COMPARATIVE GENOMIC ANALYSIS OF *Mycobacterium tuberculosis* FROM PULMONARY AND CENTRAL NERVOUS SYSTEM INFECTIONS TO DETERMINE FACTORS OF NEUROTROPISM

SAW SEOW HOON

DOCTOR OF PHILOSOPHY (MEDICAL SCIENCES)

FACULTY OF MEDICINE AND HEALTH SCIENCES
UNIVERSITI TUNKU ABDUL RAHMAN
AUGUST 2019

COMPARATIVE GENOMIC ANALYSIS OF *Mycobacterium tuberculosis* FROM PULMONARY AND CENTRAL NERVOUS SYSTEM INFECTIONS TO DETERMINE FACTORS OF NEUROTROPISM

By

SAW SEOW HOON

A thesis submitted to the Department of Pre-clinical Sciences,
Faculty of Medicine and Health Sciences,
Universiti Tunku Abdul Rahman,
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Medical Sciences
August 2019

THIS HUMBLE THESIS IS DEDICATED
TO THE
MEMORY OF MY LATE FATHER

ABSTRACT

COMPARATIVE GENOMIC ANALYSIS OF *Mycobacterium tuberculosis* FROM PULMONARY AND CENTRAL NERVOUS SYSTEM INFECTIONS TO DETERMINE FACTORS OF NEUROTROPISM

Saw Seow Hoon

Mycobacterium tuberculosis (*Mtb*) is the causative agent of tuberculosis (TB) and the major mycobacterial species associated with central nervous system (CNS) infection. Tuberculous meningitis (TBM) is a global challenge due to difficulties with diagnosis, treatment and high mortality rate. It is not clear what factors promote CNS invasion and pathology in TB, but it has been reported that *Mtb* isolates from meningitis patients show specific genetic traits not found in respiratory isolates. Thus, in this study, genomic features of *Mtb* isolates from CSF and respiratory specimens were compared to identify factors that could contribute to neurotropism in the pathogenesis of CNS TB. First, suppression subtractive hybridisation (SSH) combined with T/A cloning technology was utilized to build a subtracted cDNA library of genes showing DNA sequence differences between two strains of *Mtb* from the CSF of Malaysian patients with meningitis and two strains from sputum samples. Subsequently, the whole genome sequences (WGS) of eight CSF strains from Malaysian patients presenting with TB meningitis (TBM) were analysed against the WGS of 30 to 85 respiratory strains, including the reference strain H37Rv, downloaded from public databases. All WGS were examined for gene

rearrangements, indels, inversions, micro-variants, as well as structural and globularity changes in proteins. In addition, they were searched for homologs of genes previously reported to be associated with meningitis in *Mtb* and three other neuropathogens, *Neisseria meningitidis*, *Streptococcus pneumoniae* and *Escherichia coli* K1.

With SSH, 7 subtracted DNA fragments of various sizes were generated. With the use of BLASTN and BLASTX, these subtracted sequences were shown to be similar to those found in *Mtb* strain isolated from the brain of an Indian patient. They encode mycobacterial PE/PPE proteins which play an important role in the evasion of host immune responses, and other conserved hypothetical proteins with unknown function.

Genome-wide comparisons revealed rearrangements (translocations, inversions, insertions and deletions) and non-synonymous single nucleotide polymorphisms (ns SNPs) in the CSF-derived strains that were not observed in the respiratory *Mtb* genomes used for comparison. These rearranged segments were rich in genes for PE/PPE, transcriptional and membrane proteins. Similarly, most of the ns SNPs common in CSF strains were found in genes encoding PE/PPE proteins. Protein globularity differences were observed among mycobacteria from both CSF and respiratory sources and in proteins previously reported to be associated with TBM. Transcription factors and other transcription regulators featured prominently in these proteins. Homologs of TBM-associated genes were found in the WGS of *Mtb*, *M. bovis*, *M. leprae*, *M. lepromatosis* and two environmental non-tuberculous

mycobacteria. However, only two of these homologs were identified in all eight CSF-derived *Mtb* genomes. Three homologs of 141 genes reported to be associated with *S. pneumoniae* meningitis and two of 164 virulence genes reported in *N. meningitidis* were also found in the CSF-derived *Mtb* genomes. The detection of common meningitis-associated genes in mycobacterial and non-mycobacterial neuropathogens raises speculations on the existence of a pan-bacterial mechanism of CNS infection. However, all the genes shared by CSF *Mtb* strains and other neuropathogenic bacterial spp. were also found to be common in the comparison respiratory *Mtb* genomes examined. Hence, although some genetic traits were found in CSF-derived but not sputum-derived *Mtb* examined in this study, there are indications that neurotropic traits may not be specifically or universally found in *Mtb* causing CNS disease. Overall, the findings from this study suggest that CNS infection in TB is more likely to be directed by the expression of multiple virulence factors selected by the interaction between pathogen and host immune responses, than the presence of specific genetic traits.

ACKNOWLEDGEMENTS

First and foremost, I offer my sincere gratitude to my advisor Professor Ngeow Yun Fong, for her continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge whilst allowing me the room to work in my own way. I can never feel lucky and be thankful enough to have met a mentor like her, who has never failed to show her kindness and offered help at needy times, be it work or personal wise. My heartfelt appreciation to her can never be expressed enough in words. My sincere thanks to Professor Dr. Yap Sook Fan, my co-supervisor, Dr. Chan Kok Gan and Dr. Phelim Yong Voon Chen, for all their advices and words of wisdom in my research studies. Special thanks to the Faculty of Science at Universiti Tunku Abdul Rahman (UTAR) for providing me with a sufficient and comfortable work bench to conduct my experiments, as well as to my colleagues, laboratory officers and my dearest students of the faculty, for their empathy, assistance, cooperation and their endless support, spiritually and mentally, for making it possible for me to conduct my research here. I would like to acknowledge the program of UTAR Scholarship Schemes for sponsoring my tuition fee during my candidature in UTAR, which greatly reduced the financial burden of my family. I would also like to acknowledge the UTAR Research Fund (RF) Cycle 1 2013 (Project no: IPSR/RMC/UTARRF/C1-13/S02) which helped to finance this research project.

I am so grateful and indebted to Aunty Shirly, Sarjit, Hien Fuh, Yan Ling, Isaac, Joon Liang and Xin Yue, for their effort in ensuring a smooth sailing for my research, be it in providing clinical samples or lending materials and reagents for my research, or in giving their invaluable advice, moral support and countless help, to keep me sane during the final and writing-up months. My utmost gratitude goes to my best friends and soul mates who never failed to be there for me at difficult moments, providing me shoulders to lean on. These people are none other than Yun Li, Leon, Eddy and Wen Jie.

Life would have been different without you guys, and I thank God for the eternal friendship and great bond between us.

Last but not least, most important of all, an extra-large thank you to my family, especially my twin sister, Seow Hui, who has been a motherly figure, nurturing my needs over these years, and thank you for being such a great and caring sister cum friend to me, saying encouraging words and cheering me up at difficult and stressful times. To my dearest mom, who has suffered from Alzheimer's disease: thank you is just too small a word to describe how much I appreciate you. Your love and reassuring smile has made my life worthwhile. You are indeed my root and my foundation that I base my life on. Through life's rough patches, you still make me feel safe and secure. You have been nothing less than an angel for sure. Thanks mom! I love you, dearly.

Finally, this thesis is heartily dedicated to my dad who left for heaven before the completion of this work. Dad, we are no longer physically in touch, but I know you are watching me from above, showering me with blessings and making my wishes come true.

May you rest in peace, dad. You will be missed.

APPROVAL SHEET

This dissertation/thesis entitled **"COMPARATIVE GENOMIC ANALYSIS OF *Mycobacterium tuberculosis* FROM PULMONARY AND CENTRAL NERVOUS SYSTEM INFECTIONS TO DETERMINE FACTORS OF NEUROTROPISM"** was prepared by SAW SEOW HOON and submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy in Medical Sciences at Universiti Tunku Abdul Rahman.

Approved by:

_____
(Prof. Dr. NGEOW YUN FONG)
Date: 4th July 2019
Professor/Supervisor
Department of Pre-clinical Science
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman

_____
(Prof. Dr. YAP SOOK FAN)
Date: 4th July 2019
Professor/Co-supervisor
Department of Pre-clinical Science
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman

FACULTY OF MEDICINE AND HEALTH SCIENCES

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 4th July 2019

SUBMISSION OF ~~FINAL YEAR PROJECT /DISSERTATION/~~THESIS

It is hereby certified that **SAW SEOW HOON** (ID No: **1507058**) has completed this ~~final year project/ dissertation/~~ thesis* entitled "COMPARATIVE GENOMIC ANALYSIS OF *Mycobacterium tuberculosis* FROM PULMONARY AND CENTRAL NERVOUS SYSTEM INFECTIONS TO DETERMINE FACTORS OF NEUROTROPISM" under the supervision of Professor Dr. Ngeow Yun Fong (Supervisor) and Professor Dr. Yap Sook Fan (Co-Supervisor), from the Department of Pre-clinical Sciences, Faculty of Medicine and Health Sciences.

I understand that University will upload softcopy of my ~~final year project / dissertation/~~ thesis* in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

_____

(Saw Seow Hoon)

*Delete whichever not applicable

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

SAW SEOW HOON

Name _____

4th July 2019

Date _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND DESIGNATIONS

A_{260}	The wavelength of maximum absorption for both DNA and RNA
A_{280}	Amino acids with aromatic rings are the primary reason for the absorbance peak at 280 nm. Peptide bonds are primarily responsible for the peak at 200 nm
$A_{260/280}$	The ratio used to assess the purity of DNA and RNA
aa	Amino Acid
ABC	ATP-Binding Cassette
Ala (A)	Alanine (GCT, GCC, GCA, GCG)
Arg (R)	Arginine (CGT, CGC, CGA, CGG, AGA, AGG)
ASCII	American Standard Code for Information Interchange
Asn (N)	Asparagine (AAT, AAC)
Asp (D)	Aspartic Acid (GAT, GAC)
ATCG	Adenine; Thymine; Cytosine; Guanine
*Alu*I	Restriction enzyme recognizes AG^CT sites (an *E. coli* strain that carries the cloned *Alu*I gene from *Arthrobacter luteus*)
*Apo*I	A unique restriction endonuclease from *Arthrobacter protophormiae* which recognizes 5' RAATTY 3'
α	Alpha
Amp	Ampicillin
ß	Beta
Ba^{2+}	Barium cation
BACTEC	Becton Dickinson's fully automated system for the rapid detection of mycobacteria in clinical specimens
BALB/c	An albino, laboratory-bred strain of the house mouse

BBB	Blood Brain Barrier
BCG	Bacillus Calmette–Guérin
BCSFB	Blood-Cerebrospinal Fluid Barrier
BLAST	Basic Local Alignment Search Tool
BLASTN	BLAST program that uses nucleotide databases using a nucleotide query
BLASTP	BLAST program that searches protein subjects using a protein query
BLASTX	BLAST program that searches protein databases using a translated nucleotide query
BMEC	Brain Micro-vascular Endothelial Cells
BSL-3	Biological Safety Level-3
Buffer PE	Wash buffer for use in DNA clean-up procedures
Buffer QG	Solubilisation and binding buffer (with pH indicator), for use in DNA clean-up procedures
BWA	Burrows-Wheeler Aligner, a software package for mapping low-divergent sequences against a large reference genome, such as the human genome
$CaCl_2$	Calcium Chloride
CD	Cluster of Differentiation
CDC	Centre of Disease Control
cDNA	Complementary Deoxyribonucleic Acids
CFP	Culture Filtrate Protein
cfu	Colony forming units
$C_4H_6CaO_4$	Calcium Acetate
$C_2H_3NaO_2$	Sodium Acetate
CNS	Central Nervous System
Conc.	Concentration
CR3	Complement Receptor 3
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSF	Cerebrospinal Fluid
Cys (C)	Cysteine (TGT, TGC)

cysA1	Sulfate-transport ATP-binding protein ABC transporter
DAVID	Database for Annotation, Visualization and Integrated Discovery
DC	Dendritic Cells
DC-SIGN	Dendritic Cells Specific Intercellular Adhesion Molecule-3 Grabbing Non-Integrin
DNA	Deoxyribonucleic Acid
DNAses	Deoxyribonuclease
dNTPs	Deoxy-Nucleoside Triphosphates
DR	Direct Repeat
*Dra*I	Restriction endonuclease that recognizes the sequence TTT^AAA (isolated from *Deinococcus radiophilus*)
DVR	Direct Variable Repeats
E	Expect value
echA11	Enoyl-CoA hydratase
E. coli	*Escherichia coli*
*Eco*RI	Restriction enzyme recognizes G^AATTC sites (isolated from *E. coli*)
EDTA	Ethylene Diamine Tetraacetic Acid
e.g.	exempli gratia in Latin, which means “for example”
embR	Transcriptional regulatory protein
*end*A	Endonuclease gene
ESAT-6	Early Secretory Antigenic Target 6 kDa
Esx	ESAT-6 protein secretion systems
EtBr	Ethidium Bromide
fadD18	Fatty-acid-CoA ligase
FASTQ	Text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores
FASTQC	Quality control checks on raw sequence data coming from high throughput sequencing pipelines
Fwd	Forward

γ	Gamma
galTb	Probable galactose-1-phosphate uridylyltransferase
gDNA	Genomic DNA
GlobPlot	Exploring protein sequences for globularity and disorder
Gln (Q)	Glutamine (CAA, CAG)
Glu (E)	Glutamic Acid (GAA, GAG)
GTPase	Guanosine Triphosphate hydrolase enzyme
Gly (G)	Glysine (GGT, GGC, GGA, GGG)
H1	Adaptor 1-ligated tester DNA
H2	Adaptor 2R-ligated tester DNA
*Hae*III	Restriction enzyme recognizes GG^CC sites (An *E. coli* strain that carries the *Hae*III gene from *Haemophilus aegypticus*)
HBHA	Heparin-Binding Haemagglutinin Adhesion
HCl	Acid Hydrochloride
HIR	High Impact Research
His (H)	Histidine (CAT, CAC)
HIV	Human Immunodeficiency Virus
H_2O	Water
Hsp	Heat shocked protein
HT1	Hybridisation Buffer
i5	Index 2 adapter
i7	Index 1 adapter
ID	Identification
IDBA-UD	Algorithm that is based on the de Bruijn graph approach for assembling reads from single-cell sequencing or metagenomic sequencing technologies with uneven sequencing depths
i.e.	*id est*, a latin word means ‘that is’ (to say)
IFN	Interferon
IL	Interleukin
Ile (I)	Isoleucine (ATT, ATC, ATA)

inGAP-SV	A novel scheme to identify and visualize structural variation from paired end mapping data
IPTG	Isopropylthio-β-galactoside
IS	Insertion Element
iVEGI	*in-vivo*-expressed genomic island
JM109	Competent *E. coli* cells and it is recombination (*recA*) deficient, lack *E. coli* K restriction system which are appropriate for routine cloning applications
KO	Knockout
*lac*Z	Gene that encodes for β-galactosidase proteins
LAM	Lipoarabinomannan, the complex glycolipids in the cytoplasmic membrane of mycobacteria
LB	Luria-Bertani
LED	Light Emitting Diode
Leu (L)	Leucine (CTT, CTC, CTA, CTG, TTA, TTG)
LGT	Lateral Gene Transfer
LHP	Alternate gene name for CFP-10
LJ	Lowenstein–Jensen medium
LM	Lipomannan is a mycobacterium immune agonist
lppD	Lipid phosphate phosphatase delta
Lys (K)	Lysine (AAA, AAG)
ManLAM	Mannose-capped LAM found in the cell wall of mycobacteria
Mce	Mammalian cell entry protein
MCS	Multiple Cloning Site
MDR-TB	Multidrug-resistant tuberculosis, defined as resistance to at least isoniazid and rifampicin, the two most powerful anti-TB drugs
MEGA	Molecular Evolutionary Genetic Analysis
Met (M)	Methionine (ATG/GTG)
Mg^{2+}	Magnesium cation
$MgCl_2$	Magnesium Chloride
$MgSO_4$	Magnesium Sulphate

MGIT	Mycobacteria Growth Indicator Tube
MHC	Major Histocompatibility Cells
MIDI	96-well storage plates, round well, 0.8 ml
MIRU-VNTR	Mycobacterial Interspersed Repetitive Unit-Variable Numbers of Tandem Repeats
MLSA/MLST	Multi-Locus Sequence Analysis/Typing
MMU	Multimedia University
Mn^{2+}	Manganese cation
Mtb	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
NCBI	National Centre for Biotechnology Information
NDP	Nextera Dilution Plate
NGS	Next Generation Sequencing
NH_4OA_C	Ammonium Acetate
NPP	Nextera Pooled Plate
NST	Nextera Sample Tube
NTM	Non-tuberculous Mycobacteria
OD_{260}	Optical Density at 260 nm
ORF	Open Reading Frame
PAMP	Pathogen-associated molecular patterns
PBS	Phosphate-Buffer Saline
PCI	Phenol:Chloroform:Isoamyl Alcohol (25:24:1)
PCR	Polymerase Chain Reaction
PE/PPE/PGRS	Proline-glutamic acid/proline-proline-glutamic acid/ Proline polymorphic GC-repetitive sequences
pGEM-T Easy Vector	Linearized vectors with a single 3´-terminal thymidine at both ends
pGEM®-*luc* Vector	Cassette vector that is the source of the *luc* gene encoding firefly luciferase
PGL	Phenolic Glycolipid
pH	Negative log of the activity of hydrogen ion in an aqueous solution

Phe (F)	Phenylalanine (TTT, TTC)
PI	Phosphatidylinositol, the phospholipid in the cytoplasmic membrane of mycobacteria
PIM	Phosphatidyl Inositol Mannoside, the complex glycolipids in the cytoplasmic membrane of mycobacteria
pks	polyketide synthase gene
PLM	Phosphatidylinositol Mannoside
PMN	Polymorphonuclear Neutrophils
PR2	Incorporation buffer
Pro (P)	Proline (CCT, CCC, CCA, CCG)
PRR	Pattern Recognition Receptors
PTB	Pulmonary Tuberculosis
pUC/M13	pUC plasmids and M13 vectors used for molecular cloning purposes
ϕX174	phi X 174 bacteriophage
RacH	Protein Rho GTPase
RD	Region of Difference
RE	Restriction Enzyme
*rec*A	Recombination gene
Rev	Reverse
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNAses	Ribonuclease
rRNA	Ribosomal Ribonucleic Acid
*Rsa*I	Restriction enzyme recognizes GT^AC sites (an *E. coli* strain that carries the *Rsa*I gene from *Rhodopseudomonas sphaeroides*)
RSB	Re-suspension Buffer
RTA	Real-time Analysis
Rv	ID for the ORF/code number assigned
*Sac*I	Restriction enzyme recognizes GAGCT^C sites (isolated from *Streptomyces achromogenes*)

SAMtool	Sequence Alignment/Map format is a generic format for storing large nucleotide sequence alignments
SAS	Sub-arachnoid Space
Sau3AI	Restriction enzyme recognizes ^GATC sites (An *E. coli* strain that carries the cloned Sau3Al gene from *Staphylococcus aureus 3A*)
SAV	Sequencing Analysis Viewer
Sdn Bhd	Sendirian Berhad
SDS	Sodium Lauryl Sulfate
Ser (S)	Serine (TCT, TCC, TCA, TCG, AGT, AGC)
SL	Sulfolipid
SNP	Single Nucleotide Polymorphism
snzP	Pyridoxine biosynthesis protein
SNV	Single Nucleotide Variable
S.O.C.	Super Optimal Broth with Catabolite Repression
sp.	species
Sr^{2+}	Strontium cation
SRA	Sequence Read Archive
SSH	Suppression Subtractive Hybridisation
SSPACE	Standalone Scaffolder of Pre-assembled Contigs using paired-read data
*Sst*I	Restriction endonuclease recognize and cleave the sequence 5' GAGCTC 3' (purified from an *E. coli* strain that carries the *Sst*I gene from *Streptomyces Stanford*)
Stop	Stop codon (TAA, TAG, TGA)
STPK	Serine-Threonine Protein Kinase
T/A	A linearized “T-vector” which has single 3'-T-overhangs on both ends allows direct, high-efficiency cloning of PCR products, facilitated by complementarity between the PCR product 3'-A-overhangs and vector 3'-T-overhangs
TAE	Tris-Acetate-EDTA

TB	Tuberculosis
TBE	Tris-Borate-EDTA
TBM	Tuberculous Meningitis
T-cells	Subtype of white blood cells
TD	Tagment DNA
TDE	Tagment DNA reagent
TDM	Trehalose-dimycolate or cord factor
TDR-TB	Totally drug resistant TB, defined as resistant to both first line and second line antibiotics
TE	Transformation efficiency
TE buffer	Tris-EDTA
Th1	T-helper cells 1 drive the type-1 pathway ("cellular immunity") to fight viruses and other intracellular pathogens, eliminate cancerous cells, and stimulates delayed-type hypersensitivity (DTH) skin reactions.
Th2	T-helper cells 2 drive the type-2 pathway ("humoral immunity") and up-regulate antibody production to fight extracellular organisms
Thr (T)	Threonine (ACT, ACC, ACA, ACG)
T4	Enterobacteria phage that infects *Escherichia coli* bacteria
TLC	Thin Layer Chromatography
TLR	Toll-like Receptor
T_m	Melting temperature
tmRNA	Transfer-messenger RNA
TNF	Tumour Necrosis Factor
tRNA	Transfer RNA
Tris-HCl	Tris-Hydrochloride
Trp (W)	Tryptophan (TGG)
Tyr (Y)	Tyrosine (TAT, TAC)
UK	United Kingdom
UMMC KL	University Malaya Medical Centre Kuala Lumpur
UV	Ultraviolet
Val (V)	Valine (GTT, GTC, GTA, GTG)

XDR-TB	Extensively Drug-resistant Tuberculosis, defined as MDR-TB plus resistance to at least one fluoroquinolone and a second-line injectable
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
%	Percent
x g	Units of times gravity
°C	Degree Celsius
RT	Room temperature
bp	Base pair
kbp	Kilobase pair
kDA	Kilo Dalton
h	Hour(s)
min	Minute(s)
s	Second(s)
g	Gram
µg	Microgram
ng	Nanogram
L	Litre
ml	Millilitre
µl	Microliter
M	Molar
mM	Millimolar
µM	Micromolar
nM	Nanomolar
pM	Picomolar
mg/ml	Milligram per millilitre
µg/ml	Microgram per millilitre
ng/µl	Nanogram per microliter
units/µl	Units per microliter
Weiss units/µl	Amount of enzyme required to catalyse the exchange of 1 nmole of ^{32}P-labelled inorganic pyrophosphate into Norit adsorbable material in 20 min at 37 °C
v/v	Volume per volume

w/v	Weight per volume
V	Voltage
U	Enzyme unit; One U is defined as the amount of the enzyme that produces a certain amount of enzymatic activity, that is, the amount that catalyses the conversion of 1 µM of substrate per minute

CHAPTER 1

INTRODUCTION

1.1 Background

Tuberculosis (TB) is an ancient disease affecting almost all organs in the human body. The causative agent *Mycobacterium tuberculosis* (*Mtb*) is inhaled into the pulmonary alveoli where it is phagocytosed by alveolar macrophages. The bacilli that survive the hostile intracellular environment either cause local lesions or are disseminated to extra-pulmonary sites (Donald et al., 2005; Be et al., 2009; Bini Estela and Hernandez Pando, 2014). It is still unclear what factors (microbial or host) determine a confined pulmonary lesion (usual presentation) or dissemination to extra-pulmonary sites. In particular, the exact mechanisms of central nervous system (CNS) infection are poorly understood (Sundaram et al., 2011).

Both host susceptibility factors and specific mycobacterial genetic traits have been implicated in CNS infection. The host association has been studied extensively and numerous clinical studies have reported the greater risk of CNS infection in immune-compromised individuals (Yang et al., 2004; Vinnard and MacGregor, 2009; Elmas et al., 2011; Sáenz et al., 2013). In addition, various polymorphisms in human genes have been identified to be associated specifically with susceptibility to *Mtb* meningeal infection (Hoal-

Van Helden et al., 1999; Hawn et al., 2006; Campo et al., 2015). With respect to the mycobacterium, five *Mtb* genes (Rv0311, Rv0805, Rv0931c, Rv0986, MT3280) were identified to be associated with the invasion of CNS but not lung tissues (Be et al., 2008). In particular, the sensor domain of *Mtb pknD* (Rv0931c) was reported to be able to attack brain endothelia but not lung epithelia (Be et al., 2012). The genetics of neurotropism has also been described in other meningitis-causing agents such as HIV (Dunfee et al. 2006; Eugenin et al. 2006), *Escherichia coli* (Kim, 2006), *Neisseria meningitidis* (Tinsley and Nassif, 1996; Coureuil et al., 2012), *Mycobacterium leprae* (Shimoji et al., 1999; Singh and Cole, 2011), fungal and parasitic pathogens (Matsuura et al., 2000; Brown et al., 2014). However, data on the factors and mechanisms underlying neurotropism in mycobacterial infections is still lacking. New knowledge is needed to augment current understanding of TB meningitis and the *Mtb* strain-specific traits that are related to this pathology. Uncovering microbial genetic factors that are associated with neurotropism in *M. tuberculosis* might also lead to the future development of new diagnostics, improved therapeutics and novel vaccines against TB.

1.2 Hypothesis

In this study, it is hypothesised that certain microbial traits in *Mtb* strains predispose to neurotropism.

1.3 Research Objectives

The overall objective of this study is to compare the genomic features of *Mtb* isolates from patients with CNS and respiratory TB to identify factors that could contribute to neurotropism in the pathogenesis of CNS TB.

The specific objectives are:

1) To use suppression subtractive hybridisation (SSH) to build a subtracted cDNA library that show differences in DNA sequence in CSF and pulmonary *Mtb* isolates.
2) To use whole genome sequence analysis to provide a wider screen for genetic elements that could possibly be associated with neurotropism in *Mtb.*

CHAPTER 2

LITERATURE REVIEW

2.1 *Mycobacterium tuberculosis* (*Mtb*)

Mycobacterium tuberculosis (*Mtb*) is an acid-fast bacillus which causes the majority of tuberculosis (TB) globally. It is a member of the *Mycobacterium tuberculosis* complex (MTBC) consisting of *Mtb*, *M. africanum*, *M. bovis*, *M. canettii*, *M. microti*, *M. caprae* and *M. pinnipedii* (Smith et al., 2006; Bouakaze et al., 2010). These species differ in their host tropism, phenotype and pathogenicity. Nevertheless, all members of the MTBC are closely related genetically (Sreevatsan, 1997).

Mtb cells appear slender, slightly curved or straight rods which rarely branch measuring 0.2 to 0.7 by 1.0 to 10.0 µm in size. They are slow growing mycobacteria, aerobic, non-motile and do not have capsules. Mycobacteria have unusually high lipid content relative to other bacteria. Genomic studies have reported that large number of genes is involved in lipid metabolism in their genomes (Cole et al., 1998). These lipids form integral part of the cell wall, which is rich in long chain α-alkyl, ß-hydroxy fatty acids termed mycolic acids (Brennan and Nikaido, 1995; Minnikin et al., 2002; Forrellad et al., 2013). Mycolic acids are major and specific lipid components of the mycobacterial cell envelope and are essential for the survival of members of

the genus *Mycobacterium* that contains the causative agents of both tuberculosis and leprosy (Marrakchi et al., 2014). Mycolic acids might be the critical determinant for the emergence of multidrug-resistant (MDR), extensively drug-resistant (XDR) and totally drug-resistant (TDR) strains of *Mtb* (WHO, 2000; WHO, 2018).

TB control continues to be unsatisfactory due to multilateral migration and globalisation (Raviglione et al., 1995), as well as the co-epidemic of TB and acquired immunodeficiency syndrome (AIDS). TB causes both pulmonary TB and a wide range of extra-pulmonary diseases such as tuberculous meningitis (TBM), endometritis, lymphadenitis, and osteomyelitis (Fanning, 1999; Golden and Vikram, 2005; Marx and Chan, 2011; Thwaites et al., 2013; Willcocks and Wren, 2014). It is one of the top 10 causes of death and affects millions of people every year.

2.2 Current Status of TB Epidemic

Generally, only a small proportion (5 - 15 %) of those infected will manifest clinical TB (WHO, 2016; World Health Organization Regional Office for South-East Asia, 2017). In immunocompromised or malnourished individuals, however, the risk of clinical TB may be as high as 50 % or more (Thwaites, 2013). In 2016, TB caused 0.4 million deaths among HIV-positive patients out of more than six million TB cases reported globally (WHO, 2017).

The overall number of TB deaths world-wide in 2017 was 1.6 million in 6.4 million TB notified cases (WHO, 2018). In Malaysia, although the number of TB cases reported declined 1.3 % in the year 2018, the mortality rate due to TB increased to 6.6 per 100,000 population, compared to 6.5 per 100,000 population in 2017. Sabah recorded the highest number of deaths with 376 cases, followed by Selangor at 375 cases (Bernama, 2019).

Extra-pulmonary infections were reported in approximately 14 % of immune-competent individuals among notified TB cases world-wide; In Malaysia, 15 % of the notified cases were reported to have extra-pulmonary dissemination (WHO, 2018). TB dissemination to the brain causes tuberculous meningitis (TBM), the worst form of extrapulmonary TB. The high mortality and morbidity caused by TBM has drawn serious attention to the world to develop effective vaccines, optimal antibiotic and host-directed therapies (Marais et al., 2011). In order to address these strategies, further understanding of the pathologic mechanisms of TBM is required.

2.3 The Immune Response against *Mtb*

Phagocytic cells, such as macrophages, neutrophils and dendritic cells (DCs) serve as the first line of defence in the innate immune response to control mycobacterial infection (Davis et al., 2019). *Mtb* is transmitted by individuals with pulmonary disease who release *Mtb*-containing aerosols into the air. On inhalation, these aerosols are deposited in the alveoli where the *Mtb* are engulfed by alveolar macrophages that recognise the pathogen-

associated molecular patterns (PAMP) of *Mtb* via pattern recognition receptors (PRRs). These receptors include complement (CR3), opsonizing receptors, toll-like receptors (TLR 2, TLR 4 and TLR 9), scavenger receptor, C-type lectin receptor, mannose, immunoglobulin Rc and nucleotide oligomerisation receptors (Glickman and Jacobs, 2001; Pieters and Gatfield, 2002; Nguyen and Pieters, 2005; González-Cano et al., 2010; Kleinnijenhuis et al., 2011). TLRs recognise mycobacterial structures (Brightbill et al., 1999; Thoma-Uszynski et al., 2001) and are involved in the production of peptides and defensins against *Mtb* (Liu et al., 2006). In general, the binding of mycobacterial ligands with PRRs lead macrophages and DCs to produce various antimicrobial peptides, cytokines, including IL-1α and IL-1β, TNF-α, IL-6 and -12, chemokines, lipoxins that may stimulate necrosis and enhance the immune protection against *Mtb,* and prostaglandins that may activate apoptosis (O'Garra et al., 2013). Thus, the interactions between *Mtb* and the host may result in necrosis or apoptosis leading to host cell damage and *Mtb* dissemination or death of both host cell and *Mtb* (Lamkanfi and Dixit, 2010; Sáenz et al., 2013). In some individuals, the host innate immunity is able to kill all the inhaled bacteria and terminate the infection (Dannenberg and Rook, 1994; Dietrich and Doherty, 2009).

In the majority of cases, however, the bacteria can survive and replicate, infecting other macrophages, such as the DCs, with the influence of IL-12 and chemokines CCL-19 and -21, in which they spread via the lymphatic system to the hilar lymph nodes to activate the differentiation of T-helper 1 (Th-1) cells. Th-1 cells release IFN-γ and TNF-α at the site of

infection (Wallgren, 1948). This leads to the activation of macrophages and DCs to produce more cytokines and antimicrobial factors that contribute to containment of the TB bacillus, which results in the formation of granuloma (Dannenberg and Rook, 1994; Tufariello et al., 2003; Sasindran and Torrelles, 2011; Miranda et al., 2012; O'Garra et al., 2013). These cell-mediated immune responses happen after 2 to 8 weeks of infection.

Within the granuloma, the *Mtb* is confined (Flynn et al., 2011; Miranda et al., 2012) but is able to survive for long periods of time (Adams, 1976; Sandor et al., 2003). Cytokines play an important role in the formation of the granuloma (Kirschner et al., 2010; Cooper et al., 2011). In TNF-α–neutralized mice, granulomas tend to be disorganized (Beham et al., 2011; Olleros et al., 2015); IFN-γ knockout (KO) mice show necrotic granulomas and high susceptibility to *Mtb* infection (Cooper et al., 1993; Dalton et al., 1993). The bacteria dormant in granulomas can be reactivated by external stimuli, such as HIV infection and malnutrition. Under these conditions, the granuloma structure disintegrates and allows the release of the *Mtb* and its spread to other parts of the body including the brain and spinal cord (Thwaites et al., 2000; Thwaites, 2013).

2.4 Tuberculous Meningitis (TBM)

TB dissemination to extrapulmonary sties via bacteraemia and lymphatic drainage has been observed, especially to the sites with high concentration of oxygen such as the brain. Central nervous system (CNS)

disease caused by *Mtb* is a highly devastating although uncommon manifestation of TB, accounting for about 1 % of active TB cases worldwide (WHO, 2016; CDC, 2016). CNS TB often involves the meninges, the brain parenchyma and the spinal cord. Most patients with TB meningitis (TBM) experience permanent neurologic sequelae and considerable disability (Bishburg et al., 1986; Leonard and Des Prez, 1990; Udani, 1994). Generally, affected persons do not survive these complications, especially in drug-resistant infections and where the permeability of the blood-brain barrier (BBB) to anti-TB drugs is compromised (Gandhi et al., 2006; Shah et al., 2007).

2.5 Clinical Features of TBM

Tuberculous infection of the CNS may manifest as tuberculous meningitis (TBM), cerebral tuberculomas, tuberculous brain abscess and spinal tuberculosis (Schaller et al., 2019).

2.5.1 Tuberculous Meningitis (TBM)

TBM is the most common manifestation in CNS TB. TBM can progress to confusion, coma, and death if treatment is not started immediately. Progressively worsening headache, fever and vomiting could occur over days to weeks. Basal meningeal inflammation, fibrotic adhesions, cranial nerve palsies, cerebral infarction, and stroke are some of its neurological manifestations. In 30 % of cases, infarcts in the internal capsule and basal

ganglia occur resulting in a range of problems from paralysis to movement disorders. Intracranial pressure, hydrocephalus and reduced conscious level are the results from the obstruction of CSF flow. Seizures occur in about 30 % of children and 50 % of adults with TBM and permanent paralysis of limbs occurs in 10 % of cases as a consequence of spinal cord involvement (Chatterjee, 2011).

The diagnosis of TBM is challenging due to its non-specific symptoms and often inconclusive laboratory results (Bernaerts et al., 2003). Typical CSF analysis results include increased leukocytes (10 - 1000 x 10^3 cells/ml) and protein content (0.5 – 3.0 g/l) but decreased level in glucose (< 45 mg/dl) (Jeren and Beus, 1982; Bishburg et al., 1986; Sutlas et al., 2003; Rock et al., 2008; Thwaites et al., 2009; García-Moncó and Rodriguez-Sainz, 2018). In addition, identification of *Mtb* in CSF through both smear and culture methods, such as Ziehl-Neelsen staining for acid-fast bacilli is necessary, even though these methods are slower and the sensitivity achieved is only 5 – 58 % (Moghtaderi et al., 2009). Molecular assays are preferable as they facilitate early diagnosis (Rock et al., 2008; Ropper et al., 2014). Other than CSF, tissues from biopsies of focal lesions within or without the CNS can contribute to the diagnosis of TBM and neuroimaging is sensitive for the detection of basal exudates, hydrocephalus and infarctions in TBM patients (Garg et al., 2016).

The typical treatment of TBM is a two-month course of isoniazid (INH), rifampicin (RMP), pyrazinamide (PZA), streptomycin (SM) or

ethambutol (EMB), followed by a 6-month or 10-month course of INH and RMP (Thwaites et al., 2009; WHO, 2010). In addition, corticosteroids like dexamethasone and prednisolone have been shown to improve the survival rate in adults with TBM although the exact mechanisms are yet to be discovered (Thwaites, Bang, et al., 2004; Thwaites et al., 2007). However, treatment of TBM poses a continuous challenge especially with complications such as hydrocephalus, the rising incidence of multidrug-resistant (MDR-TB), extensively resistant (XDR-TB) or even the totally drug-resistant (TDR-TB) strains of *Mtb* (Streicher et al., 2012; Udwadia, 2012; Török, 2015).

2.5.2 Cerebral Tuberculomas

Tuberculomas can occur as space-occupying lesions anywhere in the CNS. Although an isolated seizure is the most common presentation, signs and symptoms are varied and dependent on the site of the brain affected. These granulomas can be fatal if they rupture and cause TBM (Monteiro et al., 2013). Neoplasms, primary central nervous system lymphoma, pyogenic abscess, fungal infection, cysticercosis and toxoplasmosis are the differential diagnosis for CNS tuberculoma (Luthra et al., 2007).

2.5.3 Tuberculous Brain Abscess (TBA)

TBA is similar to pyogenic brain abscess with the formation of cavity with central pus. It is larger than tuberculoma and is a rare manifestation of CNS TB (Chakraborti et al., 2009). Diagnosis is established with the demonstration of lipid peaks on magnetic resonance (MR) spectroscopy and positive microscopy and culture of pus following stereoscopic aspiration (Luthra et al., 2007).

2.5.4 Spinal Tuberculosis

TB affects all parts of the spine which classified as extradural and intradural. Extradural TB includes tuberculous spondylitis, paraspinal and epidural abscess (Jevtic, 2004), whereas intradural TB includes granulomatous leptomeningits, spinal tuberculoma and tuberculous myelitis. MRI findings allow a highly specific and sensitive differentiation of the manifestations (Baldwin and Zunt, 2014; Li et al., 2016; Weidauer et al., 2017).

2.6 From Lung to Brain: The Pathogenesis of Cerebral Tuberculosis

The alveolar epithelium is the first obstacle to the tubercle bacilli. In order to breach the epithelial layer, the bacteria have to invade and lyse the epithelial cells, or translocate across the cells, or travel within professional phagocytes. *Mtb* has been shown to spread within the epithelial layer through the necrosis and lysis of epithelial cells induced by cytotoxins (Byrd et al.,

1998; Dobos et al., 2000; Castro-Garza et al., 2002). Among the important bacterial factors produced by *Mtb* are a culture filtrate protein (CFP-10) encoded by *Rv3874* and an early secretory antigenic target 6 kDa (ESAT-6) encoded by *Rv3875*. Both genes belong to the region of difference 1 (RD1) gene cluster associated with *Mtb* virulence and the lysis of host cells (Hsu et al., 2003; Gao et al., 2004; Smith et al., 2008; Kinhikar et al., 2010).

In addition, *Mtb* are able to traverse epithelial cells without the need to pass through the extracellular compartment (Byrd et al., 1998). In this non-lytic intercellular spread, *Mtb* frees itself from the endosomal vacuole into the cytosol of the host cell via an actin-based structure called the ejectosome. This event precedes cell-to-cell passage and is governed by the RD1 locus with the involvement of secreted factors such as the ESAT-6 which has been associated with the formation of pores in macrophages (van der Wel et al., 2007; Hagedorn et al., 2009). The dissemination of *Mtb* in the alveolar epithelium is greatly facilitated by a heparin-binding haemagglutinin adhesion (HBHA) factor that enables endocytosis of the *Mtb* followed by transcytosis and subsequent entry into the bloodstream (Pethe et al., 2001; Locht et al., 2006; Menozzi et al., 2006). Table 2.6.1 provides further information on the bacterial factors linked with *Mtb* dissemination.

Table 2.6.1: Bacterial factors implicated for extra-pulmonary dissemination (Mycobrowser, 2019).

Rv number	Gene name	Product	Functional category	Relevant references
Rv3874	*esxB*	10 kDa culture filtrate antigen EsxB (LHP) (CFP10)	Unknown. Exported protein co-transcribed with *Rv3875.*	(Berthet et al., 1998; Hsu et al., 2003)
Rv3875	*esxA*	6 kDa early secretory antigenic target EsxA (ESAT-6)	Unknown. Stimulates the production of IFN-γ from memory cells in early immune response. Exported protein co-transcribed with *Rv3874.*	(Berthet et al., 1998; Hsu et al., 2003)
Rv0475	*hbhA*	Iron-regulated heparin binding haemagglutinin HBHA (adhesin)	Binds to sulphated glycoconjugates on the surface of epithelial cells to enable adherence of bacteria. Binds heparin, dextran sulfate, fucoidan and chondroitin sulfate. Promotes haemagglutination of erythrocytes of certain host species. Induces mycobacterial aggregation.	(Pethe et al., 2001; Locht et al., 2006; Menozzi et al., 2006)

Rv number is the ID for the ORF/code number assigned, the synonym used to index all tables in the TB database.

Although the interactions between *Mtb* and alveolar epithelial cells play a significant role in initiating bacterial dissemination to distant sites of the lung, professional phagocytic cells are also required. It is widely recognized that macrophages play a very important part in the intracellular survival of *Mtb* (Giacomini et al., 2001; Bermudez et al., 2002; Nguyen and Pieters, 2005; Houben et al., 2006). In addition to macrophages, dendritic cells (DC) also ingest *Mtb* but are less efficient in phagolysosomal digestion (Henderson et al., 1997; Orme and Cooper, 1999; Tascon et al., 2000; Bodnar et al., 2001; Humphreys et al., 2006). By carrying antigens to lymphoid organs, they help to prime the adaptive immune system of the host. In addition, they may also transport viable mycobacteria, providing them a safe haven where they can evade the host immune surveillance. The interaction between a DC receptor, the specific intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN), and mannosylated lipoarabinomannan (ManLAM) on the mycobacteria, results in secretion of IL-10 which inhibits antigen presentation, and expression of major histocompatibility complex (MHC) molecules, thereby deregulating the host immune mechanisms (van Kooyk and Geijtenbeek, 2003). Furthermore, the formation of DC-SIGN-ManLAM inhibits DC maturation (Jonuleit et al., 2000; Geijtenbeek et al., 2003; Tailleux et al., 2003; Fu et al., 2008). These two events act together to enhance the survival of *Mtb* within DCs and its subsequent dissemination via the lymphatic and blood circulation (Figure 2.6.1).

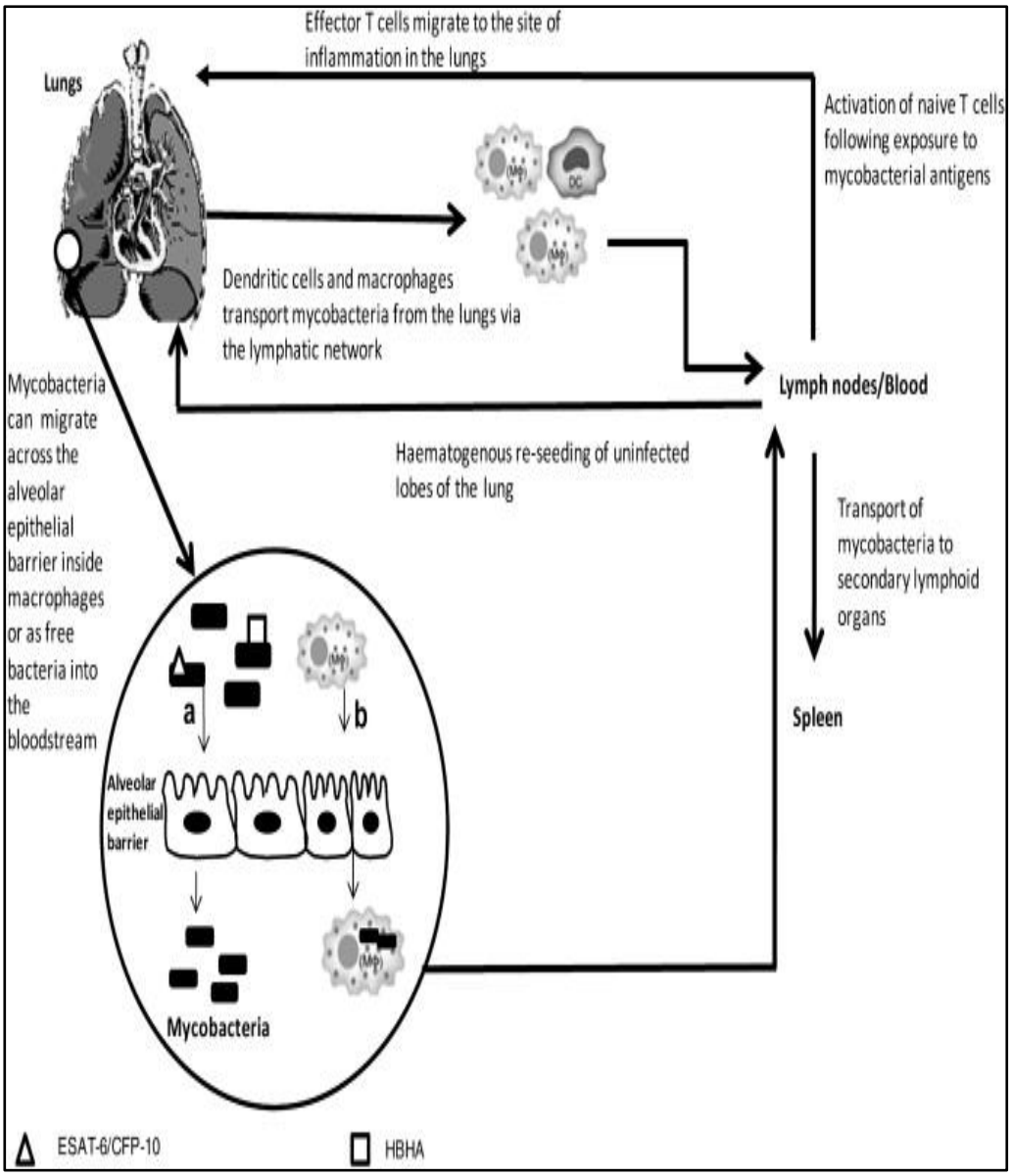


Figure 2.6.1: Mechanisms of extra-pulmonary dissemination in tuberculosis. Reproduced from Krishnan et al., (2010) with permission from Elsevier Ltd. Copyright © 2010.

To disseminate to the CNS, the *Mtb* must be capable of breaching the blood-brain barrier (BBB) and/or blood-cerebrospinal fluid barrier (BCSFB). The mechanism(s) by which the bacillus crosses this 'network' is not well studied. Although the brain micro-endothelial cells (BMEC) in the BBB are connected by tight junctions (Figure 2.6.2), *Mtb* is still able to cross the barrier,

and invade the brain, causing cerebral tissue destruction that is accompanied by excessive inflammation. Researchers have postulated that certain *Mtb* genes may encode certain microbial factors (virulence determinants) that promote BBB/BCSFB invasion and dissemination (de Viedma et al., 2006; Caws et al., 2008; Hernandez-Pando et al., 2010; Hesseling et al., 2010).

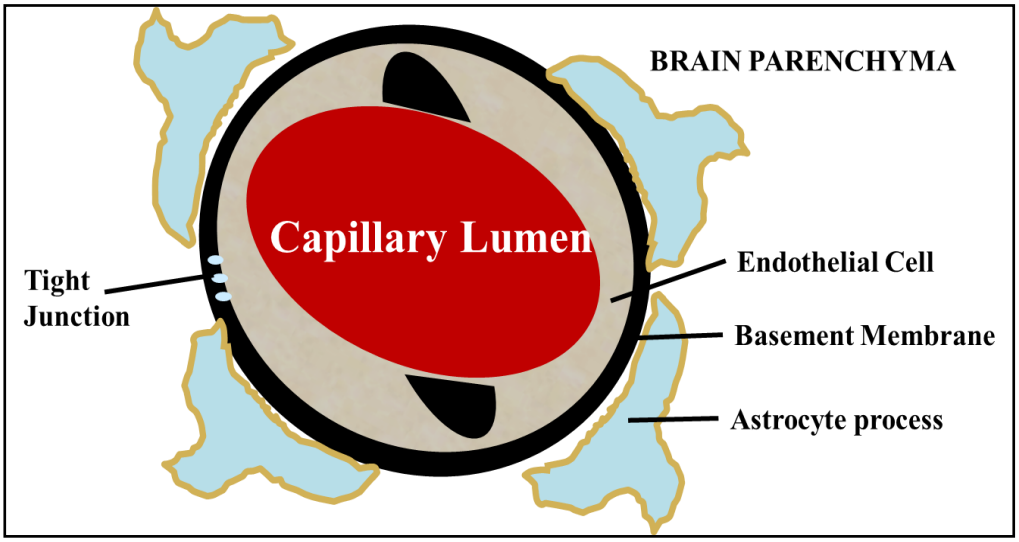


Figure 2.6.2: The BBB is composed of brain micro-vascular endothelial cells, basement membrane and astrocyte processes (Rubin and Staddon, 1999).

2.7 Meningitis-Associated Genes in *Mtb*

The ability of *Mtb* to invade and traverse the BMEC monolayers was demonstrated by Jain et al. (2006) who described the involvement of actin cytoskeletal rearrangements and bacterial translocation in the invasion process. The crossing of these monolayers was more efficient in *Mtb* compared to the less pathogenic *Mycobacterium smegmatis*. Of 33 genes involved in the bacterial translocation, 18 belonged to a set of genes (*Rv0960-Rv1001*) termed '*in-vivo*-expressed genomic island' (iVEGI) that was previously shown to be

up-regulated during mouse lung infection (Talaat et al., 2004). To further verify these genes in BBB invasion, isogenic transposon mutants of all the 33 genes were constructed and their phenotypes were examined in the BBB model (Jain et al., 2006). Mutations of three of these genes (*Rv0980c, Rv0987* and *Rv0989c*) and *Rv1801* reduced the ability of the mycobacterium to invade the BBB (Table 2.7.1).

Table 2.7.1: A summary of *Mtb* genes implicated in CNS invasion.

