008	mcq	http://www.genome.jp/kegg-bin/show_organism?org=mcq	Supply *et al*., 2013
Mycobacterium canettii CIPT 140070008	mcv	http://www.genome.jp/kegg-bin/show_organism?org=mcv	Supply *et al*., 2013
Mycobacterium canettii CIPT 140070010	mcx	http://www.genome.jp/kegg-bin/show_organism?org=mcx	Supply *et al*., 2013
Mycobacterium canettii CIPT 140070017	mcz	http://www.genome.jp/kegg-bin/show_organism?org=mcz	Supply *et al*., 2013
Mycobacteria Causing Leprosy (MCL)			
Mycobacterium leprae TN	mle	http://www.genome.jp/kegg-bin/show_organism?org=mle	Cole *et al*., 2001
Mycobacterium leprae Br4923	mlb	http://www.genome.jp/kegg-bin/show_organism?org=mlb	Monot *et al*., 2009
***Mycobacterium avium* Complex (MAC)**			
Mycobacterium avium subsp. *paratuberculosis* K-10	mpa	http://www.genome.jp/kegg-bin/show_organism?org=mpa	Li *et al*., 2005
Mycobacterium avium subsp. *paratuberculosis* MAP4	mao	http://www.genome.jp/kegg-bin/show_organism?org=mao	Bannantine *et al*., 2014
Mycobacterium avium subsp. *paratuberculosis* E1	mavi	http://www.genome.jp/kegg-bin/show_organism?org=mavi	Amin *et al*., 2015
Mycobacterium avium subsp. *paratuberculosis* E93	mavu	http://www.genome.jp/kegg-bin/show_organism?org=mavu	Amin *et al*., 2015

Species Name	Organism Code	Database Link	Reference
Mycobacterium avium 104	mav	http://www.genome.jp/kegg-bin/show_organism?org=mav	
Mycobacterium avium subsp. *avium* DJO-44271	mavd	http://www.genome.jp/kegg-bin/show_organism?org=mavd	
Mycobacterium avium subsp. *avium* 2285 (R)	mavr	http://www.genome.jp/kegg-bin/show_organism?org=mavr	
Mycobacterium avium subsp. *avium* 2285 (S)	mava	http://www.genome.jp/kegg-bin/show_organism?org=mava	
Mycobacterium intracellulare MOTT-02	mit	http://www.genome.jp/kegg-bin/show_organism?org=mit	Kim *et al.*, 2012 (a)
Mycobacterium intracellulare MOTT-64	mir	http://www.genome.jp/kegg-bin/show_organism?org=mir	Kim *et al.*, 2012 (b)
Mycobacterium intracellulare ATCC 13950	mia	http://www.genome.jp/kegg-bin/show_organism?org=mia	Kim *et al.*, 2012 (c)
Mycobacterium intracellulare 1956	mie	http://www.genome.jp/kegg-bin/show_organism?org=mie	
Mycobacterium indicus pranii	mid	http://www.genome.jp/kegg-bin/show_organism?org=mid	Saini *et al.*, 2012
Mycobacterium yongonense	myo	http://www.genome.jp/kegg-bin/show_organism?org=myo	Kim *et al.*, 2013 (e)
Mycobacterium sp. MOTT36Y	mmm	http://www.genome.jp/kegg-bin/show_organism?org=mmm	Kim *et al.*, 2012 (d)
Saprophytes (SAP)			
Mycobacterium smegmatis MC2 155	msm	http://www.genome.jp/kegg-bin/show_organism?org=msm	
Mycobacterium smegmatis MC2 155	msg	http://www.genome.jp/kegg-bin/show_organism?org=msg	Gallien *et al.*, 2009
Mycobacterium smegmatis MC2 155	msb	http://www.genome.jp/kegg-bin/show_organism?org=msb	Mohan *et al.*, 2015
Mycobacterium smegmatis INHR1	msn	http://www.genome.jp/kegg-bin/show_organism?org=msn	Mohan *et al.*, 2015
Mycobacterium smegmatis INHR2	msh	http://www.genome.jp/kegg-bin/show_organism?org=msh	Mohan *et al.*, 2015
Mycobacterium sp. JS623	msa	http://www.genome.jp/kegg-bin/show_organism?org=msa	
Mycobacterium vanbaalenii	mva	http://www.genome.jp/kegg-bin/show_organism?org=mva	
Mycobacterium gilvum PYR-GCK	mgi	http://www.genome.jp/kegg-bin/show_organism?org=mgi	
Mycobacterium gilvum Spyr1	msp	http://www.genome.jp/kegg-bin/show_organism?org=msp	Kallimanis *et al.*, 2011
Mycobacterium sp. MCS	mmc	http://www.genome.jp/kegg-bin/show_organism?org=mmc	
Mycobacterium sp. KMS	mkm	http://www.genome.jp/kegg-bin/show_organism?org=mkm	
Mycobacterium sp. JLS	mjl	http://www.genome.jp/kegg-bin/show_organism?org=mjl	
Mycobacterium rhodesiae	mrh	http://www.genome.jp/kegg-bin/show_organism?org=mrh	
Mycobacterium chubuense	mcb	http://www.genome.jp/kegg-bin/show_organism?org=mcb	
Mycobacterium neoaurum	mne	http://www.genome.jp/kegg-bin/show_organism?org=mne	Bragin *et al.*, 2013
Mycobacterium sp. VKM Ac-1817D	myv	http://www.genome.jp/kegg-bin/show_organism?org=myv	Bragin *et al.*, 2013
Mycobacterium sp. EPa45	mye	http://www.genome.jp/kegg-bin/show_organism?org=mye	Kato *et al.*, 2015

Species Name	Organism Code	Database Link	Reference
Mycobacterium goodii	mgo	http://www.genome.jp/kegg-bin/show_organism?org=mgo	Yu *et al*., 2015
Mycobacterium fortuitum	mft	http://www.genome.jp/kegg-bin/show_organism?org=mft	Costa *et al*., 2015
Non-tuberculosis Mycobacteria (NTM)			
Mycobacterium ulcerans	mul	http://www.genome.jp/kegg-bin/show_organism?org=mul	Stinear *et al*., 2007
Mycobacterium sinense	mjd	http://www.genome.jp/kegg-bin/show_organism?org=mjd	Zhang ZY *et al*., 2011
Mycobacterium marinum	mmi	http://www.genome.jp/kegg-bin/show_organism?org=mmi	Stinear *et al*., 2008
Mycobacterium liflandii	mli	http://www.genome.jp/kegg-bin/show_organism?org=mli	Tobias *et al*., 2013
Mycobacterium kansasii ATCC 12478	mkn	http://www.genome.jp/kegg-bin/show_organism?org=mkn	
Mycobacterium kansasii 662	mks	http://www.genome.jp/kegg-bin/show_organism?org=mks	
Mycobacterium kansasii 824	mki	http://www.genome.jp/kegg-bin/show_organism?org=mki	
Mycobacterium haemophilum	mhad	http://www.genome.jp/kegg-bin/show_organism?org=mhad	Tufariello *et al*., 2015
***Mycobacterium chelonae-abscessus* Complex (MCAC)**			
Mycobacterium abscessus ATCC 19977	mab	http://www.genome.jp/kegg-bin/show_organism?org=mab	Ripoll *et al*., 2009
Mycobacterium abscessus subsp. *bolletii* 50594	mabb	http://www.genome.jp/kegg-bin/show_organism?org=mabb	Kim *et al*., 2013 (f)
Mycobacterium abscessus subsp. *bolletii* GO 06	mmv	http://www.genome.jp/kegg-bin/show_organism?org=mmv	Raiol *et al*., 2012
Mycobacterium abscessus subsp. *bolletii* MA 1948	may	http://www.genome.jp/kegg-bin/show_organism?org=may	
Mycobacterium abscessus subsp. *bolletii* MC1518	mabo	http://www.genome.jp/kegg-bin/show_organism?org=mabo	
Mycobacterium abscessus subsp. *bolletii* CCUG 48898 = JCM 15300	mabl	http://www.genome.jp/kegg-bin/show_organism?org=mabl	Sekizuka *et al*., 2014
Mycobacterium abscessus subsp. *bolletii* 103	maz	http://www.genome.jp/kegg-bin/show_organism?org=maz	
Mycobacterium abscessus subsp. *abscessus*	mak	http://www.genome.jp/kegg-bin/show_organism?org=mak	
Mycobacterium abscessus DJO-44274	mys	http://www.genome.jp/kegg-bin/show_organism?org=mys	
Mycobacterium abscessus 4529	myc	http://www.genome.jp/kegg-bin/show_organism?org=myc	

3.2.2. Protein domain/function analysis

In Chapter 2, 152 genes/proteins were identified to be probably, possiblty or definitely involved in the cholesterol degradation pathway of *M. tuberculosis* H37Rv (Table 2.6 in Chapter 2). It is necessary to identify the presence or absence of the above 152 genes/proteins in different mycobacterial species in order to determine if these mycobacterial species have the capability to degrade cholesterol as a sole carbon and energy source. The following sections explain the procedures followed for protein domain/function analysis.

3.2.2.1. Protein sequences

The selected 152 proteins' (Table 2.6 in Chapter 2) sequences were retrieved from the KEGG database using the respective gene codes.

3.2.2.2. BLAST analysis

The protein sequences of 152 *M. tuberculosis* proteins were copied and pasted into the Basic Local Alignment Search Tool (BLAST) in the KEGG database (http://www.genome.jp/tools/blast/) for BLAST. The amino acid sequence was entered in the "sequence data" field, then "favorite organism code or category" was selected under the "KEGG GENES" button, "Mycobacterium" was entered in the free text field provided and the "Compute" link at the top clicked. Once the BLAST was complete, the "show all result" link at the bottom of the output data was selected. The resulting output was then copied and pasted into an Excel program (see description below) to extract the required data (organism code, enzyme code, enzyme name, identity and homology (positives)) from all of the BLAST output data, which was then tabulated under each organism's name and code (Supplementary dataset 1).

3.2.2.3. Excel program for extracting KEGG BLAST data

To extract the required data from the BLAST output data obtained from the KEGG database, an Excel program written in an Excel worksheet was used. The generated program is presented below:

Sheet/tab 1: Original Data

Cell A2-A80 000:

Copy the output data from KEGG database into cell A1 to A80 000 (depending on the size of the data).

Cell B2-B80 000:

=IF(ISNUMBER(SEARCH(">",A2)),"1",IF(ISNUMBER(SEARCH("Identities",A2)),"2",IF(ISNUMBER(SEARCH("Length",A2)),"3","N")))

Determine if the character ">", the word "Identities" or "Length" is present in the data in column A.