***Rv number**	**Gene name**	**Probable function**	**Effect of mutation in *Mtb* gene**	**References**
Rv0980c	*PE-PGRS18*	Predicted to be an outer membrane protein	Reduces ability to invade the *in vitro* BBB model.	(Jain et al., 2006)
Rv0987	*Rv0987*	Adhesion component of ABC transporter permease	Reduces ability to invade the *in vitro* BBB model.	(Jain et al., 2006)
Rv0989c	*grcC2*	Probable polyprenyl-diphosphate synthase	Reduces ability to invade the *in vitro* BBB model.	(Jain et al., 2006)
Rv1801	*PPE29*	Unknown	Reduces ability to invade the *in vitro* BBB model.	(Jain et al., 2006)
Rv0311	*Rv0311*	Unknown	Attenuates survival in mouse brain	(Be et al., 2008)
Rv0805	*CpdA*	Encodes for cyclic nucleotide phosphodiesterase	Attenuates invasion and survival in the mouse brain	(Be et al., 2008)
Rv0986	*Rv0986*	ATP-binding cassette (ABC) transporter	Attenuates invasion and survival in the mouse brain and reduces capacity	(Be et al., 2008)

			to invade macrophages and endothelial cells.	
Rv0931c	*pknD*	Serine-threonine protein kinase. Involved in signal transduction	Attenuates survival in the mouse brain. *pknD* required for brain endothelial invasion.	(Be et al., 2012)

**Rv number is the ID for the ORF/code number assigned, the synonym used to index all tables in the TB database.*

Two years later, Be and colleagues utilized the mouse model to study 28 *Mtb* genes that might be responsible for BBB invasion (Be et al., 2008). Two transposon mutants (*Rv0805* and *Rv0986*) of the clinical strain CDC1551 were observed to be associated with less efficient invasion of the mouse brain (Table 2.7.1). These genes encode for phosphodiesterase and ATP-binding cassette (ABC) transporter, respectively. The latter is part of a three-gene operon, *Rv0986-Rv0988*, two of which (*Rv0986* and *Rv0987*) were up-regulated during endothelial infection, but absent in non-pathogenic bacteria (Jain et al., 2006). Both genes (*Rv0986* and *Rv0987*) were postulated to be involved in cell adhesion and entry (Braibant et al., 2000). Three additional transposon mutants of CDC1551 (*Rv0311*, *Rv0805*, *Rv0931c*) were observed to be weakened for survival in the mouse brain (Table 2.7.1). In 2012, Be and colleagues further evaluated a panel of 398 transposon mutants of *Mtb* CDC1551 in guinea pigs. Fourteen showed CNS-specific attenuation and were unable to disseminate to the guinea pig brain. One of the 14, *pknD* (*Rv0931c*), encoding a serine-threonine protein kinase (STPK), was previously found to be particularly important in CNS dissemination in the murine model (Be et al.,

2008; Be et al., 2012). Thus, the *Mtb pknD* could be a potential target for the development of a novel vaccine against *Mtb* infection (Skerry et al., 2013).

Mtb strains causing meningitis have been reported to be associated with distinct genotypes (Arvanitakis et al., 1998; de Viedma et al., 2006; Caws et al., 2008; Hesseling et al., 2010). At the Indian National Institute of Mental Health and Neurosciences, several unique DNA sequences were shown to be present in *Mtb* isolates from cerebrospinal fluid. All these sequences have not been previously deposited in the database (of 23,000 strains from across the world) (NIMHANS, 2017). In addition, a study conducted in Columbia reported that BALB/c mice came down with meningitis when infected intra-tracheally with isolates from CSF but not with isolates from sputum (Hernandez-Pando et al., 2010).

2.8 Strain-Dependent Neurotropism of *Mtb*

It is often said that the virulence of an organism is the outcome of a complex and dynamic interaction between host and microbial properties (Casadevall and Pirofski, 2001). While it is well recognised that host factors such as old age, poor nutritional status and immunosuppressive disease, are important risk factors for TBM, the importance of bacterial factors in the pathogenesis of CNS disease, remains unclear. Many researchers have, however, observed apparent associations between *Mtb* genotypes or lineages and development of CNS TB.

Decades ago, studies in guinea pigs demonstrated that *Mtb* from the south of India were less virulent than strains from the UK (Mitchison and Wallace, 1960). More recent studies using mice and rabbits also showed variation in virulence in different *Mtb* genotypes (Dunn and North, 1995; Bishai et al., 1999; López et al., 2003; Dormans et al., 2004). With intra-cerebral and intra-cisternal injections of *Mtb* into rabbits, Tsenova et al., (2005) compared two virulent strains of the W/Beijing family (HN848 and W4) with *Mtb* CDC1551. They found a higher bacillary load at the infected site with HN878 and W4, as well as increased extra-CNS dissipation of bacilli, greater leucocytosis, and more severe clinical manifestations (Tsenova et al., 2005). The hyper-virulence of HN878 was attributed to a polyketide synthase-derived phenolic glycolipid (PGL) as it was shown that a mutant of HN878 with disrupted PGL synthesis caused fewer mice to die following infection (Reed et al., 2004).

Caws et al. (2006; 2008) showed a link between the Beijing genotype and HIV infection in Vietnamese patients with TBM and found the East Asian Beijing and Indo-Oceanic strains to be more often associated with TBM than strains from the Euro-American lineage (Caws et al., 2006; Caws et al., 2008). Reed et al. (2004) suggested that the virulence of the Beijing *Mtb* genotype is due to the presence of an intact polyketide synthase gene (*pks15/1*) that encodes a phenolic glycolipid (PGL). This glycolipid inhibits the release of some pro-inflammatory cytokines. It was thought that the *Mtb* of Euro-American lineages do not have the same propensity for extra-pulmonary TB including TBM because they do not have an intact *pks15/1* gene. However, It

was also observed that many *Mtb* of Beijing lineages do not produce PGL but are still capable of causing disseminated and CNS TB in humans (Reed et al., 2004; Sinsimer et al., 2008). Hence, the apparent link between the Beijing lineage and CNS TB remains unexplained.

Detailed characterization of selected *Mtb* strains from different lineages revealed a greater ability of East Asian/Beijing and Indo-Oceanic *Mtb* to induce TNF-α and IL-1ß from macrophages, compared to Euro-American strains (Krishnan et al., 2011). It was hypothesized that the difference in cytokine induction was related to the lipids in the *Mtb* cell wall. Further work on this hypothesis led to the discovery of lineage-associated lipid profiles. East Asian/Beijing strains were found to produce a phthiotriol dimycocerosate instead of the previously described phenolic glycolipid (PGL) (Reed et al., 2004; Sinsimer et al., 2008). Additionally, phenolphthiocerol dimycocerosate, a rare lipid produced only in *Mtb* with an intact *pks15/1*, appeared in all Indo-Oceanic strains (Krishnan et al., 2011). These findings suggest that *Mtb* lineage can affect early host-parasite interactions and thus, may influence the development of TBM and other TB manifestations in the CNS.

Other surface lipids, including the sulfolipid SL 1 (Converse et al., 2003; Domenech et al., 2004), lipoarabinomannan (Chan et al., 1991), the 19-kDa lipoprotein (Noss et al., 2001) and others (Smith, 2003), have also been studied for their possible contribution to the virulence of mycobacteria but the role they play in the pathogenesis of TB CNS disease is not known.

2.9 Genomic Variations in *Mtb*

Various DNA fingerprinting methods have been developed to study DNA polymorphism in *Mtb* (Van Soolingen, 2001; Mostrom et al., 2002). IS6110-RFLP, spoligotyping and MIRU-VNTR have been applied to show variations among different clinical isolates of *Mtb* (Brown et al., 2010; de Beer et al., 2012). For instance, Sola et al. (2001) defined the EAI lineage for the first time; Tsolaki et al. (2005) pointed out the region of deletions; Gagneux et al. (2006) suggested co-evolution and sympatry between human and bacterial populations; Filliol et al. (2006) revealed SNPs diversity; Comas et al. (2013) suggested an out-of-Africa history of TB; Gardy et al. (2011) performed the first WGS-based epidemiological study in Vancouver; Coll et al. (2014) created the wonderful PolyTB database project and provided the 62 SNPs barcode typing proposal; Merker et al. (2015) studied the Beijing lineage history; and Stucki et al. (2016) provided a referential frame for lineage 4 clinical isolates (Sola et al., 2001; Tsolaki et al., 2005; Filliol et al., 2006; Gagneux et al., 2006; Gardy et al., 2011; Comas et al., 2013; Coll et al., 2014; Merker et al., 2015; Stucki et al., 2016). These findings provided important insight into the genetic diversity and population structure of *Mycobacterium tuberculosis* complex (MTBC).

2.9.1 Strain Differentiation with Restriction Fragment Length Polymorphism (RFLP)

RFLP has been used reliably in the characterization of *Mtb* isolates and in the study of TB epidemiology (van Embden et al., 1993; Fandinho et al., 2000; Braden et al., 2001). The IS*6110*-specific sequence is the common target for RFLP analysis. It is a repetitive element of size 1.35 kb belonging to the IS*3* family of mobile insertion sequences (McAdam et al., 1990). Up to 25 IS*6110* sequences can be integrated into one *Mtb* genome to yield thousands of banding patterns among *Mtb* strains (McHugh and Gillespie, 1998). Several preferred integration loci or hotspots have been reported (Kato-Maeda et al., 2011). This fingerprinting technique has been used successfully to verify laboratory cross-contaminations (Small et al., 1993; Alland et al., 1994) and to identify small-scale outbreaks of TB and drug-resistant TB in a large variety of settings (van Rie et al., 1999; Kato-Maeda et al., 2011; Green et al., 2013).

RFLP with IS*6110* has been considered to be the gold standard in *Mtb* typing since the early 1990s. However, various studies have observed that its discriminatory power is limited for isolates with five or less copies of the insertion sequence (Yang et al., 2000; Soini et al., 2001). It is also disadvantaged by the requirement for relatively large amounts of DNA for restriction.

2.9.2 Strain Differentiation with Spoligotyping

Spoligotyping (Spacer Oligonucleotide typing) is based on the Direct Repeat (DR) locus first identified by Hermans et al., (1991) in *Mycobacterium bovis.* This locus is only found in members of the MTBC (Hermans et al., 1991). It contains many direct variant repeats (DVRs) each consisting of a 36-bp DNA sequence called the direct repeat followed by a unique spacer sequence of variable lengths (35 - 41bp) (Zhang et al., 2010). The absence or presence of DVRs make the DR locus highly polymorphic and it is this polymorphic feature that makes spoligotyping a useful tool for TB epidemiological studies (Groenen et al., 1993; Kamerbeek et al., 1997).

Spoligotyping is often used together with MIRU-VNTR to increase its discriminating power (Chaoui et al., 2014). It is also useful for the differentiation of *Mtb s*trains with insufficient IS*6110* bands for RFLP typing (Flores et al., 2010) and for phylogenetic analysis (Molina-Torres et al., 2010; Kandhakumari et al., 2015).

2.9.3 Analysis of Genomic Variations based on Mycobacterial Interspersed Repetitive Unit-Variable Numbers of Tandem Repeats (MIRU-VNTR)

In MIRU-VNTR typing, strain discrimination is based on different copies of tandem repeats in different strains (Weniger et al., 2010). The repetitive genetic elements called MIRUs, range from 40 to 100 base pairs in

size (Supply et al., 2001). Till date, 41 different MIRUs have been uploaded in the *Mtb* H37Rv chromosome database (Supply et al., 1997; Cole et al., 1998; Supply et al., 2000). For strain differentiation and for molecular epidemiologic studies as well as phylogenetic studies, 15 and 24 MIRU loci are used, respectively (Supply et al., 2006; Christianson et al., 2010; Chatterjee and Mistry, 2013; Jonsson et al., 2014). Hyper-variable VNTR targets are recommended to increase the discriminatory power for Beijing isolates (Iwamoto et al., 2007; Alonso et al., 2010). Allix-Béguec et al., (2014) found 24 loci MIRU-VNTR not sufficient to differentiate specifically Beijing isolates, but the use of seven hyper-variable MIRU-VNTR loci produced better resolution and reduced the clustering rate for the Beijing strains (Allix-Béguec et al., 2014).

Although the MIRU-VNTR method is considered another gold standard for typing *Mtb* isolates (Singh et al., 2007), it has been shown to give the same VNTR pattern for different genotypes, especially in the case of closely related isolates (Oelemann et al., 2007; Hanekom et al., 2008).

2.9.4 Genotyping based on Multi-locus Sequence Variations

Multi-locus sequence analysis (MLSA) is a DNA fingerprinting method based on the sequence analysis of up to 10 housekeeping genes (Jagielski et al., 2014). It provides insights into the phylogenetic relationships among members of a species, and is an ideal method for large-scale epidemiological studies and analyses of populations (Bougnoux and Morand,

2002; Tavanti et al., 2003; Chen et al., 2017). For instance, the use of MLSA provided a better understanding of the genetic divergence and evolution of *Mycobacterium avium* (Turenne et al., 2008) and facilitated the investigation of non-tuberculous mycobacteria (NTM) outbreaks in laboratory settings (Kim SY et al., 2017). With the use of MLSA, Hershberg and colleagues discovered a greater genetic diversity among the *Mtb* complex than generally recognized (Hershberg et al., 2008). Recently, a group of Korean scientists applied multi-locus sequence typing (MLST) to indicate lateral gene transfer events in the *rpoB* gene in members of the *M. abscessus* complex (Kim BJ et al., 2017). Unlike MIRU-VNTR and spoligotyping, MLSA/MLST derived phylogenetic relationships are not affected by homoplasy or convergence evolution (Comas et al., 2009).

Nevertheless, MLSA may not be adequate for the discrimination of members of the *Mtb* complex because of the low level of sequence polymorphisms in these bacteria (Feizabadi et al., 1996; Achtman, 2008).

2.9.5 Whole Genome Sequence (WGS) of *Mtb*

The whole genome sequencing of *Mtb* strain H37Rv in 1998 opened the door to intensive and extensive genomic analyses that led to useful knowledge for scientific investigations as well as the prevention and control of TB (Cole et al., 1998). There are now 5,181 genome assembly and annotation reports on *Mtb* in the NCBI database (National Center for Biotechnology Information, 2017).

It is now well known that *Mtb* has a circular chromosome with 4,411,529 base pairs, 4,043 genes encoding 3,993 proteins, 56 insertion elements and two prophages. The most interesting features of the genome are the large number of about 250 distinct enzymes for the generation and lysis of lipids (Riley and Labedan, 1996), and two extensive families of novel glycine-rich PE and PPE proteins. These polymorphic proteins, encoded by about 10 % of the *Mtb* genome, are distinctly different from the PE-PPE proteins in other mycobacteria (Poulet and Cole, 1995). They have been proposed to be responsible for antigenic variation in *Mtb* or the inhibition of immune responses in the host (Vega-Lopez et al., 1993; Smith et al., 1997; Mukhopadhyay and Balaji, 2011; Sampson, 2011). In addition, crucial protein families for virulence and drug resistance of *Mtb* are found, which include the ESAT-6 (Sorensen et al., 1995) and CFP-10 antigens (Ganguly et al., 2008) exclusively found in *Mtb*, cell entry (*mce*) proteins (Zhang and Xie, 2011), as well as large membrane proteins encoded by *mmpL* genes (Collins, 1996; Converse et al., 2003).

2.9.6 Strain Identification and Differentiation based on Comparative Genomic Studies

Following the genomic study of H37Rv, the complete genomes of other *Mtb* strains were sequenced, including the highly infectious clinical strain CDC1551 (Fleischmann et al., 2002), the avirulent laboratory strain H37Ra (Zheng et al., 2008), strain Haarlem (Kremer et al., 1999), strain F11

(Victor et al., 2004), and many other clinical strains (Valway et al., 1998; Kato-Maeda et al., 2001; Tsolaki et al., 2005; Kang et al., 2010). The *Mtb* genome has been described as highly conserved as there is little genetic diversity among members of the *Mtb* complex (Supply et al., 2013). However, recent advances in gene sequencing technology have enabled large-scale comparisons to be made between *Mtb* genomes from diverse origins (Zheng et al., 2008), and have created doubts about the conservation of the *Mtb* genome, while revealing crucial insights into the phylogeny and global spread of this pathogen (Willcocks and Wren, 2014; Dou et al., 2016). This is especially true when new strains emerged and displayed various diversity in virulence and in clinical outcome (Romero et al., 2008; Coscolla and Gagneux, 2010; Rosales et al., 2010; Viegas et al., 2010; Liu F et al., 2014). Numerous SNPs have been identified in clinical isolates, especially in sequences linked with vital phenotypic features such as resistance to antibiotics (Dark et al., 2009; Joshi et al., 2012; Ilina et al., 2013; Shorten et al., 2013; Stucki and Gagneux, 2013). Differences have also been detected among mycobacterial cell wall glycolipids (Rocha-Ramírez et al. 2008; Barkan et al. 2012; Loman et al. 2012) and secreted proteins such as those in the PE/PPE family (Sayes et al., 2012; Liu F et al., 2014). Thus, in the past 10 years, there has been an explosion of computational comparative genomics analyses and comparative biological studies between different isolates of *Mtb* (Gagneux and Small, 2007; Niemann et al., 2009; Sultana et al., 2011; Uplekar et al., 2011; Kohli et al., 2012; Bryant et al., 2013; Liu F et al., 2014; Guerra-Assunção et al., 2015; Periwal et al., 2015). Such comparative genomics studies have the potential to uncover

strain specific genetic differences, especially to identify genes responsible for important functions.

There have been very few studies to elucidate the genetic differences between *Mtb* strains that cause CNS TB and strains causing lung disease. Das and colleagues compared the whole genomes of five CNS *Mtb* in India with the reference *Mtb* H37Rv strain and discovered non-synonymous single nucleotide variations (SNVs) and insertion sequences (IS) (Das et al., 2013). The results indicated genomic heterogeneity between CNS and respiratory isolates.

2.9.7 Identification of Genetic Variations using Suppression Subtractive Hybridisation (SSH)

SSH is a functional genomic analysis designed to compare mRNA (cDNA) or genomic DNA for variants between the same species or closely related species exhibiting different levels of pathogenicity (Diatchenko et al., 1996; Lukyanov et al., 2007). Various research studies have suggested the applicability of SSH for molecular genetics investigations of diseases (Diatchenko et al., 1996; Townsend et al., 1998; Morrow et al., 1999). It has been used successfully to identify pathogen-specific genes in *Neisseria sp.* (Tinsley and Nassif, 1996); serovar-specific genes in *Salmonella enterica* (Emmerth et al., 1999); novel restriction-modification systems in *Helicobacter pylori* (Lin et al., 2001) as well as virulence genes and genome diversity in isolates of *Burkholderia mallei* (DeShazer et al., 2001; Deshazer, 2004).

In mycobacterial research, SSH has been applied successfully for the identification of genomic differences among *M. bovis* (Mahairas et al., 1996); *M. avium* subspecies (Tizard et al., 1998); and between *M. avium* and *M. intracellulare* (McGarvey and Bermudez, 2001). It has been used to clarify the genetic basis of virulence in *M. ulcerans,* the causative agent of Buruli ulcers (Jenkin et al., 2003). In European wild boars exposed to *M. bovis*, the differential gene expression in cells from tonsils and mandibular lymph nodes was investigated with SSH (Naranjo et al., 2006).

In China, SSH was used to compare virulence-related genes in the virulent strain H37Rv and attenuated strain H37Ra (Xiong et al., 2005). Two novel genes were found in H37Ra (AY534505 and AY560011) and 8 different DNA fragments in H37Rv were obtained. The latter consist of genes coding for virulence factors *mce*, *purC* synthenzyme, and PE family proteins. Later in 2009, with the aid of SSH, Huang and colleagues managed to subtract 6 different DNA fragments which were only present in Xinjiang *Mtb* strains but not in *Mtb* H37Rv. These genes were found to be important for *Mtb* survival and virulence (Huang et al., 2009). Another group of scientists from Chongqing used SSH to compare the genes from an MDR-TB clinical isolate with those in H37Rv, and built a library of cDNA that were expressed in MDR-TB but not in H37Rv. From the cDNA library, 113 positive clones were sequenced for homology analysis. The researchers found four novel genes that might contribute to the development of multidrug resistance in TB (Zhang et al., 2012).

CHAPTER 3

MATERIAL AND METHODS

This study was approved by the Medical Ethics Committee, University Malaya Medical Centre, Kuala Lumpur (UMMC KL), Malaysia (Reference no. 975.28). The overview of the research work flow is summarised in Figure 3.0.

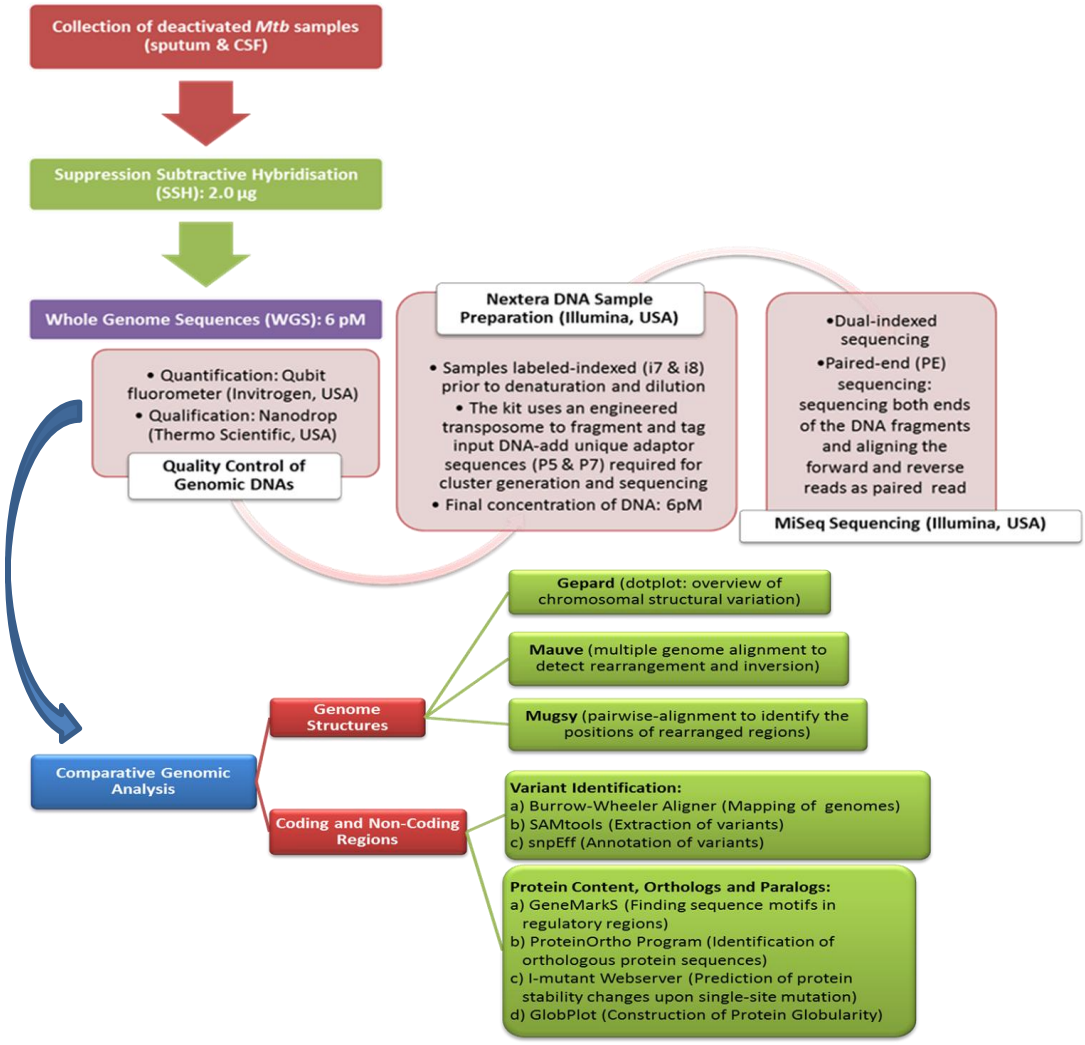


Figure 3.0: The overview of the research work flow.

3.1 Bacterial Strains

3.1.1 *Mtb* isolates for SSH

The *Mtb* used for SSH were eight CSF isolates (UM-CSF01, UM-CSF04, UM-CSF05, UM-CSF06, UM-CSF08, UM-CSF09, UM-CSF15 and UM-CSF17) and two sputum isolates (UM-Sputum B and UM-Sputum F) from patients treated for TB in the University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia (Table 4.1.1). They were recovered from routine cultures in the BACTEC MGIT 960 liquid culture system (Becton Dickinson) and identified using a reverse line probe hybridisation assay (GenoType Mycobacterium CM/AS; Hain Lifescience, Nehren, Germany). Corresponding CSF samples were collected from patients who presented with meningitis (clinical and laboratory evidence of meningeal inflammation), were sputum culture negative for *Mtb* and were not having features of pulmonary TB in their chest X-rays. On the other hand, the two sputum samples were from patients presenting as pulmonary TB with no central nervous system (CNS) manifestations. The genotypes of the clinical isolates were identified previously using the spoligotyping kit from Ocimum Biosolutions (Hyderabad, India). They were Beijing Sequence Type-1: ST1 (UM-CSF01, UM-CSF05, UM-CSF08, UM-CSF09, UM-CSF17), East African Indian: EAI_IND (UM-CSF06) and Haarlem type 3 Sequence Type-50: H3 ST50 (UM-CSF15). One isolate (UM-CSF04) was classified as “unknown”. Until required for further testing, the isolates were deep frozen at – 80 °C in Middlebrook 7H9 broth with 15 % glycerol. For suppression subtractive hybridisation (SSH), isolates

were thawed out, sub-cultured on Lowenstein-Jensen (LJ) slants, heat inactivated at 80 °C for 2 h and cooled down to RT before they were used for DNA extraction.

3.1.2 *Mtb* Genomes for WGS analysis

The eight CSF *Mtb* genomes for whole genome sequence (WGS) analysis were extracted from the same CSF *Mtb* isolates used for the SSH experiments (refer to 3.1.1). In addition, 13 *Mtb* genomes from sputum isolates (UMMC KL) were used for comparison with the CSF genomes.

3.2 Culture Media and Reagents

All media and reagents used are listed in Appendix 1 (refer to 3.2.1 and 3.2.2).

3.3 Genomic DNA Extraction and Evaluation

Bacterial genomic DNA was extracted using the Phenol/Chloroform/Isoamylalcohol (PCI) method (Sambrook et al., 1989) to obtain a high yield of DNA.

The heat-inactivated *Mtb* culture was centrifuged to a maximum speed (12 054 *x g*) for 15 min. The pellet obtained was lysed by incubation in 10.0 mg/ml of lysozyme overnight at 37 °C, followed by the addition of 10 % SDS.

10.0 mg/ml Proteinase K was then added to inactivate DNases and RNases. The mixture was vortexed and incubated at 55 °C for 5 h.

Before the purification of nucleic acid, 5.0 M of sodium chloride was added, followed by the addition of phenol/chloroform/isoamyl alcohol (25:24:1) (ROTI, Germany) to remove all proteins. Phenol denatures proteins efficiently while chloroform laced with isoamyl alcohol removes phenol from the nucleic acid preparation (Kirby, 1957; Kirby, 1964). In addition, isoamyl alcohol also reduces foaming. The mixture was then centrifuged at maximum speed (12 054 *x g*). The aqueous supernatant containing DNA was transferred to a micro-centrifuge tube and the purification step was repeated.

Finally, nucleic acids were recovered from the aqueous solution with ethanol precipitation using 3.0 M of sodium acetate and ice-cold isopropanol, and overnight incubation of the mixture at - 20 °C. The pellet was washed with 80 % ethanol and dried at RT. The required DNA precipitate was dissolved and diluted with autoclaved distilled water.

The concentration and purity of the recovered DNA were measured in a spectrophotometer at OD_{260} (Nanophotometer, Implen USA). Agarose gel electrophoresis was used to detect the presence of the genomic DNA. The extracted DNA was used to perform SSH and WGS analyses.

3.4 Oligonucleotide Primers and Adaptors for SSH

The sequences and sizes of the oligonucleotide primers and adaptors used for SSH are listed in Table 3.4. These synthetic nucleotides were provided at 10.0 µM (Clontech Laboratories, USA). The T_m values were estimated using IDT OligoAnalyzer 3.1 software (https://sg.idtdna.com/calc/analyzer). The sequences similarities of the adaptors and PCR primers are showed in Appendix 2: Figure A2.

Table 3.4: The list of adaptors and primers used in the study.

No	Primers	Sequences	Size (bp)	T_m (°C)
1	Adaptor 1	5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3'	44	73.1
2	Adaptor 2R	5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT-3'	42	70.9
3	23S RNA Forward	5'-CTACCTTAGGACCGTTATAGTTAC-3'	24	51.8
4	23S RNA Reverse	5'-GAAGGAACTAGGCAAAATGGT-3'	21	52.9
5	PCR primer 1	5'-CTAATACGACTCACTATAGGGC-3'	22	51.7
6	Nested PCR primer 1	5'-TCGAGCGGCCGCCCGGGCAGGT-3'	22	74.8
7	Nested PCR primer 2R	5'-AGCGTGGTCGCGGCCGAGGT-3'	20	69.2
8	pUC/M13 Forward	5'-GTTTTCCCAGTCACGAC-3'	17	50.6
9	pUC/M13 Reverse	5'-CAGGAAACAGCTATGAC-3'	17	47.0
10	Control Primer 5'	5'-GCAACTGCAGGAAGAGCAAGAAATGCA-3'	27	61.9
11	Control Primer 3'	5'-TGGCACGGCCATAAGAGGTAGATGTCA-3'	27	63.0

3.5 Suppression Subtractive Hybridisation (SSH)

SSH was performed using the PCR-Select^{TM} Bacterial Genome Subtraction Kit (Clontech Laboratories Inc., USA). The steps of the assay which include digestion, joint connections, subtractive hybridisation and nested PCR were carried out according to the manufacturer's instructions (Figure 3.5). The paired tester and driver samples used in this study consisted of CSF *Mtb* UM-CSF01 with UM-Sputum F (1st pair) and UM-CSF04 with UM-Sputum B (2nd pair).

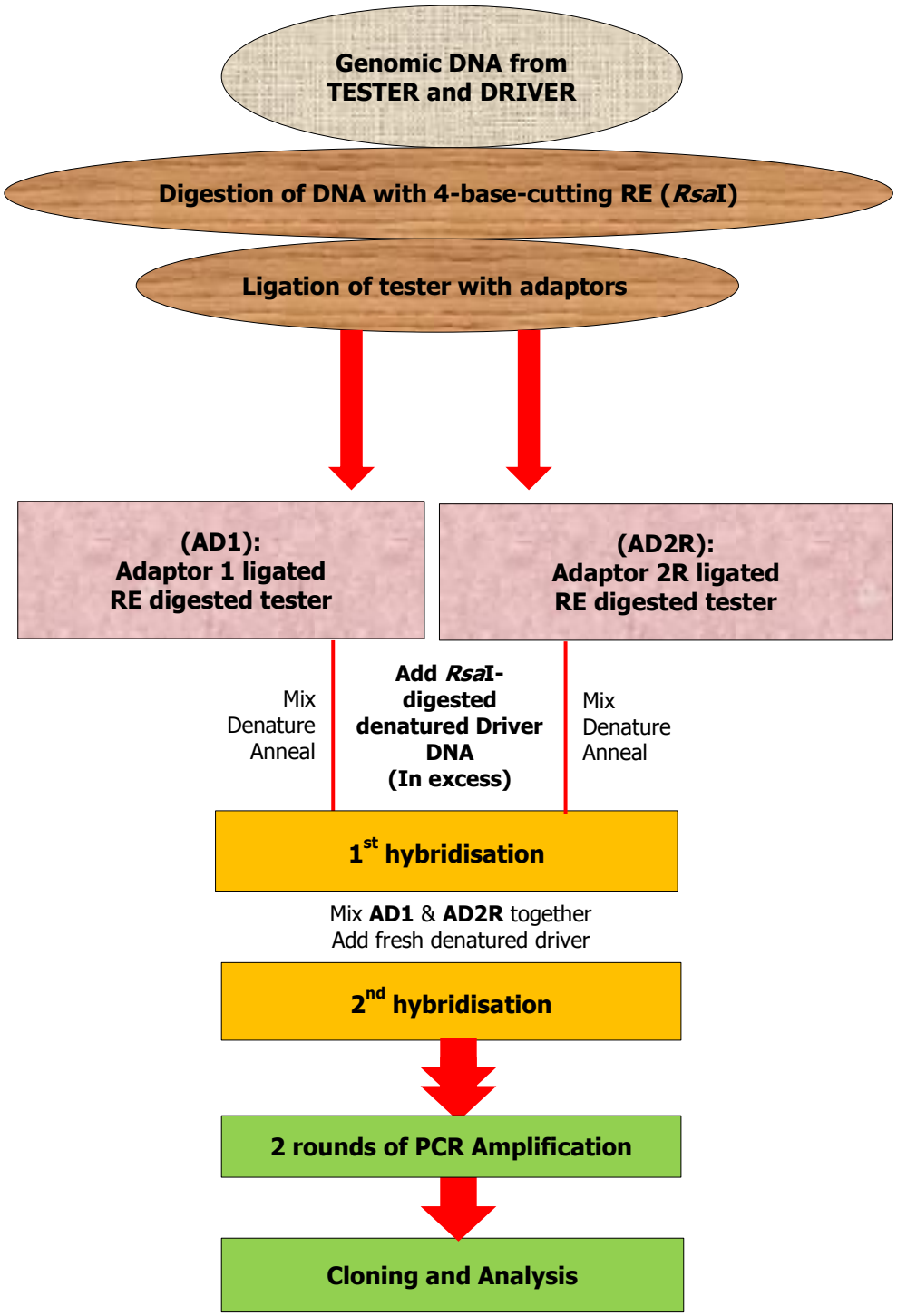


Figure 3.5: General work flow of SSH. Genomic DNA sample with the sequence of interest is referred to as tester sample (CSF *Mtb:* UM-CSF01 and UM-CSF04) and the reference sample is referred to as driver sample (sputum *Mtb:* UM-Sputum F and UM-Sputum B) (Clontech Laboratories, 2008).

3.5.1 Restriction Enzyme (RE) Digestion and Termination

3.5.1.1 RE Digestion of Tester, Driver and Control *Escherichia coli* Genomic DNA

The volume used for restriction enzyme (RE) digestion was calculated to achieve a total of 2.0 µg of DNA. *Rsa*I RE from different suppliers (Clontech and Invitrogen) was compared to obtain the best restriction cutting pattern for the extracted DNA. Firstly, 5 tubes of 50.0 µl of restriction mixture were prepared: each consisting of 2.0 µg of genomic DNA (2 sputum *Mtb* isolates as driver samples – UM-Sputum F and UM-Sputum B; 2 CSF *Mtb* isolates as tester samples – UM-CSF01 and UM-CSF04; 1 control DNA – *E. coli* genomic DNA), 5.0 µl of 10 X *Rsa*I Restriction Buffer, 1.5 µl of *Rsa*I enzyme (10.0 units/µl) and deionized water. The mixture in each tube was vortexed and centrifuged before it was incubated overnight at 37 °C. After incubation, 5.0 µl of the RE mixture was used to assess the digestion efficiency.

3.5.1.2 Termination of RE Digestion of DNA

Termination of the digestion reaction in each tube was performed by adding 2.5 µl of 0.2 M EDTA, followed by the addition of 50.0 µl of phenol:chloroform:isoamyl alcohol (25:24:1). The mixture was vortexed and centrifuged at 12 054 *x g* for 10 min at RT to separate the mixture into upper/aqueous and bottom/white precipitation layers. The top aqueous layer of

the mixture containing DNA was carefully collected and placed in a fresh 0.5 ml micro-centrifuge tube. 50.0 µl of chloroform:isoamyl alcohol (24:1) was added, vortexed and centrifuged at 12 054 *x g* for 10 min at RT. Again, the top aqueous layer of the mixture was removed and 0.5 volume of 4.0 M NH_4OA_C and 2.5 volume of 95 % ethanol were added. The solution was vortexed and centrifuged at 12 054 *x g* for 30 min to pellet DNA. The supernatant was removed. The pellet was overlaid with 200.0 µl of 80 % ethanol and centrifuged for 5 min at 12 054 *x g*. Finally, the supernatant was removed. The pellet of DNA was air-dried for 15 min and then dissolved in 6.5 µl of H_2O. It was stored at - 20 °C. The final concentration of the DNA was 300.0 ng/µl.

3.5.2 Adaptors Ligation

Figure 3.5.2 shows the experimental flow-chart for preparing adaptor-ligated experimental tester, un-subtracted experimental tester, and control tester genomic DNA. The ligated un-subtracted experimental tester is used as a control in later PCR assays.

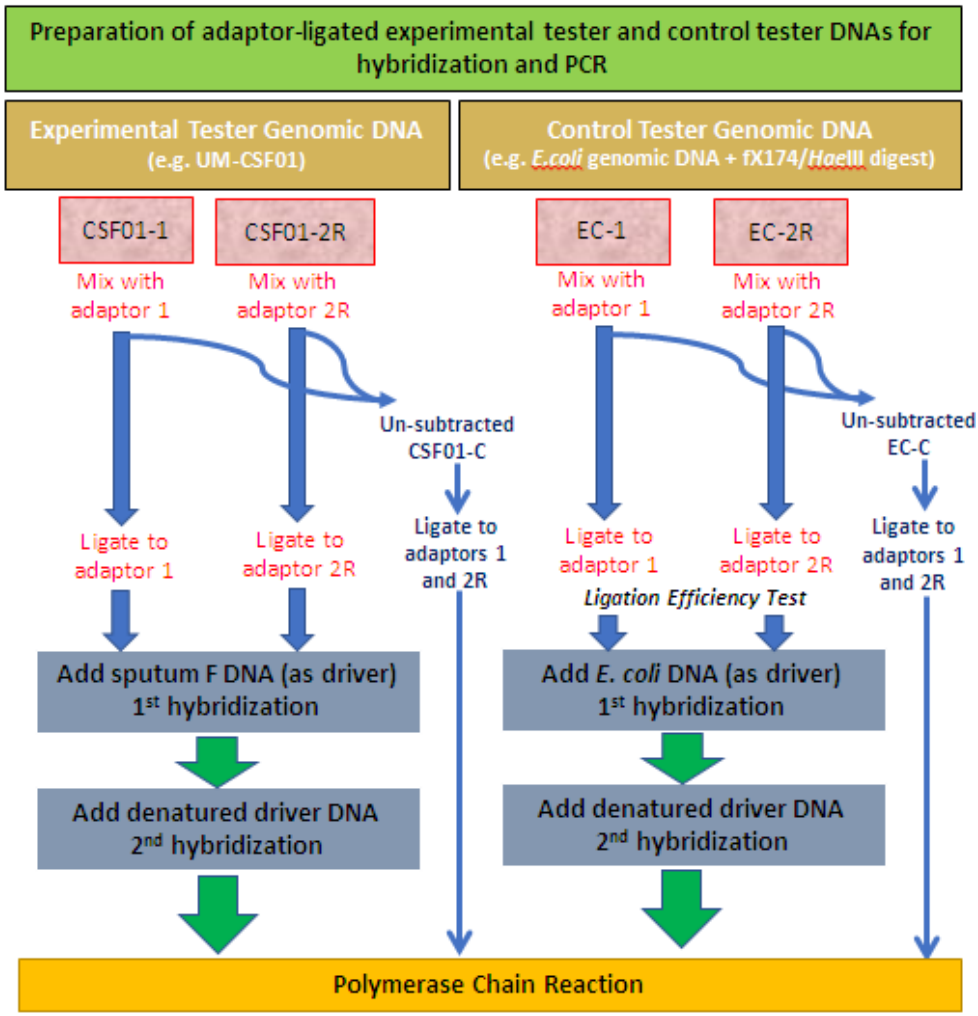


Figure 3.5.2: Experimental flow chart showing the steps for preparing adaptor-ligated experimental tester and control tester DNA for hybridisation and PCR. Each tester DNA (UM-CSF01 and UM-CSF04) must be ligated to the appropriate adaptors (adaptor 1 and adaptor 2R). The drivers DNA used in the study were UM-Sputum F and UM-Sputum B, which were paired with UM-CSF01 and UM-CSF05, respectively. Left panel shows the subtraction experiment (e.g. UM-CSF01) and right panel shows the control subtraction (*E. coli* genomic DNA).

3.5.2.1 Preparation of Control Tester DNA (*Hae*III-digested ϕX174 DNA)

The control tester DNA was prepared by diluting the *Hae*III-digested ϕX174 DNA (provided in the kit) from 3.0 ng/µl to 0.2 ng/µl with sterile water. Then, 1.2 µl of the RE-digested control *E. coli* genomic DNA (1.0 mg/ml: refer to 3.5.1) was mixed with 1.8 µl of the diluted *Hae*III-digested ϕX174 DNA (0.2 ng/µl). For the ligation to adaptors, a ligation master mix consisting of 2.0 µl of 5 X ligation buffer, 1.0 µl of T4 DNA ligase (400 units/µl) and 4.0 µl of sterile water was added to the control tester DNA prepared (Table 3.5.2.1).

Table 3.5.2.1: Setting up for ligation reactions for positive control of tester DNA subtraction.

Tube	**Positive Control**
Component (µl)	
Diluted *Hae*III-digested ϕX174 DNA (0.2 ng/µl)	1.8
***Rsa*I-digested *E. coli* genomic DNA (1.0 mg/ml)**	1.2
Ligation Master Mix	7.0
Final Volume (µl)	**10.0**

3.5.2.2 Preparation of UM-CSF Tester DNA

Similarly, 1.2 µl of each RE-digested experimental tester DNA (100 ng/µl: UM-CSF01 and UM-CSF04, refer to 3.5.1) was diluted with 1.8 µl of sterile water. Next, a ligation master mix which consisted of 2.0 µl of 5 X

ligation buffer, 1.0 µl of T4 DNA ligase (400 units/µl) and 4.0 µl of sterile water was prepared (Table 3.5.2.2). Each diluted tester DNA sample (e.g. UM-CSF01) was aliquoted into two separate tubes: one aliquot (1.0 µl) was ligated with Adaptor 1 (e.g. UM-CSF01-1), and the second (1.0 µl) was ligated with Adaptor 2R (e.g. UM-CSF01-2R). The procedure was repeated for UM-CSF04.

Table 3.5.2.2: Setting up for ligation reactions using adaptor 1 and adaptor 2R, respectively. Example used here was UM-CSF01.

Tube	**UM-CSF01-1**	**UM-CSF01-2R**
Component (µl)		
Diluted tester DNA	1.0	1.0
Adaptor 1 (10 µM)	2.0	-
Adaptor 2R (10 µM)	-	2.0
Master Mix	7.0	7.0
Final Volume (µl)	**10.0**	**10.0**

3.5.2.3 Preparation of Un-subtracted UM-CSF Tester Control

Subsequently, in a fresh micro-centrifuge tube, 1.5 µl of the Adaptor 1-tester DNA (e.g. UM-CSF01-1) and 1.5 µl of the Adaptor 2R-tester DNA (e.g. UM-CSF01-2R) were mixed and centrifuged (Figure 3.5.2). This tube served as an un-subtracted experimental tester control (e.g. UM-CSF01-C), which acted as a negative control for subtraction.

3.5.2.4 Ligation Reaction

All the samples (refer to 3.5.2.1, 3.5.2.2, 3.5.2.3) were incubated overnight at 16 °C to allow for ligation to take place. After ligation, approximately 1/3 of the DNA molecules in each un-subtracted tester control tube had two different adaptors. These ligation processes were repeated for the experimental tester UM-CSF04 DNA and the control tester *E. coli* genomic DNA. The adaptors were not ligated to the RE-digested driver DNA (UM-Sputum F and UM-Sputum B).

3.5.2.5 Ligation Termination

After the completion of ligation, 1.0 µl of 0.2 M EDTA was added into the ligation mixtures and was heated at 72 °C for 5 min to inactivate the ligase. 1.0 µl from the un-subtracted tester control (refer to 3.5.2.3) was removed and diluted into 500.0 µl of sterile water. These ligated samples were ready for PCR amplification. The samples were stored at - 20 °C prior usage.

3.5.2.6 Ligation Efficiency Analysis

The ligation efficiency analysis was performed on the control tester *E. coli* genomic DNA to verify that at least 25.0 % of the DNA fragments have adaptors on both ends. 1.0 µl of each *E. coli* control-ligated DNA was diluted into 100.0 µl of sterile water. The ligation analysis was set up in four separate reaction tubes as shown in Table 3.5.2.6.1. Two sets of different primers were

used to amplify fragments that span the adaptor/DNA junctions of *E.coli* 1 and *E.coli* 2R. The primer sets were (a) primer 23S RNA Forward and Primer 1 (one gene-specific primer), which yielded a PCR product of 374 bp size; and (b) primers 23S RNA Forward and 23S RNA Reverse (two gene-specific primers), which yielded a PCR product of 270 bp size. The sequences for these primers are shown in Table 3.4. PCR Primer 1 is designed to be able to bind with adaptor 1 and adaptor 2R during PCR amplification. Appendix 3: Figure A3 shows the position of the primers in the *E. coli* chromosome.

Table 3.5.2.6.1: Setting up for ligation efficiency analysis for control tester *E. coli* genomic DNA.1 to 4 indicates the number of tubes used.

Component (tube/µl)	**1**	**2**	**3**	**4**
Tester *EC-1 (ligated to Adaptor 1)	1.0	1.0	-	-
Tester *EC-2 (ligated to Adaptor 2R)	-	-	1.0	1.0
23S RNA Forward Primer (10 µM)	1.0	1.0	1.0	1.0
23S RNA Reverse Primer (10 µM)	-	1.0	-	1.0
PCR Primer 1 (10 µM)	1.0	-	1.0	-
Total Volume (µl)	**3.0**	**3.0**	**3.0**	**3.0**
Expected PCR Product Size (bp)	**374**	**270**	**374**	**270**
Occurrence of Ligation Reaction	**+**	**-**	**+**	**-**

*EC=*E. coli*

Next, a master mix was prepared for all the reaction tubes plus one additional tube to perform the PCR amplification for ligation efficiency analysis (Table 3.5.2.6.2). 22.0 µl of the master mix was aliquoted into each of

the reaction tubes after it was well mixed and centrifuged. 25.0 µl of the reaction mix was incubated at 72 °C for 2 min in a thermal cycler for an extension of adaptors and thermal cycling profiles were conducted (Table 3.5.2.6.3) immediately. The PCR products were run on a 2.0 % agarose gel electrophoresis and viewed under UV trans-illuminator using a gel documentation system (SynGene BioImaging, UK).

Table 3.5.2.6.2: PCR master mix preparation for ligation efficiency analysis.

Component	**Per reaction (µl)**	**5 reactions mix (µl)**	**Concentration**
sterile water	18.5	92.5	
10 X PCR buffer	2.5	12.5	1 X
dNTP mix (10 mM)	0.5	2.5	0.2 mM
50 X Advantage 2 Polymerase Mix	0.5	2.5	1 X
Total Volume	**22.0**	**110.0**	

Table 3.5.2.6.3: Thermal cycler profiles for ligation analysis.

Temperature (°C)	**Time (s)**	**Cycle**
94	30	30 x
65	30	
68	60	

3.5.3 First and Second Hybridisation

3.5.3.1 First Hybridisation

2.0 µl of driver DNA (600.0 ng) was added to each of the adaptor-ligated tester DNA (2:1 ratio) together with 1.0 µl of 4 X hybridisation buffer (Table 3.5.3.1). This first hybridisation was performed at 98 °C for 1.5 min in a thermal cycler (for denaturing), followed by heating at 63 °C for 1.5 h (for annealing).

Table 3.5.3.1: Setting up for first hybridisation. The example used here was the sputum F driver sample and the UM-CSF01 tester sample.

Component	**Hybridisation samples (µl)**	
	H1	**H2**
***Rsa*I-digested Sputum F (driver)**	2.0	2.0
Adaptor 1-ligated CSF 1 (tester)	1.0	-
Adaptor 2R-ligated CSF 1 (tester)	-	1.0
4 X Hybridisation Buffer	1.0	1.0
Final Volume	**4.0**	**4.0**

3.5.3.2 Denaturation of Driver DNA for Second Hybridisation

Prior to the second hybridisation, the driver DNA was denatured at 98 °C for 1.5 min in a thermal cycler, in the presence of 2 X hybridisation buffer (Table 3.5.3.2).

Table 3.5.3.2: Setting up for denaturation of driver DNA. The example used here was the Sputum F driver sample.

Reagents	**Volume (µl)**
***Rsa*I-digested Sputum F (driver)**	1.0
2 X Hybridisation Buffer	1.0
Total	**2.0**

3.5.3.3 Second Hybridisation

After the first hybridisation, the hybridisation samples of adaptor 1-ligated tester DNA (H1) and adaptor 2R-ligated tester DNA (H2) were combined (refer to 3.5.3.1). For this procedure, a micro-pipette was set at 15.0 µl and sample H2 was drawn into the pipette tip, followed by a small amount of air, creating a slight air space below the droplet of sample. Then, 300.0 ng of fresh heat-denatured driver DNA (refer to 3.5.3.2) was drawn into the same tip. The pipette tip then contained both H2 and denatured driver, separated by a small pocket of air. The entire mixture was transferred to the H1 tube. This mixture was further mixed by pipetting up and down and then allowed to hybridise for an additional 16 h at 63 °C (note: Both H1 and H2 contain more than just the adaptor-ligated tester DNA).

These procedures ensure that the two hybridisation samples H1 and H2 mix together only in the presence of the driver. In addition, H1 and H2 are mixed together without prior denaturing to ensure that only the remaining

equalized and subtracted single stranded tester DNA can re-associate to form new type *e* hybrids (double-stranded tester molecules with adaptor 1 and adaptor 2R ligated to opposite ends of each strand) for PCR amplification (Figure 5.2.3.1 and Figure 5.2.3.2).

3.5.3.4 Pre-PCR

10 µl of the final hybridisation mixture (refer to 3.5.3.3) was diluted in 100 µl of dilution buffer, followed by heating at 63 °C for 7 min to eliminate non-specific hybridisation. The sample was stored at - 20 °C prior to PCR. A standard PCR control obtained from Advantage® 2 PCR Kit (Clontech Laboratories, USA) was used to verify the efficiency of the polymerase enzyme and the PCR. Control DNA template (100 ng/µl) derived from calf thymus DNA and the control primer 5' and 3' by the kit were utilized (Table 3.4).

3.5.4 Suppression PCR Amplification

Two sequential PCRs were carried out.