For ">", return the number "1" (a "1" will populate the next column).

For "Identities", return the number "2" (a "2" will populate the next column).

For "Length", return the number "3" (a "3" will populate the next column).

Else, if none of the above is present, return the letter "N" (an "N" will populate the next column).

Cell C2-C80 000:

=IF(B2="1",IF(B3="3",1,IF(B4="3",2,IF(B5="3",3,IF(B6="3",4,"many")))),"")

Determine the number of rows the specific record consists of, up to a maximum of four lines. If the record is longer than four lines, the word "many" will be displayed, indicating that manual investigation and entry are required.

Cell D2-D80 000:

=IF(B2="1", A2, "")

If valid data to be filtered contains “>” in Column A, then display the data in this row.

Cell E2:E80 000:

=IF(B2="2", A2, "")

If valid data to be filtered contains “Identifies” in Column A, then display the data in this row.

Cell F2-F80 000:

=IF(ISNUMBER(SEARCH(">",D2)),LEFT(D2, LEN(D2)-5),D2)

If this row in Column D includes the character “>”, then it displays the first characters, less 5 than the length of Column D.

Cell G2-G80 000:

=IF(C2=1,F2,IF(C2=2,(F2&""&A3),IF(C2=3,(F2&""&A3&" "&A4),IF(C2=4,(F2&""&A3&" "&A4&" "&A5),""))))

Combines the data into one cell to include all separate rows with valid data.

Cell H2 - O80 000:

Find the position of the first “:”, first “ “, total length, first “(“, first “%”, second “(“, second “%” to determine the “Org Code”, “Code”, “Name”, “Identity” and “Positive” for *Final Data Blanks* sheet/tab (Columns O to S).

Sheet/tab 2: Final Data Blanks

“Unformatted” column is populated if valid data is present in the corresponding cell in the Original Data sheet/tab.

“Org Code” obtained from Original Data sheet: Column O.

“Code” obtained from Original Data: Column P.

“Name” obtained from Original Data: Column Q and multiple spaces filtered.

“Identity” obtained from Original Data: Column R.

“Positives” obtained from Original Data: Column S.

Sheet/tab 3: Final Data

The data in columns B to F in the "Final Data Blanks" sheet/tab is copied and pasted into the "Final Data" sheet/tab and sorted alphabetically according to the first column (Org Code) – this just makes the copying into the BLAST sheet easier. The data was also copied and transpose pasted to display it horizontally to make copying easier.

Validation

Before using this program, validation was performed with three different datasets and continuous checks were carried out throughout the use of the program.

3.2.2.4. Data collection and domain/function analysis

All the top hit protein sequences in 93 mycobacterial species were collected (Supplementary dataset 2) and submitted to NCBI CDD database (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Based on the NCBI CDD results, proteins belonging to the same family/superfamily were identified (Supplementary dataset 3). For some proteins no results were obtained at NCBI CDD. Thus, the KEGG database was searched for possible functions or domains to determine whether they belonged to the same group (Supplementary dataset 1).

3.2.3. Data-processing methodology

The superfamilies as per the NCBI CDD output were considered to determine whether the genes/proteins from the 92 mycobacterial species were a match to those from *M. tuberculosis* H37Rv. Where no data on superfamilies was available in the NCBI database, a secondary review was performed of the KEGG BLAST output data, by looking at the percentage identity, percentage homology and the name (thus also the function) of each of the genes/proteins. However, the presence or absence of some proteins in different mycobacterial species was determined based on the information below:

The Rv3512 gene/protein homolog was not identified in many species in the KEGG BLAST output. This may be due to annotation errors, as *M. tuberculosis* H37Rv (1998) (*Mycobacterium* code mtu) and *M. tuberculosis* H37Rv (2012) (*Mycobacterium* code mtv) yielded different results. Furthermore, this gene is not proven to be essential for cholesterol degradation. Thus, this gene was omitted from the analysis.

For Rv1906, more than 40% identity to *M. tuberculosis* H37Rv was taken as positive across all the categories, as the proteins are hypothetical. From this, the negative species are *Mycobacterium abscessus* ATCC 19977 (mab); *Mycobacterium abscessus* subsp. *bolletii* 50594 (mabb), *Mycobacterium abscessus* subsp. *bolletii* GO 06 (mmv), *Mycobacterium abscessus* subsp. *bolletii* MA 1948 (may), *Mycobacterium abscessus* subsp. *bolletii* MC1518 (mabo), *Mycobacterium abscessus* subsp. *bolletii* CCUG 48898 = JCM 15300 (mabl), *Mycobacterium abscessus* subsp. *bolletii* 103 (maz), *Mycobacterium abscessus* subsp. *abscessus* MM1513 (mak), *Mycobacterium abscessus* DJO-44274 (mys), and *Mycobacterium abscessus* 4529 (myc).

For Rv3566A, more than 40% identity to *M. tuberculosis* H37Rv was taken as positive across all the categories as the proteins are hypothetical.

For Rv3572, more than 40% identity to *M. tuberculosis* H37Rv was taken as positive across all the categories as the proteins are hypothetical.

For Rv3527, more than 40% identity to *M. tuberculosis* H37Rv was taken as positive across all the categories as the proteins are hypothetical.

The results were tabulated per complex by colour-coding the cells according to the following criteria: green = gene homolog present; black = gene homolog absent.

The following section provides a detailed discussion of the results.

3.3. Results and Discussion

3.3.1. Determining the cholesterol-degrading ability of 39 MTBC species

Detailed analysis of 151 gene homologs across 39 MTBC species revealed the absence of the following cholesterol-degrading genes (Tables 3.2 and 3.3):

Table 3.2. Analysis of 151 gene homologs across 39 MTBC species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code / Gene/Protein	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Rv0009																																							
Rv0153c										■																													
Rv0202c																																							
Rv0244c																																							
Rv0362																																							
Rv0391																																							
Rv0450c																																							
Rv0485										■									■			■																	
Rv0495c																																							
Rv0655																																							
Rv0693																																							
Rv0694																																							
Rv0695																			■																				
Rv0696																																							
Rv0761c																																							
Rv0805										■	■		■			■																■							
Rv0876c										■												■																	
Rv1071c																																							
Rv1084									■										■			■																	
Rv1096																						■																	
Rv1129c																						■																	
Rv1130																			■																■				

Organism Code / Gene/Protein	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Rv1131																																							
Rv1183																																							
Rv1193																																							
Rv1428c																																							
Rv1432																			■																				
Rv1608c																																							
Rv1626																																							
Rv1627c																																							
Rv1798																																							
Rv1906c																																							
Rv1919c											■		■																										
Rv1963c																																							
Rv2048c																																							
Rv2118c																																							
Rv2206																																■							
Rv2239c																																							
Rv2416c										■									■			■																	
Rv2462c																																							
Rv2506																																							
Rv2668																																							
Rv2681										■									■	■																			
Rv2684																																							
Rv2710																																							
Rv2799									■																														
Rv2914c																																							

Organism Code Gene/Protein	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Rv2985																																							
Rv3050c																																							
Rv3274c																																							
Rv3409c																																							
Rv3419c																																							
Rv3421c																																							
Rv3492c																																							
Rv3493c																																							
Rv3494c																																							
Rv3495c																																							
Rv3496c																																							
Rv3497c																																							
Rv3498c																																							
Rv3499c																																							
Rv3500c																																							
Rv3501c																																							
Rv3502c																																							
Rv3504																																							
Rv3505																																							
Rv3506																																							
Rv3515c																																							
Rv3516																																							
Rv3522																																							
Rv3523																																							
Rv3526																																							

Organism Code	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Gene/Protein																																							
Rv3531c																																							
Rv3534c																																							
Rv3535c																																							
Rv3536c																																							
Rv3537																																							
Rv3538																																							
Rv3540c																																							
Rv3542c																																							
Rv3543c																																							
Rv3544c																																							
Rv3545c																																							
Rv3546																																							
Rv3548c																																							
Rv3549c																																							
Rv3551																																							
Rv3553																																							
Rv3559c																																							
Rv3560c																																							
Rv3561																																							
Rv3563																																							
Rv3564																																							
Rv3567c																																							
Rv3568c																																							
Rv3569c																																							
Rv3570c																																							

Organism Code	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Gene/Protein																																							
Rv3571																																							
Rv3572																																							
Rv3573c																																							
Rv3575c																																							
Rv3779																																							
Rv3820c																																							
Rv3824c																																							
Rv3825c																																							
Rv3911																																							
Rv1106c																																							
Rv3503c																																							
Rv3507																																							
Rv3508																																							
Rv3509c																																							
Rv3510c																																							
Rv3511																																							
Rv3513c																																							
Rv3514																																							
Rv3517																																							
Rv3518c																																							
Rv3519																																							
Rv3520c																																							
Rv3521																																							
Rv3524																																							
Rv3525c																																							

Organism Code Gene/Protein	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Rv3527																																							
Rv3528c																																		■					■
Rv3529c																																							
Rv3530c																																							
Rv3532																																							
Rv3533c																																							
Rv3539																																							
Rv3541c																																							
Rv3547																																							
Rv3550																																							
Rv3552																																							
Rv3554																																							
Rv3555c			■																																				
Rv3556c																																							
Rv3557c																																							
Rv3558																																							
Rv3562																																							
Rv3565																																							
Rv3566A					■	■	■	■			■		■			■	■		■	■			■								■	■						■	■
Rv3574																																							
Rv3566c																																■							
Rv2590																																							
Rv0099																																							
Rv1346																																							
Rv3061c																																							

Organism Code Gene/Protein	mtu	mtv	mtc	mra	mtf	mtb	mtk	mtz	mtg	mti	mte	mtur	mtl	mto	mtd	mtn	mtj	mtub	mtuc	mtue	mtx	mtuh	mtul	mtut	mtuu	mtq	mbo	mbb	mbt	mbm	mbk	mbx	mbz	maf	mce	mcq	mcv	mcx	mcz
Rv3140																																							
Rv3139																																							
Rv0468																																							
Rv1715																																							

Table 3.3. List of cholesterol-degrading homologs that are not found in MTBC species.