3.5.4.1 Primary PCR

For the first PCR, the reaction mix contained 1 µl of genomic DNA (refer to 3.5.3), 1 µl of PCR primer 1 (10 µM), and 23 µl of a PCR master mix from the Advantage 2 PCR kit (Clontech Lab Incorporation, USA), making a

total volume of 25 µl (Table 3.5.4.1.1). The reaction mix was incubated at 72 °C for 2 min in a thermal cycler to extend the adaptors, and then subjected to 30 cycles at 94 °C for 30 s, 66 °C for 30 s and finally, 72 °C for 1.5 min (Table 3.5.4.1.2). The amplified products were then diluted to 20-fold in distilled water, and 1 µl of the diluted sample was used for the second PCR.

Table 3.5.4.1.1: Setting up for primary PCR.

Component	**Per reaction (µl)**	**Concentration**
Sterile water	19.5	
10 X PCR buffer	2.5	1 X
dNTP mix (10 mM)	0.5	0.2 mM
PCR Primer 1 (10 µM)	1.0	0.4 µM
50 X Advantage 2 Polymerase Mix	0.5	1 X
Template DNA (refer to 3.5.3)	1.0	
Total Volume	**25.0**	

Table 3.5.4.1.2: Thermal cycler profiles for primary PCR amplification.

Temperature (°C)	**Duration**	**Cycle**
94	30 sec	30 x
66	30 sec	
72	1.5 min	

3.5.4.2 Secondary PCR

In the second PCR, 1 µl of nested PCR primer 1 (10 µM) and 1 µl of nested PCR primer 2R (10 µM) were added to the PCR mix to make a total volume of 25 µl (Table 3.5.4.2.1). Amplification was done with 15 cycles of denaturation at 94 °C for 30 s and annealing at 68 °C for 30 s and a final extension at 72 °C for 1.5 min (Table 3.5.4.2.2). To determine the efficiency of SSH, 1 µl each of the diluted primary and secondary PCR products (20 X) was analysed on a 2.0 % agarose/EtBr gel run in 1 X TBE buffer.

Table 3.5.4.2.1: Setting up for secondary PCR.

Component	**Per reaction (µl)**	**Concentration**
Sterile water	18.5	
10 X PCR buffer	2.5	1 X
Nested Primer 1 (10 µM)	1.0	0.4 µM
Nested Primer 2R (10 µM)	1.0	0.4 µM
dNTP Mix (10 mM)	0.5	0.4 mM
50 X Advantage 2 Polymerase Mix	0.5	1 X
Diluted Primary PCR product	1.0	
Total Volume	**25.0**	

Table 3.5.4.2.2: Thermal cycler profiles for secondary PCR amplification.

Temperature (°C)	Duration	Cycle
94	30 sec	15 x
68	30 sec	
72	1.5 min	

3.5.5 Gel Extraction and Purification of the Subtracted DNA

3.5.5.1 Gel Electrophoresis

Products from the secondary PCR were then electrophoresed on 2.0 % w/v agarose gels at 45 V for 150 min, to ensure distinct separation of the DNA fragments. 20.0 µl of the mixture was analysed on the gel with 250.0 ng 1 kb Plus DNA Ladder (Invitrogen, USA). The fragments of the DNA were visualized under UV transilluminator (ChemiDoc XRS Biorad, USA) after electrophoresis.

3.5.5.2 Extraction and Purification

The desired PCR products were purified using the QIAquick Gel Extraction Kit (Qiagen, Netherlands). All centrifugation steps in the purification procedure were performed at 12 054 *x g* for 1 min. Subtracted DNA excised from the agarose gel were weighed out. Three volumes of buffer QG were added to 1 volume of gel in a 1.5 ml micro-centrifuge tube. The sample was then incubated at 50 °C for 10 min. After adding one gel volume

of isopropanol to the dissolved gel suspension, the mixture was transferred to a QIAquick spin column to collect DNA on the membrane. Then, 500.0 µl of buffer QG was added to the column. After centrifugation, the flow-through was discarded and 750.0 µl of buffer PE was added to the column for 5 min of centrifugation to wash the column. The flow-through was discarded, and centrifugation was repeated. The column with the membrane was transferred to a 1.5 ml Eppendorf tube. Finally, 30.0 µl of PCR-grade water was applied to the QIAquick membrane and incubated for 1 min before it was centrifuged to elute the DNA.

After the purification steps were completed, 1 volume of 6 X loading dye was added to 5 volumes of purified DNA for agarose gel analysis. The concentration and the quality of the purified subtracted DNA were performed before it was further used for molecular cloning and DNA sequencing.

3.5.6 Preparation of Competent *Escherichia coli* JM109 Cells

3.5.6.1 Culture of *E. coli* JM109 Cells

JM109 cells are free of endonuclease (*end*A) to preserve the quality of miniprep DNA and recombination (*rec*A) deficient to improve insert stability. In addition, JM109 cells contain the *lacIq ZΔM15* gene on the F´ episome, which allows blue-white screening for recombinant plasmids (Promega Corporation, 2009). For molecular cloning, competent *E. coli* JM109 cells were prepared using the $CaCl_2$ method (Sambrook and Russell, 2006a). In the

primary inoculation step, a colony of *E. coli* JM109 was grown overnight in 10 ml LB broth at 37 °C with shaking at 4.46 *x g*. 400.0 µl of the primary inoculation culture was transferred into 40.0 ml of LB broth for shaking at 4.46 *x g* for approximately 2.5 - 3.0 h or until the cell growth reached the log phase (OD_{600} = 0.3 - 0.4) (secondary inoculation step).

3.5.6.2 Calcium Chloride ($CaCl_2$) Treatment

The *E. coli* LB broth culture was centrifuged at (15 941 *x g* for 10 min at 4 °C). The cell pellet obtained was re-suspended in 20.0 ml of ice-cold 0.1 M $CaCl_2$ and incubated on ice for more than 30 min to render the cell membranes more permeable to DNA. The cells were then centrifuged again (15 941 *x g* for 10 min at 4 °C) and re-suspended in 4.0 ml of $CaCl_2$ solution (85.0 mM, 15.0 % glycerol). 200.0 µl of the competent cells were then aliquoted into chilled micro-centrifuge tubes and stored at - 80 °C until further usage.

3.5.6.3 Evaluation of Transformation Efficiency (TE)

The competency of *E. coli* JM109 was tested to measure transformation efficiency, which is defined as the efficiency by which cells can take up and express extracellular DNA (Figure 3.5.6.3). The Control Insert DNA supplied with the pGEM®-T Easy Systems is a 542 bp fragment from pGEM®-*luc* Vector DNA (Promega, USA). The sequence of this fragment has been mutated to contain multiple stop codons in all six reading frames, to

ensure a low background of blue colonies for the control reaction. For transformation, 0.8 ng/µl of control insert DNA (4.0 ng/µl; provided by the kit) was ligated with 5.0 ng/µl of pGEM-T Easy Vector (50 ng/µl) (refer to 3.5.7). The *E. coli* JM109 transformation efficiency (TE) was calculated using the formula shown in Figure 3.5.6.3.

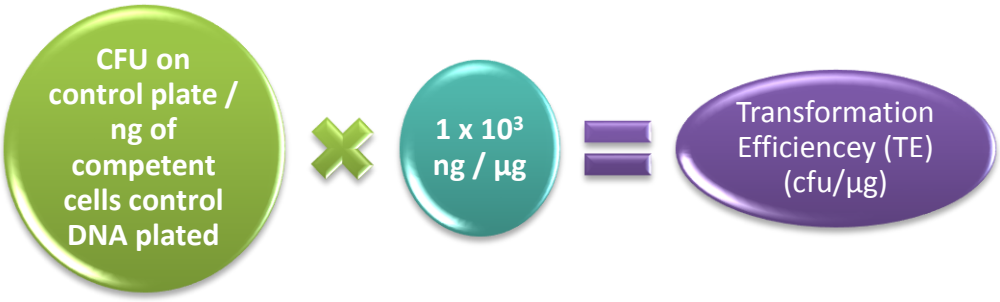


Figure 3.5.6.3: The equation for transformation efficiency (TE) (cfu/µg). TE is defined as the number of colony forming units (cfu) produced by 1 µg of competent cells control DNA (plasmid DNA). It is measured by the number of cfu formed per µg of DNA in a control transformation reaction conducted with a known quantity (0.1ng) of DNA (Promega Corporation, 2010).

3.5.7 Molecular Cloning and Transformation of Subtracted DNA into *E. coli* JM109 cells

3.5.7.1 Blue-White Screening

Molecular cloning using blue-white screening was performed on the subtracted sequences. Purified subtracted DNA was ligated into multiple cloning regions of pGEM®-T Easy Vectors (Promega, USA) according to the manufacturer's protocol. The multiple cloning regions are located within the α-peptide coding region of the enzyme ß-galactosidase; insertional inactivation of the *α*-peptide allows recombinant clones to be directly identified by blue-white screening on X-gal indicator plates (Appendix 4: Figure A4).

X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside) is much used in molecular biology to test for the presence of an enzyme, *β*-galactosidase, and yield insoluble blue compounds as a result of enzyme-catalysed hydrolysis in this case by *β*-galactosidase. *β*-galactosidase is a protein encoded by the *lac*Z gene of the lac operon, and it exists as a homo-tetramer in its active state. The blue/white screening method works by disrupting this α-complementation process. The vector carries within the *lac*Z*α* sequence an internal multiple cloning site (MCS). This MCS within the *lac*Z*α* sequence can be cut by RE so that the foreign DNA may be inserted within the *lac*Z*α* gene. In cells containing the vector with an insert, no functional *β*-galactosidase may be formed. The presence of an active *β*-galactosidase can be detected by X-gal

within the agar plate. X-gal is cleaved by *β*-galactosidase to form 5-bromo-4-chloro-indoxyl, which then spontaneously dimerises and oxidizes to form a bright blue insoluble pigment 5,5'-dibromo-4,4'-dichloro-indigo. This results in a characteristic blue colour in cells containing a functional *β*-galactosidase. Blue colonies therefore show that they may contain a vector with an uninterrupted *lacZα* (therefore no insert), while white colonies, where X-gal is not hydrolysed, indicate the presence of an insert.

3.5.7.2 pGEM-T Ligation

Ligation was being optimised using a 3:1 molar ratio of the control insert DNA to the vectors. The purified subtracted DNA was ligated into a pGEM®-T vector plasmid using a T/A cloning kit (Promega Corp., USA). The optimising insert:vector molar ratios has been recommended to be at 3:1 for good initial parameters (Promega Corporation, 2010). Calculation for the appropriate amount of the purified subtracted DNA (insert) to include in the ligation reaction is as the equation shown in Figure 3.5.7.2. Ligation volume was then calculated based on the concentration of the insert(s).

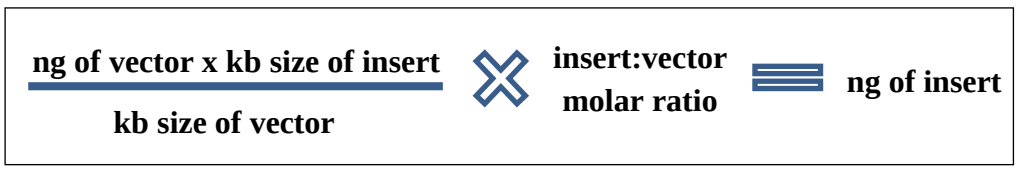


Figure 3.5.7.2: Insert:vector ratio calculation. The concentration of PCR product should be estimated by comparison to DNA mass standards on a gel. The pGEM®-T Easy Vectors are approximately 3.0 kb in size and are supplied at 50.0 ng/µl of concentration.

The ligation mix consisted of a final concentration of 15.0 - 100.0 ng of purified subtracted DNA, 5.0 ng/µl of pGEM®-T Easy vector (50 ng/µl), 1 X rapid ligation buffer (2 X), and 0.3 Weiss units/µl of T4 DNA ligase (3 Weiss units/µl). The reactions were incubated at 4 °C overnight.

3.5.7.3 Transformation into Competent Cells

The reaction mixture (refer to 3.5.7.2) was then transformed into 200.0 µl competent JM109 *E. coli* cells (refer to 3.5.6) by heat-shock treatment at 42 °C for 90 s and bringing back to ice for 2 min. Subsequently, 800.0 µl of Super Optimal Culture (S.O.C.) medium was added, and the tubes were incubated at 37 °C while shaking at 2.85 *x g* for 2.5 hr. 200.0 µl of each transformed competent cell population was transferred onto an LB agar plate containing 100 µl/ml of ampicillin, spread with 100 µl of 100 mM of Isopropylthio-β-galactoside (IPTG) and 20 µl of 50 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), for blue-white screening. The plates were then incubated overnight at 37 °C. White colonies containing inserts were randomly selected for colony PCR analysis.

3.5.8 Isolation, Purification and Verification of Positive Clones

3.5.8.1 Colony PCR

Colony PCR was used to quickly screen for plasmid inserts from *E. coli* white colonies. A 50.0 µl of PCR mix (final concentration) was prepared with 38.0 µl of sterile distilled water, 5.0 µl of 10 X PCR buffer (1 X), 2.0 µl of 25.0 mM $MgCl_2$ (2.0 mM), 2.0 µl of 10.0 mM dNTP (200.0 µM), 1.0 µl each of 10.0 µM pUC/M13 forward and reverse primers (0.2 µM each), 0.5 µl of HotStar Taq DNA Polymerase (2.5 units/reaction), and finally, 0.5 µl of the plasmid template. A small amount of white colony (plasmid template with insert) was picked (using a sterile toothpick) and added into each of the PCR tubes containing the PCR mixture. The tubes were mixed well to achieve complete cell lysis and high yields. The PCR conditions used were as follows: 1 cycle at 95 °C for 2 min; 30 cycles at 95 °C for 1 min, 50 °C for 30 s, 72 °C for 1 min, and ending with 72 °C for 5 min. Thermal cycling was carried out in a Peltier Thermal Cycler-200 from MJ Research (USA). The PCR products were separated by 2.0 % w/v agarose gel electrophoresis.

3.5.8.2 Isolation of Positive Clones

The positive clones (with subtracted DNA-inserts) were cultured overnight in LB broth with ampicillin. The DNA with insert were extracted, purified and verified before sending for sequencing. Isolation and purification were performed using the Wizard® Plus SV Minipreps Kit (Promega, USA)

according to the manufacturer's protocol. Briefly, the procedures are a) resuspension of harvested bacterial cultures in 250.0 µl of cell re-suspension solution b) cells were lysed by the addition of a mixture of 250 µl cell lysis solution, 10 µl alkaline protease solution and 350 µl of neutralization solution c) centrifugation of the bacterial lysate at 12 054 *x g* m for 10 min at RT d) transfer of about 850 µl of clear lysate to a prepared spin column e) centrifugation before adding 750 µl of column wash solution f) elution of the plasmid DNA using 50 µl of nuclease-free water and centrifuging at 12 054 *x g* for 1 min at RT.

3.5.8.3 Verification Using *EcoR*I RE Digestion

Verification of recombinant plasmids was performed using *EcoR*I RE digestion analysis. The *EcoR*I site is found at the flanking site of the inserts in pGEM®-T. Staggered cuts were made in the two DNA strands, generating fragments with a single-stranded 'tail' at both ends. These tails are complementary to those on all other fragments generated by the same restriction enzyme. The reaction mix contained 2.0 µl of 10 X NE Buffer 4, 0.5 µl of *EcoR*I (NEB, England) and deionized water. The reaction was incubated at 37 °C overnight, after which agarose gel electrophoresis was carried out at 85 V for 45 min to verify the presence of the subtracted DNA. The linearized pGEM®-T vectors (3.0 kb) and the fragment DNA (subtracted sequence) were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Netherlands) (refer to 3.5.5). The purified subtracted DNA was sent to First Base Laboratory for sequencing.

3.5.9 Nucleotide Sequence Analysis

The resulting DNA sequences were analysed, filtered and trimmed with the aid of chromatograms using Finch TV (Geospiza Inc., USA). The inserted nucleic acid sequences were identified while remaining primer and vector sequences were removed using a process termed manual base calling. After trimming, nucleic acid sequences were analysed using tools available at the NCBI website (https://www.ncbi.nlm.nih.gov/). The BLAST software was used for DNA sequence and homology analysis. BLASTN and BLASTX were used to search ~~for~~ nucleotide and protein databases, respectively. The results obtained were classified by the function and homologous scores of between 91 and 100, and between 81 and 90, respectively.

3.6 Whole Genome Sequence (WGS)

WGS for the 8 CSF isolates was performed using the Illumina MiSeq System (Illumina, USA), located in the High Impact Research (HIR) building, University of Malaya, Kuala Lumpur. The library preparations and sequencing procedures were performed with the help of research officers maintaining the Illumina system in the HIR laboratory.

3.6.1 Library Preparation and Sequencing

The extracted genomic DNA of *Mtb* (refer to 3.3) was quantified using the Qubit® fluorometer (Invitrogen, USA) and Qubit dsDNA High sensitivity kit, prior to NGS. Briefly, two assay tubes for standards (standard 1 and 2) were set up along with a separate tube for the sample. 200.0 µl of Qubit working solution (Qubit reagent: Qubit buffer in 1:200) was added to the standard and sample tubes. All tubes were vortex mixed for 2 - 3 s and incubated for 2 min at RT. Calibration was performed using standards 1 and 2 before the sample was read in the sample chamber. All genomic DNA was diluted to a range of 2.0 - 2.5 ng/µl and 20.0 µl of each diluted genomic DNA was used (50.0 ng of genomic DNA/sequencing) for the preparation of DNA sequencing libraries. The Nextera™ DNA Sample Preparation kit (Illumina, USA) was used to prepare the Dual-Indexed Paired-End Library on CSF samples. The Nextera DNA library preparation work flow is shown in Figure 3.6.1.

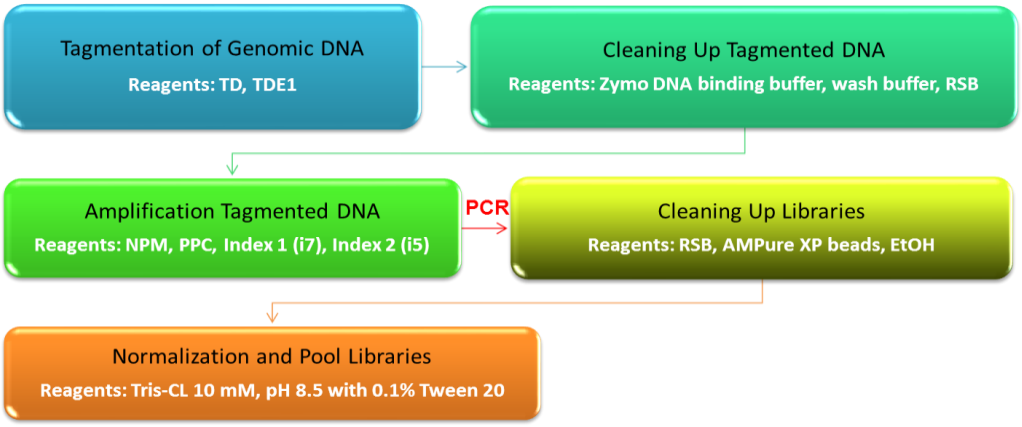


Figure 3.6.1: Work flow using a Nextera DNA Library Prep Kit (Illumina, 2016).

3.6.1.1 Tagmentation of DNA

Going by the manufacturer's protocol, 50.0 ng of the genomic DNA was added to 0.2 µl of PCR tube (Bio-Rad, USA) and was labelled as "NST1-UM-CSF01" (Nextera Sample Tube 1: UM-CSF01). Then, 25.0 µl of Tagment DNA (TD) buffer and 5.0 µl of TDE1 reagents (TDE1) (provided) were added. The mixture was mixed by pipetting up and down 10 X and centrifuged at 280 *x g* at 20 °C for 1 min. NST1 was placed in a thermo-cycler and limited-cycle PCR was performed at 55 °C for 5 min.

3.6.1.2 Purification of Tagmented DNA

The tagmented DNA was purified using the Zymo^{TM} Purification Kit (Zymo Research, USA). A new PCR tube was labelled as "NST2-UM-CSF01" (Nextera Sample Tube 2: UM-CSF01). 180.0 µl of Zymo DNA binding buffer was added into the tube and 50.0 µl of the PCR products from "NST1-UM-CSF01" were transferred into "NST2-UM-CSF01". The mixture was pipetted up and down for 10 X and then centrifuged at 1300 *x g* at 20 °C for 2 min. The flow-through was discarded and 300 µl of Zymo wash buffer was added for another centrifugation. This washing procedure was repeated a few times to ensure that there is no residual wash buffer left in the tube. Then, 25.0 µl of RSB (Re-suspension buffer) was added to the tube and the mixture was incubated at RT for 2 min. It was then centrifuged prior to usage for limited-cycle PCR.

3.6.1.3 Amplification of Tagmented DNA

In order to perform PCR amplification, 5.0 µl of index 2 primer (i5: white cap), 5.0 µl of index 1 primer (i7: orange cap), 15.0 µl of NPM (Nextera PCR Master Mix) and 5.0 µl of PPC (PCR Primer Cocktail) was added into a new PCR tube labelled as "NAT1-UM-CSF01" (Nextera Amplification Tube 1: UM-CSF01). 20.0 µl of the purified tagmented DNA from "NST2-UM-CSF01" was transferred to "NAT1-UM-CSF01" and was gently pipetted up and down for 3 to 5 X to combine the DNA with the PCR mix. The mixture was centrifuged at 280 *x g* at 20 °C for 1 min after which PCR amplification was performed with these parameters: 72 °C for 3 min, 98 °C for 30 s, 5 cycles of 98 °C for 10 s, 63 °C for 30 s, and 72 °C for 3 min.

3.6.1.4 Purification of the Libraries

A new tube was labelled as "NAT2-UM-CSF01" (Nextera Amplification Tube 2-UM-CSF01) and 50.0 µl of the centrifuged PCR products was transferred into it. 30.0 µl AMPure XP beads (Beckman Coulter, USA) was added in and the mixture was pipetted up and down for 10 X, followed by incubation at RT for 5 min. Prior to usage, AMPure XP beads were vortex mixed for 30 s to ensure even dispersal of beads. The tube was placed on a magnetic stand for 2 min and the supernatant was carefully removed and discarded. The beads were then washed with 200.0 µl of 80 % ethanol and incubated for 30 s. Re-suspension of the beads was not allowed

for this step. The supernatant was removed and discarded. The washing step was repeated again. After the beads were air-dried for 15 min, 32.5 µl RSB was added to "NAT2-UM-CSF01". The mixture was re-suspended with gentle up and down pipetting for 10 X and incubated for 2 min at RT. Again, a new PCR tube was labelled as "NLT-UM-CSF01" (Nextera Library Tube: UM-CSF01) and 30 µl of the supernatant from "NAT2-UM-CSF01" was transferred into "NLT-UM-CSF01".

Tagmentation, PCR amplification and clean-up steps were repeated for all other CSF samples and the tubes were labelled accordingly.

3.6.1.5 Normalisation and Pool Libraries

In the next step, 10 µl of the sample library was transferred to the corresponding well of a new MIDI plate labelled with NDP (Nextera Dilution Plate) barcode. The concentration of the sample library was normalized in a sample well to 4 nM using 10 mM of Tris-Cl, pH 8.5 and 0.1 % Tween 20. The plate was sealed and shaken on a micro-plate shaker at 71.33 *x g* for 2 min and centrifuged at 280 *x g* for 1 min. 5 µl of the samples was transferred to the same column of the NPP (Nextera Pooled Plate) in order to generate pooled samples of multiplexed paired-end libraries. Again, NPP was sealed and shaken on a micro-plate shaker at 231 *x g* for 2 min and centrifuged at 280 *x g* for 1 min. The denaturation solution (0.2 M NaOH) was prepared by combining 800 µl laboratory-grade water with 200 µl of 1 M NaOH in a micro-centrifuge tube and the mixture was inverted several times to mix.

Hybridisation Buffer (HT1) was used to dilute the denatured libraries before loading the libraries onto the reagent cartridge for sequencing (Appendix 5: Table A5). Next, 4 nM of library denaturation buffer was prepared by adding 5 µl of 4 nM sample DNA to 5 µl of 0.2 M NaOH. The sample solution was vortex mixed briefly and centrifuged at 280 *x g* for 1 min. The mixture was then incubated for 5 min at RT to denature the DNA into single strands. 990 µl of pre-chilled HT1 was added into the tube with 10 µl of denatured DNA to make a 20 pM denatured library in 1 mM NaOH. The denatured DNA was placed on ice. Next, 180 µl of 20 pM denatured DNA was further diluted into 420 µl of pre-chilled HT1 to give a total volume of 600 µl and the DNA solution was centrifuged. DNA (6 pM) was loaded into the pre-filled reagent cartridge in position 17 and the sequencing was performed on the Illumina MiSeq platform (Illumina, USA).

3.6.1.6 Pre-Sequencing

The reagent cartridge (Appendix 5: Table A5) was thawed in a water bath containing enough RT deionized water to submerge the base of the cartridge for 60 to 90 min, before use. Then, it was inverted 10 X to mix the thawed reagents. Positions 1, 2 and 4 were visually inspected on the cartridge to make sure they were fully mixed and free of precipitates. 'Sequence' button was selected from the software interface to start the run setup in order to transfer the content of the reagent cartridge to the flow cell. The PR2 bottle (with incorporation buffer) was loaded and the waste bottle (0.5 % Tween 20) was emptied. The run could be monitored from a computer using Sequencing

Analysis Viewer (SAV) and a post-run wash was performed after the run by the button 'Start Wash'. After the wash was completed, the used flow cell, wash tray and wash bottle containing the remaining wash solution were left in the instrument.

3.6.2 MiSeq® Sequencing

MiSeq Illumina is one of the most popular options for the laboratory with low to high work flow. This second generation sequencing technology is easy to use, can be used on the DNA from all types of organisms, and allows complete DNA sequencing and analysis, including variant and base calling, alignment and reporting, with every run (Illumina, 2019). Thus, MiSeq was selected to generate the whole genome sequencing (WGS) for the *Mtb* isolates.

Illumina uses a green laser to sequence G/T and a red laser to sequence A/C. At each cycle at least one of two nucleotides for each colour channel was needed to be read to ensure a proper registration. Thus, careful selection of indexes in this study is crucial to obtain an accurate sequencing result (Table 3.6.2).

For sequencing on the MiSeq instrument, prepared samples were placed in the reagent cartridge and loaded on the instrument along with the flow cell. All subsequent steps were performed on the instrument, including cluster generation, paired-end sequencing and primary data analysis. For *de novo* sequencing, the MTB genomic DNA library using Nextera library prep

was sequenced using a 2 X 150 read length on MiSeq. Base call quality scores were written to FASTQ files in an encoded ASCII format (the value + 33). Phasing and pre-phasing calculations were performed using statistical averaging over many clusters and a mixture of different sequences. No quality values were considered for de-multiplexing step. Primary run metrics of the analysis of MiSeq was completed in more than 20 hours.

Table 3.6.2: Sequences for Index 1 and Index 2 Adapters. The dual indexing strategy uses two 8 base indices, index 1 (i7) adjacent to the P7 sequence, and index 2 (i5) adjacent to the P5 sequence. Dual indexing is enabled by adding a unique index 1 (i7) and index 2 (i5) to each sample from 12 different index 1 (i7) adapters (N701-N712) and 8 different index 2 (i5) adapters (N501-N508). In the index adapter name, the N refers to Nextera library preparation, 7 or 5 refers to Index 1 (i7) or Index 2 (i5), respectively, and 01-12 refers to the Index number (Illumina, 2014).

Sample	Index 1 (i7)	Index 2 (i5)
UM-CSF01	N704	N508
UM-CSF04	N702	N507
UM-CSF05	N703	N504
UM-CSF06	N712	N504
UM-CSF08	N709	N506
UM-CSF09	N712	N507
UM-CSF15	N709	N507
UM-CSF17	N703	N517

3.6.3 Read Quality Assessment, Assembly and Annotation

Following cluster generation, clusters were imaged using LED and filter combinations specific to each of the four fluorescently labelled dideoxynucleotides. After the imaging of one tile was completed, the flow cell was moved into place to expose the next tile. This process was repeated for each cycle of sequencing. In this study, 12 tiles of the flow cell were used. During the sequencing run, Real-Time Analysis (RTA) generated data files (FASTQ) that include primary analysis metrics include base calling, filtering and intensity. The raw data was generated in FASTQ format, which is a text format used to represent sequences. The quality of raw sequences generated from MiSeq was checked using FastQC. Raw reads were trimmed at Phred probability score of 30 and were *de novo* assembled using CLC Genomic Workbench 5.1 (Qiagen Inc., Netherlands). Trimmed sequences were assembled with length fraction of 0.8 and similarity fraction of 0.8. All assemblies were evaluated based on statistical assessment, focusing on genome size, sequence continuity and number of contigs. The genomes were further screened for contamination against common contaminants databases and then used for downstream analyses. To decrease the possibility of inaccurate assembly, the assembly and scaffolding of the genomes in IDBA-UD, a *de novo* assembler of NGS data (Peng et al., 2012) and SSPACE, a stand-alone program for scaffolding pre-assembled contigs using NGS paired-read data (Boetzer and Pirovano, 2014), were repeated, respectively.

The Whole Genome Shotgun project for UM-CSF01 and UM-CSF05 has been deposited at DDBJ/ENA/GenBank under the accessions LLXF00000000, LLXG00000000, respectively, and that for UM-CSF04, UM-CSF06, UM-CSF08, UM-CSF09, UM-CSF15 and UM-CSF17 under the accessions LXGA00000000, LXGB00000000, LXGC00000000, LXGD00000000, LXGE00000000 and LXGF00000000, respectively.

3.7 Comparative Genomic Analysis

This analysis was conducted under the guidance of Dr. Tan Joon Liang, bioinformaticist from the Multimedia University (MMU), Melaka, Malaysia.

To reduce the likelihood of observing traits peculiar to local strains, respiratory strains from Malaysian patients were included in the study for comparison. Hence, in addition to the UM-CSF strains, 13 other *Mtb* sputum isolates of patients managed in UMMC over the same time period (2009 to 2011) were *de novo* sequenced (using Illumina Miseq technology), and their genomes were reconstructed in IDBA-UD 1.0.9.

3.7.1 Genome Rearrangements

Gepard (Krumsiek et al., 2007) and Mauve (Darling et al., 2004) were applied to gain an overview of chromosomal structural variation between UM-CSF strains, H37Rv reference strain, 16 genomes (Appendix 6: Table A6)

downloaded on 6th August 2015 from Genome Database (one representative from each group as defined in the database) (https://www.ncbi.nlm.nih.gov/genome/genomegroups/166?) and 13 local sputum strains. This was followed by pairwise-alignment and comparisons using Mugsy with default parameters, to identify the positions of rearranged regions in UM-CSF strains (Angiuoli and Salzberg, 2011).

Multiple alignments of genomic sequences were performed by using Mauve multiple alignment software and the progressive alignment option. The output file produced by Mauve was parsed by using a custom Perl script to retrieve multiple aligned sequences for Indel loci. For each genome-wide Ilumina sequence dataset, the sequence reads were aligned against the reference genome sequence using Burrow-Wheeler Aligner (BWA) 0.6.2 (Li and Durbin, 2010). Following this, SAMtools 0.1.18 (Li et al., 2009), which is based on a Bayesian model for Indel calling was used to perform the analysis using the default Indel detection parameters, with a small increase in the coverage threshold (−D 200); Indels from paired-end mapping data were identified and visualized with a (Qi and Zhao, 2011), which uses read depth and read pair data to detect and visualize large and complex sequence variation.

3.7.2 Identification of Polymorphisms

For this analysis, the sequencing reads for 56 *Mtb* genomes were downloaded from NCBI's SRA database. These strains are reported to be from

sputum strains (Appendix 7: Table A7). These reads and those from UM-CSF and respiratory strains were mapped to H37Rv using BWA. Gene variants were extracted using mpileup of SAMtools (Li et al., 2009) and annotated with snpEff (Cingolani et al., 2012). The variants identified were filtered based on the following criteria: minimum number of good quality read of 3 (DP4 > = 3); minimum mapping quality of 25 (MQ > = 25); SNP and Indel quality of 20 and 60 respectively.

3.7.3 Amino Acid Comparisons

The assembled genomes of UM-CSF strains were annotated using the self-training annotation algorithm in GeneMarkS (Besemer et al., 2001). Orthologous protein sequences were identified in the ProteinOrtho program, with *e*-value of $1x10^{-5}$ (Lechner et al., 2011). The effect of amino acid substitution was evaluated using the I-mutant webserver (https://folding.biofold.org/cgi-bin/i-mutant2.0.cgi) (Capriotti et al., 2005) for change in protein stability, the GlobPlot standalone python script (https://globplot.embl.de/) (Linding et al., 2003) for globularity and the ProTherm database for the calculation of Gibbs free energy change (https://www.abren.net/protherm/).

3.7.4 Identification of Meningitis-Associated Genes

Scientific literature with reports on meningitis-associated genes in *Mtb* were extensively reviewed and collected (Appendix 8: Table A8). Sixty-three

putative genes were revealed in articles from Av-Gay and Everett (2000), Pethe et al (2001), Tsenova et al (2005), Jain et al. (2006), Be et al. (2008), Be et al. (2012) and Haldar et al (2012). The amino acids of these 63 putative genes were compiled and homology studies (refer to 3.7.3) were performed in five other mycobacteria associated with neuropathology, namely *M. leprae*, *M. lepromatosis, M. bovis*, *M. ilatzerense* and *M. immunogenum*. Homology analysis (refer to 3.7.3) was also performed for meningitis-associated proteins reported in *Streptococcus pneumoniae*, *Escherichia coli* K-1 and *Neisseria meningitidis.* These comprised 141 proteins of *Streptococcus pneumoniae* reported by Orihuela et al (2004), Molzen et al (2011) and Mahdi et al (2012) (Appendix 9: Table A9); 164 virulence genes of *Neisseria meningitidis* reported by Hao et al (2011) (Appendix 10: Table A10) and 73 neurotropic genes reported by Pouttu et al (1999), Huang et al (2001) and Yao et al (2006) which confer tissue tropism in *Neisseria meningitidis* and *E. coli K*1 (Appendix 11: Table A11).

3.8 Verification of Putative Findings from WGS analysis

To verify the presence of putative genes detected by the *in silico* study, polymerase chain reaction (PCR) was performed, followed by Sanger sequencing.

3.8.1 Polymerase Chain Reaction (PCR) and Sequencing

Six putative genes identified in UM-CSF strains were selected to be verified using PCR-sequencing (Table 3.8.1). The PCR primers used in this study are listed in Table 3.8.2. These primers were designed based on the sequence of the reference H37Rv genome retrieved from the Mycobrowser website (https://mycobrowser.epfl.ch/). PCR primer pairs were designed with similar melting temperatures (T_m ± 2 °C), low probability of forming thermo-stable secondary structures and homo-dimers in particular, under PCR annealing conditions as determined by Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), and synthesized by 1st BASE Laboratories (Singapore).

Table 3.8.1: Six putative genes selected for PCR verification.

No	Rv number	Gene name	Products	Predictions in CSF isolates							
				1	4	5	6	8	9	15	17
1	Rv3425	PPE57	Unknown				Del*				
2	Rv1141c	echA11	unsaturated acyl-CoA hydratase						Del*		
3	Rv0311[1]	Rv0311	Unknown	√	√	√	√	√	√	√	√
4	RV3344c[2]	PE_PGRS49	PE-PGRS family protein	√	√	√	√	√	√	√	√
5	Rv2606c	snzP	Possible pyridoxine biosynthesis protein SnzP	√	√	√	√	√	√	√	√
6	Rv2397c	cysA1	Sulfate-transport ATP-binding	√	√	√	√	√	√	√	√

		protein ABC transporter CysA1			

* Deletion causing gene truncation.

[1] SNP found at aa position 119: changes Glutamic acid (GAG) to Aspartic acid (GAT).

[2] Deletion found at aa position 164 to 173.

√ Gene predicted

Table 3.8.2: Oligonucleotide primers used to verify 6 putative meningitis-associated genes found in the *in silico* study.

No	Primer	Sequences	Size (bp)	Tm (°C)	Amplicon size (bp)
1	Rv3425 Fwd	ATGCATCCAATGATACCAGCG	21	58.8	531
2	Rv3425 Rev	CTACCCGCCCCTGTAGATC	19	58.7	
3	Rv1141c Fwd	ATGCCAGATTCCGGGATTGC	20	60.8	807
4	Rv1141c Rev	TCAGGAACCGGTGAAGTTGG	20	59.9	
5	Rv3344c Fwd	GCACAGGCCAGTCCGGCG	18	66.4	1455
6	Rv3344c Rev	TCAGGGTGTTGCGCCGGC	18	65.9	
7	Rv0311 Fwd	AGCTGGCAGTTCTGTTACCC	20	57.4	556
8	Rv0311 Rev	CAGGTGGCAGCTTTGGTTTC	20	57.1	
9	Rv2606c Fwd	ATGGATCCTGCAGGTAACCC	20	56.8	900
10	Rv2606c Rev	TCACCAGCCGCGCTGGGCGA	20	69.6	
11	Rv2397c Fwd	TCGGATTCGTCTTCCAGCAC	20	57.2	471
12	Rv2397c Rev	AGAAGGACATCACGAAGGCG	20	57.2	

Each PCR consisted of 12.5 µL master mix (GoTaq Green Master Mix; Promega), 2 µL of each forward and reverse primers, 1 µL DNA template and distilled water in a total volume of 25 µL. The thermal cycle procedure was based on a three-step cycling procedure: initiated with pre-denaturation at 95 °C for 2 min, followed by 30 cycles of denaturation at 95 °C for 45 s,

primer annealing temperature, 56.1 °C to 60.2 °C for 30 s, and extension at 72 °C for 2 min. Finally, the PCR process was ended by a final extension at 72 °C for 5 min. When the reaction was completed, 5 μL of the PCR amplicons were analysed by agarose gel electrophoresis. The amplicons were then purified and subjected to DNA sequencing using forward and reverse primers.

CHAPTER 4

RESULTS

4.1 Suppression Subtractive Hybridisation (SSH)

SSH was used to survey the differences in the gene content of pulmonary and CNS *Mtb* isolates that might be related to neurotropism. The results at various steps in the SSH protocol are described in 4.1.1 to 4.1.5.

4.1.1 Quantity and Quality of Genomic DNA

Genomic DNA was extracted from two strains of *Mtb* isolated from sputum (DNA samples B and F) and eight strains isolated from CSF (DNA samples 01, 04, 05, 06, 08, 09, 15 and 17). The quantities of the extracted genomic DNA were between 29.1 ng/µl to 762.0 ng/µl (Table 4.1.1). The ratio of the absorbance at 260 nm and 280 nm ($A_{260/280}$) was 1.61 to 1.83, which is a little below the recommended ratio of 1.8 to 2.0 (Promega Corporation, 2015). As the A_{280} nm represents maximum light absorption by proteins and phenol, about 60 % of the samples could have been contaminated with residual phenol or other reagents associated with the extraction protocol. Nevertheless, these slightly lower ratios are still generally acceptable for most applications (Wilfinger et al., 1997; Matlock, 2015; Promega Corporation, 2015).

The quantity and quality of the extracted DNA was also evaluated by agarose gel electrophoresis. Electrophoresis on a 2.0 % agarose gel revealed clearly visible genomic DNA bands (Figure 4.1.1). By comparing the band intensity of samples to that of a DNA quantitation standard (Sambrook et al., 1989; Lee et al., 2012), the concentration of UM-CSF05, UM-CSF06, UM-CSF09 and UM-Sputum F appeared more than 200 ng/µl, while that of UM-CSF08, UM-CSF15, UM-CSF17 and UM-Sputum B appeared to be lower than 60 ng/µl.

Table 4.1.1: Concentration and purity of the extracted and purified DNA from 8 CSF and 2 sputum isolates.

Sample	Date (samples received from UMMC)	Name	*LN	Sample type	Conc. (ng/µl)	A260/A280
B	02/07/2013	T. S. P.	1112443082	Sputum	42.5	1.60
F	30/07/2012	Unknown	1092420059	Sputum	418.0	1.80
01	20/04/2012	T. T. L.	2567521	CSF	150.0	1.61
04	27/07/2012	K. A. Y.	1042343702	CSF	343.0	1.70
05	27/07/2012	Unknown	3001987	CSF	762.0	1.80
06	27/07/2012	V.	1062370273	CSF	385.0	1.80
08	02/09/2013	N. F.	3050172074	CSF	29.1	1.83
09	13/10/2011	X. J. L.	3128028	CSF	226.5	1.60
15	20/11/2013	F.	3080247132	CSF	57.7	1.74
17	20/11/2013	T. S. L.	3090261031	CSF	52.5	1.64

*LN=laboratory number

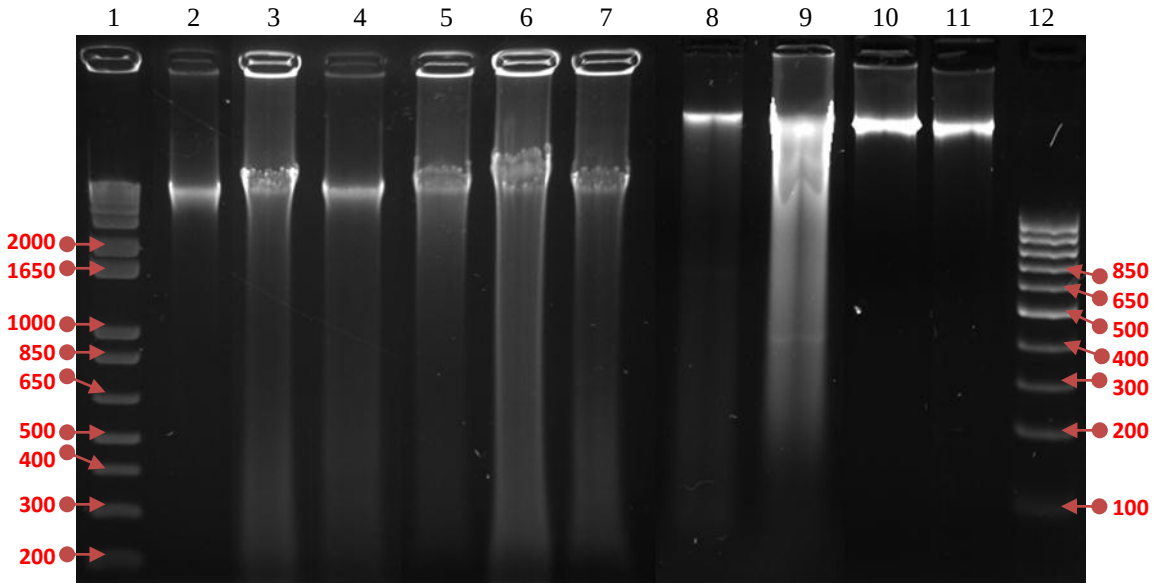


Figure 4.1.1: Agarose gel electrophoresis (2.0 %) of the extracted genomic DNA, performed at 80 V for 50 min. Lane 1 and 12: 1 kb Plus DNA Ladder (Invitrogen, USA); Lane 2: Sputum B; Lane 3: Sputum F: Lane 4: UM-CSF01; Lane 5: UM-CSF04; Lane 6: UM-CSF05; Lane 7: UM-CSF06; Lane 8: UM-CSF08; Lane 9: UM-CSF09; Lane 10: UM-CSF15; Lane 11: UM-CSF17.

4.1.2 Analysis of *Rsa*I Digestion

In the comparison of *Rsa*1 from Invitrogen and Clontech, the Invitrogen enzyme gave better restriction patterns (results not shown). Hence, this enzyme was used to digest the DNA extracted from two sputum isolates and eight CSF isolates. From the results of the digestion, DNA samples of sputum B and F together with CSF 01 and 04 were chosen for SSH as they were more completely digested compared to the other samples. In the gel electrophoresis (Figure 4.1.2), the undigested genomic DNA which should appear as a high-molecular weight band at the top of the gel, is not seen. In its

place, the digested DNA, which is much smaller in size, appears as a 0.1 to 0.2 kb smear. This gel pattern is consistent with that obtained for the control genomic *E. coli* DNA. The DNA size distribution for control *E. coli* genomic DNA, driver sputum F sample, tester UM-CSF01 and UM-CSF04 samples were identical, indicating complete digestion for these samples.

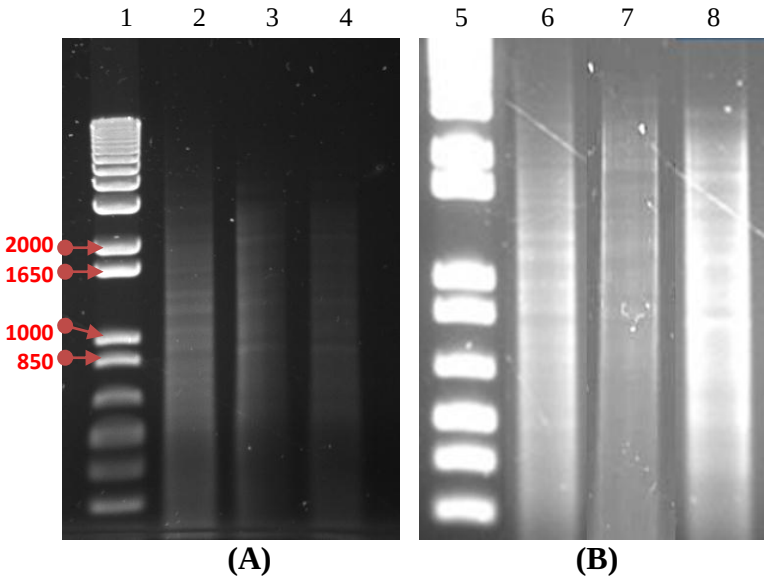


Figure 4.1.2: Agarose gel electrophoresis (1.0 %) of *Rsa*I digested genomic DNA from sputum and CSF *Mtb* isolates; (A) Lane 1: 1 kb Plus DNA Ladder (Invitrogen, USA); Lane 2: *E.coli* genomic DNA (0.2 µg); Lane 3: UM-CSF01; Lane 4: Sputum F; (B) Lane 5: 1 kb Plus DNA Ladder (Invitrogen, USA); Lane 6: *E.coli* genomic DNA (0.2 µg); Lane 7: UM-CSF04; Lane 8: Sputum B.

4.1.3 Analysis of Ligation Efficiency

Since the ends of the adaptors 1 and 2R are un-phosphorylated, only one strand of each adaptor can attach to the 5' ends of the DNA. The two adaptors have stretches of identical sequence, which allows annealing of the

same PCR primer to both ends once the recessed ends have been filled in. These features allow the analysis to determine how efficiently the adaptors 1 and 2R are ligated with the genomic *E. coli* DNA. Both PCR products from 2 different primer sets (23S RNA F-Primer 1 and 23S RNA F-23S RNA R) should show the same intensity on gel electrophoresis. A more than 4-fold difference in band intensity for the PCR products indicates less than 25% completion of ligation, a factor that could reduce subsequent subtraction efficiency.

As shown in Figure 4.1.3, the band intensity of the PCR product generated using one gene-specific primer (23S RNA Forward Primer) and PCR Primer 1 is about the same as that for the PCR product amplified using two gene-specific primers (23S RNA Forward and Reverse Primers); hence, the ligation is considered to be more than 25 % complete and adequate for subsequent subtraction efficiency.

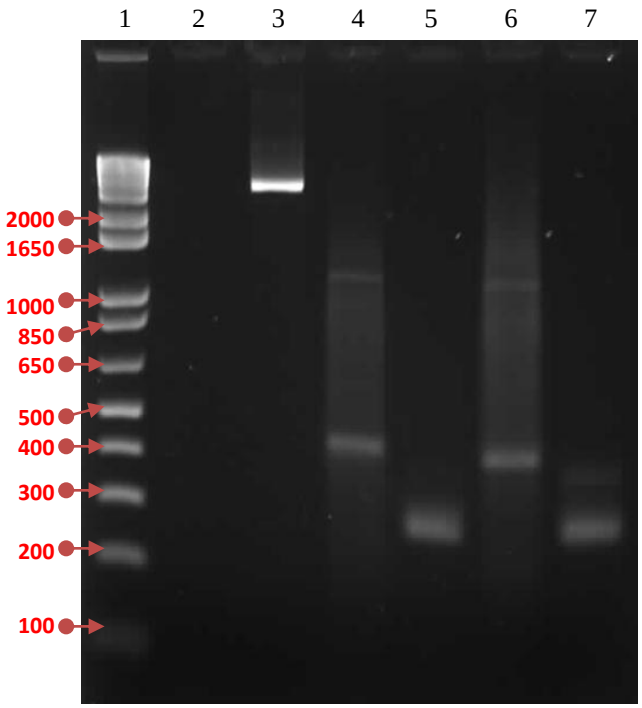


Figure 4.1.3: Gel electrophoresis analysis of ligation efficiency for the control *E. coli* genomic DNA. The primers, provided by the kit, were designed to amplify fragments that span the adaptor/DNA junctions of tester samples. Lane 1: 1 kb Plus DNA marker (Invitrogen, USA); Lane 2: PCR non-template control; Lane 3: PCR positive control (3.5 kb); Lane 4: *E. coli* control adaptor 1-ligated DNA, 23S RNA Fwd-Primer 1 (374 bp); Lane 5: *E. coli* control adaptor 1-ligated DNA, 23S RNA Fwd-23S RNA Rev (270 bp); Lane 6: *E. coli* control adaptor 2-ligated DNA, 23S RNA Fwd-Primer 1 (374 bp); Lane 7: *E. coli* control adaptor 2-ligated DNA, 23S RNA Fwd-23S RNA Rev (270 bp). Similar band intensities of PCR products in lanes 4 and 5 as well as 6 and 7 indicate at least 25 % completion of the ligation reactions. Some non-specific amplification was observed for *E. coli* control adaptor 1-ligated DNA (lane 4) and *E. coli* control adaptor 2-ligated DNA (lane 6), amplified with 23S RNA Forward Primer and Primer 1.

4.1.4 The Subtracted Sequences

After two sequential hybridisations and two suppression PCR amplifications, common genes present in both tester and driver samples should be subtracted out, whereas genes that are present only in the tester samples should be selectively amplified. Using CSF01 and CSF04 as testers and sputum F and B as their corresponding drivers, the subtracted DNA of CSF01 and CSF04 showed some major bands in the gel electrophoresis, which appeared after 30 cycles of the secondary PCR. This indicates that the DNA fragments in the tester samples but not in the driver samples were effectively enriched and amplified. Primary PCR subtraction products appeared as a 0.2 to 2.0 kb smear, with some distinct bands within the smear (Figure 4.1.4).