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv0153c	mti	Genes predicted to be specifically required for growth on cholesterol
Rv0485	mti; mtuc; mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv0695	mtuc	Genes predicted to be specifically required for growth on cholesterol
Rv0805	mti; mte; mtl; mtn; mbx	Genes predicted to be specifically required for growth on cholesterol
Rv0876c	mti; mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv1084	mtg; mtuc; mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv1096	mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv1129c	mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv1130	mtuc; mce	Genes predicted to be specifically required for growth on cholesterol
Rv1432	mtuc	Genes predicted to be specifically required for growth on cholesterol
Rv1919c	mte; mtl	Genes predicted to be specifically required for growth on cholesterol
Rv2206	mbx	Genes predicted to be specifically required for growth on cholesterol
Rv2416c	mti; mtuc; mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv2681	mti; mtuc; mtue	Genes predicted to be specifically required for growth on cholesterol
Rv2799	mtg	Genes predicted to be specifically required for growth on cholesterol
Rv3526	mti	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3531c	mti; mtuh	Genes predicted to be specifically required for growth on cholesterol
Rv3536c	mtuc	Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3779	mtuc	Genes predicted to be specifically required for growth on cholesterol
Rv3517	mcz	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3521	mtuc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv3528c	maf; mcz	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3555c	mtc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566A	mtf; mtb; mtk; mtz; mte; mtl; mtn; mtj; mtuc; mtue; mtul; mbk; mbx; mcx; mcz	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566c	mbx	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

Note: (i) For Rv0495c, homolog proteins were identified based on percentage identity as the NCBI CDD database did not assign proteins to a particular superfamily. The percentage identity was resourced from KEGG and it ranged from 100% to 99%. (ii) For Rv0805, homolog proteins in mti (*Mycobacterium tuberculosis* RGTB423) and mbx (*Mycobacterium bovis* BCG ATCC 35743) were not identified, as NCBI CDD did not yield any results. Furthermore, the KEGG database showed only 49% identity compared to other species' homolog proteins that showed 100% identity. Based on this, it was concluded that mti and mbx do not have Rv0805 homolog(s). (iii) For Rv1432 there was no hit data for mtuc (*Mycobacterium tuberculosis* CAS/NITR204) and KEGG data revealed a different dehydrogenase hit. Thus, it was concluded that the homolog is absent. (iv) Upon review of Rv2416c, it was found that the homolog protein sequence for mtuh (*Mycobacterium tuberculosis* Haarlem/NITR202) is truncated and present as 28 amino acids, compared to the other species' homologs with more than 360 amino acids, and therefore it was judged to be absent.

3.3.2. Determining the cholesterol-degrading ability of 10 MCAC species

Detailed analysis of 151 gene homologs across 10 MCAC species revealed the absence of the following cholesterol-degrading genes (Tables 3.4 and 3.5):

Table 3.4. Analysis of 151 gene homologs across 10 MCAC species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code Gene/Protein	mab	mabb	mmv	may	mabo	mabl	maz	mak	mys	myc
Rv0009	green	green	green	green	green	green	green	green	green	green
Rv0153c	green	green	green	green	green	green	green	green	green	green
Rv0202c	green	green	green	green	green	green	green	green	green	green
Rv0244c	green	green	green	green	green	green	green	green	green	green
Rv0362	green	green	green	green	green	green	green	green	green	green
Rv0391	green	green	green	green	green	green	green	green	green	green
Rv0450c	green	green	green	green	green	green	green	green	green	green
Rv0485	green	green	green	green	green	green	green	green	green	green
Rv0495c	green	green	green	green	green	green	green	green	green	green
Rv0655	green	green	green	green	green	green	green	green	green	green
Rv0693	green	green	green	green	green	green	green	green	green	green
Rv0694	green	green	green	green	green	green	green	green	green	green
Rv0695	green	green	green	green	green	green	green	green	green	green
Rv0696	green	green	green	green	green	green	green	green	green	green
Rv0761c	green	green	green	green	green	green	green	green	green	green
Rv0805	green	green	green	green	green	green	green	green	green	green
Rv0876c	black	black	black	green	green	black	green	green	green	green
Rv1071c	green	green	green	green	green	green	green	green	green	green
Rv1084	green	green	green	green	green	green	green	green	green	green
Rv1096	green	green	green	green	green	green	green	green	green	green
Rv1129c	green	green	green	green	green	green	green	green	green	green
Rv1130	green	green	green	green	green	green	green	green	green	green
Rv1131	green	green	green	green	green	green	green	green	green	green
Rv1183	green	green	green	green	green	green	green	green	green	green
Rv1193	green	green	green	green	green	green	green	green	green	green
Rv1428c	green	green	green	green	green	green	green	green	green	green
Rv1432	green	green	green	green	green	green	green	green	green	green
Rv1608c	green	green	green	green	green	green	green	green	green	green
Rv1626	green	green	green	green	green	green	green	green	green	green
Rv1627c	green	green	green	green	green	green	green	green	green	green
Rv1798	green	green	green	green	green	green	green	green	green	green
Rv1906c	black	black	green	black	black	black	black	black	green	green
Rv1919c	green	green	green	green	green	green	green	green	green	green
Rv1963c	green	green	green	green	green	green	green	green	green	green
Rv2048c	green	green	green	green	green	green	green	green	green	green
Rv2118c	green	green	green	green	green	green	green	green	green	green
Rv2206	green	green	green	green	green	green	green	green	green	green
Rv2239c	green	green	green	green	green	green	green	green	green	green

Organism Code	mab	mabb	mmv	may	mabo	mabl	maz	mak	mys	myc
Gene/Protein										
Rv2416c										
Rv2462c										
Rv2506										
Rv2668										
Rv2681										
Rv2684										
Rv2710										
Rv2799										
Rv2914c										
Rv2985										
Rv3050c										
Rv3274c										
Rv3409c										
Rv3419c										
Rv3421c										
Rv3492c										
Rv3493c										
Rv3494c										
Rv3495c										
Rv3496c										
Rv3497c										
Rv3498c										
Rv3499c										
Rv3500c										
Rv3501c										
Rv3502c										
Rv3504										
Rv3505										
Rv3506										
Rv3515c										
Rv3516										
Rv3522										
Rv3523										
Rv3526										
Rv3531c										
Rv3534c										
Rv3535c										
Rv3536c										
Rv3537										
Rv3538										
Rv3540c										
Rv3542c										
Rv3543c										

Organism Code	mab	mabb	mmv	may	mabo	mabl	maz	mak	mys	myc
Gene/Protein										
Rv3544c										
Rv3545c										
Rv3546										
Rv3548c										
Rv3549c										
Rv3551										
Rv3553										
Rv3559c										
Rv3560c										
Rv3561										
Rv3563										
Rv3564										
Rv3567c										
Rv3568c										
Rv3569c										
Rv3570c										
Rv3571										
Rv3572										
Rv3573c										
Rv3575c										
Rv3779										
Rv3820c										
Rv3824c										
Rv3825c										
Rv3911										
Rv1106c										
Rv3503c										
Rv3507										
Rv3508										
Rv3509c										
Rv3510c										
Rv3511										
Rv3513c										
Rv3514										
Rv3517										
Rv3518c										
Rv3519										
Rv3520c										
Rv3521										
Rv3524										
Rv3525c										
Rv3527										
Rv3528c										

Organism Code	mab	mabb	mmv	may	mabo	mabl	maz	mak	mys	myc
Gene/Protein										
Rv3529c										
Rv3530c										
Rv3532										
Rv3533c										
Rv3539										
Rv3541c										
Rv3547										
Rv3550										
Rv3552										
Rv3554										
Rv3555c										
Rv3556c										
Rv3557c										
Rv3558										
Rv3562										
Rv3565										
Rv3566A										
Rv3574										
Rv3566c										
Rv2590										
Rv0099										
Rv1346										
Rv3061c										
Rv3140										
Rv3139										
Rv0468										
Rv1715										

Table 3.5. List of cholesterol-degrading homologs that are not found in MCAC species.

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv0876c	mab; mabb; mmv; mabl	Genes predicted to be specifically required for growth on cholesterol
Rv1906c	mab; mabb; may; mabo; mabl; maz; mak	Genes predicted to be specifically required for growth on cholesterol
Rv2684	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes predicted to be specifically required for growth on cholesterol
Rv3575c	mabb; mmv; mabl; mak; mys; myc	Genes predicted to be specifically required for growth on cholesterol
Rv3507	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3508	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3511	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3514	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3517	mabl	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3519	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes proven to be essential for survival in macrophage cells and in murine infection
Rv3524	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3528c	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566A	mab; mabb; mmv; may; mabo; mabl; maz; mak; mys; myc	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

Note: As reported earlier, with review of Rv1906, more than 40% identity to *M. tuberculosis* H37Rv was taken as positive across all the categories, as the proteins are hypothetical. From this, 10 negative species were identified – all from the MCAC category: *Mycobacterium abscessus* ATCC 19977 (mab); *Mycobacterium abscessus* subsp. *bolletii* 50594 (mabb), *Mycobacterium abscessus* subsp. *bolletii* GO 06 (mmv), *Mycobacterium abscessus* subsp. *bolletii* MA 1948 (may), *Mycobacterium abscessus* subsp. *bolletii* MC1518 (mabo), *Mycobacterium abscessus* subsp. *bolletii* CCUG 48898 = JCM 15300 (mabl), *Mycobacterium abscessus* subsp. *bolletii* 103 (maz), *Mycobacterium abscessus* subsp. *abcessus* MM1513 (mak), *Mycobacterium abscessus* DJO-44274 (mys), and *Mycobacterium abscessus* 4529 (myc).