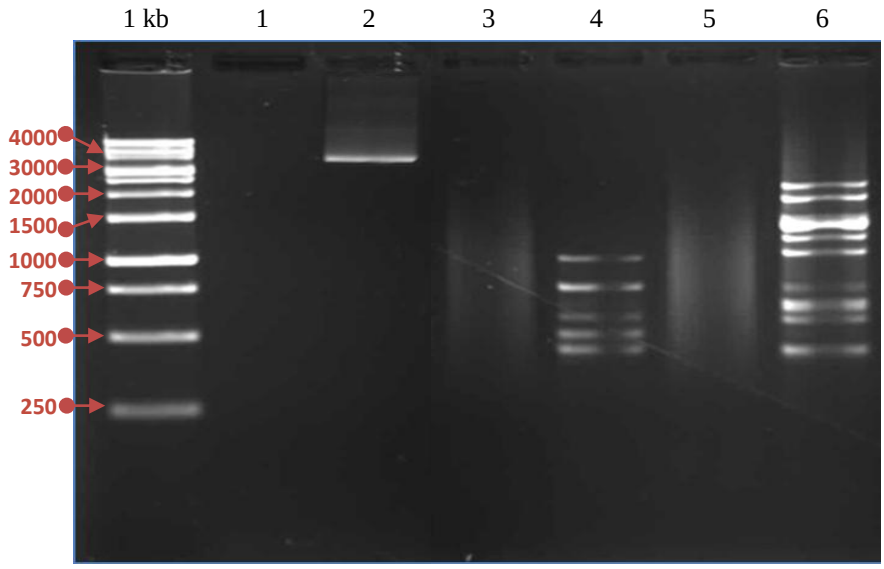


Figure 4.1.4: Agarose gel electrophoresis (2.0 %) of primary (un-subtracted) and secondary PCR products using 1 kb of DNA marker (Bioron Life Science, Germany). Lane 1: PCR negative control; Lane 2: PCR positive control (3.5 kb); Lane 3: Primary PCR for tester DNA (Sputum F: UM-CSF01); Lane 4: Secondary PCR/subtracted tester DNA (Sputum F: UM-CSF01); Lane 5:

Primary PCR for tester DNA (Sputum B: UM-CSF04); Lane 6: Secondary PCR/subtracted tester DNA (Sputum B: UM-CSF04).

A total of 15 subtracted sequences were generated from these 2 pairs of tester-driver DNA. The pair of sputum F: UM-CSF01 yielded 6 subtracted sequences FI-1 to F1-6 while the sputum B: UM-CSF04 pair yielded 9 subtracted sequences B4-1 to B4-9. Their concentrations and purities were measured and tabulated in Table 4.1.4.

Table 4.1.4: Concentration, purity and size of the purified subtracted DNA. The DNA was obtained using QIAquick Gel Extraction Kit (Qiagen, Netherlands). The sizes of the subtracted sequences are estimates.

No	Subtracted sequences	Size (bp)	Concentration (ng/µl)	A260/A280
1	**F1-1**	1000	10.000	1.333
2	**F1-2**	800	7.500	1.500
3	**F1-3**	600	10.000	1.000
4	**F1-4**	500	5.000	1.000
5	**F1-5**	450	17.500	2.333
6	**F1-6**	300	10.000	1.333
7	**B4-1**	2000	10.000	1.333
8	**B4-2**	1600	15.000	1.500
9	**B4-3**	1300	20.000	1.600
10	**B4-4**	1100	10.000	1.000
11	**B4-5**	950	5.000	---
12	**B4-6**	700	15.000	1.333
13	**B4-7**	600	10.000	0.500
14	**B4-8**	500	5.000	1.000
15	**B4-9**	400	7.500	0.400

F1-1 to F1-6, subtracted sequences from sputum F: UM-CSF01.

B4-1 to B4-9, subtracted sequences from sputum B: UM-CSF04.

4.1.5 Molecular Transformation of the Competent *E. coli* JM109 cells with Plasmid pGEM®-T Ligated-Purified Subtracted DNA

After adding 800 µl of S.O.C. medium to 200 µl of competent cells that have been transformed with 0.8 ng controls insert DNA, 200 µl (equivalent to 0.2 ng DNA) to 800 µl of S.O.C. medium was transferred and plated on the agar (equivalent to 0.04 ng). 60 colonies were detected on the LB/Amp/IPTG/X-Gal agar plate.

Transformation efficiency of the competent *E. coli* strain JM109 cells

= [Colony count/Amount (ng) of DNA plated] x 1000 ng/µg

= (60/0.04) x 1000 ng/µg

= $\underline{1.5 \times 10^6 \text{ cfu/µg}}$

As efficient competent cells should have transformation efficiencies in between 5 x 10^6 to 2 x 10^7 transformed colonies/µg of supercoiled plasmid DNA (Nishimura et al., 1990), the transformation efficiency of *E. coli* strain JM109 was about 25.0 % lower than expected. Nevertheless, these cells were used to perform the transformation for the 15 subtracted sequences.

The optimal insert:vector molar ratio at 3:1 was used for the insert ligation in the study (Promega Corporation, 2010). Table 4.1.5 summarizes the amount of insert needed (ng) (calculation formula from Figure 3.5.7.2) and the amount of ligation volumes used for the cloning of 15 purified subtracted

DNA into the pGEM®-T vector. The final amount of the purified subtracted DNA was between 15.0 ng to 100.0 ng. There were 46 clones (F1-1 to F1-6) from the pair of Sputum F: UM-CSF01 and 27 clones (B4-1 to B4-9) from the pair of Sputum B: UM-CSF04, after transformation into the host cells. Only 43.5 % clones from the F1 pairs were successful in the ligation process and only 14.8 % clones from the B4 pairs showed white (transformed) colonies. The transformation efficiency for all the subtracted sequences, calculated using the formula shown in Figure 3.5.6.3, was from 0.0 to 6.8 x 10^4 cfu/µg. These values were 10 times lower than expected. However, the experiments were continued by picking the white colonies for further studies on the selected subtracted sequences.

Table 4.1.5: Blue-White screening results for transformed *E. coli* JM109 cells.

Type	Samples	ng of insert	Ligation volume* (µl)	Number of colonies			
				Blue	Half Blue	White^	Total
Sputum F: CSF1	**F1-1**	50.0	5.0	3	3	5	11
	F1-2	40.0	5.0	2	0	1	3
	F1-3	30.0	3.0	0	5	3	8
	F1-4	25.0	5.0	7	2	8	17
	F1-5	22.5	1.0	1	2	2	5
	F1-6	15.0	2.0	0	1	1	2
Total				**13**	**13**	**20**	**46**
Sputum B: CSF 4	**B4-1**	100.0	10.0	0	2	0	2
	B4-2	80.0	5.5	1	2	0	3
	B4-3	65.0	4.0	2	2	0	4
	B4-4	55.0	5.0	0	0	0	0
	B4-5	47.5	9.5	0	0	2	2

	B4-6	35.0	3.0	0	2	0	2
	B4-7	30.0	3.5	2	1	0	3
	B4-8	25.0	5.5	1	0	1	2
	B4-9	20.0	3.0	2	6	1	9
Total				**8**	**15**	**4**	**27**

* Calculated by size in kb and concentration of the insert(s).

^ Colonies with the desired insert.

4.1.6 Identification of Positive Clones via *EcoR*I Digestion

White *E. coli* transformants were screened by digesting the purified plasmids with *EcoR*I which cleaves the nucleotides flanking the cloning site of pGEM-T Easy. After screening 73 possibly transformed colonies with colony PCR, only 17 (18.7 %) were found to be positive for inserts. These clones were pGEM-T-F1-1 clones 1, 2, 4, 22; pGEM-T-F1-4 clones 24, 26, 28, 30; pGEM-T-F1-5 clones 34, 37, 39; pGEM-T-F1-6 clone 41 (Figure 4.1.6.1 (A)) and pGEM-T-B4-5 clones 50, 51; pGEM-T-B4-7 clone 56 and pGEM-T-B4-9 clones 65 and 67 (Figure 4.1.6.1 (B)). However, only 11 clones were selected randomly for verification by *EcoR*I enzyme digestion. Some were duplicates from the same plates (Figure 4.1.6.2). These were pGEM-T-F1-1 colony 2; pGEM-T-F1-4 colony 24 and 30; pGEM-T-F1-5 colony 37 and 39; pGEM-T-F1-6 colony 41; and all the pGEM-T-B4 colonies (Figure 4.1.6.2). *EcoR*I produced 2 distinct bands: the linearized pGEM®-T vectors (3.0 kb) and the fragment DNA (subtracted sequences). These clones were also sent for DNA sequencing for further verification (Table 4.1.6).

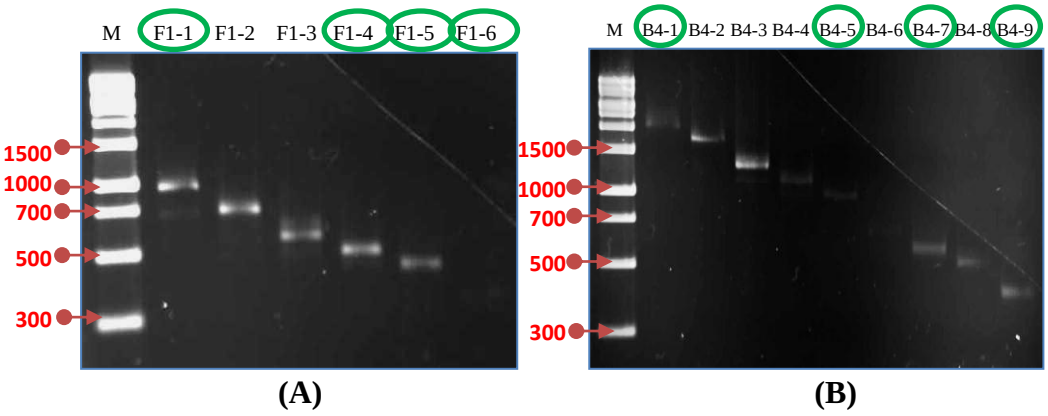


Figure 4.1.6.1: Agarose gel electrophoresis (2.0 %) of subtracted DNAs from paired sputum and CSF isolates. Lane M: HighRanger 1 kb DNA Ladder (Norgen Biotek Corporation, Canada) (A) Lane 1-6: subtracted sequences from F1-1 to F1-6 (size ranges from 1000 bp, 800 bp, 600 bp, 500 bp, 450 bp and 300 bp); successfully cloned samples circled in green; (B) Lane 1-9: subtracted sequences from B4-1 to B4-9 (size ranges from 2000 bp, 1600 bp, 1300 bp, 1100 bp, 950 bp, 650 bp, 600 bp, 500 bp and 400 bp); successfully cloned samples circled in green.

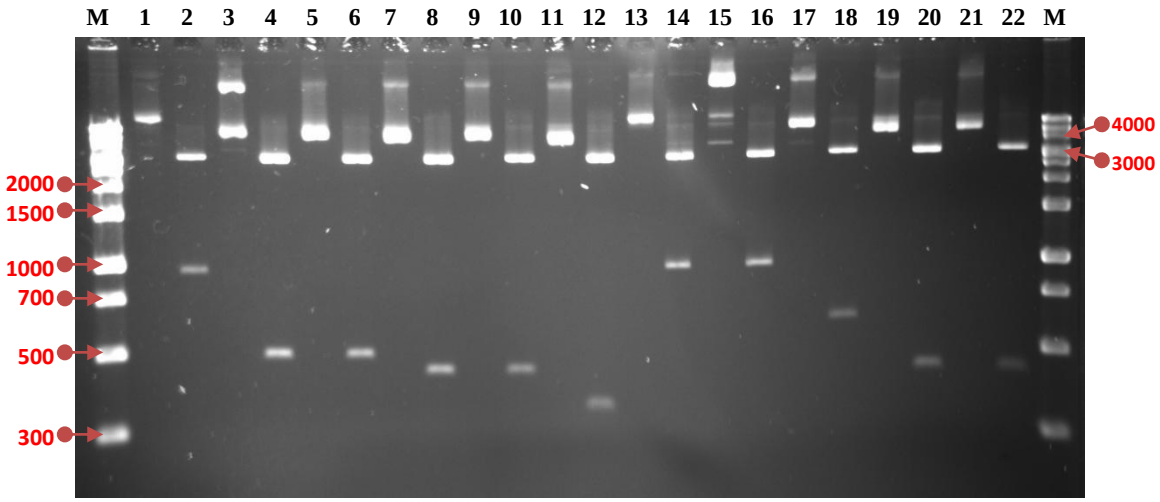


Figure 4.1.6.2: Agarose gel electrophoresis (2.0 %) of *EcoR*I digested plasmid DNA, with the size of subtracted sequences indicated (bp). Lane M: HighRanger 1 kb DNA Ladder (Norgen Biotek Corporation, Canada); Lane 1:

pGEM-T-F1-1 (2) undigested; Lane 2: pGEM-T-F1-1 clone 2 digested (1000 bp); Lane 3: pGEM-T-F1-4 clone 24 undigested; Lane 4: pGEM-T-F1-4 clone 24 digested (500 bp); Lane 5: pGEM-T-F1-4 clone 30 undigested; Lane 6: pGEM-T-F1-4 clone 30 digested (500 bp); Lane 7: pGEM-T-F1-5 clone 37 undigested; Lane 8: pGEM-T-F1-5 clone 37 digested (400 bp); Lane 9: pGEM-T-F1-5 clone 39 undigested; Lane 10: pGEM-T-F1-5 clone 39 digested (400 bp); Lane 11: pGEM-T-F1-6 clone 41 undigested; Lane 12: pGEM-T-F1-6 clone 41 digested (300 bp); Lane 13: pGEM-T-B4-5 clone 50 undigested; Lane 14: pGEM-T-B4-5 clone 50 digested (950 bp); Lane 15: pGEM-T-B4-5 clone 51 undigested; Lane 16: pGEM-T-B4-5 clone 51 digested (950 bp); Lane 17: pGEM-T-B4-7 clone 56 undigested; Lane 18: pGEM-T-B4-7 clone 56 digested (≈ 600 bp); Lane 19: pGEM-T-B4-9 clone 65 undigested; Lane 20: pGEM-T-B4-9 clone 65 digested (≈ 400 bp); Lane 21: pGEM-T-B4-9 clone 67 undigested; Lane 22: pGEM-T-B4-9 clone 67 digested (≈ 400 bp).

Table 4.1.6: Summary of the insert-positive clones verified and sequenced.

No	pGEM-T-insert	Size (bp)	Verified positive clones	Clones sequenced
1	F1-1	900	4	1
2	F1-2	700	NA	NA
3	F1-3	600	NA	NA
4	F1-4	>500	4	2
5	F1-5	<500	3	2
6	F1-6	>300	1	1
7	B4-1	2000	NA	NA
8	B4-2	>1500	NA	NA
9	B4-3	<1500	NA	NA
10	B4-4	>1000	NA	NA

11	B4-5	<1000	2	2
12	B4-6	<700	NA	NA
13	B4-7	>500	1	1
14	B4-8	500	NA	NA
15	B4-9	>300	2	2
		TOTAL	**17**	**11**

*NA: Not available.

4.1.7 Sequencing and Homology Analysis of the Inserts

From the 11 clones verified with RE digestion, seven inserts (subtracted sequences) were identified by their amplicon size and Sanger sequencing. After trimming and filtering to remove vector and primer sequences, the sizes of the 7 subtracted sequences were: 886 bp (F1-1-2), 441 bp (F1-4-24/30), 375 bp (F1-5-37/39), 245 bp (F1-6-41), 722 bp (B4-5-50/51), 581 bp (B4-7-56) and 370 bp (B4-9-65/67). These gene sequences are listed in Appendix 12: Table A12. With NCBI-BLASTN the subtracted sequences were found to be 99 - 100 % homologous with sequences reported in a strain of *Mtb* isolated from the brain of a patient in India (Table 4.1.7.1). The Expect value (E) for all the subtracted sequences was zero or closer to zero, which meant the match with *Mtb* strain C3 was significant. The Expect value describes the number of hits one can 'expect' to see by chance when searching a database of a particular size. It decreases exponentially as the Score (S) of the match increases. On the other hand, BLASTX analysis showed the subtracted proteins to be classified as a PPE protein, integral membrane protein, secretion system protein, regulatory protein, *Mtb* virulence- and

resistance-related proteins, as well as a hypothetical protein (Table 4.1.7.2). E-values were almost near zero and the identities ranged from 80 - 100 %.

Table 4.1.7.1: BLASTN results of subtracted DNAs.

No	Positive clones	Score	E-value	Identity (%)	Description
S1	pGT-F1-1-2	1624	0.0	99 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S2	pGT-F1-4-24&30	815	0.0	100 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S3	pGT-F1-5-37&39	693	0.0	100 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S4	pGT-F1-6-41	448	1e-122	99 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S5	pGT-B4-5-50&51	1334	0.0	100 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S6	pGT-B4-7-56	1068	0.0	99 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*
S7	pGT-B4-9-65&67	678	0.0	99 %	*Mycobacterium tuberculosis* strain C3 (Brain: India)*

*Bacteria available from Pathology Lab, Central India Institute of Medical Sciences, Nagpur, India (Husain et al., 2017).

Table 4.1.7.2: BLASTX results of 7 subtracted sequences.

No	Positive clones	Score	*E*-value	Identity (%)	Homologous gene
S1	pGT-F1-1-2	195	3e-58	82	PPE family protein
S2	pGT-F1-4-24/30	241	1e-79	100	Divalent cation-transport integral membrane protein
S3	pGT-F1-5-37/39	257	3e-86	100	Esx conserved component EccA5

S4	pGT-F1-6-41	182	1e-53	100	Uncharacterized protein
S5	pGT-B4-5-50/51	211	9e-67	100	Cationic amino acid transport integral membrane protein
S6	pGT-B4-7-56	201	2e-61	99	Beta lactamase
S7	pGT-B4-9-65/67	251	4e-84	100	dTDP-4-dehydrorhamnose reductase

4.2 Whole Genome Sequencing (WGS)

Whole genome sequencing (WGS) was used to broaden the comparison of genomic features between *Mtb* isolates from extra-pulmonary tuberculosis and *Mtb* isolates from pulmonary tuberculosis.

The eight CSF isolates sequenced were: UM-CSF01, UM-CSF04, UM-CSF05, UM-CSF06, UM-CSF08, UM-CSF09, UM-CSF15 and UM-CSF17. The respiratory genomes used for comparison were: 13 from sputum isolates of Malaysian patients, 72 downloaded from NCBI databases (16 from NCBI Genome database and 56 from NCBI SRA database).

4.2.1 Genome Data Trimming and Assembly

Plots descriptions of data trimming and results of UM-CSF01 are shown as an example in Appendix 13: Table A13. The genomes of UM-CSF strains showed approximately 55X to 92X sequencing coverage. The detailed statistical measurements of the genomes after assembly are shown in Table

4.2.1. Details of whole genome sequences including the accession number and descriptions for the eight UM-CSFs are listed in Appendix 14: Table A14.

Table 4.2.1: Statistical measurements of the UM-CSF genomes.

Strain	N_{50}	No. contigs	No. scaffolds	Reads used (%)	Scaffold genome Size (bp)	No. protein CDS
UM-CSF01	36,496	315	234	96.67	4,282,569	4271
UM-CSF04	98,670	140	136	93.09	4,392,768	4323
UM-CSF05	74,553	182	138	97.56	4,338,921	4190
UM-CSF06	83,364	135	127	87.01	4,371,105	4310
UM-CSF08	91,928	174	156	93.77	4,352,163	4353
UM-CSF09	122,404	133	103	89.71	4,355,856	4291
UM-CSF15	91,928	129	126	94.05	4,374,121	4313
UM-CSF17	122,408	131	111	90.99	4,356,783	4308

4.2.2 Comparative Genomic Analysis

Compared to H37Rv and the other 29 respiratory *Mtb* genomes (16 downloaded from NCBI and 13 extracted from Malaysian isolates), the UM-CSF strains appeared to have a larger number of PE/PPE/PGRS proteins but fewer gene duplications. There is no notable difference between the two groups in the structure of 16S rRNA and tmRNAs and in the number of tRNAs (20 - 30 for CSF strains and 19 - 33 for respiratory strains) (Appendix 15: Figure A15 and Appendix 16: Figure A16).

4.2.3 Genome Rearrangements

Genome rearrangement is a phenomenon which involves large-scale changes in the structure of the chromosome, such as insertion, deletion, duplication and translocation. These genomic events could cause differences in gene order. It is well known that the gene order in bacteria is poorly conserved during evolution (Mushegian and Koonin, 1996; Casjens, 1998; Huynen and Bork, 1998). For instance, many homologous genes shared by *Escherichia coli*, *Haemophilus influenza* and *Helicobacter pylori* are found in very different locations or positions in each genome (Tatusov et al., 1996; Kolstø, 1997; Tamames et al., 1997; Casjens, 1998). The complete genomes of closely related bacteria such as *Mycobacterium* species allow determination of the processes and mechanisms involved in genome evolution.

The UM-CSF strains showed structural differences from H37Rv and the 29 respiratory *Mtb* genomes used for comparison. Rearrangement analysis by Gepard and Mauve gave identical results (Appendix 17: Figure A17.1 to A17.6). Both analyses indicated sequence fragments that could have undergone rearrangement events in six of the eight UM-CSF strains (UM-CSF01, UM-CSF05, UM-CSF06, UM-CSF09, UM-CSF15 and UM-CSF17). The affected regions ranged from 223 to 500,211 bp in one to eight contigs and involved one to 492 genes (Appendix 18: Table A18). Based on the annotation of H37Rv, many of the genes affected belonged to PE/PPE/PGRS and mammalian cell entry gene families that are known to be associated with

virulence properties in *Mtb*. Also affected were genes encoding transcriptional factors, metabolic enzymes and toxin-antitoxin proteins.

In UM-CSF01 and UM-CSF05, the recombination sites for translocations and inversions were not within genes and hence, did not affect adjacent gene sequences (Appendix 19: Figure A19.1 to A19.2). In the remainder four affected UM-CSF strains, the rearrangements caused deletions resulting in gene truncations (Appendix 20: Figure A20.1 to A20.4). The affected genes are Rv3425 encoding PPE57 in UM-CSF06; Rv1141c encoding enoyl-CoA hydratase EchA11 in UM-CSF09; Rv1587c coding for a hypothetical protein in UM-CSF15, and Rv3513c encoding a fatty-acid-CoA ligase FadD18 in UM-CSF17. The function of PPE57 is unknown but both EchA11 and FadD18 are suspected to be involved in lipid metabolism, the former in fatty acid oxidation and the latter in lipid degradation.

4.2.4 Micro-variants

Micro-variants are short tandem repeats with fractional value arisen from a deletion, mutational loss of one or more nucleotides, or insertions. They are sometimes referred to as micro-alleles, fractional markers or partial repeats (ISOGG Wiki, 2017). Against the H37Rv genome, 737 to 2578 micro-variants were identified in the UM-CSF strains. To identify only those micro-variants possibly associated with cerebrospinal invasion, the UM-CSF strains were compared with another 69 respiratory *Mtb* genomes (56 sets of raw reads downloaded from the NCBI SRA database (Appendix 7: Table A7) and 13

local respiratory *Mtb* genomes). In this comparison, 63 to 534 micro-variants were found specific to the eight UM-CSF strains, within protein coding regions. None were in any of the regions of difference (RD1 to RD16) where many *Mtb* strain-specific features and virulence factors are normally found. No variant was shared by all eight strains but 36 variants involving 10 genes (PE-PGRS10, PPE58, PE_PGRS49, lppD, PE_PGRS21, Rv0278c, embR, PE_PGRS19, PPE53 and PPE24) were found in at least four of the strains. The variants in eight of these genes led to amino acid changes but only two altered genes have known functions: PE_PGRS19, a putative outer membrane protein (Song et al., 2008) and embR which is involved in transcription cell wall arabinan synthesis and ethambutol resistance (Table 4.2.4).

Table 4.2.4: Eight genes found in at least four of the UM-CSF strains, with gene variants that led to amino acid (aa) changes. Only two altered genes (PE_PGRS19 and embR) have known functions.

Gene	Rv no.	Amino acid position	Affected genes
embR	Rv1267c	372	Cys changed to Gly
embR	Rv1267c	376	Phe changed to Leu
lppD	Rv1899c	between 50 and 51	Insertion - Ala,Ser,Thr,Ala,Arg,Pro,Ala,Ala,Thr,Ala,Leu,Pro,Ala,Val,Ala
PE_PGRS10	Rv0747	447	Ala changed to Asp
PE_PGRS10	Rv0747	450	Ala changed to Asp
PE_PGRS10	Rv0747	277	Asp changed to Asn
PE_PGRS10	Rv0747	362	Ile changed to Leu
PE_PGRS10	Rv0747	457	Leu changed to Phe
PE_PGRS10	Rv0747	287	Ser changed to Ala
PE_PGRS19	Rv1067c	271	Ala changed to Val

PE_PGRS21	Rv1087	between 711 and 712	Insertion - Ala,Gly,Gly,Gly,Gly,Gly,Ala,Gly,Gly,Ile,Gly,Gly,Asp,Gly,Gly
PE_PGRS49	Rv3344c	from 164 to 173	Del - Ala,Thr,Asn,Pro,Gly,Ser,Gly,Ser,Arg,Gly
PPE58	Rv3426	107	Thr changed to Ala
PPE58	Rv3426	from 217 to 218	Del - Met,Val
Rv0278c	Rv0278c	346	Ala changed to Ser
Rv0278c	Rv0278c	842	Ala changed to Gly
Rv0278c	Rv0278c	344	Met changed to Thr
Rv0278c	Rv0278c	841	Ser changed to Tyr

4.2.5 Protein Globularity Changes

In UM-CSF strains, 1,556 amino acid substitutions were noted in 1,084 orthologs of proteins in H37Rv. With GlobPlot analysis, 646 of the substitutions were predicted to be in disordered segments, and 910 in globular segments. Figure 4.2.5 shows an example of the amino acid and globularity differences between UM-CSF strains and H37Rv. All proteins showing different globularity between UM-CSF strains and H37Rv were subjected to enrichment analysis using DAVID (https://david.ncifcrf.gov/) (Dennis Jr et al., 2003) (Appendix 21: Table A21). The single largest group of the enriched proteins involved transcription factors and transcription regulators. Four of the enriched proteins were each found in three different CSF strains. These were Rv0144 (probable transcriptional regulatory protein with Leu to Asp substitution), Rv3730c (hypothetical protein with Thr to Asp substitution), Rv0802c (possible succinyltransferase with Ser to Pro substitution) and

Rv2034 (transcriptional regulatory protein with Ala to Thr substitution). All four showed higher propensity to disorder, and all, with the exception of Rv0802c, were accompanied by a decrease in stability (Appendix 22: Figure A22). These changes are consistent with previous observations that amino acid substitutions leading to the acquisition of new protein functions are often accompanied by a loss of thermodynamic stability which may be compensated subsequently by the stabilizing effect of other unrelated mutations (Shoichet et al., 1995; Wang et al., 2002; Tokuriki et al., 2008; Studer et al., 2013). Furthermore, all four proteins were enriched (49 – 53 %) in the disorder-promoting amino acids (A, R, G, Q, S, P, E, K) and depleted (28 – 34%) of order-promoting amino acids (W, C, F, I, Y, V, L, N) as described by Dunker et al. (2001). Hydrophobic amino acids form less than 50 % of the amino acid sequence in each of these disordered proteins (Appendix 23: Figure A23). As many disordered proteins or protein domains are functionally related to disease, it is possible that some of those identified in UM-CSF strains might contribute to an enhanced ability to cause TBM. However, in a widen comparison with respiratory strains, many of the disordered proteins were also found in respiratory strains.

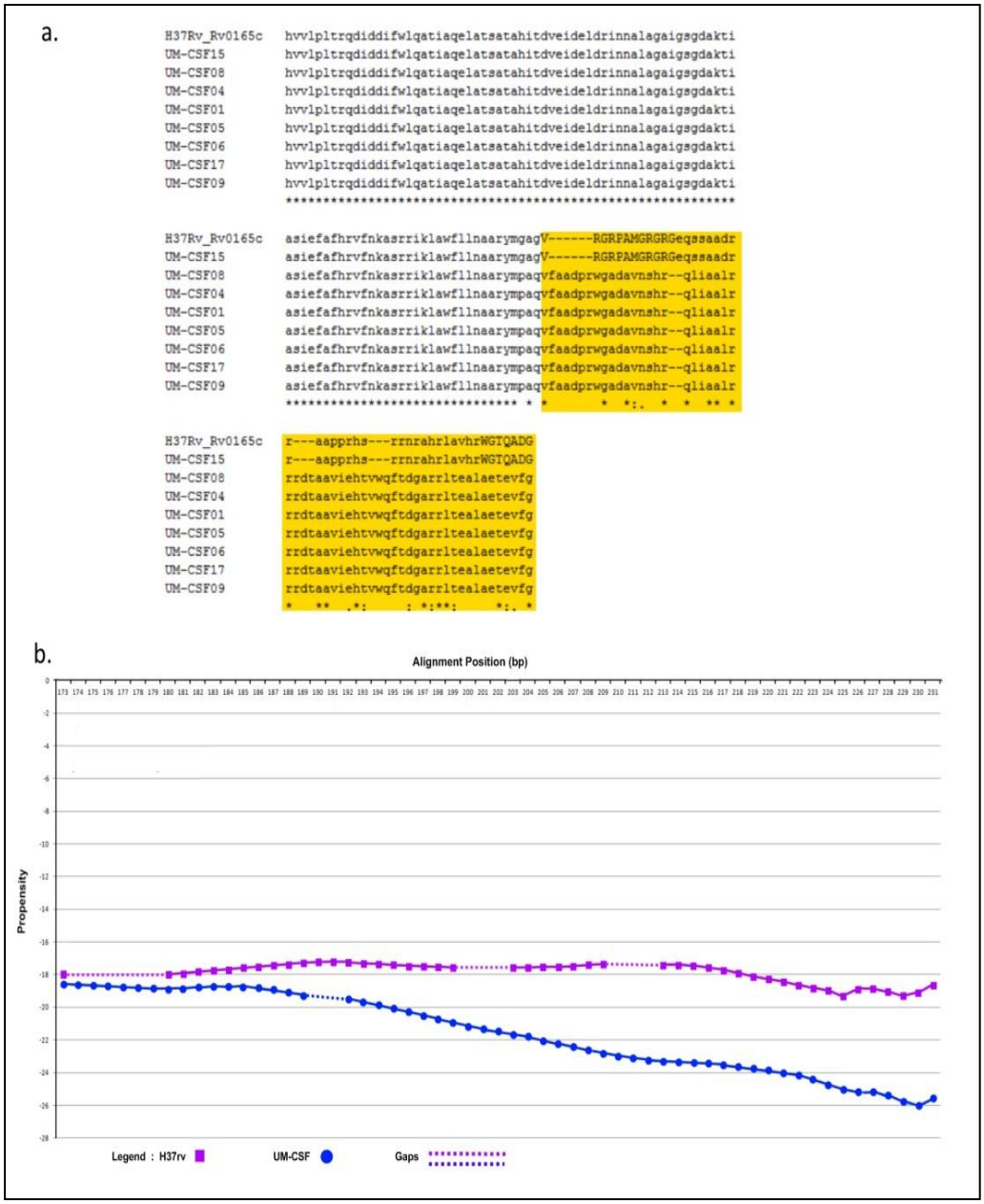


Figure 4.2.5: (A) Alignment of Rv0165c sequences in eight UM-CSF strains and H37Rv, showing amino acid changes. (B) Differences in propensity scores for amino acids in UM-CSF strains and H37Rv. Propensity score is used to predict protein secondary structure. It is derived from looking at the amino acids residue of the accessible surface of the protein and the interface which enables interactions between other proteins. Propensity is the ratio of the probability of the residue in the interface to the probability of the residue on the surface (Huang, 2014).

4.2.6 Proteins Reported to be Associated with Meningitis in *Mtb*

Of 63 proteins reported for *Mtb* (Av-Gay and Everett, 2000; Pethe et al., 2001; Tsenova et al., 2005; Jain et al., 2006; Be et al., 2008; Be et al., 2012) (Appendix 8: Table A8), homologs of 56–60 were found in UM-CSF strains but only two, Rv0311 encoding a hypothetical protein and Rv0619 encoding a probable galactose-1-phosphateuridylyltransferase GalTb), were found in all eight CSF strains (Table 4.2.6). Compared to H37Rv, both proteins showed amino acid and globularity differences. The Rv0311 protein in five of the eight UM-CSF strains had Asp (GAT) instead of Glu (GAG) at aa position 119 and was predicted to be disordered. Rv0619 in all eight UM-CSF strains had a substitution of Ala (GCC) for Thr (ACC) at position 174 and appeared as a globular segment with a change from polarity to hydrophobicity (Appendix 24: Figure A24).

Additionally, a similar search was performed in five other mycobacteria that are associated with neuropathology. These comprised *M. leprae* and *M. lepromatosis* that cause different forms of leprosy, *M. bovis* that is usually linked with extrapulmonary TB, and two rapid-growers *M. ilatzerense* and *M. immunogenum* that had been isolated from a case of brain abscess (Greninger et al., 2015). Fifty-six homologs of the 63 meningitis related genes from *Mtb* (Appendix 8: A8) were identified in *M. bovis*, followed by 16 in *M. leprae*, 15 in *M. lepromatosis*, 14 in *M. ilatzerense*, and 11 in *M. immunogenum*. The UM-CSF strains shared four of the 63

meningitis-related genes (Rv0014c, Rv1837c, Rv2176 and Rv0984) with all five of these mycobacterial species and five other genes (Rv1273c, Rv2318, Rv0983, Rv0966c and Rv0805) with the three slow-growing mycobacteria (Table 4.2.6). The Rv2947c (*pks* 15/1) gene was found in the UM-CSF strains, *M. leprae* and *M. ilatzerense*. In the Beijing genotype of *Mtb*, an intact *pks* 15/1 is believed to be responsible for virulence and extrapulmonary disease (Reed et al., 2004). Consistent with their extrapulmonary (CNS) location in the host, five of the eight UM-CSF strains were genotyped as Beijing ST1 and each carried an intact *pks* 15/1 gene.

Table 4.2.6: Genes shared by eight UM-CSF strains and *Mtb*, other mycobacteria associated with neuropathology, *S. pneumoniae* and *N. meningitidis*, respectively.

Common hosts	Rv no.	Gene name	Product	Function
1a) UM-CSF strains and *Mtb*	Rv0311	Rv0311	Unknown protein	Unknown
1b) UM-CSF strains and *Mtb*	Rv0619	galTB	Probable galactose-1-phosphate uridylyltransferase GalTb	Galactose metabolism (leloir pathway) catalytic activity
2a) UM-CSF strains and Mycobacteria, *S pneumoniae*	Rv1699	pyrG	CTP synthase PyrG	Pyrimidine biosynthesis catalytic activity

2b) UM-CSF strains and Mycobacteria, *S pneumoniae*	Rv2606c	snzP	Pyridoxine biosynthesis protein SnzP	Biosynthesis of pyridoxine/pyridoxal 5-phosphate
2c) UM-CSF strains and Mycobacteria, *S pneumoniae*	Rv0357c	purA	Probable adenylosuccinate synthetase purA	AMP biosynthesis
3) UM-CSF strains and Mycobacteria, *N meningitidis*	Rv2457c	clpX	ATP-dependent CLP protease ATP-binding subunit ClpX	Directs the CLP protease to specific substrates; performs chaperone functions in the absence of CLP protease
4) UM-CSF strains and *N meningitidis*	Rv2397c	cysA1	Probable sulfate-transport ATP-binding protein ABC transporter cysA1	Active transport of multiple sulphur-containing compounds across the cell membrane; energy coupling to the transport system
5a) UM-CSF strains and Mycobacteria	Rv1837c	glcB	Malate synthase G GlcB	glyoxylate bypass; alternative to TCA cycle
5b) UM-CSF strains and Mycobacteria	Rv0014c	pknB	Transmembrane serine/threonine-protein kinase B PknB	Signal transduction (via phosphorylation)
5c) UM-CSF strains and Mycobacteria	Rv2176	pknL	Probable transmembrane serine/threonine-protein kinase L PknL	Signal transduction (via phosphorylation)
5d) UM-CSF	Rv0984	moaB2	Possible pterin-4-alpha-	Molybdopterin

strains and Mycobacteria			carbinolamine dehydratase MoaB2	biosynthesis
6a) UM-CSF strains and SG Mycobacteria	Rv1273c	Rv1273c	Probable drugs-transport transmembrane ATP-binding protein ABC transporter	Active transport of drugs across the membrane (export)
6b) UM-CSF strains and SG Mycobacteria	Rv2318*	uspC	Probable periplasmic sugar-binding lipoprotein UspC	Active transport of sugar across the membrane (import)
6c) UM-CSF strains and SG Mycobacteria	Rv0983	pepD	Probable serine protease PepD (serine proteinase)	Unknown; possibly hydrolyses peptides and/or proteins
6d) UM-CSF strains and SG Mycobacteria	Rv0966**	Rv0966	Conserved protein	Unknown
6e) UM-CSF strains and SG Mycobacteria	Rv0805	Rv0805	Class III cyclic nucleotide phosphodiesterase	Hydrolyses cyclic nucleotide monophosphate to nucleotide monophosphate
7) UM-CSF strains and *M. leprae, M. ilatzerense*	Rv2947c	pks15	Probable polyketide synthase Pks15	Polyketide synthase possibly involved in lipid synthesis

Mtb, *Mycobacterium tuberculosis*

Mycobacteria, *M.bovis, M.leprae, M.lepromatosis, M.ilatzerense, M.immunogenum*

SG Mycobacteria, *M.bovis, M.leprae, M.lepromatosis*

*Rv2318 is not found in CSF08

** Rv0966 is not found in CSF04 and CSF06

4.2.7 Meningitis-Associated Genes from Other Bacterial Pathogens

Streptococcus pneumoniae, *Escherichia coli* K-1 and *Neisseria meningitidis* are pathogens known to cause meningitis in humans. Of 141 proteins reported to be associated with *S. pneumoniae* meningitis (Orihuela et al., 2004; Molzen et al., 2011; Mahdi et al., 2012) (Appendix 9: Table A9), three, Rv1699 (CTP synthase PyrG), Rv2606c (pyridoxine biosynthesis protein SnzP) and Rv0357c (adenylosuccinate synthetase PurA) were found in the UM-CSF strains (Table 4.2.6). These genes showed 51 – 68 % sequence similarity with their homologs in *S. pneumoniae* but were identical in all UM-CSF strains and H37Rv, in protein sequence as well as globularity (Figure 4.2.7.1). When compared against 164 *N. meningitidis* virulence genes reported by Hao et al. (2011) (Appendix 10: Table A10), UM-CSF strains shared two virulence homologs with this neuropathogen: Rv2457c, encoding ATP-dependent CLP protease ATP-binding subunit clpX and Rv2397c, encoding sulfate-transport ATP-binding protein ABC transporter CysA1 (Appendix 25: Table A25). The genes for cell surface outer membrane Opa and Opc proteins that were previously reported to confer tissue tropism in *N. meningitidis* (Virji et al., 1993) were not found, neither were homologs of previously reported *E. coli* K1 neurotropic genes such as *IbeA*, *IbeB*, *AslA*, *YijP*, and *OmpA* (Pouttu et al., 1999; Huang et al., 2001; Yao et al., 2006) (Appendix 11: Table A11), in UM-CSF strains. However, the five genes associated with *S. pneumoniae* and *N. meningitidis* meningitis was also found in many respiratory *Mtb* (Appendix 26: Table A26).

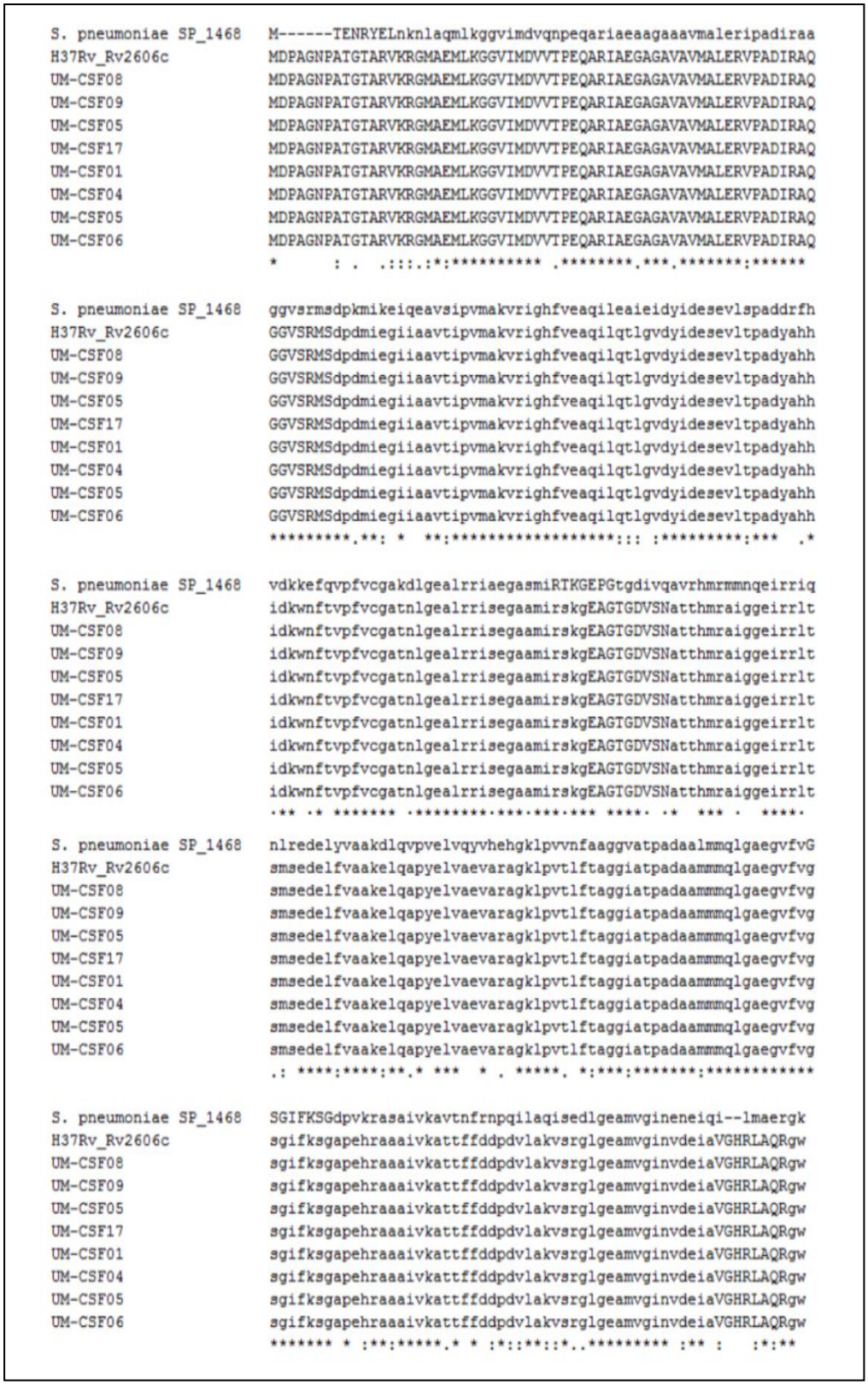


Figure 4.2.7.1: Alignment of meningitis-associated protein (Rv2606c) from *S. pneumoniae*, H37Rv and eight UM-CSF strains.

Figure 4.2.7.2 summarizes the NGS findings of 17 meningitis-associated genes that are found in the eight CSF *Mtb* strains, *M. leprae* and *M. lepromatosis*, *M. bovis* and two rapid growers *M. ilatzerense* and *M. immunogenum*, together with pathogens known to cause meningitis in humans, *Streptococcus pneumoniae* and *Neisseria meningitidis*.

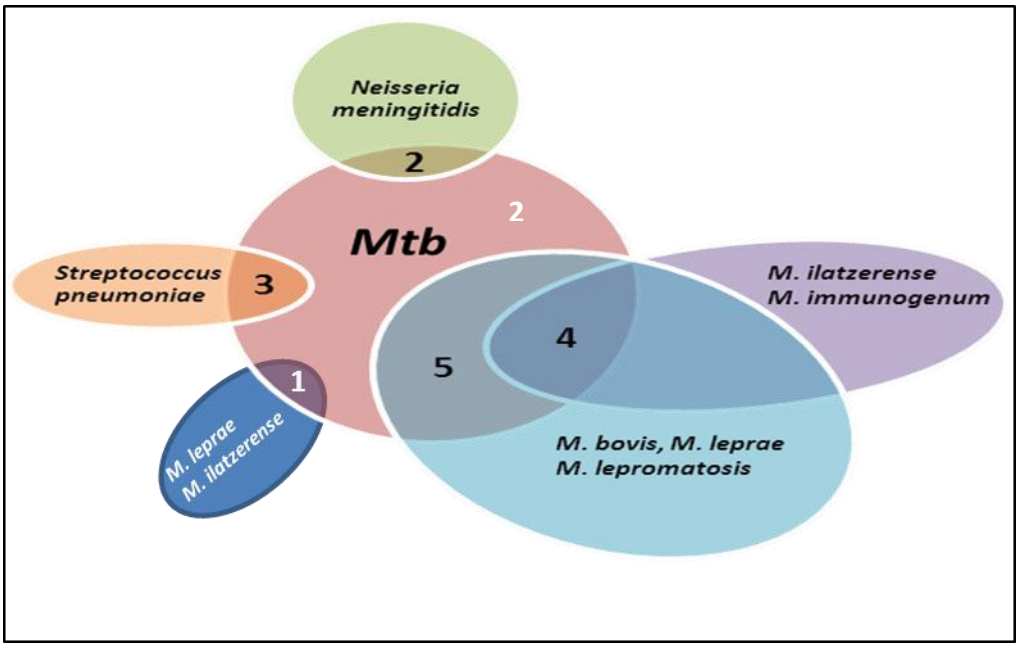


Figure 4.2.7.2: Venn diagram showing the number of meningitis-associated genes that are shared between the CSF *Mtb* strains, mycobacterial and non-mycobacterial neuro-pathogens.

4.3 Verification of Putative Findings in UM-CSF Strains

A summary of all the putative genes identified in UM-CSF strains (35 genes in total) are tabulated in Appendix 27: Table A27 and their characteristics in terms of size and function are summarized in Appendix 28: Table A28. Six of these were selected for verification by PCR followed by

Sanger sequencing. They were three genes (Rv3425, Rv1141c, Rv3344c) with putative deletions, one gene (Rv0311) with an nsSNP, and two homologues genes (Rv2606c, Rv2397c) reported to be associated with *S. pneumoniae* and *N. meningitidis* meningitis. The results of PCR amplification and sequence analysis are summarized in Table 4.3.

Table 4.3: Summary of verification results for selected genes.

Gene	Size (bp)/Functions	Description	PCR and sequencing result
Rv3425 (PPE57)	531; Unknown	Deletion (3842239…3842769) in UM-CSF 06	Gene amplified in UM-CSF 06 but no predicted deletion seen; instead multiple SNPs were observed
Rv1141c (echA11)	807; Could possibly oxidize fatty acids using specific components	Deletion (1268203…1269009) in UM-CSF 09	Gene amplified in UM-CSF09 but no deletion found
Rv3344c (PE_PGRS49)	1455; Unknown	Micro-variants and deletion of nucleotides from position 490 to 529 (amino acid position 164 to 173) resulting in amino acid changes	Gene amplified in UM-CSF 01, with smaller PCR product (~825 bp) and multiple SNPs and deletions including the predicted deletion
Rv0311	1230; Unknown	G to T SNP at position 357 (amino acid position 119) resulted in a replacement of glutamic acid by aspartic acid	Gene amplified in all UM-CSF samples. G to T SNP was found at position 357 in 5 isolates
Rv2606c (snzP)	900; Involved in the biosynthesis of pyridoxine/pyridoxal 5-	Reported in *S. pneumoniae* CNS infection	Amplified in all UM-CSF samples and H37Rv with identical gene

	phosphate biosynthesis		sequence
Rv2397c (cysA1)	1056; Involved in the active transport of multiple sulfur-containing compounds across the membrane; responsible for energy coupling to the transport system.	Reported in *N. meningitidis* CNS infection	Amplified in all UM-CSF samples and H37Rv with identical sequence

4.3.1 Verification of Putative Deletions in Rv3425 (UM-CSF06), Rv1141c (UM-CSF09) and Rv3344c (UM-CSF01)

The amplicons obtained for Rv3425, Rv1141c and Rv3344c are shown in Figure 4.3.1.1 and the corresponding DNA sequencing results are shown in Figures 4.3.1.2, 4.3.1.3 and 4.3.1.4, respectively. Rv3344c was only amplified in UM-CSF01 (Figure 4.3.1.1). The PCR product was approximately 825 bp in size instead of the expected 1455 bp. This is likely to be the result of multiple deletions including the deletion of 30 nucleotides (position 490 to 529) identified in the WGS analysis (Figure 4.3.1.4). BLASTN and BLASTX analysis (Table 4.3.1) shows the 825 bp sequence to be a putative PE family protein similar to a sequence reported from a *Mycobacterium tuberculosis* isolate from the human brain (Husain et al., 2017). Numerous micro-variants were also seen.

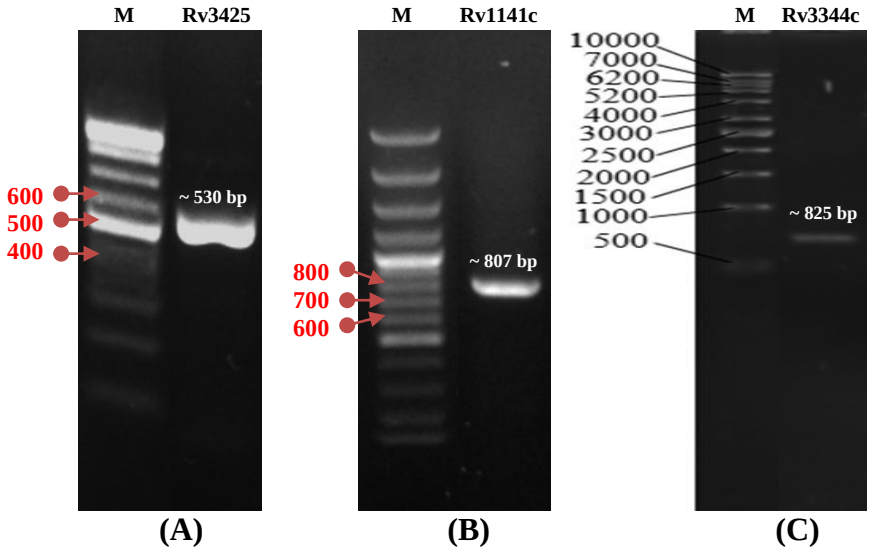


Figure 4.3.1.1: Agarose gel electrophoresis of purified PCR products: (A) Rv3425 (531 bp) in UM-CSF06; (B) Rv1141c (807 bp) in UM-CSF09 and (C) Rv3344c in UM-CSF01 (~ 825 bp). Lane M: GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, USA).

Rv3425 (UM-CSF06)

```
H37Rv           ATGCATCCAATGATACCAGCGGAGTATATCTCCAACATAATATATGAAGGCCCGGGCGCT  60
Rv3425 (CSF 6) ------------------------------------------------------------GCT  3
                                                                           ***

H37Rv           GACTCATTGTTTTTCGCCTCCGGGCAATTGCGAGAATTGGCTTACTCAGTTGAAACGACG 120
Rv3425 (CSF 6) GACTCATTGTCTGCCGCCGCCGAGCAATTGCGACTAATGTATAACTCAGCTAACATGGCG  63
               ********** *  **** *** **********  * **  * ****** * * * * **

H37Rv           GCTGAGTCGCTCGAGGACGAGCTCGACGAGCTGGATGAGAACTGGAAAGGTAGTTCGTCG 180
Rv3425(CSF6)    GCTAAGTCGCTCACCGACAGGCTCGGCGAGCTGCAGGAGAACTGGAAAGGTAGTTCGTCG 123
               *** ********   ***  ***** ******* * ************************

H37Rv           GACTTGTTGGCCGACGCGGTTGAGCGGTATCTCCAATGGCTGTCTAAACACTCCAGTCAG 240
Rv3425(CSF6)    GACTTGTTGGCCGACGCGGCTGGGCGGTATCTCGACTGGCTGACTAAACACTCTCGTCAA 183
               ****************** ** ********** * ****** **********  ****

H37Rv           CTTAAGCATGCCGCCTGGGTGATCAACGGCCTCGCGAACGCCTATAACGACACACGTCGG 300
Rv3425 (CSF 6) ATTCTGGAAACCGCCTACGTGATCGACTTCCTCGCATACGTCTATGAGGAGACACGTCAC 243
                **  * *  ******  ****** **  ******  *** **** * ** *******

H37Rv           AAGGTGGTACCCCCGGAGGAGATCGCCGCCAACCGCGAGGAGAGGCGCAGGCTGATCGCG 360
```

```
Rv3425 (CSF 6) AAGGTGGTACCCCCGGCGACTATCGCCAACAACCGCGAGGAGAGGCGCAGGCTGATCGCG 303
               *************** *   ******  ******************************

H37Rv          AGCAACGTGGCCGGGGTAAACACTCCAGCAATCGCAGACCTCGATGCACAATACGACCAG 420
Rv3425 (CSF 6) AGCAACGTGGCCGGGGTAAACGCTCCGGCAATCGCAGACCTCGATGCACAATACGACCAG 363
               ******************** **** ********************************

H37Rv          TACCGGGCCCGCAATGTCGCTGTAATGAACGCCTATGTAAGTTGGACCCGATCTGCGCTA 480
Rv3425 (CSF 6) TACCGGGCCCGCAATGTCGCTGTAATGAACGCCTATGTAAGTTGGACCCGATCTGCGCTA 423
               ************************************************************

H37Rv          TCGGATCTGCCCCGGTGGCGGGAACCGCCGCAGATCTACAGGGGCGGGTAG 531
Rv3425 (CSF 6) TCGGATCTGCCCCGGTGGCGGGAACCGCCG---------------------- 453
```

Figure 4.3.1.2: Nucleic acid alignment of Rv3425 gene sequence in UM-CSF06 sample and H37Rv strain. The letters in red show the position of the expected deletion.

Rv1141c (UM-CSF09)

```
H37Rv      ATGCCAGATTCCGGGATTGCCGCATTAACGCCGGTCACAGGCCTCAACGTCACCCTGACC 60
CSF9       ------GATTCCGGGATTGCCGCATTAACGCCGGTCACAGGCCTCAACGTCACCCTGACC
                 ******************************************************

H37Rv      GACAGAGTGTTGTCGGTGCGCATCAACCGCCCTAGCAGTCTCAACTCGCTGACCGTGCCA 120
CSF9       GACAGAGTGTTGTCGGTGCGCATCAACCGCCCTAGCAGTCTCAACTCGCTGACCGTGCCA
           ************************************************************

H37Rv      ATCCTGACGGGGATCGCCGACACGCTGGAGCGCGCGGCGGCCGATCCCGTGGTCAAGGTG 180
CSF9       ATCCTGACGGGGATCGCCGACACGCTGGAGCGCGCGGCGGCCGATCCCGTGGTCAAGGTG
           ************************************************************

H37Rv      GTGCGCCTAGGCGGGGTGGGCCGCGGTTTCAGCTCCGGAGTGTCTATGTCTGTGGACGAT 240
CSF9       GTGCGCCTAGGCGGGGTGGGCCGCGGTTTCAGCTCCGGAGTGTCTATGTCTGTGGACGAT
           ************************************************************

H37Rv      GTGTGGGGCGGAGGGCCGCCGACCGCCATCGTCGAAGAGGCCAACCGCGCAGTACGCGCC 300
CSF9       GTGTGGGGCGGAGGGCCGCCGACCGCCATCGTCGAAGAGGCCAACCGCGCAGTACGCGCC
           ************************************************************

H37Rv      GTGGCCGCGCTACCGCACCCGGTTGTAGCTGTCGTTCAAGGACCAGCGGTCGGCGTCGCT 360
CSF9       GTGGCCGCGCTACCGCACCCGGTTGTAGCTGTCGTTCAAGGACCAGCGGTCGGCGTCGCT
           ************************************************************

H37Rv      GTCTCGCTAGCGCTGGCGTGTGACTTCATATTGGCTTCTGATAGTGCATTTTTCATGCTC 420
CSF9       GTCTCGCTAGCGCTGGCGTGTGACTTCATATTGGCTTCTGATAGTGCATTTTTCATGCTC
```

```
                ************************************************************

H37Rv           GCCAACACCAAGGTAGCGTTGATGCCCGACGGCGGCGCATCGGCGTTAGTCGCCGCGGCC 480
CSF9            GCCAACACCAAGGTAGCGTTGATGCCCGACGGCGGCGCATCGGCGTTAGTCGCCGCGGCC
                ************************************************************

H37Rv           ACCGGCCGGATCCGGGCGATGCGGCTGGCGCTGCTGGCCGAGCAACTGCCGGCCCGCGAG 540
CSF9            ACCGGCCGGATCCGGGCGATGCGGCTGGCGCTGCTGGCCGAGCAACTGCCGGCCCGCGAG
                ************************************************************

H37Rv           GCACTGGCCTGGGGCCTGATCAGCGCGGTATATCCGGACAGCGACTTCGAGGCCGAGGTG 600
CSF9            GCACTGGCCTGGGGCCTGATCAGCGCGGTATATCCGGACAGCGACTTCGAGGCCGAGGTG
                ************************************************************

H37Rv           GACAAGGTGATTTCACGGTTGCTGGCCGGCCCGGCGCTGGCGTTCGCCCAGGCCAAAAAC 660
CSF9            GACAAGGTGATTTCACGGTTGCTGGCCGGCCCGGCGCTGGCGTTCGCCCAGGCCAAAAAC
                ************************************************************

H37Rv           GCCATCAATGCAGCCGCCCTCACCGAATTGGAACCCACGTTCGCGCGCGAATTGGATGGA 720
CSF9            GCCATCAATGCAGCCGCCCTCACCGAATTGGAACCCACGTTCGCGCGCGAATTGGATGGA
                ************************************************************

H37Rv           CAGGAAGTCCTGCTGCGAACACACGACTTCGCCGAGGGCGCAGCGGCGTTCCTGCAACGC 780
CSF9            CAGGAAGTCCTGCTGCGAACACACGACTTCGCCGAGGGCGCAGCGGCGTTCCTGCAACGC
                ************************************************************

H37Rv           CGCACCCCCAACTTCACCGGTTCCTGA 840
CSF9            CGCACCCCCAACTTCA-----------
                ***************
```

Figure 4.3.1.3: Nucleic acid alignment of Rv1141c genes in UM-CSF09 sample and H37Rv strain. No deletion mutation was observed in the CSF sample. The expected deletion is indicated by the letters in red colour.

Rv3344c (UM-CSF01)

```
H37Rv           TGGGTTCGGTGGCACCGGGGGCAGCGGCAGCGGCATTGGCGGCGGCGCCGGCGGCAACGG
CSF01           ---ATTCGGTGTTGCGGTATGCAGTCGCGCCAGTTTT-----------------------
                   *******   * *   ****  **  * *  **

H37Rv           CGGCAACGGCGGCGCCGGCGGCACCGGCGTCGTGCTTGGCGGCAAGGGCGGCGACGGGGG
CSF01           ---CAACAG----------------GTTGACGAACTCGTCATGAAACGCGGAGGCCTGGG
                   **** *               *  * **  ** * *   **  ****   *  ***

H37Rv           CAACGGTGACCACGGTGGGCCTGCCACCAACCCGGGCAGCGGCAGCCGCGGCGGCGCCGG
CSF01           CGCTGAGAACC----------T----------------------------GGCACTGT
```

```
           *   *   ***          *                              *** * *

H37Rv      CGGCTCCGGCGGCAACGGTGGCGCCGGGGGTAACGCCACCGGCTCAGGCGGCAAGGGCGG
CSF01      CGAC-----------CGTAGGCGCCGAA--------TATCGAC-----------------
           ** *           **  *******           * ** *

H37Rv      CGCCGGTGGCAATGGCGGTGATGGGAGCTTCGGCGCTACCAGCGGCCCCGCCTCCATCGG
CSF01      --------GCAATCGCGGT-----------------------CGACACCTCGTCC-TCGG
                  ***** *****                       ** * ** * *** ****
```

Figure 4.3.1.4: Nucleic acid alignment of Rv3344v in UM-CSF01. The red letters in H37Rv (position 490 to 529) were predicted to be deleted in UM-CSF01.

Table 4.3.1: BLASTN and BLASTX results of Rv3344c gene amplified from UM-CSF01 sample.

Gene/BLAST variants	BLASTN			BLASTX		
	Description	E-value	Identity	Description	E-value	Identity
Rv3344c in UM-CSF01	*Mtb* strain C3 (Brain: India)	0.0	99 %	PE family protein	3e-87	100 %

4.3.2 Verification of Putative SNP in Rv0311

Rv0311 was amplified in all UM-CSF strains (Figure 4.3.2.1). The nucleic acid alignment of Rv0311 from all CSF samples and H37Rv is shown in Figure 4.3.2.2. Five samples (UM-CSF01, 05, 08, 09 and 17) showed a G to T change at nucleotide position 357, resulting in the replacement of glutamic acid (GAG) by aspartic acid (GAT) as predicted in the WGS analysis.

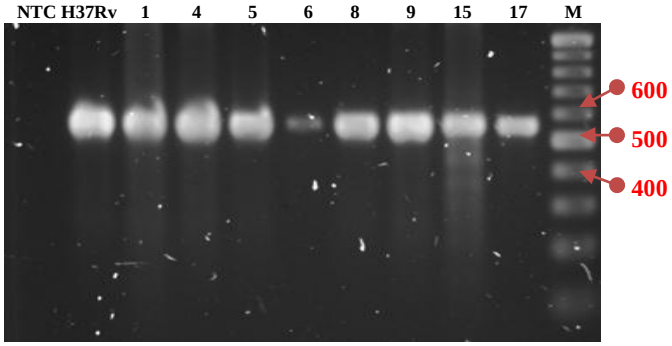


Figure 4.3.2.1: Agarose gel electrophoresis of PCR products Rv0311 (556 bp) in all CSF samples. Lane NTC: non-template control; Lane H37Rv: positive control; Lane M: GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, USA).

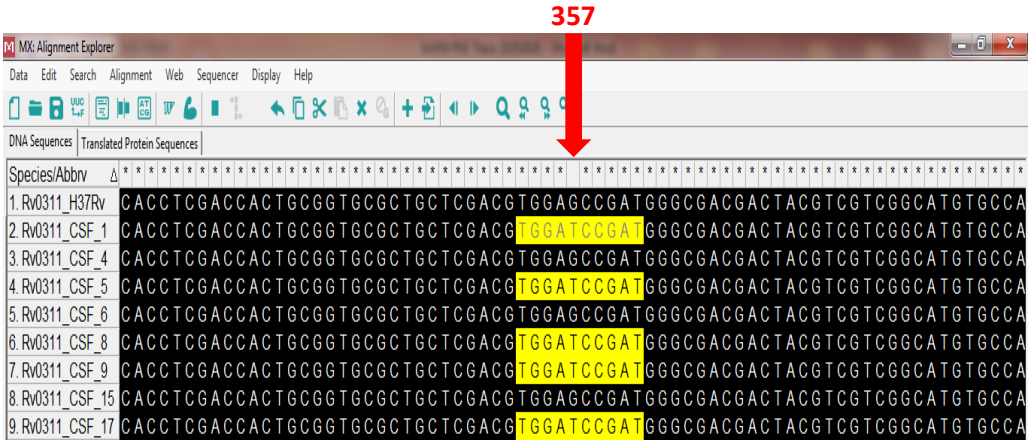


Figure 4.3.2.2: Nucleic acid alignment of Rv0311 in all CSF samples. G to T SNP was observed in nucleotide position 357 in 5 samples: UM-CSF01, 05, 08, 09 and 17.

4.3.3 Verification of Pneumococcal and Meningococcal Meningitis-Associated Genes, Rv2606c and Rv2397c, in CSF Samples

UM-CSF15 was not included in this analysis as it was no longer available. Amplification was observed for Rv2606c and Rv2397c genes in all

the other 7 UM-CSF strains (Figure 4.3.3.1). The amplicons were sent for DNA sequencing using forward and reverse primers. Nucleic acid alignment results (Figure 4.3.3.2 and 4.3.3.3) showed identical sequences in UM-CSF strains and the respiratory control H37Rv.

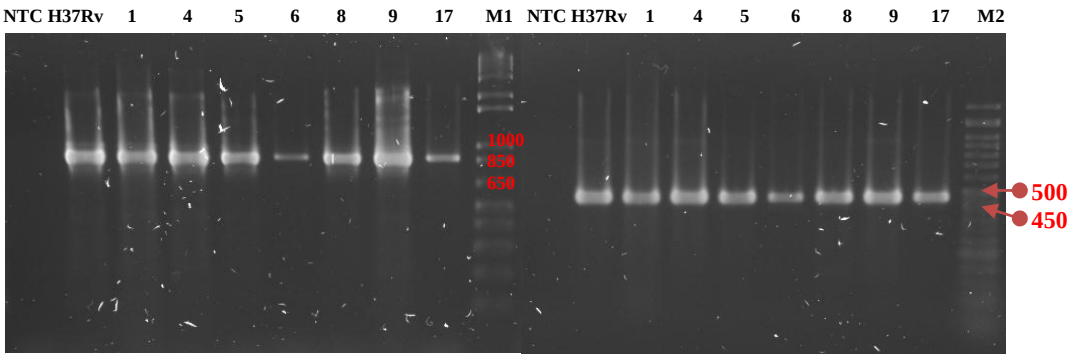


Figure 4.3.3.1: Agarose gel electrophoresis of PCR products: Rv2606c (900 bp) (left), and Rv2397c (471 bp) (right). Lane NTC: Non-template control; Lane H37Rv: positive control; Lane M1: GeneRuler 1 kb Plus DNA Ladder; Lane M2: GeneDireX 50 bp DNA Ladder.

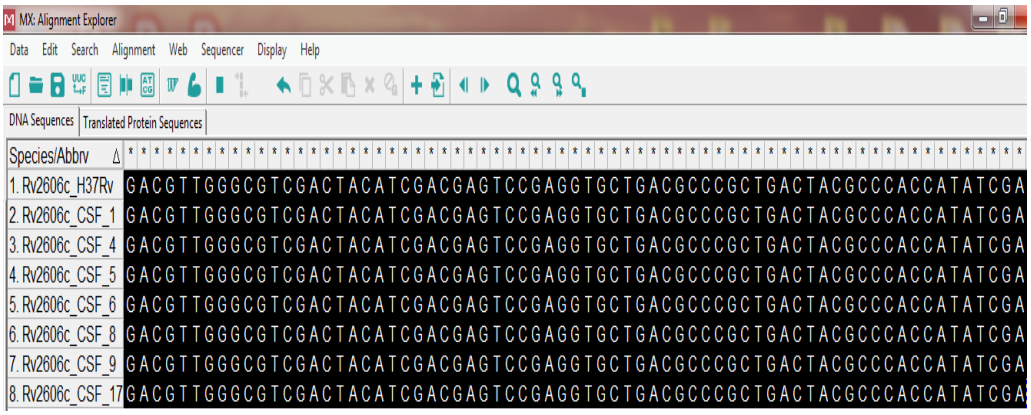


Figure 4.3.3.2: Nucleic acid alignment of Rv2606c in 7 CSF samples. The sequences were identical in all CSF samples and the respiratory control (H37Rv).

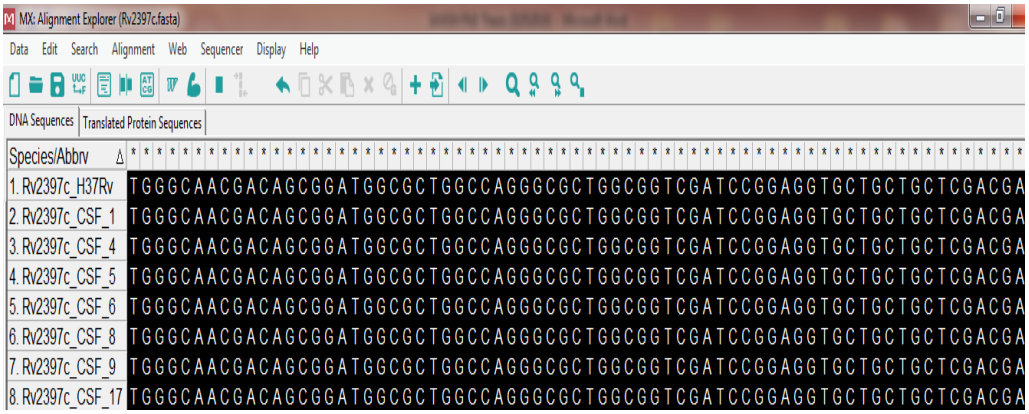


Figure 4.3.3.3: Nucleic acid alignment of Rv2397c in 7 CSF samples. The sequences were identical for all CSF samples and the respiratory control (H37Rv).

The sequences of Rv2606c and Rv2397c from UM-CSF01 were used to search NCBI genome databases using BLASTN. The aim was to identify the relative frequency of these 2 genes in respiratory and CSF strains. The results are tabulated in Appendix 26: Table A26. Of 26 strains isolated from different countries at different times, all were found to carry both genes. Twenty-one strains were isolated from respiratory secretions while only 5 were from CSF.

CHAPTER 5

DISCUSSION

Very few clinical or laboratory studies have been conducted on extra-pulmonary TB (EPT) in Malaysia (Swarna, 2014). The launch of the National Tuberculosis Control Programme (NTBCP) in Malaysia in 1961 led to the reduction of PTB incidence from 350 cases per 100, 000 to less than 100 cases per 100,000 in the 1980's (Ministry of Health, 2011). However, the number of EPT cases remained constant in the capital city of Kuala Lumpur while TBM was found to be the presentation of 48.8 % of TB cases in Kota Kinabalu, Sabah, East Malaysia (Lee et al., 2016). It is critical to understand the pathogenesis of CNS TB in order to look for appropriate and effective methods of prevention and treatment. Although it is well-known that various host factors such as age, malnutrition and co-infection with HIV have an apparent influence on the development of TBM, the role of bacterial genetic variation remains unclear. Hence, in this study, attempts were made to find mycobacterial genetic polymorphisms that might extend existing knowledge on CNS tropism in clinical TB. In order to achieve this objective, SSH was first applied to look for genetic differences between two CNS and two pulmonary *Mtb* isolates. As various constraints prohibited the extension of SSH to more paired CSF/sputum samples, NGS was utilized to examine a larger sample of *Mtb* genomes for genetic traits that could be related to neurotropism in TB.

5.1 Technical Issues Related to *Mtb* DNAs

5.1.1 Deactivation of *Mtb* Culture

Since *Mtb* is an airborne pathogen classified in Risk Group 3 based on the international biosafety classification, specific safety equipment and work practices are necessary to better define the containment level in the diagnostic and research laboratory (Herman et al., 2006; WHO, 2012). TB diagnostics have to be carried out under Biosafety Level 2 (BSL-2) containment with BSL-3 safety equipment and work practices. In this study, owing to the lack of BSL level 3 facilities, heat inactivated *Mtb* cultures from a collaborator's laboratory (UMMC TB laboratory) were used for DNA extraction. Heat inactivation was carried out by heating one loop-full of *Mtb* growth from a solid Lowenstein-Jensen (LJ) culture slope in a water bath at 80 °C for 20 minutes (van Embden et al., 1993; Doig et al., 2002). Doig and colleagues (2002) had proven this protocol to be effective, as, after 20 weeks of incubation at 37 °C, heat-treated *Mtb* cultures from both LJ slants and liquid media did not exhibit any visible growth (Doig et al., 2002). Importantly, this protocol was shown not to affect the DNA integrity for subsequent molecular investigations such as polymerase chain reactions (PCR) (Djelouagji and Drancourt, 2006). The reason for using a loop-full of *Mtb* is that mycobacteria survive better when heated in a larger population (Zwadyk et al., 1994). A water bath was used for heating because sporadic growth of various mycobacteria has been observed during heating in a dry heat block set at

95 °C (Zwadyk et al., 1994). This incomplete inactivation in a dry bath might be due to the shape of the heat block wells not achieving a close fit to the shape of the sample tubes. Complete immersion of the sample bottle in a water bath is crucial, as some viable mycobacteria might still be trapped in the lid of the screw-capped glass bottle. Bemer-Melchior and Drugeon (1999) reported about 50 - 65 % survival of the *Mtb* cultures heated at 80 °C for 20 min in bottles that were not fully immersed in the water bath (Bemer-Melchior and Drugeon, 1999). The application of these sample deactivation protocols in this study resulted in the collection of intact genomic DNA with a little smearing (Figure 4.1.1) which could have been due to the large amount of samples used for DNA extraction.

5.1.2 Extraction of *Mtb* DNA and Determination of DNA Quality

A high yield of *Mtb* DNA is critical to ensure successful subtraction (1.5 - 2.0 µg of DNA per subtraction) and quality WGS analysis (3.0 µg of DNA per sample) (Chomczynski and Sacchi, 1987; Clontech Laboratories, 2011; Gowda, 2013). However, the resistant mycobacterial cell wall makes DNA extraction difficult. Several studies had evaluated methods for mycobacterial cell wall lysis and DNA extraction for sputum and extra-pulmonary samples such as (Thwaites, Caws, et al., 2004; Desai et al., 2006). CSF is a precious sample with limited amounts available for diagnostic and research purposes. Hence, the optimisation of protocols for extracting *Mtb* DNA from both sputum and CSF samples was necessary.

In this study, since *Mycobacterium tuberculosis* multiplies extremely slowly and thus produces only small amounts of starting material, *Mtb* cultures were pooled from a few MGIT tubes for DNA extraction. This small modification increased the DNA concentration to more than 20 times compared to the use of only a single culture tube (Table 4.1.1).

A preliminary evaluation showed that a combined physical, enzymatic and chemical treatment for DNA extraction was more efficient than the use of a commercial kit (GE Illustra Bacteria Genomic Prep Mini Spin Kit, USA) which uses a lysis solution containing proteinase K to release gDNA and a silica membrane to enhance the purity of the gDNA. Hence, DNA was extracted from the heat-inactivated and frozen stock *Mtb* by a quick thaw to weaken the mycobacterial cell wall (Jain et al., 2002; Aldous et al., 2005; Thakur et al., 2011) followed by treatment with lysozyme (Nakatani et al., 2004; Amaro et al., 2008; Geneaid, 2017), proteinase K (Thakur et al., 2011) and sodium dodecyl sulfate (SDS) (Banavaliker et al., 1998; Bhatia, 2017; Promega Corporation, 2017; Thermo Fisher Scientific, 2017) to further disrupt the cell wall and improve the yield of the extracted DNA.

In the PCI protocol for separating DNA from proteins and other cellular debris, residual contaminating phenol and chloroform could have contributed to the lower DNA purity (A_{260}/A_{280} ratio of 1.60 to 1.83; Table 4.1.1) and DNA smearing in the gel electrophoresis (Figure 4.1.1).

5.2 Suppression Subtractive Hybridisation (SSH)

SSH involves a combination of subtraction and suppression polymerase chain reaction (PCR) that allows the identification of sequences which are present only in one genome (tester/tracer) but absent in the other genome (driver) of two closely related organisms. Adequate quality control is required for each step in the procedure. One important limitation is that not all genes are subtracted evenly. For instance, subtraction is less efficient for common genes that are 2-fold higher in tester samples. This means there will be common DNA fragments left in the subtracted sample. This background, if not adequately removed, will affect the results of all downstream procedures.

5.2.1 Analysis of Restriction Enzyme (RE)

The tester samples have to be restriction enzyme digested before the amplification and subtraction steps. This is because large fragments will not be efficiently amplified (Ko, 1990) and thus will not be well represented in the product pool. On the other hand, while small fragments (< 200 bp) are otherwise preferentially amplified and provide better representation of individual genes, they may promote the formation of pan-handle structures which are stable enough to reduce the amplification efficiency (Lukyanov et al., 2007). Therefore, it is crucial to have restriction fragment sizes ranging between 0.2 to 1.2 kb.

For efficient subtraction, the restriction enzyme chosen has to be able to digest a bacterial genomic DNA into fragments of optimal length. 4-base-cutting REs including *Rsa*I, *Alu*I, *Apo*I, *Dra*I, *Sau3*AI, and *Hae*III are commonly used for digestion. A combination of several 6-base-cutting REs such as *EcoR*I, *Sac*I and *Sst*I has also been applied. The use of multiple RE is to increase coverage of fragments of optimal length (Agron et al., 2001; Agron et al., 2002). *Rsa*I (a 4-base-cutting RE) was selected in this study to generate short, blunt-ended DNA fragments that are optimal for subtraction and adaptor ligation (Figure 4.1.2). Quality control experiments showed that it generated the largest average size of fragments at approximately 600 bp and was able to achieve complete digestion of both driver and tester samples (Appendix 29: Figure A29).

5.2.2 Analysis of Adaptors Ligation

The analysis of ligation efficiency is designed to amplify fragments that span the adaptor-DNA junctions (Clontech Laboratories, 2008). In this study, the 23S RNA of *E. coli* was used for the estimation of ligation efficiency. The 23S RNA is a 2904 bp component of the large ribosomal subunit (50S) of *E. coli* and is a conserved RNA (Weinberg et al., 2007). The amount of adaptor-ligated DNA obtained with two gene-specific primers (23S RNA Forward and Reverse Primers) was compared with that obtained with one gene-specific primer (23S RNA Forward Primer and PCR Primer 1 which binds to both adaptors). The PCR product generated using two gene-specific primers (270 bp) was observed to have about the same intensity as the PCR

product amplified using just one gene-specific primer (374 bp) (Figure 4.1.3). This is the typical result of ligation efficiency analysis that indicates more than 25% completion of ligation. A 574 bp band in addition to the expected 374 bp band in the PCR product generated with one gene-specific primer would indicate incomplete digestion of DNA and the *Rsa*I digestion would then have to be repeated. However, there are disadvantages with the use of an *E. coli* tester as the control for the ligation efficiency analysis of *Mtb* DNA fragments. In this case, successful adaptor ligation was observed with the *E. coli* but only assumed for *Mtb*. Impurities in the experimental DNA from UM-CSF01 and UM-CSF04 might inhibit the ligase for adaptor ligation. Thus, for a more accurate determination, the ligation efficiency analysis should have been performed on the experimental (*Mtb*) DNA instead of DNA from a convenient *E. coli* substitute (Figure 4.1.3). This, however, would require additional funding for more reagents and further optimisation steps.

5.2.3 Analysis of First, Second Hybridisation and Suppression PCR

The serial hybridisation and suppression steps were in place to enrich and amplify the tester-specific sequences. After the first hybridisation, four types of molecules were generated (Figure 5.2.3.1). The tester-specific DNA (Type A), which were not affected by 'driver-pressure' remained as single-stranded (ss)-tester molecules. It appeared as smears on gel electrophoresis (Figure 4.1.4). In the second hybridisation, the remaining ss DNA (type A) were enriched for tester-specific sequences, because non-target DNA present in the testers (UM-CSF01 and UM-CSF04) formed hybrids with their

respective drivers (sputum F and sputum B) (Figure 5.2.3.1 and Figure 5.2.3.2) (Britten and Davidson, 1986). In the first PCR following the second hybridisation, only ds DNA with different adaptor sequences on each end are exponentially amplified. This was followed by a second nested PCR to further reduce background and enrich for tester-specific sequences (Figure 4.1.4). The result of the whole hybridisation and suppression PCR procedure was six specific DNA fragments subtracted from UM-CSF01 and nine from UM-CSF04.

For a successful subtraction, the hybridisation temperature is of critical importance and should be optimised for different bacteria. 63 °C is the optimal temperature for genomic DNA with an average GC content of 40.0 - 51.0 % as is found in *E. coli*, *Yersinia spp.*, *S. typhimurium*, and *D. nodosus* (Hacker et al., 1997). *Mtb* possesses a high GC content with an average of 65.6 %. Thus, the hybridisation temperatures for the first and second hybridisation were set to be 5 - 7 % higher, at 66 °C and 68 °C, respectively.

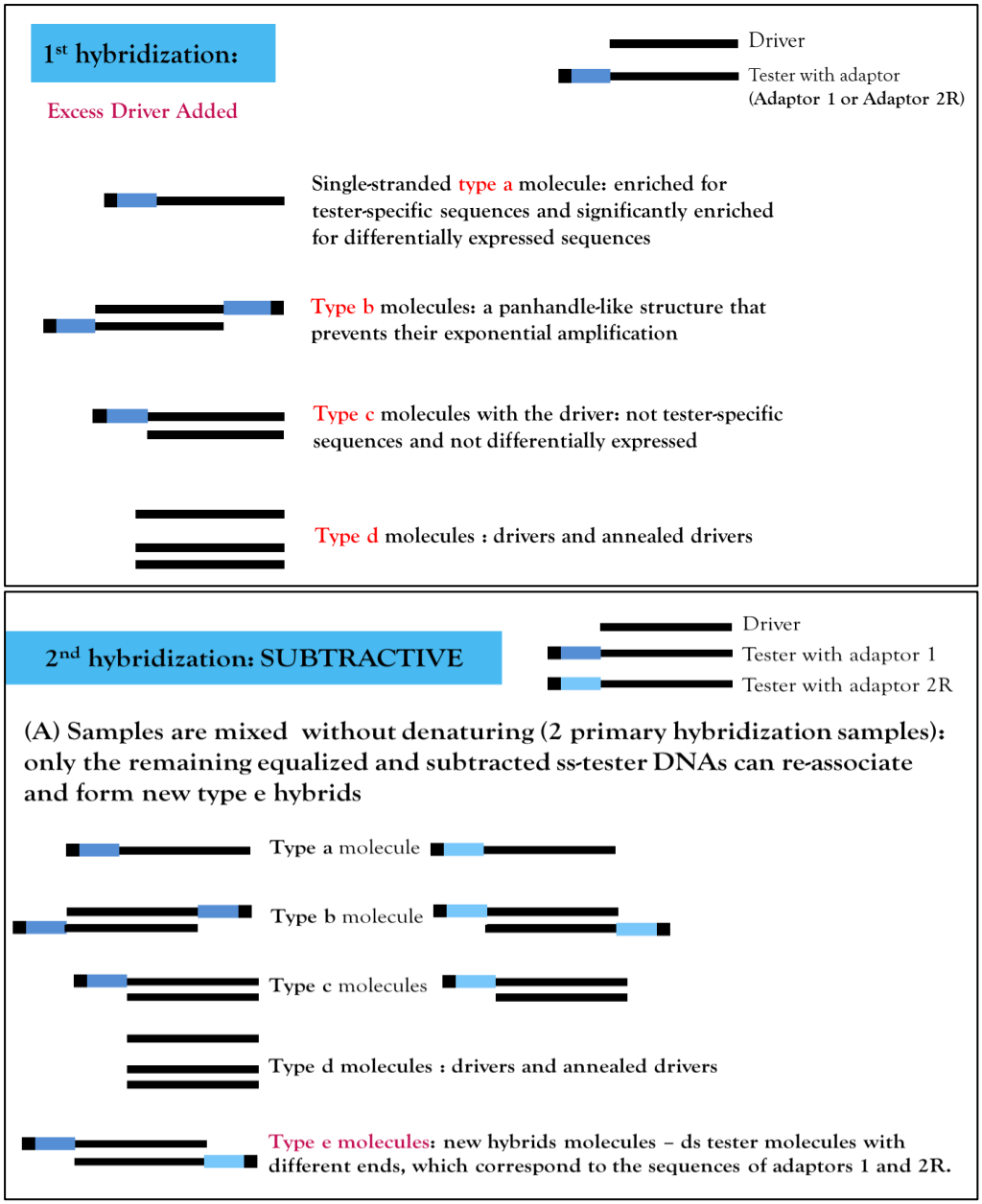


Figure 5.2.3.1: The schematic diagram of PCR-Select bacterial genome subtraction (Part I). Type a, b, c, and d molecules are generated on the first hybridization, after an excess of driver is added (Top). The ss type a molecules are significantly enriched for tester-specific sequences. The two primary hybridization samples are mixed together without denaturing on the second hybridization. This ensures that only the remaining equalized and subtracted ss tester DNAs can reassociate to form new type e hybrids. These new hybrids

are ds tester molecules with different ends, which correspond to the sequences of Adaptors 1 and 2R (Bottom).

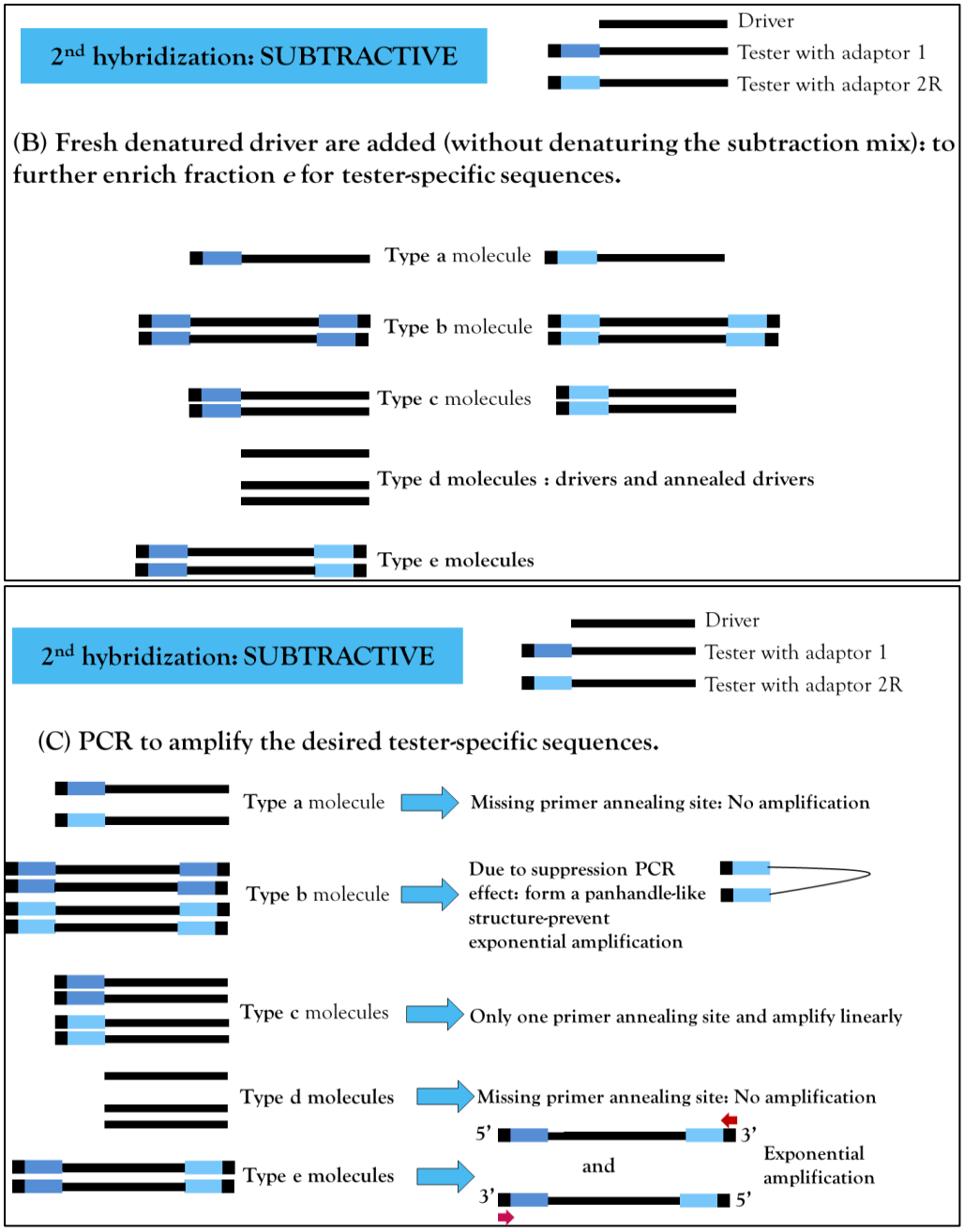


Figure 5.2.3.2: The schematic diagram of PCR-Select bacterial genome subtraction (Part II). Fresh denatured driver DNA is added again to further enrich type e molecules (Top). PCR is performed to amplify the desired tester-specific sequences. Type a and d molecules are missing primer-annealing sites,

and thus cannot be amplified. Due to the suppression PCR effect, most type b molecules form a panhandle-like structure that prevents their exponential amplification. Type c molecules have only one primer annealing site and amplify linearly. Only type e molecules, the differentially expressed tester-specific with two different adaptors, will amplify exponentially (Bottom) (Britten and Davidson, 1985; Clontech Laboratories, 2008).

The normalisation step is introduced in the subtraction procedure to increase the enrichment of rare DNA sequences (Siebert et al., 1995). A new software tool: SUBTRACT, is available to allow the selection of optimal strategy and conditions to perform SSH (Ermolaeva and Wagner, 1995). In addition, a novel mathematical model has been proposed to allow a routine comparison of genomes and products of genome expression (Ermolaeva et al., 1996). These models suggest that rare sequences can be enriched more than 1000 times during the first round of subtraction. This observation was supported in a model experiment with artificial targets added to DNA libraries (Siebert et al., 1995; Diatchenko et al., 1996). The level of enrichment for a particular sequence depends greatly on various factors including the original abundance of the sequence and the ratio of its concentration in the driver and tester. The normalisation step enables the highest level of enrichment to be obtained for sequences present in low abundance and/or in very different levels in the tester and driver DNA populations.

5.2.4 Quality Evaluation of the Subtracted DNA sequences

DNA fragments from an amplification reaction or restriction enzyme digestion have to be treated to remove contaminants such as enzymes, reaction buffers, primer dimers, PCR additives and amplification primers. This step is crucial as the presence of impurities can interfere with downstream molecular applications such as ligation, cloning, and PCR (Sambrook and Russell, 2006b). In this study, the subtracted DNA fragments seen in the gel electrophoresis were purified with a commercial kit. However, only fragments at concentrations between 5.0 and 20.0 ng/µl could be recovered (Table 4.1.4/Figure 4.1.6.1). Thus, the recovery (12.5 – 50.0 % of starting material) was lower than that expected with the use of the extraction kit which should be 50.0 - 80.0 % of DNA of more than 100 bp ion size (Dworkin, 2007). This low recovery might be due to the intrinsic limitations of the commercial kit, the efficiency of which is known to be affected by the concentration of the product, the final elution volume for purification and the concentration of agarose used for the electrophoresis (Yang et al., 2010; Sun et al., 2012). Furthermore, the purity of the recovered DNA fragments (A260/A280 ratio < 1.5) was not entirely satisfactory, probably due to the introduction of guanidine and ethanol during DNA purification (New England Biolabs, 2017).

The inadequate amount and purity of the subtracted DNA fragments could have affected the insert-pGEM®-T ligation efficiency. Other contributory factors could be the presence of high salt concentration in the elution, the presence of residual ethanol, agarose, primer-dimers, or denatured

ssDNA which appeared as a smaller smeared band in the gel electrophoresis (Figure 4.1.6.1). Unfortunately, there were insufficient left-over samples to repeat the experiments.

5.2.5 Technical Issues of Molecular Cloning

5.2.5.1 Transformation of *E. coli* JM109 cells

JM109 cells were chosen to be the host for the process of transformation in this study because it is an ideal host for blue/white screening (Yanisch-Perron et al., 1985). Sub-cloning Efficiency (Invitrogen) at greater than 10^7 cfu/μg, prepared according to a modified procedure of Hanahan (Hanahan, 1985), was purchased. However, the purchased *E. coli* JM109 had to be induced to be competent again using the $CaCl_2$ method (Cohen et al., 1972; Bergmans et al., 1981; Sambrook et al., 1989) as transformation efficiency was demonstrated to be very low in a trial run (Figure 4.1.5). This low efficiency might have been due to the competent cells (which are very sensitive to temperature) being transferred from one freezer to another during transportation from the supplier to the university. In addition, prolonged storage could have caused deterioration in the cell competence. However, despite the repeat induction leading to an improved transformation efficiency of 1.5×10^6 cfu/μg, the number of colonies obtained with the 15 subtracted sequences was disappointing (Table 4.1.5). This could be due to multiple factors including the strain type, plasmid concentration, and incubation time,

which have been reported to affect the transformation efficiency of nucleic acid.

$CaCl_2$ was used for the preparation of competent cells in this study as it was reported to be superior to other bivalent ions such as Sr^{2+}, Mn^{2+}, Mg^{2+} and Ba^{2+}. The procedure is cost-effective and simple to follow without any specialized equipment (Sambrook and Russell, 2006a). It has been reported that competent cells formed by this method could achieve transformation efficiencies of about 5.0×10^6 to 2.0×10^7 transformed colonies/µg of supercoiled plasmid DNA (Nishimura et al., 1990). This efficiency was achieved in a control transformation of competent *E. coli* JM109 cells using a supercoiled plasmid DNA (Promega Corporation, 2009). However, transformation with the subtracted DNA ligated plasmid produced much fewer colonies (Table 4.1.5). The reason for this could be that the ligated plasmid is not as compact as the supercoiled plasmid, the concentration of the plasmid was too high, the ice bath time was too short, or the cells were heat-pulsed for too long at 42 °C (Sratagene Technical Service, 2003; Chan et al., 2013; Liu X et al., 2014).

5.2.5.2 The Positive Clones

Although the purified PCR products (subtracted sequences) were sequenced, cloning was still used to increase the sensitivity of the detection of sequence mutations and to utilise the proof reading mechanism in the bacterial DNA replication for error reduction (Sanger, 1988). Unfortunately, the

number of desired recombinants obtained was very small (Table 4.1.6). The reasons could be incomplete ligation due to the subtracted sequences being impure and low in concentration (Table 4.1.4), incorrect insert:vector molar ratio, or the absence of 3'-A-overhangs in the PCR-amplified subtracted sequence. The pGEM®-T Easy Vector is linearized with a single 3'-terminal thymidine at both ends. These T-overhangs which are insertion sites for PCR products, improve ligation efficiency by preventing vector recircularisation. However, not all PCR products are 3'-A-tailed fragments. The Advantage 2 Polymerase Mix used in this study to generate PCR products comprises Titanium® *Taq* DNA Polymerase, a nuclease-deficient N-terminal deletion of *Taq* DNA polymerase plus TaqStart® Antibody to provide automatic hot-start PCR (Kellogg et al., 1994), and a minor amount of a proofreading polymerase. Titanium *Taq* PCR products are compatible with T/A cloning methods. However, for optimal ligation efficiency, PCR products have to be ligated immediately. If this were not possible, an additional 10 min extension at 70 °C is recommended, followed by immediate cloning or freezing of the PCR product and not storing at 4 °C. These steps will help ensure the incorporation and preservation of 3'-A-overhangs (Takara Bio USA, 2016). Even freezing at – 20 °C for an unduly long period can cause the deterioration of the 3'-A-overhangs of the subtracted sequences.

The recombinants obtained for the subtracted sequences (F1-1 to F1-6) from the Sputum F: UM-CSF01 pair was approximately 30.0 % higher than those (B4-1 to B4-9) from the Sputum B: UM-CSF04 pair (Table 4.1.6). This observation could be explained by the size range of the subtracted sequences.

Theoretically, smaller inserts (F1-1 to F1-6, ranging from 300 bp to 1 kb) are cloned more efficiently than longer inserts (B4-1 to B4-9, ranging from 400 bp to 2 kb). Although Promega Corporation (USA) has shown that the pGEM®-T Easy Vector Systems are able to give a high number of recombinants across a broad range of insert sizes (0.5 – 3kb), some researchers have demonstrated a much poorer performance with larger (1 – 3kb) insert sizes (Litterer, 2009) (Figure 4.1.6.2).

5.2.6 Analysis and Characterisation of Subtracted DNAs

The seven subtracted sequences generated after the process of trimming and filtering ranged in size from 240 bp to 900 bp. BLASTN analysis showed their sequences to be almost identical to sequences reported for a strain of *Mycobacterium tuberculosis* (strain C3) isolated from the brain of a patient in India (Table 4.1.7.1) (Husain et al., 2017). Hence, there is a possibility that these subtracted sequences could play a role in mycobacterial invasion of the CNS.

The BLASTX results (Table 4.1.7.2) identified the first subtracted sequence of sputum F: UM-CSF01 (**S1**) to be 82.0 % homologous to the PPE family protein. PE/PPE genes have only been described in *Mycobacterium* (Cole, 1998). They are so named because they encode proteins with Pro-Glu (PE) and Pro-Pro-Glu (PPE) signature motifs (Gordon et al., 2001). There are approximately 107 PE (classified into PE and PE_PGRS subgroups) and 69 PPE proteins which show many different structures even among strains of the

same *Mycobacterium* species (Sultana et al., 2011). The encoding genes are prone for mutations which are the major source of polymorphism in the *Mtb* complex. For instance, there is a deletion of 29 codons and insertion of 46 codons in the PGRS domain in the PE_PGRS9 gene (*Rv0746*) in *M. bovis* BCG that is not present in H37Rv (Brennan and Delogu, 2002). The propensity to mutations in PG_PGRS genes is due to their high number of repetitive sequences, association with intergenic or intragenic recombination events, or the consequences of the strand slippage during recombination between PE and PPE multi-gene families (Karboul et al., 2008).

The high number of single-base substitutions in these genes generate multiple antigenic variations (Cole et al., 1998; Fleischmann et al., 2002). They are found on the cell surface and are reported to be antigenic in nature. The corresponding genes are involved in the regulation of pathogenesis and virulence, and in the modulation of host immune response (Nair et al., 2011; Sampson, 2011; Mukhopadhyay and Balaji, 2011; Akhter et al., 2012; Sayes et al., 2012). For example, PGRS member *Rv1759*, a fibronectin-binding protein, was shown to play a role in eliciting antibody response (Abou-Zeid et al., 1991). The antigenic potential of PE_PGRS33 and several PE_PGRS proteins was demonstrated in murine models (Balaji et al., 2007). The level of PE_PGRS62-specific IgG was observed to be elevated in TB patients when compared to treated TB patients and mycobacterium-reactive TB contacts without latent infection (Koh et al., 2009).

There is, however, limited information describing the role of PPE/PE proteins in the dissemination of TB to the CNS. Some of these proteins have been shown to be immunoreactive at different stages of TB (early macrophage infection or later intracellular survival) and some are upregulated in CNS TB, suggesting that PE proteins might be involved in cerebral endothelial invasion (Jain et al., 2006; Narayana et al., 2007).

The third subtracted sequence (**S3**) showed complete homology with an *esx*-related secretion system. Bacterial protein secretory systems translocate molecules across the cytoplasmic membrane or the bacterial cell wall. In mycobacterial spp., however, the mycolic acid cell wall impedes the extracellular transport of proteins. Hence, to export effector proteins, mycobacteria have to use specialised secretion systems called type VII or ESX systems (Houben et al., 2014). The secretion of virulence-associated proteins helps virulent mycobacteria to replicate and spread or compete with other microorganisms in their environment (Abdallah et al., 2006). Of special interest are recent findings suggesting that the innate immune system of the host can detect and respond to the presence of bacterial protein secretion systems.

The second subtracted sequence (**S2**) from sputum F: UM-CSF01 and the fifth subtracted sequence (**S5**) from sputum B: UM-CSF04, were found by protein homology search to be 100.0 % similar to a manganese transporter which might play a role as an integral membrane protein. Manganese is necessary for microbial survival both in the environment and in a living host.

Even though the manganese transporter has not been described as a virulence factor in mycobacteria, it might still be important for the bacterium's survival in a host cell (Forrellad et al., 2013).

The fourth subtracted sequence (**S4**) showed complete homology with an uncharacterized protein which is conserved in *Mtb,* and has a highly conserved arginine and histidine residue that might be important for a yet unknown catalytic activity.