3.3.3. Determining the cholesterol-degrading ability of 15 MAC species

Detailed analysis of 151 gene homologs across 15 MAC species revealed the absence of the following cholesterol-degrading genes (Tables 3.6 and 3.7):

Table 3.6. Analysis of 151 gene homologs across 15 MAC species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code	mpa	mao	mavi	mavu	mav	mavr	mavd	mava	mit	mir	mia	mie	mid	myo	mmm
Gene/Protein															
Rv0009															
Rv0153c		■	■				■								
Rv0202c															
Rv0244c															
Rv0362															
Rv0391															
Rv0450c															
Rv0485															
Rv0495c															
Rv0655															
Rv0693															
Rv0694															
Rv0695															
Rv0696															
Rv0761c															
Rv0805															
Rv0876c															
Rv1071c															
Rv1084			■												
Rv1096															
Rv1129c															
Rv1130															
Rv1131															
Rv1183															
Rv1193															
Rv1428c															
Rv1432															
Rv1608c															
Rv1626															
Rv1627c															

Organism Code	mpa	mao	mavi	mavu	mav	mavr	mavd	mava	mit	mir	mia	mie	mid	myo	mmm
Gene/Protein															
Rv1798															
Rv1906c															
Rv1919c															
Rv1963c															
Rv2048c															
Rv2118c															
Rv2206															
Rv2239c															
Rv2416c															
Rv2462c															
Rv2506															
Rv2668															
Rv2681															
Rv2684															
Rv2710															
Rv2799															
Rv2914c															
Rv2985															
Rv3050c															
Rv3274c															
Rv3409c															
Rv3419c															
Rv3421c															
Rv3492c															
Rv3493c															
Rv3494c															
Rv3495c															
Rv3496c															
Rv3497c															
Rv3498c															
Rv3499c															
Rv3500c															
Rv3501c															
Rv3502c															
Rv3504															
Rv3505															
Rv3506															
Rv3515c															
Rv3516															
Rv3522															
Rv3523															
Rv3526															
Rv3531c															

Organism Code	mpa	mao	mavi	mavu	mav	mavr	mavd	mava	mit	mir	mia	mie	mid	myo	mmm
Gene/Protein															
Rv3534c															
Rv3535c															
Rv3536c															
Rv3537															
Rv3538															
Rv3540c															
Rv3542c															
Rv3543c															
Rv3544c															
Rv3545c															
Rv3546															
Rv3548c															
Rv3549c															
Rv3551															
Rv3553															
Rv3559c															
Rv3560c															
Rv3561															
Rv3563															
Rv3564															
Rv3567c															
Rv3568c															
Rv3569c															
Rv3570c															
Rv3571															
Rv3572															
Rv3573c															
Rv3575c															
Rv3779															
Rv3820c															
Rv3824c															
Rv3825c															
Rv3911															
Rv1106c															
Rv3503c															
Rv3507															
Rv3508															
Rv3509c															
Rv3510c															
Rv3511															
Rv3513c															
Rv3514															
Rv3517															

Organism Code	mpa	mao	mavi	mavu	mav	mavr	mavd	mava	mit	mir	mia	mie	mid	myo	mmm
Gene/Protein															
Rv3518c															
Rv3519															
Rv3520c															
Rv3521															
Rv3524															
Rv3525c															
Rv3527															
Rv3528c															
Rv3529c															
Rv3530c															
Rv3532															
Rv3533c															
Rv3539															
Rv3541c															
Rv3547															
Rv3550															
Rv3552															
Rv3554															
Rv3555c															
Rv3556c															
Rv3557c															
Rv3558															
Rv3562															
Rv3565															
Rv3566A															
Rv3574															
Rv3566c															
Rv2590															
Rv0099															
Rv1346															
Rv3061c															
Rv3140															
Rv3139															
Rv0468															
Rv1715															

Table 3.7. List of cholesterol-degrading homologs that are not found in MAC species.

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv0153c	mao; mavi; mavd	Genes predicted to be specifically required for growth on cholesterol
Rv1084	mavi	Genes predicted to be specifically required for growth on cholesterol
Rv3779	mav	Genes predicted to be specifically required for growth on cholesterol
Rv3519	mit	Genes proven to be essential for survival in macrophage cells and in murine infection
Rv3528c	mpa; mao; mavi; mavu; mav; mavr; mavd; mava; mit; mir; mia; mie; mid; myo; mmm	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566A	mpa; mao; mavi; mavu; mav; mavr; mavd; mava; mit; mir; mia; mie; mid; myo; mmm	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

3.3.4. Determining the cholesterol-degrading ability of 2 MCL species

Detailed analysis of 151 gene homologs across 2 MCL species revealed the absence of the following cholesterol-degrading genes (Tables 3.8 and 3.9):

Table 3.8. Analysis of 151 gene homologs across 2 MCL species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code	mle	mlb
Gene/Protein		
Rv0009		
Rv0153c		
Rv0202c		
Rv0244c		
Rv0362		
Rv0391		
Rv0450c		
Rv0485		
Rv0495c		
Rv0655		
Rv0693		
Rv0694		
Rv0695		
Rv0696		
Rv0761c		
Rv0805		
Rv0876c		
Rv1071c		
Rv1084		
Rv1096		
Rv1129c		
Rv1130		
Rv1131		
Rv1183		
Rv1193		
Rv1428c		
Rv1432		
Rv1608c		
Rv1626		
Rv1627c		
Rv1798		
Rv1906c		
Rv1919c		
Rv1963c		
Rv2048c		
Rv2118c		
Rv2206		
Rv2239c		

Organism Code	mle	mlb
Gene/Protein		
Rv2416c		
Rv2462c		
Rv2506		
Rv2668		
Rv2681		
Rv2684		
Rv2710		
Rv2799		
Rv2914c		
Rv2985		
Rv3050c		
Rv3274c		
Rv3409c		
Rv3419c		
Rv3421c		
Rv3492c		
Rv3493c		
Rv3494c		
Rv3495c		
Rv3496c		
Rv3497c		
Rv3498c		
Rv3499c		
Rv3500c		
Rv3501c		
Rv3502c		
Rv3504		
Rv3505		
Rv3506		
Rv3515c		
Rv3516		
Rv3522		
Rv3523		
Rv3526		
Rv3531c		
Rv3534c		
Rv3535c		
Rv3536c		
Rv3537		
Rv3538		
Rv3540c		
Rv3542c		
Rv3543c		

Organism Code	mle	mlb
Gene/Protein		
Rv3544c		
Rv3545c		
Rv3546		
Rv3548c		
Rv3549c		
Rv3551		
Rv3553		
Rv3559c		
Rv3560c		
Rv3561		
Rv3563		
Rv3564		
Rv3567c		
Rv3568c		
Rv3569c		
Rv3570c		
Rv3571		
Rv3572		
Rv3573c		
Rv3575c		
Rv3779		
Rv3820c		
Rv3824c		
Rv3825c		
Rv3911		
Rv1106c		
Rv3503c		
Rv3507		
Rv3508		
Rv3509c		
Rv3510c		
Rv3511		
Rv3513c		
Rv3514		
Rv3517		
Rv3518c		
Rv3519		
Rv3520c		
Rv3521		
Rv3524		
Rv3525c		
Rv3527		
Rv3528c		

Organism Code	mle	mlb
Gene/Protein		
Rv3529c		
Rv3530c		
Rv3532		
Rv3533c		
Rv3539		
Rv3541c		
Rv3547		
Rv3550		
Rv3552		
Rv3554		
Rv3555c		
Rv3556c		
Rv3557c		
Rv3558		
Rv3562		
Rv3565		
Rv3566A		
Rv3574		
Rv3566c		
Rv2590		
Rv0099		
Rv1346		
Rv3061c		
Rv3140		
Rv3139		
Rv0468		
Rv1715		

Table 3.9. List of cholesterol-degrading homologs that are not found in MCL species.

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv0153c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv0485	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv0693	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv0695	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv1084	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv1129c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv1130	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv2416c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv2668	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv2799	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv3492c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv3493c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv3523	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3526	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3531c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv3535c	mle; mlb	Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3536c	mle; mlb	Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3540c	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol
Rv3551	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection;

M. tuberculosis Gene/Protein	Species Code	Role
		Genes predicted to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol
Rv3553	mle; mlb	Genes predicted to be specifically required for growth on cholesterol
Rv3568c	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3571	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection; Genes predicted to be specifically required for growth on cholesterol; Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3503c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3510c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3517	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3519	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection
Rv3521	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3524	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3527	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection
Rv3528c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3529c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3552	mle; mlb	Genes proven to be essential for survival in macrophage cells and in murine infection
Rv3554	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3555c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566A	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

***M. tuberculosis* Gene/Protein**	**Species Code**	**Role**
Rv3566c	mle; mlb	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

The results for *Mycobacterium leprae* TN (mle) and *Mycobacterium leprae* Br4923 (mlb) are consistent with research done over the years, confirming that species from the MCL category lack the genes/proteins necessary for being able to break down cholesterol. Sequencing of the *M. leprae* genome has revealed that, compared to the *M. tuberculosis* H37Rv genome that can potentially encode 3 924 genes (Cole *et al*., 1998), the *M. leprae* genome encodes only 1 604 proteins and contains 1 116 pseudogenes (Cole *et al*., 2001). Research on ChoD of *M. tuberculosis* showed that although 50% of the *M. leprae* genome contains truncated genes, the ortholog ML0389 is complete, but some other sterol catabolic genes in *M. leprae* are encoded by pseudogenes (Brzostek *et al*., 2007).

Laboratory research done by Marques *et al*. (2015) generated results confirming the *in silico* prediction that *M. leprae* is unable to use cholesterol in central carbon metabolism and energy production.

Results from this study further confirm that because of the lack of quite a large number of cholesterol-degrading genes (Table 3.8 and 3.9), MCL species are unable to use cholesterol as a carbon and energy source.

3.3.5. Determining the cholesterol-degrading ability of 8 NTM species

Detailed analysis of 151 gene homologs across 8 NTM species revealed the absence of the following cholesterol-degrading genes (Tables 3.10 and 3.11):

Table 3.10. Analysis of 151 gene homologs across 8 NTM species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code	mul	mjd	mmi	mli	mkn	mks	mki	mhad
Gene/Protein								
Rv0009								
Rv0153c								
Rv0202c								
Rv0244c								
Rv0362								
Rv0391								
Rv0450c								
Rv0485								
Rv0495c								
Rv0655								
Rv0693								
Rv0694								
Rv0695								
Rv0696								
Rv0761c								
Rv0805								
Rv0876c								
Rv1071c								
Rv1084								
Rv1096								
Rv1129c								
Rv1130								
Rv1131								
Rv1183								
Rv1193								
Rv1428c								
Rv1432								
Rv1608c								
Rv1626								
Rv1627c								
Rv1798								
Rv1906c								
Rv1919c								
Rv1963c								
Rv2048c								
Rv2118c								
Rv2206								
Rv2239c								

Organism Code	mul	mjd	mmi	mli	mkn	mks	mki	mhad
Gene/Protein								
Rv2416c								
Rv2462c								
Rv2506								
Rv2668								
Rv2681								
Rv2684								
Rv2710								
Rv2799								
Rv2914c								
Rv2985								
Rv3050c								
Rv3274c								
Rv3409c								
Rv3419c								
Rv3421c								
Rv3492c								
Rv3493c								
Rv3494c								
Rv3495c								
Rv3496c								
Rv3497c								
Rv3498c								
Rv3499c								
Rv3500c								
Rv3501c								
Rv3502c								
Rv3504								
Rv3505								
Rv3506								
Rv3515c								
Rv3516								
Rv3522								
Rv3523								
Rv3526								
Rv3531c								
Rv3534c								
Rv3535c								
Rv3536c								
Rv3537								
Rv3538								
Rv3540c								
Rv3542c								
Rv3543c								