The subtracted sequences of **S6** and **S7**, both from sputum B: UM-CSF04, showed near 100.0 % homology for beta-lactamase and dTDP-4-dehydrorhamnose reductase, respectively. Beta-lactamase proteins are produced by many mycobacterial spp. including *Mtb,* making them intrinsically resistant to beta-lactam antibiotics. Since there are numerous types of beta-lactamases, it is not surprising that sequence differences were found in the enzymes from **S6**, **S7** and sputum B; hence, these differences may have no relevance to neurotropism. On the other hand, drug resistance-related mutations are known to affect bacterial pathogenicity. In tuberculosis, specifically, drug resistance has been associated with an increased fitness in the *Mtb* as well as an increased ability to cross the blood-brain barrier (Jain et al., 2006). This is consistent with the need for microbes to overcome immune defence systems and antimicrobial pressure in the host environment. Virulence and resistance factors can be co-transferred by horizontal gene transfer and co-selected in the host (Beceiro et al., 2013).

The dTDP-4-dehydrorhamnose reductase is involved in mycobacterial cell wall integrity. It causes the reduction of dTDP-6-deoxy-L-lyxo-4-hexulose to dTDP-L-rhamnose which is a component of the mycolic acid-peptidoglycan-arabinogalactan structure in the cell wall (Daffe et al., 1990; McNeil et al., 1991; Ma et al., 2002). Since rhamnosyl residues are important for the viability and growth of mycobacteria, two attributes that help the pathogen to resist the harsh environment such as the tight junction between the layer of BBB/BCSFB and the astrocyte processes in the CNS, it is possible that mutations in the reductase gene could be linked with the difference in neuropathogenicity between respiratory and CNS *Mtb* strains.

5.3 Comparative Genomic Analysis

Whole genome sequence (WGS)-based comparative analysis gained popularity owing to the availability of next generation DNA sequencing platforms, large datasets easily accessible from public databases and advanced bioinformatics tools (Morozova and Marra, 2008; MacLean et al., 2009; Metzker, 2010). Genetic information from WGS analysis is often used to assess population natural variation and predict host-pathogen relationships including virulence, immune modulation and response to therapy (Ford et al., 2012). Primary analysis such as filtering and trimming the sequences was performed to improve the data quality. Data metrics were generated to measure the data quality (Table 4.2.1). These metrics were included in the secondary analysis using CLC Genomics Workbench (Qiagen Inc.,

Netherlands). Then, the genomes of CSF and respiratory *Mtb* isolates were compared for features possibly related to neurotropism in TB.

5.3.1 Chromosomal Rearrangements, Micro-variants and Protein Structural Changes

Although *Mtb* exhibits less genetic diversity than other bacteria, comparative genomic analyses have successfully revealed SNPs, large sequence polymorphisms and variations in mobile elements among *Mtb* isolates from different sources (Fleischmann et al., 2002; Tsolaki et al., 2004; Ilina et al., 2013; Liu F et al., 2014). In addition, large-scale chromosomal rearrangements have been reported, such as the large inversions detected in the KwaZulu-Natal (KZN) strains from South Africa (Okumura et al., 2015) and the B0/W148 strains from Russia. In the UM-CSF strains, bioinformatics analysis revealed translocations, inversions, indels and SNPs that were not detected in the respiratory strains used for comparison. Many of the genes involved were putative virulence factors that included abundant PE and PPE proteins. This finding is consistent with that of Faksri and colleagues (2018) who also found many micro-variants in the PPE/PE-family genes in all the *Mtb* meningitis strains they examined (Faksri et al., 2018). The prominence of PE/PPE proteins here is not surprising as these hyper-variable cell surface antigens form about 10 % of the total coding sequences in the *Mtb* genome (Cole et al., 1998) and their association with large sequence polymorphisms (Talarico et al., 2007) has been reported. It is believed that these uniquely mycobacterial proteins are secreted by the ESX apparatus in the cell

membrane (Abdallah et al., 2006) and are involved in many aspects of the infection process, such as promoting *Mtb* entry into macrophages and evasion of host immune responses, resulting in *Mtb* dissemination and pathology in different organs and tissues. On the other hand, in a recent study, mutations in *ppe38* were reported to block PE_PGRS secretion and increase the virulence of *Mtb* from the Beijing lineage. A deletion of ppe38 was found to be shared by Beijing outbreak strains worldwide, suggesting that this deletion might have contributed to the success and global distribution of the 'Modern' Beijing strains (Ates et al., 2018). Unfortunately, all the PE/PPE proteins identified in this study (Table 4.2.4) have no known functions except for PE-PGRS30 (Rv1651c) which has been listed as a virulence factor (Iantomasi et al., 2011; Brennan, 2017).

Also abundant in the rearranged fragments are TA and Mce proteins. The TA systems in *Mtb* are up-regulated following bacterial entry into macrophages (Ramage et al., 2009). The VapBC family which made up most of the TA proteins in the UM-CSF strains was shown to control the persistence of uro-pathogenic *E. coli* within host tissues in cases of sepsis, meningitis and urinary tract infections (Norton and Mulvey, 2012). It is possible that these proteins also contribute to the persistence of *Mtb* in TBM. Similarly, Mce proteins may enhance cellular invasion and persistence of *Mtb* in the CNS by facilitating the uptake and utilisation of cholesterol from host cells during infection (Pieters and Gatfield, 2002).

Serine-threonine proteins are kinases that phosphorylate serine or threonine residues. In bacteria, they form the STPK family that is important for signal recognition and co-ordination of cellular responses including the disruption of host immune responses. Hence, they are essential virulence factors. Be and colleagues identified *Mtb pkn*D (Rv0931c) as essential for CNS tropism (Be et al., 2012). The 11 genes for these putative kinases (*pkn*A-*pkn*L) in *Mtb* were identified in the UM-CSF strains but only *pkn*G (Rv0410c) was in a rearranged fragment in UM-CSF05. The PknG protein mediates survival of mycobacteria within macrophages by blocking lysosomal activity. Its role as a virulence factor is reflected in the observation that while this protein is efficiently translated in the pathogenic *Mycobacterium bovis*, its translation is blocked in the non-pathogenic *Mycobacterium smegmatis* (Houben et al., 2009). In addition, PknG controls glutamate metabolism which mediates the transfer of signals sensing nutritional stress in *Mtb* and translates them into metabolic adaptation (Cowley et al., 2004). Its relevance to neurotropism has not yet been described.

The amino acid sequence of a protein determines its folding into three-dimensional structures which may be ordered (globular with high hydrophobicity) or disordered (unstructured, typically with low hydrophobicity). Importantly, the folding of proteins affects their function. In the DAVID analysis of proteins showing different globularity between UM-CSF strains and H37Rv, transcription factors and other transcription regulators predominate (Figure 4.2.5). It is expected that changes in transcription-related proteins would have profound effects on gene expression resulting in wide-

ranging changes in *Mtb* behaviour including altered tropism and virulence. An interesting finding in this study is the enrichment of proteins involved in androgen and oestrogen metabolism. This could be related to the role of sex hormones in the modulation of bacterial-host interactions (Klein, 2000; Hughes and Sperandio, 2008; Cabrera-Munoz et al., 2012; García-Gómez et al., 2013). It is known that, while sex hormones from the host can affect the metabolism and virulence of bacteria, they are also degraded by bacteria to be used as carbon and energy sources (Neyrolles and Quintana-Murci, 2009; García-Gómez et al., 2013). It is plausible that a change in protein domain globularity in the *Mtb* proteins involved in sex hormone metabolism could upset the usual bacterial-host interactions to facilitate CNS invasion in TBM.

Among other enriched proteins of interest is ABC transporters, glycerophospholipids and proteins involved in DNA repair. ABC transporters transfer molecules across cell membranes and are implicated in *Mtb* traversal across the blood-brain barrier (Jain et al., 2006). Glycerophospholipids are found in neural membranes (Farooqui et al., 2000). A functional change in these lipids could increase membrane permeability to *Mtb*. Bacterial DNA can be damaged by reactive oxygen species and reactive nitrogen intermediates generated from cellular metabolism and the host immune response (Davidsen et al., 2007). Among DNA repair mechanisms are nucleotide excision repair (NER) controlled by the UvrABCD endonuclease enzyme complex, base excision repair and DNA mismatch repair. Darwin & Nathan (2005) showed that the *uvrB* gene is required for *Mtb* to resist reactive oxygen and nitrogen molecules *in vivo* (Darwin and Nathan, 2005). This implies that a change in

the *uvr* genes in *Mtb* might increase susceptibility of *Mtb* DNA to damage by reactive intermediates. In UM-CSF strains, the *uvrC* gene is in a rearranged fragment in UM-CSF01 and homologs of mismatch repair genes are among the enriched proteins with globularity changes. Further investigations, for instance with computer modelling of biological data (Couce et al., 2016; Hicks et al., 2019) will be required to elucidate the relevance of the enriched proteins in TB infection. This is especially important because some of the proteins were CNS also found in the respiratory *Mtb* genomes used for comparison.

5.3.2 Meningitis-Associated Genes Reported for Mycobacteria and Other Bacterial Neuro-pathogens

In the search for 63 genes previously reported to be associated with TBM, 56 to 60 identified by other researchers were found in the UM-CSF strains but only two were found in all eight strains. This suggests that these possibly neurotropic genes are not universally present in *Mtb* causing CNS disease. Nevertheless, a few of these 63 genes were also found in *M. bovis, M. leprae, M. lepromatosis,* and two environmental rapid-growers *M. ilatzerense* and *M. immunogenum* isolated from a brain abscess (Figure 4.2.7.2). This seems to suggest that there might be common genetic traits in mycobacteria that are responsible for diseases of the nervous system (central and peripheral) in human.

The common genes identified in UM-CSF strains and the other mycobacterial spp. associated with CNS infection is mostly involved in cell

membrane transport, signal transduction, nucleotide biosynthesis, and energy metabolism (Table 4.2.6). Although these functions are common to all microbial cells, most of the genes have been previously associated with TB CNS pathology. For instance, Rv2318, Rv0983, Rv0984, and Rv0966c were found to be up-regulated in TB of the human brain microvascular endothelium (Jain et al., 2006); Rv1837c was reported to be expressed in the early stages of TBM in children (Haldar et al., 2012); the transmembrane serine/threonine-protein kinase PknD (Rv0931c) was found by Be and colleagues (2012) to be required for the invasion of brain endothelia, and Rv0805 was found to be important for invasion and survival within brain tissue in a murine model (Be et al., 2008).

Homologs of meningitis-associated genes identified in *S. pneumoniae*, and *N. meningitidis* were found in all UM-CSF genomes (Figure 4.2.7.1). The genes shared with *S. pneumoniae* are those involved in the biosynthesis of pyrimidine (*pyr*G, Rv1699), purine (*pur*A, Rv0357c) and pyridoxine (Rv2606c). Both *pur*A (adenylosuccinate synthetase) and *pyr*G (CTP synthase) have been reported to be associated with attenuated *S. pneumoniae* replication during experimental meningitis (Molzen et al., 2011) and pyridoxine is important for the functioning of nerves. Of the two homologs shared with *N. meningitidis.* one (Rv2457c) is a chaperone of the ATP-dependent CLP protease that is required for *in vitro* and *in vivo* growth of *Mtb* (Raju et al., 2012) and the other (Rv2397c), is a part of the ABC transporter complex involved in sulfate/thiosulfate import (Szklarczyk et al., 2015). The relevance

of these genes to *Mtb* neurotropism is not clear as they were also found in H37Rv and other respiratory *Mtb* genomes.

The detection of common homologs in neuro-pathogens from different bacterial taxa raises speculations on the existence of a pan-bacterial mechanism of CNS infection. Although it was disappointing to find that many of the possibly neurotropic traits in the UM-CSF strains were also found in respiratory *Mtb*, this finding is consistent with the observation by other workers that many virulence genes are conserved in non-pathogenic bacteria. For instance, all four *mce* operons in the genome of *Mtb* (Kumar et al., 2003) have been found in mycobacteria of both human and environmental origin (Chitale et al., 2001; Haile et al., 2002). CLP proteases on the whole, are common in many bacterial spp. (De Mot et al., 1999). The ABC transporter complex involved in sulfate/thiosulfate import is found in pathogens as well as environmental bacteria (Szklarczyk et al., 2015). Many designated virulence genes in *N. meningitidis* were also found to be present in non-pathogenic species such as *N. lactamica* (Snyder and Saunders, 2006). From these and other observations, Forrellad et al. (2013) concluded that free-living microbes can become intracellular pathogens with little change in genetic determinants of virulence (Forrellad et al., 2013). Hence, by analogy, it can also be hypothesised that CNS tropism in TB is not driven by the presence of specific genetic traits but by the expression of multiple virulence factors, probably elicited in response to host immune defences. This conclusion is supported by a study from Indonesia that compared the WGS of 106 TBM isolates against

216 PTB isolates and did not find any genetic variant that is specifically associated with either TBM or PTB (Ruesen et al., 2018).

5.3.3 Verification of Putative Genes Identified by WGS Analysis

While next-generation sequencing technologies have brought many advantages to genomic research, the chain-termination method of DNA sequencing remains widely used, especially for the validation of WGS findings. Sanger sequencing is preferred for the sequencing of single genes as it is cost-effective. It is used to verify sequences for site-directed mutagenesis or the presence of cloned inserts and it is less error-prone than NGS (Thermo Fisher Scientific, 2014; Biotech, 2018). Thus, PCR-sequencing assays were used to verify the existence of six putative genes identified in the *Mtb* genomes from CSF (Table 4.3; Figure 4.3.1.1). These genes represent different types of sequence variations predicted.

Of the six genes studied, two (Rv3425 and Rv1141c) did not show the genetic features (deletions) predicted by the WGS analysis (Figure 4.3.1.2; Figure 4.3.1.3). This could be due to the inadequacy of the Illumina sequencing technology which gives only a draft genome (Gardner et al., 2005). In the generation of a whole genome sequence, both sequencing errors (affecting individual nucleotides) and assembly errors (involving sequence blocks) can occur. It has been said that with next-generation sequencing methods such as Illumina, it may be impossible to sequence and assemble all nucleotides in the genome (Ellegren, 2014), and the assembly and alignment

of DNA segments, especially those in repetitive regions, gene duplications and chromosomal rearrangements, are particularly prone to errors (Brosch et al., 2002; Tsolaki et al., 2004).

The verification of Rv3344c and Rv0311 variations in UM-CSF01 was more successful. The sequence obtained for Rv3344c indicated that the smaller than predicted PCR product size (825 bp instead of 1455bp) was probably due to multiple deletions, although only one (the longest) of the deletions was detected in the WGS analysis (Figure 4.3.1.4). The result of the Rv0311 verification showed the converse. The G-T SNP in the gene was predicted in all eight CSF strains but was confirmed only in five of them, indicating a possible sequencing error in the other three strains (Figure 4.3.2.1; Figure 4.3.2.2).

The meningitis-associated genes in *N. meningitidis* and *S. pneumoniae*, Rv2397c and Rv2606c, were verified in all UM-CSF isolates and H37Rv, as predicted (Figure 4.3.3.1). For both genes, the DNA sequence was identical in all UM-CSF isolates and H37Rv (Figure 4.3.3.2; Figure 4.3.3.3).

5.4 The Combination Analysis of SSH and NGS

From the SSH study, seven subtracted sequences were obtained from two UM-CSF strains (UM-CSF01 and UM-CSF-05). From the WGS analysis of eight UM-CSF strains (including UM-CSF01 and UM-CSF05), 35 putative genes were identified to be possibly related to neurotropism (Appendix 28:

Table A28). A comparison of the subtracted sequences from SSH with the verified genes from the WGS showed S1 to be 34.0 % homologous with PPE family proteins Rv3425 and Rv3426. S2 shared only 8.0 % and 7.0 % homology with integral membrane proteins Rv1087 and Rv0805, respectively. S3 was 7.0 % homologous with Rv1837c which codes for glcB glyoxylate protein. S4 showed 9.0 % and 8.0 % similarity to Rv2947 (encoding *pks15* protein) and Rv1587c (hypothetical protein) respectively; S5 was 6.0 % similar to Rv0278, which codes for PE-PGRS family proteins, and S6 and S7 shared 4.0 % and 22.0 % similarity with Rv2318 and Rv1067c, respectively. Rv2318 encodes lipoproteins and Rv1067c encodes PE_PGRS proteins (Appendix 30: Table A30).

Although the SSH and WGS results did not correlate significantly, all the genes and sequences identified in both comparative analyses as possibly different in CNS and respiratory TB are related to the family of PPE proteins which are involved in mycobacterial cell wall hydrophobicity and mycobacteria-host immune system interactions (Sampson, 2011). Sreevatsan et al. (1997) reported that new subclones generated by genomic rearrangements impact on bacterial virulence and resistance to drugs (Sreevatsan, 1997). Faksri et al. (2018) compared the whole genome sequences of 73 TBM and 220 pulmonary TB (PTB) isolates and found structural variants in several PE/PPE gene families to be associated with TBM (Faksri et al., 2018). Further in-depth analysis of these proteins will provide more information on the role they play in mycobacterial survival and adaptation to environment.

5.5 Limitations

The SSH is a complex DNA subtraction method for the identification of tester-specific genes. Although this method has been widely used, its efficacy and limitations have not been evaluated adequately. Each one of the multiple steps in the protocol requires stringent quality control and optimisation. Ji and colleagues (2002) identified the main factor that influences the efficacy of SSH to be the concentration ratio (R) of a target gene between two DNA preparations, which should be more than five-fold (Ji et al., 2002). As the driver:tester ratio used in this study was only 2:1, the efficiency of the target gene enrichment could have been compromised.

Another difficulty with the SSH is the generation of unwanted background owing to the random annealing of adapters to non-target DNA molecules. This background can cause the production of false positive clones, particularly in cases where there are only small numbers of differential sequences, as in the highly conserved *Mtb* genome. Elimination of background is difficult and laborious. Methods of elimination include mirror-orientation selection (MOS) based on the elimination of adaptor-derived primer in the secondary PCR (Lukyanov et al., 2007) and Southern blot analysis. In the latter, more commonly applied method, ^{32}P-labelled tester and driver probes are used to hybridise with the subtracted sequences. The sequences that hybridise to both tester and driver probes (background) are then discarded. However, there must be sufficient tester and driver DNA to make the hybridisation probes. Unfortunately, this was not the case in this study; hence,

non-specific background, if present, was not eliminated and subtracted sequences from the electrophoresis gel were used directly for cloning.

Next-generation sequencing (NGS) was used in this study because it allows the detection of both rare and multiple forms of genetic change. With only a low input of DNA, the whole genome sequence can be achieved. However, there are still problems with reliability and accuracy, inconsistent results from different platforms and subjectivity in the interpretation of results (Dubchak and Pachter, 2002; Alkan et al., 2011).

The number of local *Mtb* genomes (13 respiratory and eight CSF) analysed in this study was limited by the high cost of Illumina sequencing and the scarcity of *Mtb* isolates from CSF samples. A search for *Mtb* genomes in public databases did not turn up any comparable raw genome data that could be used to augment the number of CSF genomes for analysis. The number of useful respiratory genomes, however, increased from 13 to 85. This enabled the eight CSF genomes to be compared with 30 to 85 respiratory genomes in the analyses for genome rearrangements and micro-variants. The occurrence of protein globularity and meningitis-associated genes, however, were compared only between CSF genomes and H37Rv. In these comparisons, the problem of false positive findings derived from the use of small sample numbers became evident; the disordered proteins identified in UM-CSF genomes but not in H37Rv were found in many of the respiratory genomes downloaded from public databases. Similarly, the meningitis-associated genes reported for *S. pneumoniae* and *N. meningitidis* were found in UM-CSF,

H37Rv and many other respiratory *Mtb* genomes (Appendix 8: Table A8 to Appendix 11: Table A11). These observations clouded the hypothesis that the protein structural and globularity changes found in UM-CSF strains might be genetic traits important for the pathogenesis of TBM, and that the genes associated with meningitis in mycobacteria, *S. pneumoniae* and *N. meningitidis* might represent pan-bacterial neurotropic genes.

The main limitation in this study pertinent to both SSH and WGS analyses is the limited supply of inactivated *Mtb* that has prevented further investigation of promising early findings. This shortage is because the laboratory in which this thesis was carried out does not have biosafety level 3 facilities for work with live *Mtb*. Without *Mtb* propagation, the stock of extracted *Mtb* DNA could not be replenished for use in repeat assays when results were not entirely satisfactory. Neither could extra procedures be conducted to confirm or disprove observations made. In addition, although care was taken to exclude patients with both CSF and sputum positive TB, most patients were not followed up to exclude the development of TB meningitis as a complication of pulmonary infection. Hence, some of the CSF-derived isolates in this study could have been respiratory TB and not strains with neurotropic attributes that enable them to invade the CNS directly.

As clinical outcomes are always the consequence of host-parasite interactions, a more complete description of patient and *Mtb* characteristics would greatly improve the discussion and conclusions made in this study. Unfortunately, no patient information was available other than standard

demographic and routine diagnostic data. Moreover, the information on the *Mtb* strains used for comparison with local strains was limited by what was freely available on-line from international databases. Hence, it was not possible to include detailed host and pathogen features such as the immune status of all hosts and the lineage of all *Mtb* strains into the deliberation on possible neurotropism in tuberculosis.

5.6 Summary

The main findings in this study are listed below:

1) With SSH using two UM-CSF strains as testers and two sputum-derived strains as the drivers, seven subtracted sequences were obtained.

2) With WGS analysis of eight UM-CSF strains (including the two used for SSH), large-scale genomic rearrangements, indels, gene truncations and micro-variants were found that were not detected in 30 - 85 respiratory *Mtb* genomes used for comparison.

3) PE/PPE proteins and other virulence factors predominated in the subtracted sequences and genetic traits identified by WGS analysis.

4) Homologs of meningitis-associated genes previously reported in *Mtb*, other mycobacterial spp., *S. pneumoniae* and *N. meningitidis* were identified in all eight UM-CSF strains.

5) Many protein globularity changes and homologs of meningitis-associated genes predicted in the UM-CSF strains were also present in the comparison respiratory strains.

CHAPTER 6

CONCLUSION

Many genetic traits have been described for bacterial pathogens causing CNS infections. Those related to *Mtb* CNS disease however, have yet to be thoroughly investigated. In this study, the genetic differences between *Mtb* causing meningitis and *Mtb* causing pulmonary tuberculosis were studied using SSH and WGS. With SSH, at least seven subtracted DNA sequences were identified and with WGS analysis, various genetic rearrangements, indels, nsSNPs and protein globularity changes were predicted that could be related to CNS disease. In addition, homologs of meningitis-associated genes previously reported in other neuro-pathogenic bacteria were found in *Mtb* isolated from CSF samples. Unfortunately, many of these features are either not CSF-specific or not consistently present in all CSF strains. Hence, further investigations are necessary with larger numbers of *Mtb* genomes from CNS isolates to be compared with respiratory isolates, and functional studies to explore the effect of the genetic differences identified in CNS but not respiratory strains of *Mtb*.

CHAPTER 7

ACCOMPLISHMENTS

The findings of this research work were successfully presented in conferences and published in peer-reviewed articles, as listed below, in chronological order:

1) Saw, S.H., Yong, V.C.and Ngeow, Y.F. 2012. Application of suppression subtractive hybridization (SSH) to generate differentially regulated genes in *Mycobacterium tuberculosis (Mtb)* isolates from cerebrospinal fluid (CSF) and respiratory secretions. *17th Biological Science Graduate Congress*, 8 - 10 December 2012 Chulalongkorn University, Bangkok, Thailand.

2) Saw, S.H., Yong, V.C. and Ngeow, Y.F. 2014. Generation of differentially regulated genes in *Mycobacterium tuberculosis* isolates from cerebrospinal fluid and respiratory secretions using suppression subtractive hybridization. BMC Infectious Diseases, 14(Suppl 3), p.O16. Available at: http://www.biomedcentral.com/1471-2334/14/S3/O16%0AORAL. *(Saw et al., 2014)*

3) Saw, S.H., Tan, J.L., Chan, X.Y., Chan, K.G. and Ngeow, Y.F. 2016. Chromosomal rearrangements and protein globularity changes in

Mycobacterium tuberculosis isolates from cerebrospinal fluid. PLoS ONE, 4(e2484), pp.1–21. Available at: https://peerj.com/articles/2484/. *(Saw et al., 2016)*

4) Saw, S.H., Tan, J.L., Yap, S.F.and Ngeow, Y.F. 2017. The occurrence of meningitis-associated genes in respiratory isolates of *Mycobacterium tuberculosis*. *Inaugural FMHS Scientific Meeting "Delivering Values to the Elders"*, 25 - 26 May 2017 UTAR Sungai Long, Selangor Malaysia.

5) Saw, S.H., Yap, S.F. and Ngeow, Y.F. 2019. Meningitis-associated genes in *Mycobacterium* species. RSU International Research Conference. 2019 Research Institute of Rangsit University, Pathum Thani Thailand, pp. 14–24. *(Saw et al., 2019)*

BIBLIOGRAPHY

Abdallah, A.M. et al., 2006. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. *Molecular Microbiology*, 62(3), pp.667–679. Available at: https://doi.org/10.1111/j.1365-2958.2006.05409.x.

Abou-Zeid, C. et al., 1991. Genetic and immunological analysis of Mycobacterium tuberculosis fibronectin-binding proteins. *Infection and Immunity*, 59(8), pp.2712–2718. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC258077/pdf/iai00044-0208.pdf.

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology*, 62, pp.53–70. Available at: https://doi.org/10.1146/annurev.micro.62.081307.162832.

Adams, D.O., 1976. The granulomatous inflammatory response A review. *American Journal of Pathology*, 84(1), pp.164–192. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2032357/.

Agron, P.G. et al., 2001. Identification by subtractive hybridization of sequences specific for Salmonella enterica Serovar Enteritidis. *Applied and Environmental Microbiology*, 67(11), pp.4984–4991. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC93261/.

Agron, P.G. et al., 2002. Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. *FEMS Microbiology Letters*, 211(2), pp.175–182. Available at: https://doi.org/10.1111/j.1574-6968.2002.tb11221.x.

Akhter, Y., Ehebauer, M.T., Mukhopadhyay, S. and Hasnain, S.E., 2012. The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? *Biochimie*, 94(1), pp.110–116. Available at: https://doi.org/10.1016/j.biochi.2011.09.026.

Aldous, W.K., Pounder, J.I., Cloud, J.L. and Woods, G.L., 2005. Comparison of six methods of extracting Mycobacterium tuberculosis DNA from processed sputum for testing by quantitative real-time PCR. *Journal of Clinical Microbiology*, 43(5), pp.2471–2473. Available at: https://doi.org/10.1016/j.biochi.2011.09.026.

Alkan, C., Sajjadian, S. and Eichler, E.E., 2011. Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), pp.61–65. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3115693/.

Alland, D. et al., 1994. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med*, 330, pp.1710–1716. Available at: https://www.nejm.org/doi/pdf/10.1056/NEJM199406163302403?articleTools=true.

Allix-Béguec, C. et al., 2014. Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit–variable-number tandem-repeat loci for subtyping of Mycobacterium tuberculosis Beijing isolates Moser, S.A., (ed.). *Journal of Clinical Microbiology*, 52(1), pp.164–172. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3911419/.

Alonso, M. et al., 2010. Characterization of Mycobacterium tuberculosis Beijing isolates from the Mediterranean area. *BMC Microbiology*, 10(1), pp.151–173. Available at: https://doi.org/10.1186/1471-2180-10-151.

Amaro, A. et al., 2008. Comparison of three DNA extraction methods for Mycobacterium bovis, Mycobacterium tuberculosis and Mycobacterium avium subsp. avium. *Letters in Applied Microbiology*, 47(1), pp.8–11. Available at: https://doi.org/10.1111/j.1472-765X.2008.02372.x.

Angiuoli, S. V. and Salzberg, S.L., 2011. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics*, 27(3), pp.334–342. Available at: https://doi.org/10.1093/bioinformatics/btq665.

Arvanitakis, Z. et al., 1998. Mycobacteria tuberculosis molecular variation in CNS infection: evidence for strain-dependent neurovirulence. *Neurology*, 50(6), pp.1827–1832. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9633735.

Ates, L.S. et al., 2018. Mutations in ppe38 block PE_PGRS secretion and increase virulence of Mycobacterium tuberculosis. *Nature Microbiology*, 3(2), pp.181–188. Available at: https://doi.org/10.1038/s41564-017-0090-6.

Av-Gay, Y. and Everett, M., 2000. The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. *Trends in Microbiology*, 8(5), pp.238–244. Available at: https://doi.org/10.1016/S0966-842X(00)01734-0.

Balaji, K.N. et al., 2007. Apoptosis triggered by Rv1818c, a PE family gene from Mycobacterium tuberculosis is regulated by mitochondrial intermediates in T cells. *Microbes and Infection*, 9(3), pp.271–281. Available at: https://www.sciencedirect.com/science/article/pii/S1286457906004096?via%3Dihub [Accessed: 1 June 2019].

Baldwin, K.J. and Zunt, J.R., 2014. Evaluation and treatment of chronic meningitis. *The Neurohospitalist*, 4(4), pp.185–195. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4212414/pdf/10.1177_1941874414528940.pdf.

Banavaliker, J.N. et al., 1998. Identification of Mycobacterium tuberculosis by

polymerase chain reaction in clinical specimens. *Indian Journal of Tuberculosis*, 45, pp.15–18. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.485.7137&rep=rep1&type=pdf.

Barkan, D. et al., 2012. Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice. *Infection and Immunity*, 80(6), pp.1958–1968. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3370573/.

Be, N.A. et al., 2008. Murine model to study the invasion and survival of Mycobacterium tuberculosis in the central nervous system. *The Journal of Infectious Diseases*, 198(10), pp.1520–1528. Available at: http://jid.oxfordjournals.org/content/198/10/1520.abstract.

Be, N.A., Bishai, W.R. and Jain, S.K., 2012. Role of Mycobacterium tuberculosis pknD in the pathogenesis of central nervous system tuberculosis. *BMC Microbiology*, 12(7), pp.1–12. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3322341/.

Be, N.A., Kim, K.S., Bishai, W.R. and Jain, S.K., 2009. Pathogenesis of central nervous system tuberculosis. *Current Molecular Medicine*, 9(2), pp.94–99. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19275620.

Beceiro, A., Tomás, M. and Bou, G., 2013. Antimicrobial resistance and virulence: A successful or deleterious association in the bacterial world? *Clinical Microbiology Reviews*, 26(2), pp.185–230. Available at: https://cmr.asm.org/content/26/2/185.long.

de Beer, J.L. et al., 2012. First worldwide proficiency study on variable-number tandem-repeat typing of Mycobacterium tuberculosis complex strains. *Journal of Clinical Microbiology*, 50(3), pp.662–669. Available at: http://jcm.asm.org/content/50/3/662.abstract.

Beham, A.W. et al., 2011. A TNF-regulated recombinatorial macrophage immune receptor implicated in granuloma formation in tuberculosis. *PLoS Pathogens*, 7(11: e1002375), pp.1–16. Available at: https://doi.org/10.1371/journal.ppat.1002375.

Bemer-Melchior, P. and Drugeon, H.B., 1999. Inactivation of Mycobacterium tuberculosis for DNA typing analysis. *Journal of Clinical Microbiology*, 37(7), pp.2350–2351. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC85159/.

Bergmans, H.E., van Die, I.M. and Hoekstra, W.P., 1981. Transformation in Escherichia coli: Stages in the process. *Journal of Bacteriology*, 146(2), pp.564–570. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC216999/.

Bermudez, L.E. et al., 2002. The efficiency of the translocation of

Mycobacterium tuberculosis across a bilayer of epithelial and endothelial cells as a model of the alveolar wall is a consequence of transport within mononuclear phagocytes and invasion of alveolar epithelial cells. *Infection and Immunity*, 70(1), pp.140–146. Available at: https://iai.asm.org/content/70/1/140.long.

Bernaerts, A. et al., 2003. Tuberculosis of the central nervous system: overview of neuroradiological findings. *European Radiology*, 13(8), pp.1876–1890. Available at: https://link.springer.com/article/10.1007%2Fs00330-002-1608-7.

Bernama, 2019. Health D-G: Number of TB cases in Malaysia on the decline. *Malay Mail*, p.1. Available at: https://www.malaymail.com/news/malaysia/2019/03/26/health-d-g-number-of-tb-cases-in-malaysia-on-the-decline/1736720.

Berthet, F. et al., 1998. A Mycobacterium tuberculosis operon encoding ESAT = 6 and a novel low-molecular- mass culture filtrate protein (CFP-10). *Microbiology*, 144(11), pp.3195–3203. Available at: https://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-144-11-3195#tab2.

Besemer, J., Lomsadze, A. and Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 29(12), pp.2607–2618. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55746/.

Bhatia, P., 2017, *What are the functions of SDS in DNA extraction?* [Online]. Available at: https://www.quora.com/What-are-the-functions-of-SDS-in-DNA-extraction [Accessed: 21 November 2017].

Bini Estela, I. and Hernandez Pando, R., 2014. Pathogenesis and immune response in Tuberculous Meningitis. *Malaysian Journal of Medical Sciences*, 21(1), pp.4–10. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3952336/.

Biotech, G., 2018, *Sanger Sequencing* [Online]. Available at: https://www.gatc-biotech.com/en/expertise/sanger-sequencing.html [Accessed: 21 March 2018].

Bishai, W.R. et al., 1999. Virulence of Mycobacterium tuberculosis CDC1551 and H37Rv in rabbits evaluated by Lurie's pulmonary tubercle count method. *Infection and Immunity*, 67(9), pp.4931–4934. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC96831/.

Bishburg, E., Sunderham, G., Reihman, L.B. and Kapila, R.A., 1986. Central nervous system tuberculosis with the acquired immunodeficiency syndrome and its related complex. *Annals of Internal Medicine*, 105(2), pp.210–213. Available at: http://annals.org/aim/article-abstract/700648/central-nervous-

system-tuberculosis-acquired-immunodeficiency-syndrome-its-related-complex?volume=105&issue=2&page=210.

Bodnar, K.A., Serbina, N. V and Flynn, J.L., 2001. Fate of Mycobacterium tuberculosis within murine dendritic cells. *Infection and Immunity*, 69(2), pp.800–809. Available at: https://iai.asm.org/content/69/2/800.long.

Boetzer, M. and Pirovano, W., 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, 15(211), pp.1–9. Available at: https://doi.org/10.1186/1471-2105-15-211.

Bouakaze, C. et al., 2010. Identification and genotyping of Mycobacterium tuberculosis complex species by use of a SNaPshot minisequencing-based assay. *Journal of Clinical Microbiology*.

Bougnoux, M.E. and Morand, S.D.C., 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. *Journal of Clinical Microbiology*, 40(4), pp.1290–1297. Available at: https://10.0.4.104/JCM.40.4.1290-1297.2002.

Braden, C.R. et al., 2001. Simultaneous infection with multiple strains of Mycobacterium tuberculosis. *Clinical Infectious Diseases*, 33(6), pp.e42–e47. Available at: https://doi.org/10.1086/322635.

Braibant, M., Gilot, P. and Content, J., 2000. The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis. *FEMS Microbiology Reviews*, 24(4), pp.449–467. Available at: https://doi.org/10.1111/j.1574-6976.2000.tb00550.x.

Brennan, M.J., 2017. The Enigmatic PE/PPE multigene family of Mycobacteria and tuberculosis vaccination. *Infection and Immunity*, 85(6), pp.e00969-16. Available at: https://doi.org/10.1128/IAI%0A.00969-16.

Brennan, M.J. and Delogu, G., 2002. The PE multigene family: a 'molecular mantra' for mycobacteria. *Trends in Microbiology*, 10(5), pp.246–249. Available at: https://doi.org/10.1016/S0966-842X(02)02335-1.

Brennan, P.J. and Nikaido, H., 1995. The envelope of Mycobacteria. *Annual Review of Biochemistry*, 64(1), pp.29–63. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.bi.64.070195.000333.

Brightbill, H.D. et al., 1999. Host defense mechanisms triggered by microbial lipoproteins through Toll-Like Receptors. *Science*, 285(5428), pp.732–736. Available at: http://www.sciencemag.org/content/285/5428/732.abstract.

Britten, R.L. and Davidson, E.H., 1985. Gene regulation for higher cells: a theory. *Science*, 165(3891), pp.349–357.

Britten, R.L. and Davidson, E.H., 1986. *Nucleic acid hybridization: a*

practical approach IRL Press. Higgins, S.J. and Hames, B.D., (eds.), Wiley, Oxford, Washington, DC.

Brosch, R. et al., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *PNAS*, 99(6), pp.3684–3689. Available at: http://10.0.4.49/pnas.052548299.

Brown, J.C.S. et al., 2014. Unraveling the biology of a fungal meningitis pathogen using chemical genetics. *Cell*, 159(5), pp.1168–1187. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4243055/.

Brown, T., Nikolayevskyy, V., Velji, P. and Drobniewski, F., 2010. Associations between Mycobacterium tuberculosis strains and phenotypes. *Emerging Infectious Disease*, 16(2), pp.272–280. Available at: http://dx.doi.org/10.3201/eid1602.091032.

Bryant, J.M. et al., 2013. Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. *BMC Infectious Diseases*, 13(1), pp.110–110. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23446317.

Byrd, T.F., Green, G.M., Fowlston, S.E. and Lyons, C.R., 1998. Differential growth characteristics and Streptomycin susceptibility of virulent and avirulent Mycobacterium tuberculosis strains in a novel fibroblast-mycobacterium microcolony assay. *Infection and Immunity*, 66(11), pp.5132–5139. Available at: https://iai.asm.org/content/66/11/5132.long.

Cabrera-Munoz, E., Hernandez-Hernandez, O.T. and Camacho-Arroyo, I., 2012. Role of estradiol and progesterone in HIV susceptibility and disease progression. In: *Mini Reviews in Medicinal Chemistry*. Bentham Science Publishers, United Arab Emirates, p. 1049–1054(6).

Campo, M. et al., 2015. Common polymorphisms in the CD43 gene region are associated with tuberculosis disease and mortality. *American Journal of Respiratory Cell and Molecular Biology*, 52(3), pp.342–348. Available at: https://10.0.4.141/rcmb.2014-0114OC.

Capriotti, E., Fariselli, P. and Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, 33(Web Server issue), pp.W306–W310. Available at: https://doi.org/10.1093/nar/gki375.

Casadevall, A. and Pirofski, L.A., 2001. Host-pathogen interactions: the attributes of virulence. *The Journal of Infectious Diseases*, 184(3), pp.337–344. Available at: https://doi.org/10.1086/322044.

Casjens, S., 1998. The diverse and dynamic structure of bacterial genomes. *Annual Review of Genetics*, 32(1), pp.339–377. Available at: https://doi.org/10.1146/annurev.genet.32.1.339.

Castro-Garza, J., King, C.H., Swords, W.E. and Quinn, F.D., 2002. Demonstration of spread by Mycobacterium tuberculosis bacilli in A549 epithelial cell monolayers. *FEMS Microbiology Letters*, 212(2), pp.145–149. Available at: https://doi.org/10.1111/j.1574-6968.2002.tb11258.x.

Caws, M. et al., 2006. Beijing genotype of Mycobacterium tuberculosis is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. *Journal of Clinical Microbiology*, 44(11), pp.3934–3939. Available at: https://jcm.asm.org/content/44/11/3934.long.

Caws, M. et al., 2008. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS Pathogens*, 4(3: e1000034), pp.1–9. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2268004/.

CDC, 2016, *Basic TB Facts* [Online]. Available at: www.cdc.gov/TB/topic/basics/default.htm [Accessed: 16 December 2018].

Chakraborti, S. et al., 2009. Clinicopathological study of tuberculous brain abscess. *Pathology, research and practice*, 205(12), pp.815–822. Available at: https://doi.org/10.1016/j.prp.2009.05.012.

Chan, J. et al., 1991. Lipoarabinomannan, a possible virulence factor involved in persistence of Mycobacterium tuberculosis within macrophages. *Infection and Immunity*, 59(5), pp.1755–1761. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC257912/.

Chan, W.T., Verma, C.S., Lane, D.P. and Gan, S.K.E., 2013. A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. *Bioscience Reports*, 33(6: e00086), pp.931–944. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3860579/.

Chaoui, I. et al., 2014. Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent Mycobacterium tuberculosis genotypes infecting tuberculosis patients in Morocco. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 21(January), pp.463–471. Available at: https://doi.org/10.1016/j.meegid.2013.05.023.

Chatterjee, A. and Mistry, N., 2013. MIRU–VNTR profiles of three major Mycobacterium tuberculosis spoligotypes found in western India. *Tuberculosis*, 93(2), pp.250–256. Available at: https://doi.org/10.1016/j.tube.2012.10.004.

Chatterjee, S., 2011. Brain tuberculomas, tubercular meningitis, and post-tubercular hydrocephalus in children. *Journal of Pediatric Neurosciences*, 6(Suppl. 1), pp.S96–S100. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3208909/?report=reader.

Chen, H. et al., 2017. Mycobacterium tuberculosis lineage distribution in Xinjiang and Gansu Provinces, China. *Scientific Reports*, 7(1), pp.1068–1090. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5430859/.

Chitale, S. et al., 2001. Recombinant Mycobacterium tuberculosis protein associated with mammalian cell entry. *Cellular Microbiology*, 3(4), pp.247–254. Available at: https://doi.org/10.1046/j.1462-5822.2001.00110.x.

Chomczynski, P. and Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162(1), pp.156–159. Available at: https://doi.org/10.1016/0003-2697(87)90021-2.

Christianson, S. et al., 2010. Evaluation of 24 locus MIRU-VNTR genotyping of Mycobacterium tuberculosis isolates in Canada. *Tuberculosis*, 90(1), pp.31–38. Available at: https://doi.org/10.1016/j.tube.2009.12.003.

Cingolani, P. et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w 1118; iso-2; iso-3. *Fly*.

Clontech Laboratories, I., 2008. *PCR-Select Bacterial Genome Subtraction Kit User Manual* Company, A.T.B., (ed.), Clontech Laboratories, Inc, U.S.A.

Clontech Laboratories, I., 2011. PCR-Select cDNA Subtraction Kit User Manual Company, A.T.B., (ed.). , pp.1–36. Available at: www.clontech.com.

Cohen, S.N., Chang, A.C. and Hsu, L., 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. *PNAS*, 69(8), pp.2110–2114. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC426879/.

Cole, S.T., 1998. Comparative mycobacterial genomics. *Current Opinion in Microbiology*, 1(5), pp.567–571. Available at: https://doi.org/10.1016/S1369-5274(98)80090-8.

Cole, S.T. et al., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), pp.537–544. Available at: http://dx.doi.org/10.1038/31159.

Coll, F. et al., 2014. PolyTB: A genomic variation map for Mycobacterium tuberculosis. *Tuberculosis (Edinburgh, Scotland)*, 94(3), pp.346–354. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4066953/.

Collins, D.M., 1996. In search of tuberculosis virulence genes. *Trends in Microbiology*, 4(11), pp.426–430. Available at: https://doi.org/10.1016/0966-842X(96)10066-4.

Comas, I. et al., 2013. Out-of-Africa migration and Neolithic co-expansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics*, 45(10),

pp.1176–1182. Available at: http://dx.doi.org/10.1038/ng.2744.

Comas, I., Homolka, S., Niemann, S. and Gagneux, S., 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in mycobacterium tuberculosis highlights the limitations of current methodologies. *PLoS ONE*, 4(11: e7815), pp.1–11. Available at: https://doi.org/10.1371/journal.pone.0007815.

Converse, S.E. et al., 2003. MmpL8 is required for sulfolipid-1 biosynthesis and Mycobacterium tuberculosis virulence. *PNAS*, 100(10), pp.6121–6126. Available at: https://doi.org/10.1073/pnas.1030024100.

Cooper, A.M. et al., 1993. Disseminated tuberculosis in interferon gamma gene-disrupted mice. *The Journal of Experimental Medicine*, 178(6), pp.2243–2247. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2191280/.

Cooper, A.M., Mayer-Barber, K.D. and Sher, A., 2011. Role of innate cytokines in mycobacterial infection. *Mucosal Immunology*, 4(3), pp.252–260. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3294290/pdf/nihms-360359.pdf.

Coscolla, M. and Gagneux, S., 2010. Does M. tuberculosis genomic diversity explain disease diversity? *Drug Discovery Today. Disease mechanisms*, 7(1), pp.e43–e59. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2976975/.

Couce, A., Rodríguez-Rojas, A. and Blázquez, J., 2016. Determinants of genetic diversity of spontaneous drug resistance in bacteria. *Genetics*, 203(3), pp.1369–1380. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937460/pdf/1369.pdf.

Coureuil, M. et al., 2012. Mechanism of meningeal invasion by Neisseria meningitidis. *Virulence*, 3(2), pp.164–172. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3396695/.

Cowley, S. et al., 2004. The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Molecular Microbiology*, 52(6), pp.1691–1702.

Daffe, M., Brennan, P.J. and McNeil, M., 1990. Predominant structural features of the cell wall arabinogalactan of Mycobacterium tuberculosis as revealed through characterization of oligoglycosyl alditol fragments by gas chromatography/mass spectrometry and by 1H and 13C NMR analyses. *Journal of Biological Chemistry*, 265(12), pp.6734–6743. Available at: http://www.jbc.org/content/265/12/6734.abstract.

Dalton, D.K. et al., 1993. Multiple defects of immune cell function in mice with disrupted interferon-gamma genes. *Science*, 259(5102), pp.1739–1742.

Available at: http://science.sciencemag.org/content/259/5102/1739.

Dannenberg, A.M.J. and Rook, J.A., 1994. Pathogenesis of pulmonary tuberculosis: an interplay between tissue-damaging and macrophage-activating immune responses: dual mechanisms that control bacillary multiplication. In: Bloom, B.R., (ed.) *Tuberculosis: pathogenesis, protection, and control*. American Society for Microbiology (ASM), Washington, D. C., pp. 459–483.

Dark, M.J. et al., 2009. Conservation in the face of diversity: multistrain analysis of an intracellular bacterium. *BMC Genomics*, 10(16), pp.1–12. Available at: http://www.biomedcentral.com/1471-2164/10/16.

Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 14(7), pp.1394–1403. Available at: http://dx.doi.org/10.1101/gr.2289704.

Darwin, K.H. and Nathan, C.F., 2005. Role for nucleotide excision repair in virulence of Mycobacterium tuberculosis. *Infection and Immunity*, 73(8), pp.4581–4587. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1201236/.

Das, S. et al., 2013. Genetic heterogeneity revealed by sequence analysis of Mycobacterium tuberculosis isolates from extra-pulmonary tuberculosis patients. *BMC Genomics*, 14(404), pp.1–14. Available at: http://www.biomedcentral.com/1471-2164/14/404.

Davidsen, T. et al., 2007. Genetic interactions of DNA repair pathways in the pathogen Neisseria meningitidis. *Journal of Bacteriology*, 189(15), pp.5728–5737. Available at: http://jb.asm.org/content/189/15/5728.abstract.

Davis, A.G. et al., 2019. The pathogenesis of tuberculous meningitis. *Journal of Leukocyte Biology*, 105(2), pp.267–280. Available at: https://doi.org/10.1002/JLB.MR0318-102R.

Dennis Jr, G. et al., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(9), p.R60.1-R60.11. Available at: http://genomebiology.com/2003/4/9/R60.

Desai, D. et al., 2006. Utility of the polymerase chain reaction in the diagnosis of tuberculous meningitis. *Research in Microbiology*, 157(10), pp.967–970. Available at: https://doi.org/10.1016/j.resmic.2006.08.002.

Deshazer, D., 2004. Genomic diversity of Burkholderia pseudomallei clinical isolates: subtractive hybridization reveals a Burkholderia mallei-specific prophage in B. pseudomallei 1026b. *Journal of Bacteriology*, 186(12), pp.3938–3950. Available at: https://jb.asm.org/content/186/12/3938.long.

DeShazer, D., Waag, D.M., Fritz, D.L. and Woods, D.E., 2001. Identification of a Burkholderia mallei polysaccharide gene cluster by subtractive

hybridization and demonstration that the encoded capsule is an essential virulence determinant. *Microbial Pathogenesis*, 30(5), pp.253–269. Available at: https://doi.org/10.1006/mpat.2000.0430.

Diatchenko, L. et al., 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *PNAS*, 93(12), pp.6025–6030. Available at: https://doi.org/10.1073/pnas.93.12.6025.

Dietrich, J. and Doherty, T.M., 2009. Interaction of Mycobacterium tuberculosis with the host: consequences for vaccine development. *Acta Pathologica, Microbiologica et Immunologica Scandinavica*, 117(5–6), pp.440–457. Available at: https://doi.org/10.1111/j.1600-0463.2009.02458.x.

Djelouagji, Z. and Drancourt, M., 2006. Inactivation of cultured Mycobacterium tuberculosis organisms prior to DNA extraction. *Journal of Clinical Microbiology*, 44(4), pp.1594–1595. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1448686/.

Dobos, K.M., Spotts, E.A., Quinn, F.D. and King, C.H., 2000. Necrosis of lung epithelial cells during infection with Mycobacterium tuberculosis is preceded by cell permeation. *Infection and immunity*, 68(11), pp.6300–6310. Available at: https://iai.asm.org/content/68/11/6300.long.

Doig, C., Seagar, A.L., Watt, B. and Forbes, K.J., 2002. The efficacy of the heat killing of Mycobacterium tuberculosis. *Journal of Clinical Pathology*, 55(10), pp.778–779. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1769777/.

Domenech, P. et al., 2004. The role of MmpL8 in sulfatide biogenesis and virulence of Mycobacterium tuberculosis. *Journal of Biological Chemistry*, 279(20), pp.21257–21265. Available at: http://www.jbc.org/content/279/20/21257.long.

Donald, P.R., Schaaf, H.S. and Schoeman, J.F., 2005. Tuberculous meningitis and miliary tuberculosis: the Rich focus revisited. *Journal of Infection*, 50(3), pp.193–195. Available at: http://www.sciencedirect.com/science/article/pii/S0163445304000489.

Dormans, J. et al., 2004. Correlation of virulence, lung pathology, bacterial load and delayed type hypersensitivity responses after infection with different Mycobacterium tuberculosis genotypes in a BALB/c mouse model. *Clinical and Experimental Immunology*, 137(3), pp.460–468. Available at: https://doi.org/10.1111/j.1365-2249.2004.02551.x.

Dou, H.Y. et al., 2016. Genomics study of Mycobacterium tuberculosis strains from different ethnic populations in Taiwan. *Evolutionary Bioinformatics Online*, 12(3), pp.213–221. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5040422/.

Dubchak, I. and Pachter, L., 2002. The computational challenges of applying comparative-based computational methods to whole genomes. *Briefings in Bioinformatics*, 3(1), pp.18–22. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12002220.