Organism Code	mul	mjd	mmi	mli	mkn	mks	mki	mhad
Gene/Protein								
Rv3544c								
Rv3545c								
Rv3546								
Rv3548c								
Rv3549c								
Rv3551								
Rv3553								
Rv3559c								
Rv3560c								
Rv3561								
Rv3563								
Rv3564								
Rv3567c								
Rv3568c								
Rv3569c								
Rv3570c								
Rv3571								
Rv3572								
Rv3573c								
Rv3575c								
Rv3779								
Rv3820c								
Rv3824c								
Rv3825c								
Rv3911								
Rv1106c								
Rv3503c								
Rv3507								
Rv3508								
Rv3509c								
Rv3510c								
Rv3511								
Rv3513c								
Rv3514								
Rv3517								
Rv3518c								
Rv3519								
Rv3520c								
Rv3521								
Rv3524								
Rv3525c								
Rv3527								
Rv3528c								

Organism Code	mul	mjd	mmi	mli	mkn	mks	mki	mhad
Gene/Protein								
Rv3529c								
Rv3530c								
Rv3532								
Rv3533c								
Rv3539								
Rv3541c								
Rv3547								
Rv3550								
Rv3552								
Rv3554								
Rv3555c								
Rv3556c								
Rv3557c								
Rv3558								
Rv3562								
Rv3565								
Rv3566A								
Rv3574								
Rv3566c								
Rv2590								
Rv0099								
Rv1346								
Rv3061c								
Rv3140								
Rv3139								
Rv0468								
Rv1715								

Table 3.11. List of cholesterol-degrading homologs that are not found in NTM species.

M. tuberculosis Gene/Protein	Species Code	Role
Rv1130	mhad	Genes predicted to be specifically required for growth on cholesterol
Rv2416c	mul	Genes predicted to be specifically required for growth on cholesterol
Rv2462c	mks; mki	Genes predicted to be specifically required for growth on cholesterol
Rv3534c	mhad	Genes proven to be essential for survival in macrophage cells and in murine infection Genes predicted to be specifically required for growth on cholesterol Genes predicted to be involved in cholesterol catabolism compiled from annotation of RHA1, H37Rv and BCG genes assigned to cholesterol pathway
Rv3575c	mjd	Genes predicted to be specifically required for growth on cholesterol
Rv3517	mul	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3528c	mul; mjd; mmi; mli; mkn; mks; mki; mhad	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566A	mul; mjd; mmi; mli; mkn; mks; mki; mhad	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

3.3.6. Determining the cholesterol-degrading ability of 19 SAP species

Detailed analysis of 151 gene homologs across 19 SAP species revealed the absence of the following cholesterol-degrading genes (Tables 3.12 and 3.13):

Table 3.12. Analysis of 151 gene homologs across 19 SAP species. Colour codes: green colour indicates presence of the homolog and black colour indicates absence of the homolog. Mycobacterial species codes were listed in Table 3.1.

Organism Code / Gene/Protein	msm	msg	msb	msn	msh	msa	mva	mgi	msp	mmc	mkm	mjl	mrh	mcb	mne	myv	mye	mgo	mft
Rv0009																			
Rv0153c																			
Rv0202c																			
Rv0244c																			
Rv0362																			
Rv0391																			
Rv0450c																			
Rv0485																			
Rv0495c																			
Rv0655																			
Rv0693																			
Rv0694																			
Rv0695																			
Rv0696																			
Rv0761c																			
Rv0805	■	■	■	■	■		■	■	■	■	■	■	■		■	■		■	
Rv0876c																	■	■	
Rv1071c																			
Rv1084									■										
Rv1096																			
Rv1129c																			
Rv1130						■	■	■	■	■	■	■	■	■			■		
Rv1131																			
Rv1183																			
Rv1193																			
Rv1428c																			
Rv1432																			
Rv1608c																			
Rv1626																			
Rv1627c																			
Rv1798																			
Rv1906c																			
Rv1919c									■										
Rv1963c																			
Rv2048c																			
Rv2118c																			
Rv2206																			
Rv2239c																			

Organism Code	msm	msg	msb	msn	msh	msa	mva	mgi	msp	mmc	mkm	mjl	mrh	mcb	mne	myv	mye	mgo	mft
Gene/Protein																			
Rv2416c														■			■		
Rv2462c																			
Rv2506																			
Rv2668																			
Rv2681																			
Rv2684																			
Rv2710																			
Rv2799																			
Rv2914c																			
Rv2985																			
Rv3050c																			
Rv3274c																			
Rv3409c																			
Rv3419c																			
Rv3421c																			
Rv3492c									■										
Rv3493c				■															
Rv3494c																			
Rv3495c																			
Rv3496c																			
Rv3497c																			
Rv3498c																			
Rv3499c																			
Rv3500c																			
Rv3501c																			
Rv3502c																			
Rv3504																			
Rv3505																			
Rv3506																			
Rv3515c																			
Rv3516																			
Rv3522																			
Rv3523																			
Rv3526																			
Rv3531c																			
Rv3534c																			
Rv3535c																			
Rv3536c																			
Rv3537																			
Rv3538																			
Rv3540c																			
Rv3542c																			
Rv3543c																			
Rv3544c																			

Organism Code / Gene/Protein	msm	msg	msb	msn	msh	msa	mva	mgi	msp	mmc	mkm	mjl	mrh	mcb	mne	myv	mye	mgo	mft
Rv3545c																			
Rv3546																			
Rv3548c																			
Rv3549c																			
Rv3551																			
Rv3553																			
Rv3559c																			
Rv3560c																			
Rv3561																			
Rv3563																			
Rv3564																			
Rv3567c																			
Rv3568c																			
Rv3569c																			
Rv3570c																			
Rv3571																			
Rv3572																			
Rv3573c																			
Rv3575c																			
Rv3779																			
Rv3820c																			
Rv3824c																			
Rv3825c																			
Rv3911																			
Rv1106c																			
Rv3503c																			
Rv3507																			
Rv3508																			
Rv3509c																			
Rv3510c																			
Rv3511																			
Rv3513c																			
Rv3514																			
Rv3517																			
Rv3518c																			
Rv3519																			
Rv3520c																			
Rv3521																			
Rv3524																			
Rv3525c																			
Rv3527																			
Rv3528c																			
Rv3529c																			
Rv3530c																			

Organism Code	msm	msg	msb	msn	msh	msa	mva	mgi	msp	mmc	mkm	mjl	mrh	mcb	mne	myv	mye	mgo	mft
Gene/Protein																			
Rv3532																			
Rv3533c																			
Rv3539																			
Rv3541c																			
Rv3547																			
Rv3550																			
Rv3552																			
Rv3554																			
Rv3555c																			
Rv3556c																			
Rv3557c																			
Rv3558																			
Rv3562																			
Rv3565																			
Rv3566A																			
Rv3574																			
Rv3566c																			
Rv2590																			
Rv0099																			
Rv1346																			
Rv3061c																			
Rv3140																			
Rv3139																			
Rv0468																			
Rv1715																			

Table 3.13. List of cholesterol-degrading homologs that are not found in SAP species.

M. tuberculosis Gene/Protein	Species Code	Role
Rv0805	msm; msg; msb; msn; msh; mva; mgi; msp; mmc; mkm; mjl; mrh; mne; myv; mgo	Genes predicted to be specifically required for growth on cholesterol
Rv0876c	mye; mgo	Genes predicted to be specifically required for growth on cholesterol
Rv1084	msp	Genes predicted to be specifically required for growth on cholesterol
Rv1130	msa; mva; mgi; msp; mmc; mkm; mjl; mrh; mcb; mye	Genes predicted to be specifically required for growth on cholesterol
Rv1919c	msp	Genes predicted to be specifically required for growth on cholesterol
Rv2416c	mcb; mye	Genes predicted to be specifically required for growth on cholesterol
Rv3492c	msp	Genes predicted to be specifically required for growth on cholesterol
Rv3493c	msn	Genes predicted to be specifically required for growth on cholesterol
Rv3572	msm; msg; msb; msn; msh; mva; mgi; msp; mmc; mkm; mjl; mrh; mne; myv; mgo; mft	Genes predicted to be specifically required for growth on cholesterol
Rv3779	msm; msg; msb; msn; msh; mva; mgi; msp; mmc; mkm; mjl; mrh; mne	Genes predicted to be specifically required for growth on cholesterol
Rv3507	msm; msg; msb; msn; msh; msa; mva; mgi; msp; mmc; mkm; mjl; mrh; mcb; mne; myv; mye; mgo; mft	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3508	msm; msg; msb; msn; msh; msa; mva; mgi; msp; mmc; mkm; mjl; mrh; mcb; mne; myv; mye; mgo; mft	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3511	msm; msg; msb; msn; msh; msa; mva; msp; mmc; mkm; mjl; mcb; mne; myv; mye; mgo; mft	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3514	msm; msg; msb; msn; msh; msa; mva; mgi; msp; mmc; mkm; mjl; mrh; mcb; mne; myv; mgo; mft	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3517	msa; mva; mne; mye	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3528c	msm; msg; msb; msn; msh; msa; mva; mgi; msp; mmc; mkm; mjl;	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

M. tuberculosis Gene/Protein	Species Code	Role
	mrh; mcb; mne; myv; mye; mgo; mft	
Rv3566A	msm; msg; msb; msn; msh; mva; mgi; msp; mmc; mkm; mjl; mrh; mcb; mne; mye; mgo	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data
Rv3566c	mcb; mye	Genes involved in cholesterol degradation by *M. tuberculosis* H37Rv but not confirmed or predicted as essential, as per published data

Based on the above results, a detailed analysis of the presence or absence of 151 homologs of *M. tuberculosis* H37Rv in 92 mycobacterial species and the species' ability to degrade cholesterol is shown in Table 3.14.

Table 3.14. *In silico* analysis of cholesterol-degrading genes/proteins in mycobacterial species.