Dunfee, R. et al., 2006. Mechanisms of HIV-1 neurotropism. *Current HIV Research*, 4(3), pp.267–278. Available at: http://www.eurekaselect.com/57601/article.

Dunker, A.K. et al., 2001. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1), pp.26–59. Available at: http://www.sciencedirect.com/science/article/pii/S1093326300001388.

Dunn, P.L. and North, R.J., 1995. Virulence ranking of some Mycobacterium tuberculosis and Mycobacterium bovis strains according to their ability to multiply in the lungs, induce lung pathology, and cause mortality in mice. *Infection and Immunity*, 63(9), pp.3428–3437. Available at: https://iai.asm.org/content/63/9/3428.long.

Dworkin, S., 2007, *QIAquick Gel Extraction (and PCR Purification) Kit From Qiagen* [Online]. Available at: https://www.biocompare.com/Product-Reviews/40845-QIAquick-Gel-Extraction-and-PCR-Purification-Kit-From-Qiagen/ [Accessed: 15 February 2018].

Ellegren, H., 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29(1), pp.51–63. Available at: https://doi.org/10.1016/j.tree.2013.09.008.

Elmas, Ö.N., Akıncı, A. and Bilir, P., 2011. Tuberculous meningitis associated with diabetic ketoacidosis. *Journal of Clinical Research in Pediatric Endocrinology*, 3(4), pp.222–224. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245499/.

van Embden, J.D. et al., 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *Journal of Clinical Microbiology*, 31(2), pp.406–409. Available at: http://jcm.asm.org/cgi/content/abstract/31/2/406.

Emmerth, M., Goebel, W., Miller, S.I. and Hueck, C.J., 1999. Genomic subtraction identifies Salmonella typhimurium prophages, F- related plasmid sequences, and a novel fimbrial operon, stf, which are absent in Salmonella typhi. *Journal of Bacteriology*, 181(18), pp.5652–5661. Available at: https://jb.asm.org/content/181/18/5652.long.

Ermolaeva, O.D., Lukyanov, S.A. and Sverdlov, E.D., 1996. The mathematical model of subtractive hybridization and its practical application. *Proceedings International Conference on Intelligent Systems for Molecular Biology*, 4(ISMB-96), pp.52–58. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8877504.

Ermolaeva, O.D. and Wagner, M.C., 1995. SUBTRACT: a computer program for modeling the process of subtractive hydridization. *Computer Applications in the Biosciences*, 11(4), pp.457–462. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8521056.

Eugenin, E.A. et al., 2006. CCL2/monocyte chemoattractant protein-1 mediates enhanced transmigration of Human Immunodeficiency Virus (HIV)-infected leukocytes across the blood-brain barrier: a potential mechanism of HIV-CNS invasion and neuroAIDS. *The Journal of Neuroscience*, 26(4), pp.1098–1106. Available at: https://doi.org/10.1523/JNEUROSCI.3863-05.2006.

Faksri, K. et al., 2018. Comparative whole- genome sequence analysis of Mycobacterium tuberculosis isolated from tuberculous meningitis and pulmonary tuberculosis patients. *Scientific Reports*, 8(4910), pp.1–10. Available at: www.nature.com/scientificreports.

Fandinho, F.C. et al., 2000. RFLP patterns and risk factors for recent tuberculosis transmission among hospitalized tuberculosis patients in Rio de Janeiro, Brazil. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 94(3), pp.271–275. Available at: https://www.ncbi.nlm.nih.gov/pubmed/10974996.

Fanning, A., 1999. Tuberculosis: 6. Extrapulmonary disease. *Canadian Medical Association Journal*, 160(11), pp.1597–1603. Available at: http://www.cmaj.ca/content/cmaj/160/11/1597.full.pdf.

Farooqui, A.A., Horrocks, L.A. and Farooqui, T., 2000. Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders. *Chemistry and Physics of Lipids*, 106(1), pp.1–29. Available at: https://doi.org/10.1016/S0009-3084(00)00128-6.

Feizabadi, M.M., Robertson, I.D., Cousins, D. V and Hampson, D.J., 1996. Genomic analysis of Mycobacterium bovis and other members of the Mycobacterium tuberculosis complex by isoenzyme analysis and pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, 34(5), pp.1136–1142. Available at: https://jcm.asm.org/content/jcm/34/5/1136.full.pdf.

Filliol, I. et al., 2006. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *Journal of Bacteriology*, 188(2), pp.759–772. Available at: http://dx.doi.org/10.1128/JB.188.2.759-772.2006.

Fleischmann, R.D. et al., 2002. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *Journal of Bacteriology*, 184(19), pp.5479–5490. Available at: https://jb.asm.org/content/jb/184/19/5479.full.pdf.

Flores, L. et al., 2010. Comparison of Restriction Fragment Length Polymorphism with the Polymorphic Guanine-Cytosine-Rich sequence and spoligotyping for differentiation of Mycobacterium tuberculosis isolates with five or fewer copies of IS6110. *Journal of Clinical Microbiology*, 48(2), pp.575–578. Available at: http://jcm.asm.org/content/48/2/575.abstract.

Flynn, J.L., Chan, J. and Lin, P.L., 2011. Macrophages and control of granulomatous inflammation in tuberculosis. *Mucosal Immunology*, 4(3), pp.271–278. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3311958/pdf/nihms351168.pdf.

Ford, C. et al., 2012. Mycobacterium tuberculosis - heterogeneity revealed through whole genome sequencing. *Tuberculosis*, 92(3), pp.194–201. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3323677/pdf/nihms-345909.pdf.

Forrellad, M.A. et al., 2013. Virulence factors of the Mycobacterium tuberculosis complex. *Virulence*, 4(1), pp.3–66. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3544749/pdf/viru-4-3.pdf.

Fu, C.L., Chuang, Y.H., Huang, H.Y. and Chiang, B.L., 2008. Induction of IL-10 producing CD4+ T cells with regulatory activities by stimulation with IL-10 gene-modified bone marrow derived dendritic cells. *Clinical and Experimental Immunology*, 153(2), pp.258–268. Available at: https://doi.org/10.1111/j.1365-2249.2008.03689.x.

Gagneux, S. et al., 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proceeding National of Academic Science USA*, 103(8), pp.2869–2873. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1413851/.

Gagneux, S. and Small, P.M., 2007. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infectious Diseases*, 7(5), pp.328–337. Available at: https://doi.org/10.1016/S1473-3099(07)70108-1.

Gandhi, N.R. et al., 2006. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *The Lancet*, 368(9547), pp.1575–1580. Available at: https://doi.org/10.1016/S0140-6736(06)69573-1.

Ganguly, N. et al., 2008. Mycobacterium tuberculosis secretory proteins CFP-10, ESAT-6 and the CFP10: ESAT6 complex inhibit lipopolysaccharide-induced NF-jB transactivation by downregulation of reactive oxidative species (ROS) production. *Immunology and Cell Biology*, 86(1), pp.98–106. Available at: https://doi.org/10.1038/sj.icb.7100117.

Gao, L.Y. et al., 2004. A mycobacterial virulence gene cluster extending RD1

is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Molecular Microbiology*, 53(6), pp.1677–1693. Available at: https://doi.org/10.1111/j.1365-2958.2004.04261.x.

García-Gómez, E., González-Pedrajo, B. and Camacho-Arroyo, I., 2013. Role of sex steroid hormones in bacterial-host interactions. *Biomed Research International*, 2013(Article ID 928290), pp.1–10. Available at: http://dx.doi.org/10.1155/2013/928290.

García-Moncó, J.C. and Rodriguez-Sainz, A., 2018. CNS tuberculosis and other mycobacterial infections. In: García-Moncó J., (ed.) *CNS Infections*. Springer, Cham, Switzerland, pp. 157–179.

Gardner, S.N. et al., 2005. Draft versus finished sequence data for DNA and protein diagnostic signature development. *Nucleic Acids Research*, 33(18), pp.5838–5850. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1266063/.

Gardy, J.L. et al., 2011. Whole genome sequencing and social network analysis of a tuberculosis outbreak. *The New England Journal of Medicine*, 364(10), pp.730–739. Available at: https://www.nejm.org/doi/full/10.1056/NEJMoa1003176.

Garg, R.K., Malhotra, H.S. and Jain, A., 2016. Neuroimaging in tuberculous meningitis. *Neurology India*, 64(2), pp.219–227.

Garner, E. et al., 1999. Predicting binding regions within disordered proteins. *Genome Informatics. Workshop on Genome Informatics*, 10, pp.41–50. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11072341.

Geijtenbeek, T.B.H. et al., 2003. Mycobacteria target DC-SIGN to suppress dendritic cell function. *The Journal of Experimental Medicine*, 197(1), pp.7–17. Available at: http://www.jem.org/cgi/doi/10.1084/jem.20021229.

Geneaid, 2017, *Lysozyme* [Online]. Available at: http://www.geneaid.com/products/enzymes/lysozyme [Accessed: 17 November 2017].

Giacomini, E. et al., 2001. Infection of human macrophages and dendritic cells with Mycobacterium tuberculosis induces a differential cytokine gene expression that modulates T cell response. *The Journal of Immunology*, 166(12), pp.7033–7041. Available at: https://doi.org/10.4049/jimmunol.166.12.7033.

Glickman, M.S. and Jacobs, W.R., 2001. Microbial pathogenesis of Mycobacterium tuberculosis: dawn of a discipline. *Cell*, 104(4), pp.477–485. Available at: https://doi.org/10.1016/S0092-8674(01)00236-7.

Golden, M.P. and Vikram, H.R., 2005. Extrapulmonary tuberculosis: an overview. *American Family Physician*, 72(9), pp.1761–1768. Available at:

https://www.aafp.org/afp/2005/1101/p1761.pdf.

González-Cano, P. et al., 2010. Mycobacterium tuberculosis H37Rv induces ectosome release in human polymorphonuclear neutrophils. *Tuberculosis*, 90(2), pp.125–134. Available at: https://www.sciencedirect.com/science/article/pii/S147297921000003X [Accessed: 30 November 2018].

Gordon, S. V et al., 2001. Genomics of Mycobacterium bovis. *Tuberculosis (Edinburg)*, 81(1–2), pp.157–163. Available at: https://doi.org/10.1054/tube.2000.0269.

Gowda, M., 2013, *Sample submission requirements* [Online]. Available at: http://www.ccamp.res.in/sites/default/files/Sample Submission Requirements - Illumina - 300813.pdf.

Green, E. et al., 2013. IS6110 Restriction Fragment Length Polymorphism typing of drug-resistant Mycobacterium tuberculosis strains from Northeast South Africa. *Journal of Health, Population and Nutrition*, 31(1), pp.1–10. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3702353/.

Greninger, A.L. et al., 2015. Two Rapidly Growing Mycobacterial Species Isolated from a Brain Abscess: First Whole-Genome Sequences of Mycobacterium immunogenum and Mycobacterium llatzerense Carroll, K.C., (ed.). *Journal of Clinical Microbiology*, 53(7), pp.2374–2377. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4473206/.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D. and van Embden, J.D., 1993. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Molecular microbiology*, 10(5), pp.1057–1065. Available at: https://doi.org/10.1111/j.1365-2958.1993.tb00976.x.

Guerra-Assunção, J.A. et al., 2015. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. *eLife*, 4(e05166), pp.1–17. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25732036.

Hacker, J., Blum-Oehler, G., Mühldorfer, I. and Tschäpe, H., 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular Microbiology*, 23(6), pp.1089–1097. Available at: https://doi.org/10.1046/j.1365-2958.1997.3101672.x.

Hagedorn, M., Rohde, K.H., Russell, D.G. and Soldati, T., 2009. Infection by tubercular mycobacteria is spread by nonlytic ejection from their amoeba hosts. *Science*, 323(5922), pp.1729–1733. Available at: http://science.sciencemag.org/content/323/5922/1729/tab-pdf.

Haile, Y., Caugant, D.A., Bjune, G. and Wiker, H.G., 2002. Mycobacterium tuberculosis mammalian cell entry operon (mce) homologs in Mycobacterium

other than tuberculosis (MOTT). *FEMS Immunology & Medical Microbiology*, 33(2), pp.125–132. Available at: https://doi.org/10.1111/j.1574-695X.2002.tb00581.x.

Haldar, S. et al., 2012. Detection of Mycobacterium tuberculosis GlcB or HspX Antigens or devR DNA impacts the rapid diagnosis of tuberculous meningitis in children Pai, M., (ed.). *PLoS ONE*, 7(9: e44630), pp.1–9. Available at: https://doi.org/10.1371/journal.pone.0044630.

Hanahan, D., 1985. *DNA cloning: a practical approach* 2nd ed. Glover, D.M. and Hames, B.D., (eds.), Oxford University Press, Washington DC.

Hanekom, M. et al., 2008. Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of Mycobacterium tuberculosis Beijing strains in a setting of high incidence of. *Journal of Cinical Microbiology*, 46(10), pp.3338–3345. Available at: https://jcm.asm.org/content/jcm/46/10/3338.full.pdf.

Hao, W. et al., 2011. Extensive genomic variation within clonal complexes of Neisseria meningitidis. *Genome Biology and Evolution*, 3(January), pp.1406–1418. Available at: https://doi.org/10.1093/gbe/evr119.

Hawn, T.R. et al., 2006. A Polymorphism in Toll-Interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. *Journal of Infectious Diseases*, 194(8), pp.1127–1134. Available at: http://jid.oxfordjournals.org/content/194/8/1127.abstract.

Henderson, R.A., Watkins, S.C. and Flynn, J.L., 1997. Activation of human dendritic cells following infection with Mycobacterium tuberculosis. *Journal of Immunology*, 159(2), pp.635–643. Available at: http://www.jimmunol.org/content/159/2/635.

Herman, P. et al., 2006. Biosafety recommendations for the contained use of Mycobacterium tuberculosis complex isolates in industrialized countries. *Mycobacterium tuberculosis and Biosafety*, (April), pp.1–17. Available at: https://www.researchgate.net/publication/228349402.

Hermans, P.W. et al., 1991. Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infection and Immunity*, 59(8), pp.2695–2705. Available at: https://iai.asm.org/content/iai/59/8/2695.full.pdf.

Hernandez-Pando, R. et al., 2010. Specific bacterial genotypes of Mycobacterium tuberculosis cause extensive dissemination and brain infection in an experimental model. *Tuberculosis*, 90(4), pp.268–277. Available at: https://doi.org/10.1016/j.tube.2010.05.002.

Hershberg, R. et al., 2008. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS bBology*,

6(12: e311), pp.2658–2671. Available at: https://doi.org/10.1371/journal.pbio.0060311.

Hesseling, A.C. et al., 2010. Mycobacterial genotype is associated with disease phenotype in children. *The International Journal of Tuberculosis and Lung Disease*, 14(10), pp.1252–1258. Available at: https://www.ingentaconnect.com/content/iuatld/ijtld/2010/00000014/00000010/art00006%3Bjsessionid=g8dro51rqf76.x-ic-live-01#.

Hicks, M. et al., 2019. Functional characterization of 3D protein structures informed by human genetic diversity. *Proceedings of the National Academy of Sciences*, 116(18), p.201820813. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1820813116.

Hoal-Van Helden, E.G. et al., 1999. Mannose-binding protein B allele confers protection against tuberculous meningitis. *Pediatric Research*, 45(4 Pt 1), pp.459–464. Available at: https://www.nature.com/articles/pr199993.pdf.

Houben, E.N.G. et al., 2009. Differential expression of a virulence factor in pathogenic and non-pathogenic mycobacteria. *Molecular Microbiology*, 72(1), pp.41–52. Available at: https://doi.org/10.1111/j.1365-2958.2009.06612.x.

Houben, E.N.G., Korotkov, K. V and Bitter, W., 2014. Take five - type VII secretion systems of Mycobacteria. *Biochimica et Biophysica Acta*, 1843(8), pp.1707–1716. Available at: https://doi.org/10.1016/j.bbamcr.2013.11.003.

Houben, E.N.G., Nguyen, L. and Pieters, J., 2006. Interaction of pathogenic mycobacteria with the host immune system. *Current Opinion in Microbiology*, 9(1), pp.76–85. Available at: https://doi.org/10.1016/j.mib.2005.12.014.

Hsu, T. et al., 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *PNAS*, 100(21), pp.12420–12425. Available at: https://doi.org/10.1073/pnas.1635213100.

Huang, C., Li, R.Q. and Zhang, W.J., 2009. Screening of Mycobacterium tuberculosis distinctive genes by suppression subtractive hybridization technique. *Chinese Journal of Microbiology and Immunology*, 29(6), pp.507–512. Available at: https://www.researchgate.net/publication/288092193_Screening_of_Mycobacterium_tuberculosis_distinctive_genes_by_suppression_subtractive_hybridization_technique.

Huang, H.L., 2014. Propensity scores for prediction and characterization of bioluminescent proteins from sequences. *PLoS ONE*, 9(5: e97158), pp.1–15. Available at: https://doi.org/10.1371/journal.pone.0097158.

Huang, S.H. et al., 2001. Further characterization of Escherichia coli brain microvascular endothelial cell invasion gene ibeA by deletion, complementation, and protein expression. *Journal of Infectious Disease*,

183(7), pp.1071–1078. Available at: https://doi.org/10.1086/319290.

Hughes, D.T. and Sperandio, V., 2008. Inter-kingdom signalling: communication between bacteria and their hosts. *Nature Reviews Microbiology*, 6, pp.111–120. Available at: https://doi.org/10.1038/nrmicro1836.

Humphreys, I.R. et al., 2006. A role for dendritic cells in the dissemination of mycobacterial infection. *Microbes and Infection*, 8(5), pp.1339–1346. Available at: https://doi.org/10.1016/j.micinf.2005.12.023.

Husain, A. et al., 2017. Study on strain dependent pathogenesis of tuberculosis. *Unpublished*. Available at: https://www.ncbi.nlm.nih.gov/genome/annotation_prok/.

Huynen, M.A. and Bork, P., 1998. Measuring genome evolution. *Proceedings of the National Academy of Sciences*, 95(11), p.5849 LP-5856. Available at: http://www.pnas.org/content/95/11/5849.abstract.

Iantomasi, R. et al., 2011. PE_PGRS30 is required for the full virulence of Mycobacterium tuberculosis. *Cellular Microbiology*, 14(3), pp.356–367. Available at: https://doi.org/10.1111/j.1462-5822.2011.01721.x.

Ilina, E.N. et al., 2013. Comparative genomic analysis of Mycobacterium tuberculosis drug resistant strains from Russia. *PLoS ONE*, 8(2: e56577), pp.1–12. Available at: https://doi.org/10.1371/journal.pone.0056577.

Illumina, I., 2019, *A wide breadth of sequencing applications MiSeq applications include targeted gene, small genome, and amplicon sequencing, 16S metagenomics, and more* [Online]. Available at: https://sapac.illumina.com/systems/sequencing-platforms/miseq/applications.html?langsel=/my/ [Accessed: 2 June 2019].

Illumina, I., 2014, *Nextera Library Prep for the MiSeq System* [Online]. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote_miseq_nextera.pdf [Accessed: 27 July 2014].

Illumina, I., 2016, *Nextera Library Prep Reference Guide* [Online]. Available at: http://support.illumina.com/downloads/nextera_xt_sample_preparation_guide_15031942.html [Accessed: 27 July 2016].

ISOGG Wiki, 2017, *Micro-allele* [Online]. Available at: https://isogg.org/wiki/Micro-allele [Accessed: 2 June 2019].

Iwamoto, T. et al., 2007. Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on Mycobacterium tuberculosis strains predominated by the Beijing family. *FEMS Microbiology Letters*, 270(1), pp.67–74. Available at:

https://doi.org/10.1111/j.1574-6968.2007.00658.x.

Jagielski, T. et al., 2014. Current methods in the molecular typing of Mycobacterium tuberculosis and other mycobacteria. *BioMed Research International*, 2014(645802), pp.1–21. Available at: http://dx.doi.org/10.1155/2014/645802.

Jain, A., Tiwari, V., Guleria, R.S. and Verma, R.K., 2002. Qualitative evaluation of mycobacterial DNA extraction protocols for polymerase chain reaction. *Molecular Biology Today*, 3, pp.43–50. Available at: http://www.horizonpress.com/mbt/v/v3/06.pdf%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3249990&tool=pmcentrez&rendertype=abstract.

Jain, S.K. et al., 2006. Mycobacterium tuberculosis invasion and traversal across an in vitro human blood-brain barrier as a pathogenic mechanism for central nervous system tuberculosis. *The Journal of Infectious Diseases*, 193(9), pp.1287–1295. Available at: https://doi.org/10.1086/502631.

Jenkin, G.A., Stinear, T.P., Johnson, P.D.R. and Davies, J.K., 2003. Subtractive hybridization reveals a type I polyketide synthase locus specific to Mycobacterium ulcerans. *Journal of Bacteriology*, 185(23), pp.6870–6882. Available at: https://jb.asm.org/content/jb/185/23/6870.full.pdf.

Jeren, T. and Beus, I., 1982. Characteristics of cerebrospinal fluid in tuberculous meningitis. *Acta cytologica*, 26(5), pp.678–680. Available at: https://www.ncbi.nlm.nih.gov/pubmed/6959457?dopt=Abstract.

Jevtic, V., 2004. Vertebral infection. *European Radiology Supplements*, 14(3), pp.E43--E52. Available at: https://doi.org/10.1007/s00330-003-2046-x.

Ji, W. et al., 2002. Efficacy of SSH PCR in isolating differentially expressed genes. *BMC Genomics*, 3(12), pp.1–7. Available at: https://doi.org/10.1186/1471-2164-3-12.

Jonsson, J. et al., 2014. Comparison between RFLP and MIRU-VNTR genotyping of Mycobacterium tuberculosis strains isolated in Stockholm 2009 to 2011. *PLoS ONE*, 9(4: e95159), pp.1–6. Available at: https://doi.org/10.1371/journal.pone.0095159.

Jonuleit, H. et al., 2000. Induction of Interleukin 10–producing, Nonproliferating CD4+ T Cells with Regulatory Properties by Repetitive Stimulation with Allogeneic Immature Human Dendritic Cells. *Journal of Experimental Medicine*, 192(9), pp.1213–1222. Available at: http://www.jem.org/cgi/content/full/192/9/1213.

Joshi, D. et al., 2012. Single nucleotide polymorphisms in the Mycobacterium bovis genome resolve phylogenetic relationships. *Journal of Clinical Microbiology*, 50(12), pp.3853–3861. Available at: https://jcm.asm.org/content/50/12/3853.long.

Kamerbeek, J. et al., 1997. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *Journal of Clinical Microbiology*, 35(4), pp.907–914. Available at: https://jcm.asm.org/content/jcm/35/4/907.full.pdf.

Kandhakumari, G., Stephen, S., Sivakumar, S. and Narayanan, S., 2015. Spoligotype patterns of Mycobacterium tuberculosis isolated from extra pulmonary tuberculosis patients in Puducherry, India. *Indian Journal of Medical Microbiology*, 33(2), pp.267–270. Available at: http://www.ijmm.org/text.asp?2015/33/2/267/154871.

Kang, H.Y. et al., 2010. Phylogeographical particularity of the Mycobacterium tuberculosis Beijing family in South Korea based on international comparison with surrounding countries. *Journal of Medical Microbiology*, 59(Pt 10), pp.1191–1197. Available at: https://www.ncbi.nlm.nih.gov/pubmed/20576748.

Karboul, A. et al., 2008. Frequent homologous recombination events in Mycobacterium tuberculosis PE/PPE multigene families: Potential role in antigenic variability. *Journal of Bacteriology*, 190(23), pp.7838–7846. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2583619/pdf/0827-08.pdf.

Kato-Maeda, M. et al., 2001. Comparing genomes within the species Mycobacterium tuberculosis. *Genome Research*, 11(4), pp.547–554. Available at: https://genome.cshlp.org/content/11/4/547.long.

Kato-Maeda, M., Metcalfe, J.Z. and Flores, L., 2011. Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. *Future Microbiology*, 6(2), pp.203–216. Available at: https://doi.org/10.2217/fmb.10.165.

Kellogg, D.E. et al., 1994. TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *BioTechniques*, 16(6), pp.1134–1137. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8074881.

Kim, B.J. et al., 2017. Phylogenetic analysis of Mycobacterium massiliense strains having recombinant rpoB gene laterally transferred from Mycobacterium abscessus. *PLoS ONE*, 12(6: e0179237), pp.1–14. Available at: https://doi.org/10.1371/journal.pone.0179237.

Kim, K.S., 2006. Meningitis-associated Escherichia coli Donnenberg, M.S., (ed.). *EcoSal Plus*, 2(1), pp.269–286. Available at: http://www.asmscience.org/WEB-INF/jsp/journal/referenceworkarticle.jsp.

Kim, S.Y. et al., 2017. Distribution and clinical significance of Mycobacterium avium complex species isolated from respiratory specimens. *Diagnostic Microbiology and Infectious Disease*, 88(2), pp.125–137. Available at: https://doi.org/10.1016/j.diagmicrobio.2017.02.017.

Kinhikar, A.G. et al., 2010. Potential role for ESAT6 in dissemination of Mycobacterium tuberculosis via human lung epithelial cells. *Molecular Microbiology*, 75(1), pp.92–106. Available at: https://doi.org/10.1111/j.1365-2958.2009.06959.x.

Kirby, K.S., 1957. A new method for the isolation of deoxyribonucleic acids: evidence on the nature of bonds between deoxyribonucleic acid and protein. *Biochemical Journal*, 66(3), pp.495–504. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200047/.

Kirby, K.S., 1964. Isolation and fractionation of nucleic acids Davidson, J.N. and Cohn, W.E.B.T., (eds.). *Progress in Nucleic Acid Research and Molecular Biology*, 3(Supplement C), pp.1–31. Available at: https://doi.org/10.1016/S0079-6603(08)60737-0.

Kirschner, D.E., Young, D. and Flynn, J.L., 2010. Tuberculosis: global approaches to a global disease. *Current Opinion in Biotechnology*, 21(4), pp.524–531. Available at: https://doi.org/10.1016/j.copbio.2010.06.002.

Klein, S.L., 2000. The effects of hormones on sex differences in infection: from genes to behavior. *Neuroscience & Biobehavioral Reviews*, 24(6), pp.627–638. Available at: https://www.sciencedirect.com/science/article/pii/S0149763400000270?via%3Dihub [Accessed: 2 June 2019].

Kleinnijenhuis, J. et al., 2011. Innate immune recognition of Mycobacterium tuberculosis. *Clinical and Developmental Immunology*, 2011(405310), pp.1–12. Available at: http://dx.doi.org/10.1155/2011/405310.

Ko, M.S.H., 1990. An "equalized cDNA library" by the reassociation of short double-stranded cDNAs. *Nucleic Acids Research*, 18(19), pp.5705–5711. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC332303/pdf/nar00203-0102.pdf.

Koh, K.W., Soh, S.E. and Seah, G.T., 2009. Strong antibody responses to Mycobacterium tuberculosis PE-PGRS62 protein are associated with latent and active tuberculosis. *Infection and Immunity*, 77(8), pp.3337–3343. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2715692/pdf/1175-08.pdf.

Kohli, S. et al., 2012. Comparative genomic and proteomic analyses of PE/PPE multigene family of Mycobacterium tuberculosis H37Rv and H37Ra reveal novel and interesting differences with implications in virulence. *Nucleic Acids Research*, 40(15), pp.7113–7122. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3424577/pdf/gks465.pdf.

Kolstø, A.-B., 1997. Dynamic bacterial genome organization. *Molecular Microbiology*, 24(2), pp.241–248. Available at: https://doi.org/10.1046/j.1365-

2958.1997.3501715.x.

van Kooyk, Y. and Geijtenbeek, T.B., 2003. DC-SIGN: escape mechanism for pathogens. *Nature Reviews Immunology*, 3(9), pp.697–709. Available at: https://www.nature.com/articles/nri1182.

Kremer, K. et al., 1999. Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. *Journal of Clinical Microbiology*, 37(8), pp.2607–2618. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC85295/pdf/jm002607.pdf.

Krishnan, N. et al., 2011. Mycobacterium tuberculosis lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PLoS ONE*, 6(9: e23870), pp.1–16. Available at: https://doi.org/10.1371/journal.pone.0023870.

Krishnan, N., Robertson, B.D. and Thwaites, G., 2010. The mechanisms and consequences of the extra-pulmonary dissemination of Mycobacterium tuberculosis. *Tuberculosis*, 90(6), pp.361–366. Available at: https://doi.org/10.1016/j.tube.2010.08.005.

Krumsiek, J., Arnold, R. and Rattei, T., 2007. Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*, 23(8), pp.1026–1028. Available at: https://doi.org/10.1093/bioinformatics/btm039.

Kumar, A., Bose, M. and Brahmachari, V., 2003. Analysis of expression profile of mammalian cell entry (mce) operons of Mycobacterium tuberculosis. *Infection and Immunity*, 71(10), pp.6083–6087. Available at: https://iai.asm.org/content/iai/71/10/6083.full.pdf.

Lamkanfi, M. and Dixit, V.M., 2010. Manipulation of host cell death pathways during microbial infections. *Cell Host & Microbe*, 8(1), pp.44–54. Available at: https://doi.org/10.1016/j.chom.2010.06.007.

Lechner, M. et al., 2011. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, 12(1: 124), pp.1–13. Available at: http://www.biomedcentral.com/1471-2105/12/124.

Lee, H.G. et al., 2016. Tuberculous meningitis is a major cause of mortality and morbidity in adults with central nervous system infections in Kota Kinabalu, Sabah, Malaysia: an observational study. *BMC Infectious Diseases*, 16(296), pp.1–8. Available at: https://doi.org/10.1186/s12879-016-1640-x.

Lee, P.Y., Costumbrado, J., Hsu, C.Y. and Kim, Y.H., 2012. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments*, 62(e3923), pp.1–5. Available at: https://www.jove.com/video/3923.

Leonard, J.M. and Des Prez, R.M., 1990. Tuberculous meningitis. *Infectious*

Disease Clinics of North America, 4(4), pp.769–787. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2277198.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079. Available at: https://doi.org/10.1093/bioinformatics/btp352.

Li, H. and Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), pp.589–595. Available at: https://doi.org/10.1093/bioinformatics/btp698.

Li, T. et al., 2016. Diagnosing pyogenic, brucella and tuberculous spondylitis using histopathology and MRI: A retrospective study. *Experimental and therapeutic medicine*, 12(4), pp.2069–2077. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5038492/pdf/etm-12-04-2069.pdf.

Lin, L.F., Posfai, J., Roberts, R.J. and Kong, H., 2001. Comparative genomics of the restriction-modification systems in Helicobacter pylori. *PNAS*, 98(5), pp.2740–2745. Available at: https://doi.org/10.1073/pnas.051612298.

Linding, R. et al., 2003. Protein disorder prediction: Implications for structural proteomics. *Structure*, 11(11), pp.1453–1459. Available at: https://doi.org/10.1016/j.str.2003.10.002.

Litterer, L., 2009, *Comparing cloning efficiency of the pGEM®-T and pGEM®-T Easy vectors to the TOPO TA Cloning® vectors.* [Online]. Available at: http://worldwide.promega.com/resources/pubhub/enotes/comparing-cloning-efficiency-of-pgemt-and-pgemt-easy-vectors-to-topo-ta-cloning-vectors/ [Accessed: 6 March 2018].

Liu, F. et al., 2014. Comparative genomic analysis of Mycobacterium tuberculosis clinical isolates. *BMC Genomics*, 15(469), pp.1–13. Available at: https://doi.org/10.1186/1471-2164-15-469.

Liu, P.T. et al., 2006. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science*, 311(5768), pp.1770–1773. Available at: http://science.sciencemag.org/content/311/5768/1770.long.

Liu, X. et al., 2014. The study on the factors affecting transformation efficiency of E. coli competent cells. *Pakistan Journal of Pharmaceutical Sciences*, 27(3 Suppl), pp.679–684. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24816699.

Locht, C. et al., 2006. Heparin-binding hemagglutinin, from an extrapulmonary dissemination factor to a powerful diagnostic and protective antigen against tuberculosis. *Tuberculosis*, 86(3–4), pp.303–309. Available at: https://doi.org/10.1016/j.tube.2006.01.016.

Loman, N.J. et al., 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews of Microbiology*, 10(9), pp.599–606. Available at: https://www.nature.com/articles/nrmicro2850.

López, B. et al., 2003. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. *Clinical and Experimental Immunology*, 133(1), pp.30–37. Available at: https://doi.org/10.1046/j.1365-2249.2003.02171.x.

Lukyanov, S.A. et al., 2007. Selective suppression of polymerase chain reaction and its most popular applications. In: Buzdin, A.A. and Lukyanov, S.A., (eds.) *Nucleic Acids Hybridization Modern Applications*. Springer, Dordrecht, Russia, pp. 29–51.

Luthra, G. et al., 2007. Comparative evaluation of fungal, tubercular, and pyogenic brain abscesses with conventional and diffusion MR imaging and proton MR spectroscopy. *AJNR. American journal of neuroradiology*, 28(7), pp.1332–1338. Available at: https://doi.org/10.3174/ajnr.A0548.

Ma, Y., Pan, F. and McNeil, M., 2002. Formation of dTDP-rhamnose is essential for growth of mycobacteria. *Journal of Bacteriology*, 184(12), pp.3392–3395. Available at: https://jb.asm.org/content/184/12/3392.

MacLean, D., Jones, J.D.G. and Studholme, D.J., 2009. Application of "next-generation" sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, 7(4), pp.287–296. Available at: http://dx.doi.org/10.1038/nrmicro2088.

Mahairas, G.G. et al., 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. *Journal of Bacteriology*, 178(5), pp.1274–1282. Available at: https://jb.asm.org/content/178/5/1274.long.

Mahdi, L.K. et al., 2012. Identification of a novel pneumococcal vaccine antigen preferentially expressed during meningitis in mice. *Journal of Clinical Investigation*, 122(6), pp.2208–2220. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3366392/.

Marais, S. et al., 2011. Presentation and outcome of tuberculous meningitis in a high HIV prevalence setting. *PLoS ONE*, 6(5: e20077), pp.1–10. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3098272/pdf/pone.0020077.pdf.

Marrakchi, H., Laneelle, M. and Daffe, M., 2014. Mycolic acids: structures , biosynthesis , and beyond. *Chemistry & Biology Review*, 21(January), pp.67–85. Available at: http://dx.doi.org/10.1016/j.chembiol.2013.11.011.

Marx, G.E. and Chan, E.D., 2011. Tuberculous meningitis: diagnosis and

treatment overview. *Tuberculosis Research and Treatment*, 2011(798764), pp.1–9. Available at: http://dx.doi.org/10.1155/2011/798764.

Matlock, B., 2015. *Assessment of Nucleic Acid Purity*, Wilmington, MA, USA.

Matsuura, E. et al., 2000. Marked increase of matrix metalloproteinase 9 in cerebrospinal fluid of patients with funcal or tuberculous meningoencephalitis. *Journal of Neurological Sciences*, 173(1), pp.45–52. Available at: https://doi.org/10.1016/S0022-510X(99)00303-2.

McAdam, R.A. et al., 1990. Characterization of a Mycobacterium tuberculosis insertion sequence belonging to the IS3 family. *Molecular Microbiology*, 4(9), pp.1607–1613. Available at: https://doi.org/10.1111/j.1365-2958.1990.tb02073.x.

McGarvey, J.A. and Bermudez, L.E., 2001. Phenotypic and genomic analyses of the Mycobacterium avium complex reveal differences in gastrointestinal invasion and genomic composition Clements, J.D., (ed.). *Infection and Immunity*, 69(12), pp.7242–7249. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC98807/.

McHugh, T.D. and Gillespie, S.H., 1998. Nonrandom association of IS6110 and Mycobacterium tuberculosis: implications for molecular epidemiological studies. *Journal of Clinical Microbiology*, 36(5), pp.1410–1413. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC104839/.

McNeil, M., Daffe, M. and Brennan, P.J., 1991. Location of the mycolyl ester substituents in the cell walls of mycobacteria. *Journal of Biological Chemistry*, 266(20), pp.13217–13223. Available at: http://www.jbc.org/content/266/20/13217.abstract.

Menozzi, F.D. et al., 2006. Mycobacterium tuberculosis heparin-binding haemagglutinin adhesin (HBHA) triggers receptor-mediated transcytosis without altering the integrity of tight junctions. *Microbes and Infection*, 8(1), pp.1–9. Available at: https://doi.org/10.1016/j.micinf.2005.03.023.

Merker, M. et al., 2015. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. *Nature Genetics*, 47, p.242. Available at: https://doi.org/10.1038/ng.3195.

Metzker, M.L., 2010. Sequencing technologies-the next generation. *Nature Reviews Genetics*, 11(1), pp.31–46. Available at: https://www.nature.com/articles/nrg2626.

Ministry of Health, 2011. *Annual Report: Public Health*, Malaysia.

Minnikin, D.E., Kremer, L., Dover, L.G. and Besra, G.S., 2002. The methyl-branched fortifications of Mycobacterium tuberculosis. *Cell Chemical Biology*, 9, pp.545–553. Available at: https://doi.org/10.1016/S1074-5521(02)00142-4.

Miranda, S.M. et al., 2012. The tuberculous granuloma: An unsuccessful host defence mechanism providing a safety shelter for the bacteria? *Clinical and Developmental Immunology*, 2012(139127), pp.1–14. Available at: http://dx.doi.org/10.1155/2012/139127.

Mitchison, D.A. and Wallace, J.G., 1960. A comparison of the virulence in Guinea-pigs of South Indian and British tubercle bacilli. *Tubercle*, 41(1), pp.1–22. Available at: https://www.ncbi.nlm.nih.gov/pubmed/14423002.

Moghtaderi, A., Alavi-Naini, R., Izadi, S. and Cuevas, L.E., 2009. Diagnostic risk factors to differentiate tuberculous and acute bacterial meningitis. *Scandinavian journal of infectious diseases*, 41(3), pp.188–194. Available at: https://www.ncbi.nlm.nih.gov/pubmed/19235095?dopt=Abstract.

Molina-Torres, C.A. et al., 2010. Mycobacterium tuberculosis spoligotypes in Monterrey Mexico. *Journal of Clinical Microbiology*, 48(2), pp.448–455. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815641/pdf/1894-09.pdf.

Molzen, T.E. et al., 2011. Genome-wide identification of Streptococcus pneumoniae genes essential for bacterial replication during experimental meningitis. *Infection and Immunity*, 79(1), pp.288–297. Available at: https://iai.asm.org/content/iai/79/1/288.full.pdf.

Monteiro, R., Carneiro, J.C., Costa, C. and Duarte, R., 2013. Cerebral tuberculomas - a clinical challenge. *Respiratory Medicine Case Reports*, 9(Jun 3), pp.34–37. Available at: http://dx.doi.org/10.1016/j.rmcr.2013.04.003.

Morozova, O. and Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5), pp.255–264. Available at: http://www.sciencedirect.com/science/article/pii/S0888754308001651.

Morrow, B.J., Graham, J.E. and Curtiss, R., 1999. Genomic subtractive hybridization and selective capture of transcribed sequences identify a novel Salmonella typhimurium fimbrial operon and putative transcriptional regulator that are absent from the Salmonella typhi genome. *Infection and Immunity*, 67(10), pp.5106–5116. Available at: https://iai.asm.org/content/iai/67/10/5106.full.pdf.

Mostrom, P. et al., 2002. Methods used in the molecular epidemiology of tuberculosis. *Clinical Microbiology and Infection*, 8(11), pp.694–704. Available at: https://doi.org/10.1046/j.1469-0691.2002.00460.x.

De Mot, R., Nagy, I., Walz, J. and Baumeister, W., 1999. Proteasomes and other self-compartmentalizing proteases in prokaryotes. *Trends in Microbiology*, 7(2), pp.88–92. Available at: https://doi.org/10.1016/S0966-842X(98)01432-2.

Mukhopadhyay, S. and Balaji, K.N., 2011. The PE and PPE proteins of

Mycobacterium tuberculosis. *Tuberculosis*, 91(5), pp.441–447. Available at: https://doi.org/10.1016/j.tube.2011.04.004.

Mushegian, A.R. and Koonin, E. V, 1996. Gene order is not conserved in bacterial evolution. *Trends in Genetics*, 12(8), pp.289–290. Available at: https://doi.org/10.1016/0168-9525(96)20006-X.

Mycobrowser, 2019, *Mycobacterium tuberculosis H37Rv* [Online]. Available at: https://mycobrowser.epfl.ch/ [Accessed: 18 January 2018].

Nair, S., Pandey, A.D. and Mukhopadhyay, S., 2011. The PPE18 protein of Mycobacterium tuberculosis inhibits NF-κB/rel–mediated proinflammatory cytokine production by upregulating and phosphorylating suppressor of cytokine signaling 3 protein. *The Journal of Immunology*, 186(9), pp.5413–5424. Available at: https://doi.org/10.4049/jimmunol.1000773.

Nakatani, S.M. et al., 2004. Efficient method for mycobacterial DNA extraction in blood cultures aids rapid PCR identification of Mycobacterium tuberculosis and Mycobacterium avium. *European Journal of Clinical Microbiology & Infectious Diseases*, 23(11), pp.851–854. Available at: https://doi.org/10.1007/s10096-004-1236-z.

Naranjo, V. et al., 2006. Genes differentially expressed in oropharyngeal tonsils and mandibular lymph nodes of tuberculous and nontuberculous European wild boars naturally exposed to Mycobacterium bovis. *FEMS Immunology & Medical Microbiology*, 46(2), pp.298–312. Available at: https://doi.org/10.1111/j.1574-695X.2005.00035.x.

Narayana, Y. et al., 2007. Differential B-cell responses are induced by Mycobacterium tuberculosis PE antigens Rv1169c, Rv0978c, and Rv1818c. *Clinical and Vaccine Immunology*, 14(10), pp.1334–1341. Available at: https://cvi.asm.org/content/14/10/1334.

National Center for Biotechnology Information, 2018, *BLAST >> blastn suite >> Standard Nucleotide BLAST* [Online]. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome [Accessed: 26 September 2018].

National Center for Biotechnology Information, 2017, *NCBI Genome* [Online]. Available at: https://www.ncbi.nlm.nih.gov/genome/?term=mycobacterium+tuberculosis [Accessed: 20 March 2017].

New England Biolabs, 2017, *DNA gel purification* [Online]. Available at: https://www.neb.com/products/nucleic-acid-purification/dna-gel-purification/dna-gel-purification [Accessed: 8 December 2017].

Neyrolles, O. and Quintana-Murci, L., 2009. Sexual inequality in tuberculosis. *PLoS Medicine*, 6(12), pp.1–6. Available at: https://doi.org/10.1371/journal.pmed.1000199.

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. Available at: https://doi.org/10.1016/j.tcb.2005.03.009.

Niemann, S. et al., 2009. Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. *PLoS ONE*, 4(10: e7407), pp.1–7. Available at: https://doi.org/10.1371/journal.pone.0007407.

NIMHANS, 2017, *Research* [Online]. Available at: http://www.nimhans.ac.in/neuromicrobiology/research [Accessed: 21 March 2017].

Nishimura, A., Morita, M., Nishimura, Y. and Sugino, Y., 1990. A rapid and highly efficient method for preparation of competent Escherichia coli cells. *Nucleic Acids Research*, 18(20), p.6169. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC332453/pdf/nar00204-0225.pdf.

Norton, J.P. and Mulvey, M.A., 2012. Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic Escherichia coli. *PLoS Pathogens*, 8(10: e1002954), pp.1–13. Available at: https://doi.org/10.1371/journal.ppat.1002954.

Noss, E.H. et al., 2001. Toll-like receptor 2-dependent inhibition of macrophage class II MHC expression and antigen processing by 19-kDa lipoprotein of Mycobacterium tuberculosis. *The Journal of Immunology*, 167(2), pp.910–918. Available at: https://doi.org/10.4049/jimmunol.167.2.910.

O'Garra, A. et al., 2013. The immune response in tuberculosis. *Annual Review of Immunology*, 31(1), pp.475–527. Available at: https://doi.org/10.1146/annurev-immunol-032712-095939.

Oelemann, M.C. et al., 2007. Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. *Journal of Clinical Microbiology*, 45(3), pp.691–697. Available at: https://jcm.asm.org/content/45/3/691.long.

Okumura, K. et al., 2015. Construction of a virtual Mycobacterium tuberculosis consensus genome and its application to data from a next generation sequencer. *BMC Genomics*, 16(1: 218), pp.1–17. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4425900/.

Olleros, M.L. et al., 2015. Control of mycobacterial infections in mice expressing human Tumor Necrosis Factor (TNF) but not mouse TNF Ehrt, S., (ed.). *Infection and Immunity*, 83(9), pp.3612–3623. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4534677/.

Orihuela, C.J. et al., 2004. Microarray analysis of pneumococcal gene expression during invasive disease. *Infection and Immunity*, 72(10), pp.5582–5596. Available at: http://iai.asm.org/content/72/10/5582.abstract.

Orme, I.M. and Cooper, A.M., 1999. Cytokine/chemokine cascades in immunity to tuberculosis. *Immunology Today*, 20(7), pp.307–312. Available at: https://doi.org/10.1016/S0167-5699(98)01438-8.

Peng, Y., Leung, H.C.M., Yiu, S.M. and Chin, F.Y.L., 2012. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), pp.1420–1428. Available at: https://doi.org/10.1093/bioinformatics/bts174.

Periwal, V. et al., 2015. Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of Mycobacterium tuberculosis pangenome. *PLoS ONE*, 10(4), pp.1–26. Available at: https://doi.org/10.1371/journal.pone.0122979.

Pethe, K. et al., 2001. Mycobacterium smegmatis laminin-binding glycoprotein shares epitopes with Mycobacterium tuberculosis heparin-binding haemagglutinin. *Molecular Microbiology*, 39(1), pp.89–99. Available at: https://doi.org/10.1046/j.1365-2958.2001.02206.x.

Pieters, J. and Gatfield, J., 2002. Hijacking the host: survival of pathogenic mycobacteria inside macrophages. *Trends in Microbiology*, 10(3), pp.142–146. Available at: https://doi.org/10.1016/S0966-842X(02)02305-3.

Poulet, S. and Cole, S.T., 1995. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in Mycobacterium tuberculosis. *Archives of Microbiology*, 163(2), pp.87–95. Available at: https://www.ncbi.nlm.nih.gov/pubmed/7710330.

Pouttu, R. et al., 1999. Amino acid residue Ala-62 in the FimH fimbrial adhesin is critical for the adhesiveness of meningitis-associated Escherichia coli to collagens. *Molecular Microbiology*, 31(6), pp.1747–1757. Available at: https://doi.org/10.1046/j.1365-2958.1999.01311.x.

Promega Corporation, 2015. DNA purification. *Protocols & Applications Guide*, 8(9), p.9.1-9.34. Available at: http://manatee.sourceforge.net.

Promega Corporation, 2009, *E. coli competent cells* [Online]. Available at: https://www.promega.com/-/media/files/resources/protocols/technical-bulletins/0/e-coli-competent-cells-protocol.pdf?la=en%0Ahttp://www.promega.com/tbs/tb095/tb095.pdf%0Ahttps://www.promega.com/~/media/files/resources/protocols/technical bulletins/0/e coli [Accessed: 27 July 2017].

Promega Corporation, 2010. pGEM-T and pGEM-T Easy vector systems. *Technical Manual*, pp.1–28. Available at: www.promega.com [Accessed: 27 July 2017].

Promega Corporation, 2017, *SDS solution, molecular biology grade (10% w/v)* [Online]. Available at: https://worldwide.promega.com/products/biochemicals-and-labware/biochemical-buffers-and-reagents/sds-solution_-molecular-biology-grade-_10_-w_v_/?catNum=V6551 [Accessed: 17 November 2017].

Qi, J. and Zhao, F., 2011. inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. *Nucleic Acids Research*, 39(Web Server Issue), pp.W567-575. Available at: https://doi.org/10.1093/nar/gkr506.

Raju, R.M. et al., 2012. Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLOS Pathogens*, 8(2), p.e1002511. Available at: https://doi.org/10.1371/journal.ppat.1002511.

Ramage, H.R., Connolly, L.E. and Cox, J.S., 2009. Comprehensive functional analysis of Mycobacterium tuberculosis toxin-antitoxin systems: Implications for pathogenesis, stress responses, and evolution. *PLoS Genetics*, 5(12: e1000767), pp.1–14. Available at: https://doi.org/10.1371/journal.pgen.1000767.

Raviglione, M.C., Snider, D.E. and Kochi, A., 1995. Global epidemiology of tuberculosis. *The Journal of the American Medical Association (JAMA)*, 273(3), pp.220–226. Available at: http://jama.ama-assn.org/content/273/3/220.abstract.

Reed, M.B. et al., 2004. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature*, 431(7004), pp.84–87. Available at: https://www.nature.com/articles/nature02837.

van Rie, A. et al., 1999. Transmission of a multidrug-resistant Mycobacterium tuberculosis strain resembling “strain W” among noninstitutionalized, human immunodeficiency virus-seronegative patients. *The Journal of Infectious Diseases*, 180(5), pp.1608–1615. Available at: https://doi.org/10.1086/315054.

Riley, M. and Labedan, B., 1996. Escherichia coli gene products: physiological functions and common ancestries. In: Neidhardt, F.C. et al., (eds.) *Escherichia coli and Salmonella*. American Society for Microbiology, Washington, D. C., pp. 1–159.

Rocha-Ramírez, L.M. et al., 2008. Mycobacterium tuberculosis lipids regulate cytokines, TLR-2/4 and MHC class II expression in human macrophages. *Tuberculosis*, 88(3), pp.212–220. Available at: https://doi.org/10.1016/j.tube.2007.10.003.

Rock, R.B. et al., 2008. Central nervous system tuberculosis: pathogenesis and clinical aspects. *Clinical Microbiology Review*, 21(2), pp.243–261. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC96831/.

Romero, B. et al., 2008. Persistence and molecular evolution of Mycobacterium bovis population from cattle and wildlife in Doñana National Park revealed by genotype variation. *Veterinary Microbiology*, 132(1–2), pp.87–95. Available at: https://doi.org/10.1016/j.vetmic.2008.04.032.

Ropper, A., Samuels, M. and Klein, J., 2014. *Adams and Victor's principles of neurology* 10th ed., McGraw-Hill, New York, NY.

Rosales, S. et al., 2010. Molecular diversity of Mycobacterium tuberculosis isolates from patients with tuberculosis in Honduras. *BMC Microbiology*, 10(208), pp.1–9. Available at: https://doi.org/10.1186/1471-2180-10-208.