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
	***Mycobacterium tuberculosis* complex (MTBC)**					
mtu	*Mycobacterium tuberculosis* H37Rv (1998)					Positive
mtv	*Mycobacterium tuberculosis* H37Rv (2012)					Positive
mtc	*Mycobacterium tuberculosis* CDC1551				Rv3555c	Positive
mra	*Mycobacterium tuberculosis* H37Ra					Positive
mtf	*Mycobacterium tuberculosis* F11				Rv3566A	Positive
mtb	*Mycobacterium tuberculosis* KZN 1435				Rv3566A	Positive
mtk	*Mycobacterium tuberculosis* KZN 4207				Rv3566A	Positive
mtz	*Mycobacterium tuberculosis* KZN 605				Rv3566A	Positive
mtg	*Mycobacterium tuberculosis* RGTB327		Rv1084 Rv2799			Negative
mti	*Mycobacterium tuberculosis* RGTB423	Rv3526	Rv0153c Rv0485 Rv0805 Rv0876c Rv2416c Rv2681 Rv3526 Rv3531c	Rv3526		Negative
mte	*Mycobacterium tuberculosis* CCDC5079 (2012)		Rv0805 Rv1919c		Rv3566A	Negative
mtur	*Mycobacterium tuberculosis* CCDC5079 (2013)					Positive
mtl	*Mycobacterium tuberculosis* CCDC5180		Rv0805 Rv1919c		Rv3566A	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mto	*Mycobacterium tuberculosis* CTRI-2					Positive
mtd	*Mycobacterium tuberculosis* UT205					Positive
mtn	*Mycobacterium tuberculosis* Erdman = ATCC 35801		Rv0805		Rv3566A	Negative
mtj	*Mycobacterium tuberculosis* Beijing/NITR203				Rv3566A	Positive
mtub	*Mycobacterium tuberculosis* 7199-99					Positive
mtuc	*Mycobacterium tuberculosis* CAS/NITR204		Rv0485 Rv0695 Rv1084 Rv1130 Rv1432 Rv2416c Rv2681 Rv3536c Rv3779	Rv3536c	Rv3521 Rv3566A	Negative
mtue	*Mycobacterium tuberculosis* EAI5/NITR206		Rv2681		Rv3566A	Negative
mtx	*Mycobacterium tuberculosis* EAI5					Positive
mtuh	*Mycobacterium tuberculosis* Haarlem/NITR202		Rv0485 Rv0876c Rv1084 Rv1096 Rv1129c Rv2416c Rv3531c			Negative
mtul	*Mycobacterium tuberculosis* Haarlem				Rv3566A	Positive
mtut	*Mycobacterium tuberculosis* BT1					Positive
mtuu	*Mycobacterium tuberculosis* BT2					Positive

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mtq	*Mycobacterium tuberculosis* HKBS1					Positive
mbo	*Mycobacterium bovis* AF2122/97					Positive
mbb	*Mycobacterium bovis* BCG Pasteur 1173P2					Positive
mbt	*Mycobacterium bovis* BCG Tokyo 172					Positive
mbm	*Mycobacterium bovis* BCG Mexico					Positive
mbk	*Mycobacterium bovis* BCG Korea 1168P				Rv3566A	Positive
mbx	*Mycobacterium bovis* BCG ATCC 35743		Rv0805 Rv2206		Rv3566A Rv3566c	Negative
mbz	*Mycobacterium bovis* ATCC BAA-935					Positive
maf	*Mycobacterium africanum*				Rv3528c	Positive
mce	*Mycobacterium canettii* CIPT 140010059		Rv1130			Negative
mcq	*Mycobacterium canettii* CIPT 140060008					Positive
mcv	*Mycobacterium canettii* CIPT 140070008					Positive
mcx	*Mycobacterium canettii* CIPT 140070010				Rv3566A	Positive
mcz	*Mycobacterium canettii* CIPT 140070017				Rv3517 Rv3528c Rv3566A	Positive
	***Mycobacterium chelonae-abscessus* complex (MCAC)**					
mab	*Mycobacterium abscessus* ATCC 19977	Rv3519	Rv0876c Rv1906c Rv2684		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mabb	*Mycobacterium abscessus* subsp. *bolletii* 50594	Rv3519	Rv0876c Rv1906c Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
mmv	*Mycobacterium abscessus* subsp. *bolletii* GO 06	Rv3519	Rv0876c Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
may	*Mycobacterium abscessus* subsp. *bolletii* MA 1948	Rv3519	Rv1906c Rv2684		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mabo	*Mycobacterium abscessus* subsp. *bolletii* MC1518	Rv3519	Rv1906c Rv2684		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
mabl	*Mycobacterium abscessus* subsp. *bolletii* CCUG 48898 = JCM 15300	Rv3519	Rv0876c Rv1906c Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3517 Rv3524 Rv3528c Rv3566A	Negative
maz	*Mycobacterium abscessus* subsp. *bolletii* 103	Rv3519	Rv1906c Rv2684		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mak	*Mycobacterium abscessus* subsp. *abscessus*	Rv3519	Rv1906c Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
mys	*Mycobacterium abscessus* DJO-44274	Rv3519	Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
myc	*Mycobacterium abscessus* 4529	Rv3519	Rv2684 Rv3575c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3524 Rv3528c Rv3566A	Negative
	***Mycobacterium avium* complex (MAC)**					
mpa	*Mycobacterium avium* subsp. *paratuberculosis* K-10				Rv3528c Rv3566A	Positive
mao	*Mycobacterium avium* subsp. *paratuberculosis* MAP4		Rv0153c		Rv3528c Rv3566A	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mavi	*Mycobacterium avium* subsp. *paratuberculosis* E1		Rv0153c Rv1084		Rv3528c Rv3566A	Negative
mavu	*Mycobacterium avium* subsp. *paratuberculosis* E93				Rv3528c Rv3566A	Positive
mav	*Mycobacterium avium* 104		Rv3779		Rv3528c Rv3566A	Negative
mavd	*Mycobacterium avium* subsp. *avium* DJO-44271		Rv0153c		Rv3528c Rv3566A	Negative
mavr	*Mycobacterium avium* subsp. *avium* 2285 (R)				Rv3528c Rv3566A	Positive
mava	*Mycobacterium avium* subsp. *avium* 2285 (S)				Rv3528c Rv3566A	Positive
mit	*Mycobacterium intracellulare* MOTT-02	Rv3519			Rv3528c Rv3566A	Negative
mir	*Mycobacterium intracellulare* MOTT-64				Rv3528c Rv3566A	Positive
mia	*Mycobacterium intracellulare* ATCC 13950				Rv3528c Rv3566A	Positive
mie	*Mycobacterium intracellulare* 1956				Rv3528c Rv3566A	Positive
mid	*Mycobacterium indicus pranii*				Rv3528c Rv3566A	Positive
myo	*Mycobacterium yongonense*				Rv3528c Rv3566A	Positive
mmm	*Mycobacterium* sp. MOTT36Y				Rv3528c Rv3566A	Positive

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
	Mycobacteria causing leprosy (MCL)					
mle	*Mycobacterium leprae* TN	Rv3523 Rv3526 Rv3540c Rv3551 Rv3568c Rv3571 Rv3519 Rv3527 Rv3552	Rv0153c Rv0485 Rv0693 Rv0695 Rv1084 Rv1129c Rv1130 Rv2416c Rv2668 Rv2799 Rv3492c Rv3493c Rv3526 Rv3531c Rv3536c Rv3540c Rv3551 Rv3553 Rv3568c Rv3571	Rv3523 Rv3526 Rv3535c Rv3536c Rv3568c Rv3571	Rv3503c Rv3510c Rv3517 Rv3521 Rv3524 Rv3528c Rv3529c Rv3554 Rv3555c Rv3566A Rv3566c	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mlb	*Mycobacterium leprae* Br4923	Rv3523 Rv3526 Rv3540c Rv3551 Rv3568c Rv3571 Rv3519 Rv3527 Rv3552	Rv0153c Rv0485 Rv0693 Rv0695 Rv1084 Rv1129c Rv1130 Rv2416c Rv2668 Rv2799 Rv3492c Rv3493c Rv3526 Rv3531c Rv3536c Rv3540c Rv3551 Rv3553 Rv3568c Rv3571	Rv3523 Rv3526 Rv3535c Rv3536c Rv3568c Rv3571	Rv3503c Rv3510c Rv3517 Rv3521 Rv3524 Rv3528c Rv3529c Rv3554 Rv3555c Rv3566A Rv3566c	Negative
	Non-tuberculosis *Mycobacterium* (NTM)					
mul	*Mycobacterium ulcerans*		Rv2416c		Rv3517 Rv3528c Rv3566A	Negative
mjd	*Mycobacterium sinense*		Rv3575c		Rv3528c Rv3566A	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mmi	*Mycobacterium marinum*				Rv3528c Rv3566A	Positive
mli	*Mycobacterium liflandii*				Rv3528c Rv3566A	Positive
mkn	*Mycobacterium kansasii* ATCC 12478				Rv3528c Rv3566A	Positive
mks	*Mycobacterium kansasii* 662		Rv2462c		Rv3528c Rv3566A	Negative
mki	*Mycobacterium kansasii* 824		Rv2462c		Rv3528c Rv3566A	Negative
mhad	*Mycobacterium haemophilum*	Rv3534c	Rv1130 Rv3534c	Rv3534c	Rv3528c Rv3566A	Negative
	Saprophytes (SAP)					
msm	*Mycobacterium smegmatis* MC2 155 (2006)		Rv0805 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
msg	*Mycobacterium smegmatis* MC2 155 (2012)		Rv0805 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
msb	*Mycobacterium smegmatis* MC2 155 (2014)		Rv0805 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
msn	*Mycobacterium smegmatis* INHR1		Rv0805 Rv3493c Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
msh	*Mycobacterium smegmatis* INHR2		Rv0805 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
msa	*Mycobacterium* sp. JS623		Rv1130		Rv3507 Rv3508 Rv3511 Rv3514 Rv3517 Rv3528c	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mva	*Mycobacterium vanbaalenii*		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3517 Rv3528c Rv3566A	Negative
mgi	*Mycobacterium gilvum* PYR-GCK		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3514 Rv3528c Rv3566A	Negative
msp	*Mycobacterium gilvum* Spyr1		Rv0805 Rv1084 Rv1130 Rv1919c Rv3492c Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
mmc	*Mycobacterium* sp. MCS		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mkm	*Mycobacterium* sp. KMS		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
mjl	*Mycobacterium* sp. JLS		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative
mrh	*Mycobacterium rhodesiae*		Rv0805 Rv1130 Rv3572 Rv3779		Rv3507 Rv3508 Rv3514 Rv3528c Rv3566A	Negative
mcb	*Mycobacterium chubuense*		Rv1130 Rv2416c		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A Rv3566c	Negative

Organism Code	Species	Absent H37Rv Homologs Relating to Cholesterol Catabolism				Ability to Degrade Cholesterol
		Proven to be Essential	Predicted to be Essential or Specifically Required	Predicted to be Involved	Involved but not Proven or Predicted to be Essential	
mne	*Mycobacterium neoaurum*		Rv0805 Rv3572 Rv3779		Rv3507 Rv3508 Rv3511 Rv3514 Rv3517 Rv3528c Rv3566A	Negative
myv	*Mycobacterium* sp. VKM Ac-1817D		Rv0805 Rv3572		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c	Negative
mye	*Mycobacterium* sp. EPa45		Rv0876c Rv1130 Rv2416c		Rv3507 Rv3508 Rv3511 Rv3517 Rv3528c Rv3566A Rv3566c	Negative
mgo	*Mycobacterium goodii*		Rv0805 Rv0876c Rv3572		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c Rv3566A	Negative

Organism Code	**Species**	**Absent H37Rv Homologs Relating to Cholesterol Catabolism**				**Ability to Degrade Cholesterol**
		Proven to be Essential	**Predicted to be Essential or Specifically Required**	**Predicted to be Involved**	**Involved but not Proven or Predicted to be Essential**	
mft	*Mycobacterium fortuitum*		Rv3572		Rv3507 Rv3508 Rv3511 Rv3514 Rv3528c	Negative

3.4. Conclusion

In conclusion, *in silico* comparative analysis of cholesterol using genes/proteins across 93 mycobacterial species revealed the cholesterol-utilising capacity of mycobacterial species as shown in Table 3.15 below.

Table 3.15. Analysis of cholesterol-utilising capability by mycobacterial species.

Category	No of species	Ability to utilize cholesterol as carbon source	
		Negative	Positive
MTBC	39	10	29
MCAC	10	10	
MAC	15	5	10
MCL	2	2	
NTM	8	5	3
SAP	19	19	

References

- Al Rashdi, A.S., Jadhav, B.L., Deshpande, T. and Deshpande, U., 2014. Whole-genome sequencing and annotation of a clinical isolate of *Mycobacterium tuberculosis* from Mumbai, India. *Genome announcements*, *2*(2), p. 2014.
- Amin, A.S., Hsu, C.Y., Darwish, S.F., Ghosh, P., AbdEl-Fatah, E.M., Behour, T.S. and Talaat, A.M., 2015. Ecology and genomic features of infection with *Mycobacterium avium* subspecies *paratuberculosis* in Egypt. *Microbiology*, *161*(4), pp. 807-818.
- Ang, P., RattanaApiromyakij, N. and Goh, C.L., 2000. Retrospective study of *Mycobacterium marinum* skin infections. *International journal of dermatology*, *39*(5), pp. 343-347.
- Bannantine, J.P., Li, L., Mwangi, M., Cote, R., Garay, J.A.R. and Kapur, V., 2014. Complete genome sequence of *Mycobacterium avium* subsp. *paratuberculosis*, isolated from human breast milk. *Genome announcements*, *2*(1), pp. e01252-13.
- Bentley, S.D., Comas, I., Bryant, J.M., Walker, D., Smith, N.H., Harris, S.R., Thurston, S., Gagneux, S., Wood, J., Antonio, M. and Quail, M.A., 2012. The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS Negl Trop Dis*, *6*(2), p. e1552.
- Bragin, E.Y., Shtratnikova, V.Y., Dovbnya, D.V., Schelkunov, M.I., Pekov, Y.A., Malakho, S.G., Egorova, O.V., Ivashina, T.V., Sokolov, S.L., Ashapkin, V.V. and Donova, M.V., 2013. Comparative analysis of genes encoding key steroid core oxidation enzymes in fast-growing *Mycobacterium spp.* strains. *The Journal of steroid biochemistry and molecular biology*, *138*, pp. 41-53.
- Brosch, R., Gordon, S.V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., Dos Santos, S., Duthoy, S., Lacroix, C., Garcia-Pelayo, C. and Inwald, J.K., 2007. Genome

plasticity of BCG and impact on vaccine efficacy. *Proceedings of the National Academy of Sciences*, *104*(13), pp. 5596-5601.

- Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J. and Dziadek, J., 2007. Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*. *FEMS microbiology letters*, 275(1), pp. 106-112.
- Choo, S.W., Wong, Y.L., Leong, M.L., Heydari, H., Ong, C.S., Ng, K.P. and Ngeow, Y.F., 2012. Analysis of the genome of *Mycobacterium abscessus* strain M94 reveals an uncommon cluster of tRNAs. *Journal of bacteriology*, *194*(20), pp. 5724-5724.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E. and Tekaia, F., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*(6685), p. 537-544.
- Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D. and Mungall, K., 2001. Massive gene decay in the leprosy bacillus. *Nature*, 409(6823), pp. 1007-1011.
- Costa, K.C., Bergkessel, M., Saunders, S., Korlach, J. and Newman, D.K., 2015. Enzymatic degradation of phenazines can generate energy and protect sensitive organisms from toxicity. *MBio*, *6*(6), pp. e01520-15.
- Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D. and Hickey, E., 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *Journal of bacteriology*, *184*(19), pp. 5479-5490.
- Gallien, S., Perrodou, E., Carapito, C., Deshayes, C., Reyrat, J.M., Van Dorsselaer, A., Poch, O., Schaeffer, C. and Lecompte, O., 2009. Ortho-proteogenomics: multiple

proteomes investigation through orthology and a new MS-based protocol. *Genome research*, *19*(1), pp. 128-135.

- Gallivan, M., Shah, N. and Flood, J., 2015. Epidemiology of human *Mycobacterium bovis* disease, California, USA, 2003–2011. *Emerging infectious diseases*, *21*(3), p. 435.
- García- Fernández, E., Frank, D.J., Galán, B., Kells, P.M., Podust, L.M., García, J.L. and Ortiz de Montellano, P.R., 2013. A highly conserved mycobacterial cholesterol catabolic pathway. *Environmental microbiology*, *15*(8), pp. 2342-2359.
- Garnier, T., Eiglmeier, K., Camus, J.C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C. and Simon, S., 2003. The complete genome sequence of *Mycobacterium bovis*. *Proceedings of the National Academy of Sciences*, *100*(13), pp. 7877-7882.
- Griffin, J.E., Gawronski, J.D., DeJesus, M.A., Ioerger, T.R., Akerley, B.J. and Sassetti, C.M., 2011. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog*, *7*(9), p.e1002251.
- Gu, J. and Bourne, P.E. eds., 2009. *Structural bioinformatics* (Vol. 44). New Jersey: John Wiley & Sons.
- Ilina, E.N., Shitikov, E.A., Ikryannikova, L.N., Alekseev, D.G., Kamashev, D.E., Malakhova, M.V., Parfenova, T.V., Afanas'ev, M.V., Ischenko, D.S., Bazaleev, N.A. and Smirnova, T.G., 2013. Comparative genomic analysis of *Mycobacterium tuberculosis* drug resistant strains from Russia. *PloS one*, *8*(2), p. e56577.
- Isaza, J.P., Duque, C., Gomez, V., Robledo, J., Barrera, L.F. and Alzate, J.F., 2012. Whole genome shotgun sequencing of one Colombian clinical isolate of

Mycobacterium tuberculosis reveals DosR regulon gene deletions. *FEMS microbiology letters*, *330*(2), pp. 113-120.

- 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.
- Joung, S.M., Jeon, S.J., Lim, Y.J., Lim, J.S., Choi, B.S., Choi, I.Y., Yu, J.H., Na, K.I., Cho, E.H., Shin, S.S. and Park, Y.K., 2013. Complete genome sequence of *Mycobacterium bovis* BCG Korea, the Korean vaccine strain for substantial production. *Genome announcements*, *1*(2), pp. e00069-13.
- Kallimanis, A., Karabika, E., Mavromatis, K., Lapidus, A., LaButti, K.M., Liolios, K., Ivanova, N., Goodwin, L., Woyke, T., Velentzas, A.D. and Perisynakis, A., 2011. Complete genome sequence of *Mycobacterium sp.* strain (Spyr1) and reclassification to *Mycobacterium gilvum* Spyr1. *Standards in genomic sciences*, *5*(1), p. 144.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K., 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*(D1), pp. D353-D361.
- Kato, H., Ogawa, N., Ohtsubo, Y., Oshima, K., Toyoda, A., Mori, H., Nagata, Y., Kurokawa, K., Hattori, M., Fujiyama, A. and Tsuda, M., 2015. Complete genome sequence of a phenanthrene degrader, *Mycobacterium sp.* strain EPa45 (NBRC 110737), isolated from a phenanthrene-degrading consortium. *Genome announcements*, *3*(4), pp. e00782-15.

- Kim, B.J., Choi, B.S., Lim, J.S., Choi, I.Y., Lee, J.H., Chun, J., Kook, Y.H. and Kim, B.J., 2012a. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-02. *Journal of bacteriology*, *194*(10), pp. 2771-2771.
- Kim, B.J., Choi, B.S., Lim, J.S., Choi, I.Y., Kook, Y.H. and Kim, B.J., 2012b. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-64, belonging to the INT1 genotype. *Journal of bacteriology*, *194*(12), pp. 3268-3268.
- Kim, B.J., Choi, B.S., Lim, J.S., Choi, I.Y., Lee, J.H., Chun, J., Kook, Y.H. and Kim, B.J., 2012c. Complete genome sequence of *Mycobacterium intracellulare* strain ATCC 13950T. *Journal of bacteriology*, *194*(10), pp. 2750-2750.
- Kim, B.J., Choi, B.S., Choi, I.Y., Lee, J.H., Chun, J., Hong, S.H., Kook, Y.H. and Kim, B.J., 2012d. Complete genome sequence of *Mycobacterium intracellulare* clinical strain MOTT-36Y, belonging to the INT5 genotype. *Journal of bacteriology*, *194*(15), pp. 4141-4142.
- Kim, B.J., Kim, B.R., Lee, S.Y., Seok, S.H., Kook, Y.H. and Kim, B.J., 2013a. Whole-genome sequence of a novel species, *Mycobacterium yongonense* DSM 45126T. *Genome Announc*, *1*(4), pp. e00604-13.
- Kim, B.J., Kim, B.R., Hong, S.H., Seok, S.H., Kook, Y.H. and Kim, B.J., 2013b. Complete genome sequence of *Mycobacterium massiliense* clinical strain Asan 50594, belonging to the type II genotype. *Genome announcements*, *1*(4), pp. e00429-13.
- 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.

- Madhavilatha, G.K., Joseph, B.V., Paul, L.K., Kumar, R.A., Hariharan, R. and Mundayoor, S., 2012. Whole-genome sequences of two clinical isolates of *Mycobacterium tuberculosis* from Kerala, South India. *Journal of bacteriology*, *194*(16), p. 4430.
- Marques, M.A.M., Berrêdo-Pinho, M., Rosa, T.L., Pujari, V., Lemes, R.M., Lery, L.M., Silva, C.A.M., Guimarães, A.C.R., Atella, G.C., Wheat, W.H. and Brennan, P.J., 2015. The essential role of cholesterol metabolism in the intracellular survival of *Mycobacterium leprae* is not coupled to central carbon metabolism and energy production. *Journal of bacteriology*, 197(23), pp. 3698-3707.
- Miyoshi-Akiyama, T., Matsumura, K., Iwai, H., Funatogawa, K. and Kirikae, T., 2012. Complete annotated genome sequence of *Mycobacterium tuberculosis* Erdman. *Journal of bacteriology*, *194*(10), p. 2770.
- Mohan, A., Padiadpu, J., Baloni, P. and Chandra, N., 2015. Complete genome sequences of a *Mycobacterium smegmatis* laboratory strain (MC2 155) and isoniazid-resistant (4XR1/R2) mutant strains. *Genome announcements*, *3*(1), pp. e01520-14.
- Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G.M., Donoghue, H.D., Bouwman, A. and Mays, S., 2009. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nature genetics*, *41*(12), pp. 1282-1289.
- Mukhopadhyay, S., Nair, S. and Ghosh, S., 2012. Pathogenesis in tuberculosis: transcriptomic approaches to unraveling virulence mechanisms and finding new drug targets. *FEMS microbiology reviews*, *36*(2), pp. 463-485.
- Murray, P.R., Rosenthal, K.S., Kobayashi, G.S. and Pfaller, M.A., 1998. *Medical microbiology* (pp. 319). 3rd ed. Mosby Inc. New York.

- Murray, P.R., Rosenthal, K.S., Kobayashi, G.S. and Pfaller, M.A., 2003. *Medical Microbiology*. 4th ed. Mosby Inc. New York.
- Narayanan, S. and Deshpande, U., 2013. Whole-genome sequences of four clinical isolates of *Mycobacterium tuberculosis* from Tamil Nadu, South India. *Genome announc*, *1*(3), pp. e00186-13.
- Nesbitt, N.M., Yang, X., Fontán, P., Kolesnikova, I., Smith, I., Sampson, N.S. and Dubnau, E., 2010. A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infection and immunity*, *78*(1), pp. 275-282.
- Onipede, A.O., de Jong, B. and Adegbola, R.A., 2005. *Mycobacterium africanum*–A Review. *African journal of clinical and experimental microbiology*, *6*(2), pp. 167-175.
- Orduña, P., Cevallos, M.A., de León, S.P., Arvizu, A., Hernández-González, I.L., Mendoza-Hernández, G. and López-Vidal, Y., 2011. Genomic and proteomic analyses of *Mycobacterium bovis* BCG Mexico 1931 reveal a diverse immunogenic repertoire against tuberculosis infection. *BMC genomics*, *12*(1), p. 493.
- Ouellet, H., Johnston, J.B. and de Montellano, P.R.O., 2011. Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends in microbiology*, *19*(11), pp. 530-539.
- Pan, Y., Yang, X., Duan, J., Lu, N., Leung, A.S., Tran, V., Hu, Y., Wu, N., Liu, D., Wang, Z. and Yu, X., 2011. The whole genome sequence of four BCG vaccine strains. *Journal of bacteriology*.
- 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*, p. 33099.

- Prasanna, A.N. and Mehra, S., 2013. Comparative phylogenomics of pathogenic and non-pathogenic *Mycobacterium*. *PLoS One*, *8*(8), p. e71248.
- Raiol, T., Ribeiro, G.M., Maranhão, A.Q., Bocca, A.L., Silva-Pereira, I., Junqueira-Kipnis, A.P., de Macedo Brigido, M. and Kipnis, A., 2012. Complete genome sequence of *Mycobacterium massiliense*. *Journal of bacteriology*, *194*(19), pp. 5455-5455.
- Ripoll, F., Pasek, S., Schenowitz, C., Dossat, C., Barbe, V., Rottman, M., Macheras, E., Heym, B., Herrmann, J.L., Daffé, M. and Brosch, R., 2009. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. *PloS one*, *4*(6), p. e5660.
- Roetzer, A., Diel, R., Kohl, T.A., Rückert, C., Nübel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rüsch-Gerdes, S. and Supply, P., 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med*, *10*(2), p. e1001387.
- Saini, V., Raghuvanshi, S., Khurana, J.P., Ahmed, N., Hasnain, S.E., Tyagi, A.K. and Tyagi, A.K., 2012. Massive gene acquisitions in *Mycobacterium indicus pranii* provide a perspective on mycobacterial evolution. *Nucleic acids research*, *40*(21), pp. 10832-10850.
- Seki, M., Honda, I., Fujita, I., Yano, I., Yamamoto, S. and Koyama, A., 2009. Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette–Guerin (BCG) Tokyo 172: A comparative study of BCG vaccine substrains. *Vaccine*, *27*(11), pp. 1710-1716.

- Sekizuka, T., Kai, M., Nakanaga, K., Nakata, N., Kazumi, Y., Maeda, S., Makino, M., Hoshino, Y. and Kuroda, M., 2014. Complete genome sequence and comparative genomic analysis of *Mycobacterium massiliense* JCM 15300 in the *Mycobacterium abscessus* group reveal a conserved genomic island MmGI-1 related to putative lipid metabolism. *PloS one*, *9*(12), p. e114848.
- 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.
- Stinear, T.P., Seemann, T., Harrison, P.F., Jenkin, G.A., Davies, J.K., Johnson, P.D., Abdellah, Z., Arrowsmith, C., Chillingworth, T., Churcher, C. and Clarke, K., 2008. Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome research*, *18*(5), pp. 729-741.
- Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A. and Fiette, L., 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nature genetics*, *45*(2), pp. 172-179.
- Tang, B., Wang, Q., Yang, M., Xie, F., Zhu, Y., Zhuo, Y., Wang, S., Gao, H., Ding, X., Zhang, L. and Zhao, G., 2013. ContigScape: a Cytoscape plugin facilitating microbial genome gap closing. *BMC genomics*, *14*(1), p. 289.
- Tobias, N.J., Doig, K.D., Medema, M.H., Chen, H., Haring, V., Moore, R., Seemann, T. and Stinear, T.P., 2013. Complete genome sequence of the frog pathogen *Mycobacterium ulcerans* ecovar Liflandii. *Journal of bacteriology*, *195*(3), pp. 556-564.

- Tortoli, E., 2012. Phylogeny of the genus *Mycobacterium*: many doubts, few certainties. *Infection, genetics and evolution*, *12*(4), pp. 827-831.
- Tufariello, J.M., Kerantzas, C.A., Vilchèze, C., Calder, R.B., Nordberg, E.K., Fischer, J.A., Hartman, T.E., Yang, E., Driscoll, T., Cole, L.E. and Sebra, R., 2015. The complete genome sequence of the emerging pathogen *Mycobacterium haemophilum* explains its unique culture requirements. *MBio*, *6*(6), pp. e01313-15.
- Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W. and Eltis, L.D., 2007. A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the national academy of sciences*, *104*(6), pp. 1947-1952.
- Van der Geize, R., Grommen, A.W.F., Hessels, G.I., Jacobs, A.A.C. and Dijkhuizen, L., 2011. The steroid catabolic pathway of the intracellular pathogen Rhodococcus equi is important for pathogenesis and a target for vaccine development. *PLoS pathog*, *7*(8), p. e1002181.
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F. and van Sinderen, D., 2007. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews*, *71*(3), pp. 495-548.
- Yang, X., Gao, J., Smith, I., Dubnau, E. and Sampson, N.S., 2011. Cholesterol is not an essential source of nutrition for *Mycobacterium tuberculosis* during infection. *Journal of bacteriology*, *193*(6), pp. 1473-1476.

- Yu, B., Tao, F., Li, F., Hou, J., Tang, H., Ma, C. and Xu, P., 2015. Complete genome sequence of *Mycobacterium goodii* X7B, a facultative thermophilic biodesulfurizing bacterium with industrial potential. *Journal of biotechnology*, *212*, pp. 56-57.
- Zhang, Y., Chen, C., Liu, J., Deng, H., Pan, A., Zhang, L., Zhao, X., Huang, M., Lu, B., Dong, H. and Du, P., 2011. Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family. *Journal of bacteriology*, *193*(19), pp. 5591-5592.
- Zhang, Z.Y., Sun, Z.Q., Wang, Z.L., Wen, Z.L., Sun, Q.W., Zhu, Z.Q., Song, Y.Z., Zhao, J.W., Wang, H.H., Zhang, S.L. and Guo, X.K., 2011. Complete genome sequence of a novel clinical isolate, the nontuberculous *Mycobacterium* strain JDM601. *Journal of bacteriology*, *193*(16), pp. 4300-4301.
- Zheng, H., Lu, L., Wang, B., Pu, S., Zhang, X., Zhu, G., Shi, W., Zhang, L., Wang, H., Wang, S. and Zhao, G., 2008. Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PloS one*, *3*(6), p. e2375.

CHAPTER 4

CONCLUSION AND FUTURE PERSPECTIVES

In conclusion, the emergence of drug-resistant strains of *Mycobacterium tuberculosis*, along with the insufficiency of new drug targets, has necessitated research to identify novel drug targets. Research is gaining momentum on the use of cholesterol-degrading genes as drug targets against *M. tuberculosis*. In this direction this study is a first of its kind comprehensive analysis of genes/proteins involved in cholesterol degradation across 93 mycobacterial species using bioinformatic tools. This study revealed that 42 of the 93 mycobacterial species selected for this study are capable of degrading cholesterol. Fifty-one mycobacterial species were deduced to be unable to utilise cholesterol. Results from this study are based on *in silico* analysis and need to be experimentally validated.

RESEARCH OUTPUTS

1. **Van Wyk R**, Van Wyk M, Olivier D, Chen W, Mashele SS, Syed K (2017) Comprehensive comparative analysis of cholesterol degradation genes/proteins in the genus *Mycobacterium*. Scientific Reports (soon to be communicated: impact factor 5.2).

2. Rochelle van Wyk, Mari Van Wyk, Dedré Olivier, Samson Sithenni Mashele, Wanping Chen, Khajamohiddin Syed (2017) *In silico* analysis of cholesterol catabolic genes/proteins in the genus *Mycobacterium*. The Annual South African Pharmacology Conference, Faculty of Health Sciences, University of the Free State, Bloemfontein, 01 – 04 October.