Rubin, L.L. and Staddon, J.M., 1999. The cell biology of the blood-brain barrier. *Annual Review of Neuroscience*, 22(March), pp.11–28. Available at: https://doi.org/10.1146/annurev.neuro.22.1.11.

Ruesen, C. et al., 2018. Large-scale genomic analysis shows association between homoplastic genetic variation in Mycobacterium tuberculosis genes and meningeal or pulmonary tuberculosis. *BMC Genomics*, 19(122), pp.1–11. Available at: https://doi.org/10.1186/s12864-018-4498-z.

Sáenz, B. et al., 2013. The dual face of central nervous system tuberculosis: A new Janus Bifrons? *Tuberculosis*, 93(2), pp.130–135. Available at: https://doi.org/10.1016/j.tube.2012.11.011.

Sambrook, J., Fritsch, E.F. and Maniatis, T., 1989. *Molecular cloning: a laboratory manual.* 4th ed., Cold Spring Harbor Laboratory Press, New York.

Sambrook, J. and Russell, D.W., 2006a. Preparation and transformation of competent E. coli using calcium chloride. In: Irwin, N. and Janssen, K.A., (eds.) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, NY, p. 1.116-1.122.

Sambrook, J. and Russell, D.W., 2006b. Purification of PCR Products in Preparation for Cloning. *Cold Spring Harbor Protocols*, 2006(1), p.pii: pdb.prot3825. Available at: http://cshprotocols.cshlp.org/content/2006/1/pdb.prot3825.short.

Sampson, S.L., 2011. Mycobacterial PE/PPE proteins at the host-pathogen interface. *Clinical and Developmental Immunology*, 2011(497203), pp.1–11. Available at: http://dx.doi.org/10.1155/2011/497203.

Sandor, M., Weinstock, J. V. and Wynn, T.A., 2003. Granulomas in schistosome and mycobacterial infections: A model of local immune responses. *Trends in Immunology*, 24(1), pp.44–52. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12495724.

Sanger, F., 1988. Sequences, sequences and sequences. *Annual Review of Biochemistry*, 57(July), pp.1–29. Available at:

https://doi.org/10.1146/annurev.bi.57.070188.000245.

Sasindran, S.J. and Torrelles, J.B., 2011. Mycobacterium tuberculosis infection and inflammation: What is beneficial for the host and for the bacterium? *Frontiers in Microbiology*, 2(2), pp.1–16. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3109289/pdf/fmicb-02-00002.pdf.

Saw, S.H. et al., 2016. Chromosomal rearrangements and protein globularity changes in Mycobacterium tuberculosis isolates from cerebrospinal fluid. *PLoS ONE*, 4(e2484), pp.1–21. Available at: https://peerj.com/articles/2484/.

Saw, S.H., Yap, S.F. and Ngeow, Y.F., 2019. Meningitis-Associated Genes in Mycobacterium species. *RSU International Research Conference*. 2019 Research Institute of Rangsit University, Pathum Thani Thailand, pp. 14–24.

Saw, S.H., Yong, V.C. and Ngeow, Y.F., 2014. Generation of differentially regulated genes in Mycobacterium tuberculosis isolates from cerebrospinal fluid and respiratory secretions using suppression subtractive hybridization. *BMC Infectious Diseases*, 14(Suppl 3), p.O16. Available at: http://www.biomedcentral.com/1471-2334/14/S3/O16%0AORAL.

Sayes, F. et al., 2012. Strong immunogenicity and cross-reactivity of Mycobacterium tuberculosis ESX-5 type VII secretion-encoded PE-PPE proteins predicts vaccine potential. *Cell Host and Microbe*, 11(4), pp.352–363. Available at: https://doi.org/10.1016/j.chom.2012.03.003.

Schaller, M.A., Wicke, F., Foerch, C. and Weidauer, S., 2019. Central Nervous System Tuberculosis : Etiology, Clinical Manifestations and Neuroradiological Features. *Clinical neuroradiology*, 29(1), pp.3–18. Available at: https://doi.org/10.1007/s00062-018-0726-9.

Shah, N.S. et al., 2007. Worldwide emergence of extensively drug-resistant tuberculosis. *Emerging Infectious Diseases*, 13(3), pp.380–387. Available at: https://wwwnc.cdc.gov/eid/article/13/3/06-1400_article.

Shimoji, Y. et al., 1999. A 21-kDa surface protein of Mycobacterium leprae binds peripheral nerve laminin-2 and mediates Schwann cell invasion. *PNAS*, 96(17), pp.9857–9862. Available at: https://doi.org/10.1073/pnas.96.17.9857.

Shoichet, B.K., Baase, W.A., Kurokit, R. and Matrhewst, B.W., 1995. A relationship between protein stability and protein function. *PNAS*, 92(2), pp.452–456. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC42758/.

Shorten, R.J. et al., 2013. When is an outbreak not an outbreak? Fit, divergent strains of Mycobacterium tuberculosis display independent evolution of drug resistance in a large london outbreak. *The Journal of Antimicrobial Chemotherapy*, 68(3), pp.543–549. Available at: https://doi.org/10.1093/jac/dks430.

Siebert, P.D. et al., 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Research*, 23(6), pp.1087–1088.

Singh, P. and Cole, S.T., 2011. Mycobacterium leprae: genes, pseudogenes and genetic diversity. *Future Microbiology*, 6(1), pp.57–71. Available at: https://doi.org/10.2217/fmb.10.153.

Singh, U.B. et al., 2007. Genetic biodiversity of Mycobacterium tuberculosis isolates from patients with pulmonary tuberculosis in India. *Infection, Genetics and Evolution*, 7(4), pp.441–448. Available at: https://doi.org/10.1016/j.meegid.2007.01.003.

Sinsimer, D. et al., 2008. The phenolic glycolipid of Mycobacterium tuberculosis differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infection and Immunity*, 76(7), pp.3027–3036. Available at: https://iai.asm.org/content/76/7/3027.long.

Skerry, C. et al., 2013. Vaccination with recombinant Mycobacterium tuberculosis PknD attenuates bacterial dissemination to the brain in guinea pigs. *PLoS Biology*, 8(6: e66310), pp.1–6. Available at: https://doi.org/10.1371/journal.pone.0066310.

Small, P.M. et al., 1993. Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. *Journal of Clinical Microbiology*, 31(7), pp.1677–1682. Available at: https://jcm.asm.org/content/31/7/1677.long.

Smith, D.R. et al., 1997. Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome. *Genome Research*, 7(8), pp.802–819. Available at: https://genome.cshlp.org/content/7/8/802.long.

Smith, I., 2003. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clinical Microbiology Reviews*, 16(3), pp.463–496. Available at: https://cmr.asm.org/content/16/3/463.long.

Smith, J. et al., 2008. Evidence for pore formation in host cell membranes by ESX-1-secreted ESAT-6 and its role in Mycobacterium marinum escape from the vacuole. *Infection and Immunity*, 76(12), pp.5478–5487. Available at: https://iai.asm.org/content/76/12/5478.long.

Smith, N.H. et al., 2006. Ecotypes of the Mycobacterium tuberculosis complex. *Journal of Theoretical Biology*, 239(2), pp.220–225. Available at: http://www.sciencedirect.com/science/article/B6WMD-4HCDK4W-4/2/15a17c900da2fc8f03a48836dc819d16.

Snyder, L.A.S. and Saunders, N.J., 2006. The majority of genes in the pathogenic Neisseria species are present in non-pathogenic Neisseria lactamica, including those designated as “virulence genes.” *BMC Genomics*,

7(1), p.128. Available at: https://doi.org/10.1186/1471-2164-7-128.

Soini, H. et al., 2001. Transmission dynamics and molecular characterization of Mycobacterium tuberculosis isolates with low copy numbers of IS6110. *Journal of Clinical Microbiology*, 39(1), pp.217–221. Available at: https://jcm.asm.org/content/39/1/217.long.

Sola, C. et al., 2001. Spoligotype database of Mycobacterium tuberculosis: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerging infectious diseases*, 7(3), pp.390–396. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2631784/pdf/11384514.pdf.

Song, H. et al., 2008. Identification of outer membrane proteins of Mycobacterium tuberculosis. *Tuberculosis*, 88(6), pp.526–544. Available at: https://doi.org/10.1016/j.tube.2008.02.004.

Van Soolingen, D., 2001. Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. *Journal of Internal Medicine*, 249(1), pp.1–26. Available at: https://doi.org/10.1046/j.1365-2796.2001.00772.x.

Sorensen, A.L. et al., 1995. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. *Infection and Immunity*, 63(5), pp.1710–1717. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC173214/.

Sratagene Technical Service, 2003, *JM109 Competent Cells* [Online]. Available at: https://www.chem-agilent.com/pdf/strata/200235.pdf [Accessed: 27 July 2007].

Sreevatsan, S., 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *PNAS*, 94(18), pp.9869–9874. Available at: http://dx.doi.org/10.1073/pnas.94.18.9869.

Streicher, E.M. et al., 2012. Emergence and treatment of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis in South Africa. *Infection, Genetics and Evolution*, 12(4), pp.686–694. Available at: http://dx.doi.org/10.1016/j.meegid.2011.07.019.

Stucki, D. et al., 2016. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature genetics*, 48(12), pp.1535–1543. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27798628.

Stucki, D. and Gagneux, S., 2013. Single nucleotide polymorphisms in Mycobacterium tuberculosis and the need for a curated database. *Tuberculosis*, 93(1), pp.30–39. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3582841/.

Studer, R.A., Dessailly, B.H. and Orengo, C.A., 2013. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *The Biochemical Journal*, 449(3), pp.581–594. Available at: http://www.biochemj.org/content/449/3/581.long.

Sultana, R., Tanneeru, K. and Guruprasad, L., 2011. The PE-PPE domain in mycobacterium reveals a serine α/β hydrolase fold and function: an in-silico analysis. *PLoS ONE*, 6(2: e16745), pp.1–7. Available at: https://doi.org/10.1371/journal.pone.0016745.

Sun, Y., Sriramajayam, K., Luo, D. and Liao, D.J., 2012. A quick, cost-free method of purification of DNA fragments from agarose gel. *Journal of Cancer*, 3, pp.93–95. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3283835/.

Sundaram, C., Shankar, S.K., Thong, W.K. and Pardo-Villamizar, C.A., 2011. Pathology and diagnosis of central nervous system infections. *Pathology Research International*, 2011(878263), pp.1–4. Available at: http://dx.doi.org/10.4061/2011/878263.

Supply, P. et al., 2001. Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. *Journal of Clinical Microbiology*, 39(10), pp.3563–3571. Available at: https://jcm.asm.org/content/39/10/3563.long.

Supply, P. et al., 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nature Genetics*, 45(2), pp.172–179. Available at: https://www.nature.com/articles/ng.2517.pdf.

Supply, P. et al., 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 44(12), pp.4498–4510. Available at: http://dx.doi.org/10.1128/JCM.01392-06.

Supply, P. et al., 2000. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Molecular Microbiology*, 36(3), pp.762–771. Available at: https://doi.org/10.1046/j.1365-2958.2000.01905.x.

Supply, P., Magdalena, J., Himpens, S. and Locht, C., 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Molecular Microbiology*, 26(5), pp.991–1003. Available at: https://doi.org/10.1046/j.1365-2958.1997.6361999.x.

Sutlas, P.N. et al., 2003. Tuberculous meningitis in adults: review of 61 cases. *Infection*, 31(6), pp.387–391. Available at: https://www.ncbi.nlm.nih.gov/pubmed/14735380?dopt=Abstract.

Swarna, N.Y., 2014. A review of tuberculosis research in Malaysia. *Medical*

Journal of Malaysia, 69(August), pp.88–102. Available at: http://www.e-mjm.org/2014/supplement-A/tuberculosis-research.pdf.

Szklarczyk, D. et al., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43(Database issue), pp.D447-452. Available at: https://doi.org/10.1093/nar/gku1003.

Tailleux, L. et al., 2003. DC-SIGN is the major Mycobacterium tuberculosis receptor on human dendritic cells. *The Journal of Experimental Medicine*, 197(1), pp.121–127. Available at: http://jem.rupress.org/content/197/1/121.long.

Takara Bio USA, 2016. *Advantage® 2 PCR Enzyme System User Manual*, California, USA.

Talaat, A.M., Lyons, R., Howard, S.T. and Johnston, S.A., 2004. The temporal expression profile of Mycobacterium tuberculosis infection in mice. *PNAS*, 101(13), pp.4602–4607.

Talarico, S. et al., 2007. Association of Mycobacterium tuberculosis PE PGRS33 polymorphism with clinical and epidemiological characteristics. *Tuberculosis*, 87(4), pp.338–346. Available at: https://doi.org/10.1016/j.tube.2007.03.003.

Tamames, J., Casari, G., Ouzounis, C. and Valencia, A., 1997. Conserved clusters of functionally related genes in two bacterial genomes. *Journal of Molecular Evolution*, 44(1), pp.66–73. Available at: https://link.springer.com/article/10.1007/PL00006122.

Tascon, R.E. et al., 2000. Mycobacterium tuberculosis-activated dendritic cells induce protective immunity in mice. *Immunology*, 99(3), pp.473–480. Available at: https://doi.org/10.1046/j.1365-2567.2000.00963.x.

Tatusov, R.L. et al., 1996. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli. *Current Biology*, 6(3), pp.279–291. Available at: http://www.sciencedirect.com/science/article/pii/S0960982202004785.

Tavanti, A. et al., 2003. Optimization and validation of multilocus sequence typing for Candida albicans. *Journal of Clinical Microbiology*, 41(8), pp.3765–3776. Available at: https://jcm.asm.org/content/41/8/3765.long.

Thakur, R., Sarma, S. and Goyal, R., 2011. Comparison of DNA extraction protocols for Mycobacterium Tuberculosis in diagnosis of tuberculous meningitis by real-time polymerase chain reaction. *Journal of Global Infectious Diseases*, 3(4), pp.353–356. Available at: http://www.jgid.org/text.asp?2011/3/4/353/91057.

Thermo Fisher Scientific, 2014, *Sanger DNA Sequencing* [Online]. Available at: http://www.lifetechnologies.com/my/en/home/life-

science/sequencing/sanger-sequencing/sanger-dna-sequencing.html [Accessed: 27 July 2014].

Thermo Fisher Scientific, 2017, *UltraPure™ SDS Solution, 10%* [Online]. Available at: https://www.thermofisher.com/order/catalog/product/24730020 [Accessed: 17 November 2017].

Thoma-Uszynski, S. et al., 2001. Induction of direct antimicrobial activity through mammalian toll-like receptors. *Science*, 291(5508), pp.1544–1547. Available at: http://www.sciencemag.org/content/291/5508/1544.abstract.

Thwaites, G., Caws, M., et al., 2004. Comparison of conventional bacteriology with nucleic acid amplification (amplified mycobacterium direct test) for diagnosis of tuberculous meningitis before and after inception of antituberculosis chemotherapy. *Journal of Clinical Microbiology*, 42(3), pp.996–1002. Available at: https://jcm.asm.org/content/42/3/996.long.

Thwaites, G., Bang, N., et al., 2004. Dexamethasone for the Treatment of Tuberculous Meningitis in Adolescents and Adults. *New England Journal of Medicine*, 351, pp.1741–1751. Available at: https://www.nejm.org/doi/pdf/10.1056/NEJMoa040573?articleTools=true.

Thwaites, G., 2013. The pathophysiology of tuberculous meningitis. In: Christodoulides, M., (ed.) *Meningitis: Cellular and Molecular Basis*. CAB International, London UK, p. 153.

Thwaites, G. et al., 2000. Tuberculous meningitis. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(3), pp.289–299. Available at: http://dx.doi.org/10.1136/jnnp.68.3.289.

Thwaites, G.E. et al., 2009. British Infection Society guidelines for the diagnosis and treatment of tuberculosis of the central nervous system in adults and children. *The Journal of Infection*, 59(3), pp.167–187. Available at: https://doi.org/10.1016/j.jinf.2009.06.011%0A.

Thwaites, G.E. et al., 2007. Serial MRI to determine the effect of dexamethasone on the cerebral pathology of tuberculous meningitis: an observational study. *The Lancet. Neurology*, 6(3), pp.230–236. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4333204/.

Thwaites, G.E., van Toorn, R. and Schoeman, J., 2013. Tuberculous meningitis: more questions, still too few answers. *The Lancet Neurology*, 12(10), pp.999–1010. Available at: https://doi.org/10.1016/S1474-4422(13)70168-6.

Tinsley, C.R. and Nassif, A.X., 1996. Analysis of the genetic differences between Neisseria meningitidis and Neisseria gonorrhoeae: two closely related bacteria expressing two different pathogenicities. *PNAS*, 93(20), pp.11109–11114. Available at: https://www.jstor.org/stable/40291.

Tizard, M. et al., 1998. A low G+C content genetic island in Mycobacterium avium subsp. paratuberculosis and M. avium subsp. silvaticum with homologous genes in Mycobacterium tuberculosis. *Microbiology*, 144(Pt 12), pp.3413–3423. Available at: http://10.0.4.75/00221287-144-12-3413.

Tokuriki, N., Stricher, F., Serrano, L. and Tawfik, D.S., 2008. How protein stability and new functions trade off. *PLoS Computational Biology*, 4(2: e1000002), pp.1–7. Available at: https://doi.org/10.1371/journal.pcbi.1000002.

Török, M.E., 2015. Tuberculous meningitis: advances in diagnosis and treatment. *British Medical Bulletin*, 113(1), pp.117–131. Available at: http://dx.doi.org/10.1093/bmb/ldv003.

Townsend, K.M. et al., 1998. Development of PCR assays for species- and type-specific identification of Pasteurella multocida isolates. *Journal of Clinical Microbiology*, 36, pp.1096–1100.

Tsenova, L. et al., 2005. Virulence of selected Mycobacterium tuberculosis clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. *The Journal of Infectous Diseases*, 192(1), pp.98–106. Available at: https://doi.org/10.1086/430614.

Tsolaki, A.G. et al., 2004. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. *PNAS*, 101(14), pp.4865–4870. Available at: https://doi.org/10.1073/pnas.0305634101.

Tsolaki, A.G. et al., 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 43(7), pp.3185–3191. Available at: https://jcm.asm.org/content/43/7/3185.long.

Tufariello, J.A.M., Chan, J. and Flynn, J.A.L., 2003. Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. *The Lancet. Infectious Diseases*, 3(9), pp.578–590. Available at: https://doi.org/10.1016/S1473-3099(03)00741-2.

Turenne, C.Y., Collins, D.M., Alexander, D.C. and Behr, M.A., 2008. Mycobacterium avium subsp. paratuberculosis and M. avium subsp. avium are independently evolved pathogenic clones of a much broader group of M. avium organisms. *Journal of Bacteriology*, 190(7), pp.2479–2487. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2293204/.

Udani, P.M., 1994. BCG vaccination in India and tuberculosis in children: newer facets. *The Indian Journal of Pediatrics*, 61(5), pp.451–462. Available at: https://www.ncbi.nlm.nih.gov/pubmed/7744445.

Udwadia, Z.F., 2012. MDR, XDR, TDR tuberculosis: ominous progression. *Thorax*, 67(4), p.286 LP-288. Available at: http://dx.doi.org/10.1136/thoraxjnl-2012-201663.

Uplekar, S. et al., 2011. Comparative genomics of ESX genes from clinical isolates of Mycobacterium tuberculosis provides evidence for gene conversion and epitope variation. *Infection and Immunity*, 79(10), pp.4042–4049. Available at: http://iai.asm.org/content/79/10/4042.abstract.

Valway, S.E. et al., 1998. An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. *New England Journal of Medicine*, 338(10), pp.633–639. Available at: http://www.nejm.org/doi/full/10.1056/NEJM199803053381001.

Vega-Lopez, F. et al., 1993. Sequence and immunological characterization of a serine-rich antigen from Mycobacterium leprae. *Infection and Immunity*, 61(5), pp.2145–2153. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC280815/.

Victor, T.C. et al., 2004. Molecular characteristics and global spread of Mycobacterium tuberculosis with a Western Cape F11 genotype. *Journal of Clinical Microbiology*, 42(2), pp.769–772. Available at: https://jcm.asm.org/content/42/2/769.long.

de Viedma, G.D. et al., 2006. Complex clonal features in an Mycobacterium Tuberculosis infection in a two-year-old child. *The Pediatric Infectious Disease Journal*, 25(5), pp.457–459. Available at: https://insights.ovid.com/pubmed?pmid=16645515.

Viegas, S. et al., 2010. Molecular diversity of Mycobacterium tuberculosis isolates from patients with pulmonary tuberculosis in Mozambique. *BMC Microbiology*, 10(1), p.195. Available at: http://www.biomedcentral.com/1471-2180/10/195.

Vinnard, C. and MacGregor, R.R., 2009. Tuberculous meningitis in HIV-infected individuals. *Current HIV/AIDS Reports*, 6(3), pp.139–145. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3131531/.

Virji, M. et al., 1993. Meningococcal Opa and Opc proteins: their role in colonization and invasion of human epithelial and endothelial cells. *Molecular Microbiology*, 10(3), pp.499–510. Available at: https://doi.org/10.1111/j.1365-2958.1993.tb00922.x.

Wallgren, A., 1948. The time-table of tuberculosis. *Tubercle*, 29(11), pp.245–251. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18101320.

Wang, X., Minasov, G. and Shoichet, B.K., 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *Journal of Molecular Biology*, 320(1), pp.85–95. Available at: https://doi.org/10.1016/S0022-2836(02)00400-X.

Weidauer, S., Wagner, M. and Nichtweiss, M., 2017. Magnetic Resonance Imaging and Clinical Features in Acute and Subacute Myelopathies. *Clinical*

neuroradiology, 27(4), pp.417–433. Available at: https://link.springer.com/article/10.1007%2Fs00062-017-0604-x.

Weinberg, Z. et al., 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Research*, 35(14), pp.4809–4819. Available at: https://doi.org/10.1093/nar/gkm487.

van der Wel, N. et al., 2007. M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. *Cell*, 129(7), pp.1287–1298. Available at: https://doi.org/10.1016/j.cell.2007.05.059.

Weniger, T. et al., 2010. MIRU-VNTRplus: a web tool for polyphasic genotyping of Mycobacterium tuberculosis complex bacteria. *Nucleic Acids Research*, 38(Web Server issue), pp.W326–W331. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2896200/.

WHO, 2000. Anti-tuberculosis drug resistance in the world: The WHO/IUATLD Global Project on Anti-Tuberculosis Dug Resistance Surveillance. *Report 2:Prevalence and trends*.

WHO, 2012, *Global Tuberculosis Report 2012* [Online]. Available at: http://www.who.int/about/ [Accessed: 27 July 2010].

WHO, 2016, *Global Tuberculosis Report 2016* [Online]. Available at: http://www.who.int.

WHO, 2017, *Global Tuberculosis Report 2017* [Online]. Available at: https://www.who.int/csr/don/archive/year/2017/en/.

WHO, 2018. *Global tuberculosis report 2018*, Geneva, Switzerland.

WHO, 2010. *Guidelines for treatment of tuberculosis*, WHO, USA.

Wilfinger, W.W., Mackey, K. and Chomczynski, P., 1997. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity BioTechniques, (ed.). *Biotechniques*, 22(3), pp.474–481. Available at: https://doi.org/10.2144/97223st01%0A.

Willcocks, S. and Wren, B.W., 2014. Shared characteristics between Mycobacterium tuberculosis and fungi contribute to virulence. *Future Microbiology*, 9(5), pp.657–668. Available at: https://doi.org/10.2217/fmb.14.29.

World Health Organization Regional Office for South-East Asia, 2017, *Bending the curve - ending TB: annual report 2017* [Online]. Available at: http://www.who.int/iris/handle/10665/254762.

Wright, P.E. and Dyson, H.J., 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular*

Biology, 293(2), pp.321–331. Available at: https://doi.org/10.1006/jmbi.1999.3110.

Xiong, Z.H., Zhuang, Y.H. and Li, G.L., 2005. Identification of differential genomic genes of Mycobacterium tuberculosis H37Rv and attenuated strain H37Ra by suppression subtractive hybridization. *Yi Chuan Xue Bao = Acta Genetica Sinica*, 32(9), pp.979–985. Available at: https://www.ncbi.nlm.nih.gov/pubmed/16201243.

Yang, J.L. et al., 2010. Five-minute purification of PCR products by new-freeze-squeeze method. *Journal of Food Agriculture and Environment*, 8(2), pp.32–33. Available at: https://doi.org/10.1234/4.2010.1558%0A.

Yang, Z. et al., 2004. Identification of risk factors for extrapulmonary tuberculosis. *Clinical Infectious Diseases*, 38(2), pp.199–205. Available at: https://doi.org/10.1086/380644.

Yang, Z.H. et al., 2000. Spoligotyping and polymorphic GC-rich repetitive sequence fingerprinting of mycobacterium tuberculosis strains having few copies of IS6110. *Journal of Clinical Microbiology*, 38(10), pp.3572–3576. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC87438/.

Yanisch-Perron, C., Vieira, J. and Messing, J., 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33(1), pp.103–119. Available at: https://doi.org/10.1016/0378-1119(85)90120-9.

Yao, Y., Xie, Y. and Kim, K.S., 2006. Genomic comparison of Escherichia coli K1 strains isolated from the cerebrospinal fluid of patients with meningitis. *Infection and Immunity*, 74(4), pp.2196–2206. Available at: http://iai.asm.org/content/74/4/2196.abstract.

Zhang, F. and Xie, J.P., 2011. Mammalian cell entry gene family of Mycobacterium tuberculosis. *Molecular and Cellular Biochemistry*, 352(1–2), pp.1–10. Available at: http://dx.doi.org/10.1007/s11010-011-0733-5.

Zhang, J. et al., 2010. Mycobacterium tuberculosis complex CRISPR genotyping: improving efficiency, throughput and discriminative power of "spoligotyping" with new spacers and a microbead-based hybridization assay. *Journal of Medical Microbiology*, 59(Pt 3), pp.285–294. Available at: https://10.0.4.75/jmm.0.016949-0.

Zhang, Y. et al., 2012. Screening and comparison of differentially expressed genes between one MDR-TB strain and the virulent M. tuberculosis H37Rv. *Gene*, 506(1), pp.223–229. Available at: https://doi.org/10.1016/j.gene.2012.06.044.

Zheng, H. et al., 2008. Genetic basis of virulence attenuation revealed by comparative genomic analysis of Mycobacterium tuberculosis strain H37Ra versus H37Rv. *PLoS One*, 3(6: e2375), pp.1–12. Available at:

https://doi.org/10.1371/journal.pone.0002375.

Zwadyk, P.J., Down, J.A., Myers, N. and Dey, M.S., 1994. Rendering of mycobacteria safe for molecular diagnostic studies and development of a lysis method for strand displacement amplification and PCR. *Journal of Clinical Microbiology*, 32(9), pp.2140–2146. Available at: https://jcm.asm.org/content/32/9/2140.long.

APPENDICES

Appendix 1

3.2 Culture Media and Reagents

All culture media were obtained from Pronadisa (Spain) and all chemicals were obtained from Qiagen (Netherland) or Invitrogen by Life Technologies (USA), unless otherwise stated. Media and reagents were sterilised by autoclaving at 121 °C for 15 min unless otherwise indicated.

3.2.1 Culture Media

LB Broth and Agar

Purpose: *Non-selective media for growth of bacteria*

Luria-Bertani (LB) broth was prepared by dissolving 4.0 g of tryptone, 2.0 g of yeast extract, and 2.0 g of sodium chloride (NaCl), in distilled water to a final volume of 400.0 ml. The pH was adjusted to 7.4. LB agar was prepared as described for LB broth, with the addition of 1.5 % w/v agar powder.

3.2.2 Reagents

Lysozyme buffer, 10.0 mg/ml

Purpose: *For DNA isolation*

Lysozyme buffer consisting of 10.0 mM of Tris-HCl (pH 8), 1.0 mM of EDTA, 0.1 M of NaCl, and 20 % w/v of Triton X-100 was mixed together.

Sodium Lauryl Sulfate (SDS), 20 % w/v

Purpose: *To lyse the cell wall, denature chromosomal DNA and proteins as well as release plasmid DNA into the supernatant*

200.0 g of electrophoresis-grade SDS was dissolved in 900.0 ml of H_2O. It was then heated up to 68 °C with stirring to assist dissolution. The pH of the solution mix was adjusted to 7.2 by adding a few drops of concentrated hydrochloric acid (HCl). The volume was adjusted to 1.0 L with H_2O.

Proteinase K, 20.0 mg/ml

Purpose: *To digest native proteins and rapidly inactivate DNases and RNases in cell lysates, which facilitates the isolation of DNA*

Proteinase K was purchased as lyophilized powder and was dissolved at a concentration of 20.0 mg/ml in sterile 50.0 mM of Tris-HCl (pH 8.0) and 1.5 mM of calcium acetate ($C_4H_6CaO_4$). The stock solution was divided into small aliquots and stored at 4 °C.

Sodium Chloride (NaCl), 5.0 M, pH 5.2

Purpose: *To precipitate DNA and RNA*

5.0 M of NaCl was prepared by dissolving 87.7 g of NaCl crystals in 300.0 ml of distilled water. The solution was then autoclaved.

Sodium Acetate ($C_2H_3NaO_2$), 3.0 M, pH 5.2

Purpose: *To precipitate DNA*

408.3 g of $C_2H_3NaO_2$ was dissolved in 800.0 ml of distilled H_2O. The solution was adjusted to pH 5.2 with glacial acetic acid and adjusted to 1.0 L with H_2O. The solution was sterilized by autoclaving.

Ethanol, 80 % v/v

Purpose: *Acts as* a *disinfectant*

1.0 L of 80 % ethanol was prepared by mixing 800.0 ml of 99.9 % ethanol to 200.0 ml of autoclaved distilled water.

Glycerol solution, 65 % v/v

Purpose: *Long-term storage of bacterial strains at – 80 °C*

Glycerol solution was prepared by mixing 162.5 g of glycerol, 20.0 ml of 1.0 M $MgSO_4$, and 5.0 ml of 1.0 M Tris-HCl (pH 8), in a final volume of 200.0 ml.

Gel loading buffer, 5 X

Purpose: *To load and monitor migration of PCR products during gel electrophoresis*

The gel loading buffer consists of 3 different marker dyes (bromophenol blue, xylene cyanol and orange G) supplied as a 5 X concentrated solution. One volume of loading dye was added to 4 volumes of the PCR or DNA sample prior to loading the sample on an agarose gel.

Tris-hydrochloride (HCl) buffer, 1.0 M, pH 8.0

Purpose: *To inhibit enzymatic reactions*

Tris-HCl buffer was prepared by adding 121.1 g of Tris-base into 800.0 ml of distilled water. The pH was adjusted to 8.0 by adding HCl (~42.0 ml). The solution was autoclaved prior usage.

Ethylenediaminetetra-acetic acid (EDTA), 0.5 M, pH 8.0

Purpose: *To protect the nucleic acids against enzymatic degradation*

EDTA solution was prepared by adding 186.1 g of EDTA.$2H_2O$ into 800.0 ml of distilled water. While stirring vigorously with magnetic stirrer, its pH was adjusted to 8.0 with sodium hydroxide (NaOH) pellets (~80.0 g). The mixture was sterilized by autoclaving.

Tris-EDTA (TE), 1.0 M, pH 8.0

Purpose: *A buffer for slightly basic conditions to keep DNA deprotonated and soluble in water*

TE was prepared by mixing 100.0 mM of Tris-HCl with 10.0 mM of EDTA (pH 8.0) under sterile condition.

Tris/Borate/EDTA (TBE) buffer, 1 X

Purpose: *Buffer for dissolving agarose powder and gel electrophoresis*

1.0 L of 1 X TBE solution was prepared by diluting 100.0 ml of 10 X TBE solutions (Bio Basic Inc., Canada) with 900.0 ml of distilled water. The buffer solution was autoclaved prior usage.

Appendix 2

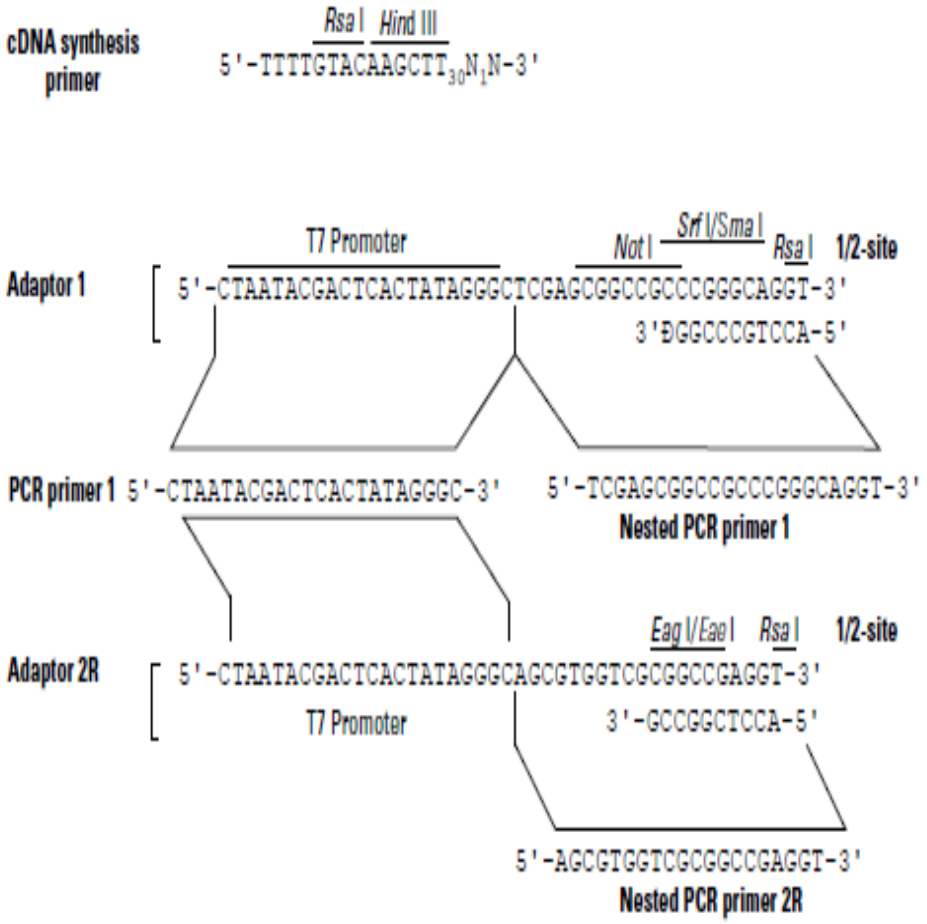


Figure A2: Sequences of the adaptor 1, adaptor 2R, PCR primer 1, nested PCR primer 1 and nested PCR primer 2R. When Adaptors 1 and 2R are ligated to *Rsa*I-digested DNA, the *Rsa*I site is restored.

Appendix 3

> CP028589.1:5035979-5038904 *Escherichia coli* strain WCHEC4533 chromosome complete genome

CTTCGGCGTTGTAAGGTTAAGCCTCACGGTTCATTAGTACCGGTTAGCTCAACGCAT
CGCTGCGCTTACACACCCGGCCTATCAACGTCGTCGTCTTCAACGTTCCTTCAGGAC
TCTCAAGGAGTCAGGGAGAACTCATCTCGGGGCAAGTTTCGTGCTTAGATGCTTTC
AGCACTTATCTCTTCCGCATTTAGCTACCGGGCAGTGCCATTGGCATGACAACCCGA
ACACCAGTGATGCGTCCACTCCGGTCCTCTCGTACTAGGAGCAGCCCCCCTCAGTTC
TCCAGCGCCCACGGCAGATAGGGACCGAACTGTCTCACGACGTTCTAAACCCAGCT
CGCGTACCACTTTAAATGGCGAACAGCCATACCCTTGGGACCTACTTCAGCCCCAG
GATGTGATGAGCCGACATCGAGGTGCCAAACACCGCCGTCGATATGAACTCTTGG
GCGGTATCAGCCTGTTATCCCCGGAGTACCTTTTATCCGTTGAGCGATGGCCCTTCC
ATTCAGAACCACCGGATCACTATGACCTGCTTTCGCACCTGCTCGCGCCGTCACGCT
CGCAGTCAAGCTGGCTTATGCCATTGCACTAACCTCCTGATGTCCGACCAGGATTA
GCCAACCTTCGTGCTCCTCCGTTACTCTTTAGGAGGAGACCGCCCCAGTCAAACTAC
CCACCAGACACTGTCCGCAACCCGGATAACGGGCCAACGTTAGAACATCAAACATT
AAAGGGTGGTATTTCAAGGTCGGCTCCATGCAGACTGGCGTCCACACTTCAAAGCC
TCCCACCTATCCTACACATCAAGGCTCAATGTTCAGTGTCAAGCTATAGTAAAGGTT
CACGGGGTCTTTCCGTCTTGCCGCGGGTACACTGCATCTTCACAGCGAGTTCAATTT
CACTGAGTCTCGGGTGGAGACAGCCTGGCCATCATTACGCCATTCGTGCAGGTCGG
AACTTACCCGACAAGGAATTTCGCTACCTTAGGACCGTTATAGTTACGGCCGCCGT
TTACCGGGGCTTCGATCAAGAGCTTCGCGTTACCGCTAACCCCATCAATTAACCTTC
CGGCACCGGGCAGGCGTCACACCGTATACGTCCACTTTCGTGTTTGCACAGTGCTG
TGTTTTTAATAAACAGTTGCAGCCAGCTGGTATCTTCGACTGATTTCAGCTCCACGA
GCAAGTCGCTTCACCTACATATCAGCGTGCCTTCTCCCGAAGTTACGGCACCATTTT
GCCTAGTTCCTTCACCCGAGTTCTCTCAAGCGCCTTGGTATTCTCTACCTGACCACCT
GTGTCGGTTTGGGGT----------ACGATTTGATGTTACCTGA
TGCTTAGAGGCTTTTCCTGGAAGCAGGGCATTTGTTGCTTCAGCACCGTAGTGCCTC
GTCATCACGCCTCAGCCTTGATTTTCCGGATTTGCCTGGAAAATCAGCCTACACGCT
TAAACCGGGACAACCGTCGCCCGGCCAACATAGCCTTCTCCGTCCCCCCTTCGCAGT
AACACCAAGTACAGGAATATTAACCTGTTTCCCATCGACTACGCCTTTCGGCCTCGC
CTTAGGGGTCGACTCACCCTGCCCCGATTAACGTTGGACAGGAACCCTTGGTCTTCC
GGCGAGCGGGCTTTTCACCCGCTTTATCGTTACTTATGTCAGCATTCGCACTTCTGA
TACCTCCAGCATACCTCACAGTACACCTTCACAGGCTTACAGAACGCTCCCCTACCC
AACAACACATAGTGTCGCTGCCGCAGCTTCGGTGCATGGTTTAGCCCCGTTACATCT
TCCGCGCAGGCCGACTCGACCAGTGAGCTATTACGCTTTCTTTAAATGATGGCTGCT
TCTAAGCCAACATCCTGGCTGTCTGGGCCTTCCCACATCGTTTCCCACTTAACCATG
ACTTTGGGACCTTAGCTGGCGGTCTGGGTTGTTTCCCTCTTCACGACGGACGTTAGC
ACCCGCCGTGTGTCTCCCGTGATAACATTCTCCGGTATTCGCAGTTTGCATCGGGTT

GGTAAGTCGGGATGACCCCCTTGCCGAAACAGTGCTCTACCCCCGGAGATGAATTC
ACGAGGCGCTACCTAAATAGCTTTCGGGGAGAACCAGCTATCTCCCGGTTTGATTG
GCCTTTCACCCCCAGCCACAAGTCATCCGCTAATTTTTCAACATTAGTCGGTTCGGTC

23S RNA Forward Primer: **CTACCTTAGGACCGTTATAGTTAC**

23S RNA Reverse Primer: **ACCATTTTGCCTAGTTCCTTC**

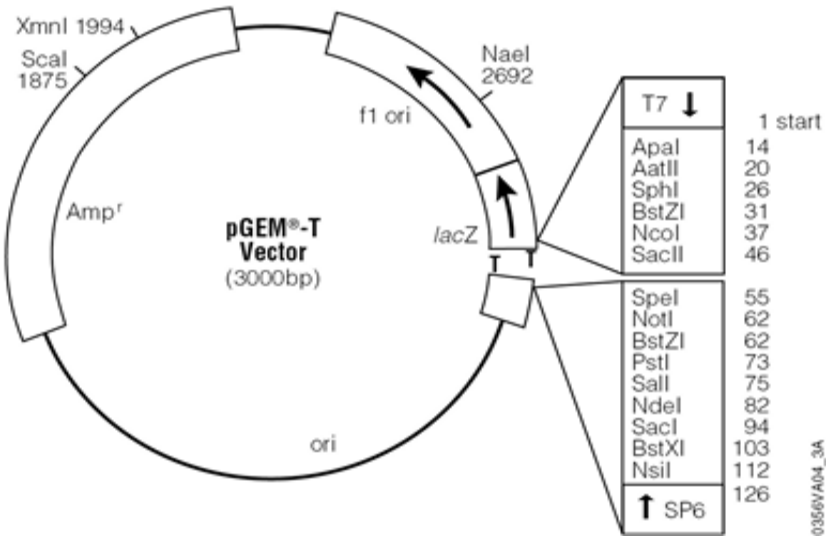
*Rsa*I restriction enzyme cutting site: GTAC/CATG

PCR Primer 1 binding site/Adaptor 1 or Adaptor 2R ligation site:

Figure A3: The partial genome sequences of *E. coli* which generate 23S rRNA gene using two gene-specific primers (23S RNA Forward and Reverse Primers). The product size is approximately 270 bp. As for the PCR product generated using one gene-specific primer (23S RNA Forward Primer) and PCR Primer 1; the product size is approximately 374 bp. Primer 1 shares similar binding sequences with adaptor 1 and adaptor 2R, which is about 40 bp in size.

Appendix 4

(I)



(II) pGEM®-T Easy Vector

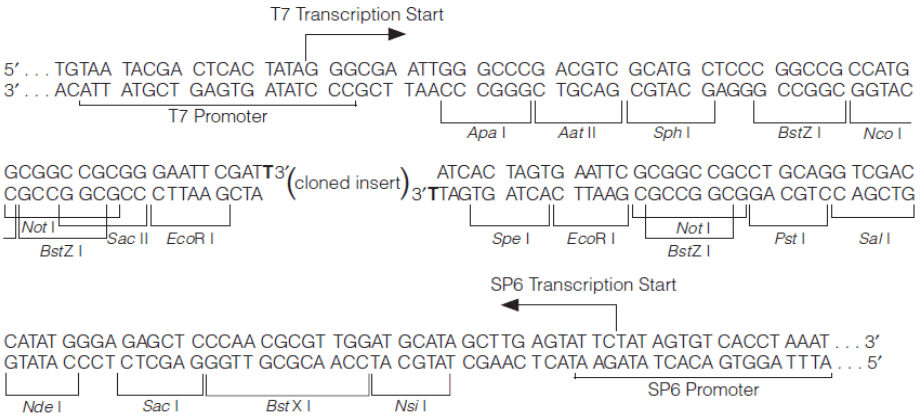


Figure A4: pGEM®-T Easy Vector. (I) pGEM®-T Vector Map and restriction enzyme points; (II) the promoter and multiple cloning sequence of the pGEM®-T Easy Vector. The top strand shown corresponds to the RNA synthesized by T7 RNA polymerase and the bottom strand corresponds to the RNA synthesized by SP6 RNA polymerase.

Appendix 5

Table A5: Reagent Cartridge Contents (Illumina, 2013).

Position	Reagent Name	Description
1	IMT	Incorporation Mix
2	USM	Scan Mix
4	CMS	Cleavage Mix
5	AMS1	Amplification Mix, Read 1
6	AMS2	Amplification Mix, Read 2
7	LPM	Linearization Premix
8	LDR	Formamide
9	LMX1	Linearization Mix
10	LMX2	Read 2 Linearization Mix
11	RMF	Resynthesis Mix
12	HP10	Read 1 Primer Mix
13	HP12	Index Primer Mix
14	HP11	Read 2 Primer Mix
15	PW1	Laboratory-grade water
16	PW1	Laboratory-grade water
17	Empty	Load Samples (Reserved for sample libraries)
18	Empty	Optional use for custom Read 1 primer
19	Empty	Optional use for custom Index Read primer
20	Empty	Optional use for custom Read 2 primer
21	PW1	Laboratory-grade water
22	Empty	Empty

Appendix 6

Table A6: Representative genomes retrieved from NCBI for comparative analysis.

NCBI Group ID	Representative Strain	Accession No.
1	H37Rv	NC_000962.3
2	0B235DS	CP008962.1
3	HN878	NZ_ADNF00000000.1
4	S96-129	NZ_AEGB00000000.1
5	98-R604 INH-RIF-EM	NZ_ABVM00000000.1
6	W-148	NZ_ACSX00000000.1
7	2074CJ	NZ_JKBN00000000.1
8	2230BH	NZ_JKBM00000000.1
9	M1734	NZ_JKMF00000000.1
10	TB_RSA57	NZ_JKVI00000000.1
11	Mtb194	NZ_AUNH00000000.1
12	OSDD071	NZ_AHHX00000000.1
13	KT-0187	NZ_JUFA00000000.1
14	Aethiop_vetus_233	CEGA00000000.1
15	A70011_4	CQFD00000000.1
16	A70136	CQFB00000000.1
17	A70448	CQFF00000000.1

Appendix 7

Table A7: Reads from SRA Depository for Variants Analysis.

Strains	SRA Number	Platform	Location
ERX009495	ERR023195	Illumina Genome Analyser II	Edinburg
ERX157998	ERR181888	Illumina HiSeq 2000	Malawi
ERX509474	ERR550405	Illumina MiSeq	Ireland
ERX662243	ERR718321	Illumina HiSeq 2000	Thailand
ERX662267	ERR718345	Illumina HiSeq 2000	Thailand
ERX662394	ERR718472	Illumina HiSeq 2000	Thailand
ERX680508	ERR736817	Illumina HiSeq 2000	Karonga Malawi
ERX680519	ERR736828	Illumina HiSeq 2000	Karonga Malawi
ERX701819	ERR757189	Illumina MiSeq	Argentina
M	ERR764904	Illumina MiSeq	Argentina
SRX473357	SRR1172814	Illumina HiSeq 2000	Stanger Hospital
SRX473469	SRR1172957	Illumina HiSeq 2000	Church of Scotland
SRX473507	SRR1173000	Illumina HiSeq 2000	Sihleza Clinic
SRX473631	SRR1173185	Illumina HiSeq 2000	Richard's Bay Clinic
SRX473638	SRR1173193	Illumina HiSeq 2000	Dundee Hospital
SRX473664	SRR1173223	Illumina HiSeq 2000	Thulasizwe Hopsital
SRX473782	SRR1173375	Illumina HiSeq 2000	St Margaret's Hospital
SRX473864	SRR1173482	Illumina HiSeq 2000	Doris Goodwin Hospital
SRX473807	SRR1173560	Illumina HiSeq 2000	Chwezi Clinic
SRX474483	SRR1174317	Illumina HiSeq 2000	Uganda
SRX475357	SRR1175470	Illumina HiSeq 2000	Romania
SRX479642	SRR1180449	Illumina HiSeq 2000	India: Tiruvallur
SRX480066	SRR1181090	Illumina HiSeq 2000	Buenaventura/Valle del Cauca/Colombia
SRX480028	SRR1184026	Illumina HiSeq 2000	Durban Chest Clinic
SRX481760	SRR1184322	Illumina HiSeq 2000	King Edward VIII Hospital

SRX481812	SRR1184378	Illumina HiSeq 2000	Thulasizwe Hospital
SRX493314	SRR1196472	Illumina HiSeq 2000	Ngwelezana Hospital
SRX493320	SRR1196478	Illumina HiSeq 2000	Richard's Bay Clinic
SRX493351	SRR1196511	Illumina HiSeq 2000	Goodwins Clinic
SRX493355	SRR1196515	Illumina HiSeq 2000	FOSA Hospital
SRX493361	SRR1196523	Illumina HiSeq 2000	Manguzi Hospital
SRX493377	SRR1196539	Illumina HiSeq 2000	Edendale Hospital
SRX493382	SRR1196544	Illumina HiSeq 2000	Dundee Hospital
SRX493388	SRR1196551	Illumina HiSeq 2000	Catherine Booth
SRX493397	SRR1196563	Illumina HiSeq 2000	Shallcross Clinic
SRX493410	SRR1196576	Illumina HiSeq 2000	Chwezi Clinic
SRX493412	SRR1196580	Illumina HiSeq 2000	M3 TB Hospital
SRX493417	SRR1196586	Illumina HiSeq 2000	FOSA Clinic
SRX493473	SRR1196677	Illumina HiSeq 2000	Westville Prison
SRX493481	SRR1196685	Illumina HiSeq 2000	Mpumuza Clinic
SRX493499	SRR1196702	Illumina HiSeq 2000	Siloah Clinic
SRX493519	SRR1196723	Illumina HiSeq 2000	King Dinuzulu Hospital
SRX493557	SRR1196768	Illumina HiSeq 2000	King Dinuzulu Hospital
SRX648745	SRR1510041	Illumina HiSeq 2500	Manila
SRX648759	SRR1510055	Illumina MiSeq	Manila
SRX648774	SRR1510070	Illumina HiSeq 2500	Manila
SRX691156	SRR1564305	Illumina HiSeq 2000	Thailand
SRX699671	SRR1573680	Illumina MiSeq	Canada Ontario
SRX699698	SRR1573707	Illumina MiSeq	Canada Ontario
SRX691501	SRR1595970	Illumina HiSeq 2000	Thailand
MT-578	SRR1640752	Illumina MiSeq	Montreal Hospital
SRX816244	SRR1722712	Illumina HiSeq 2000	Russian Academy of Sciences
SRX816309	SRR1722867	Illumina HiSeq 2000	Russian Academy of Sciences
SRX347313	SRR974840	Illumina HiSeq 2000	South Africa: Western Cape
SRX347319	SRR974846	Illumina HiSeq 2000	Netherlands: Haarlem
SRX347317	SRR974849	Illumina HiSeq 2000	China

Appendix 8

Table A8: The 63 proteins reported in scientific literature for *Mtb* associated with meningitis.

Locus	Reference
Rv0311	Be et al. 2008
Rv0805	Be et al. 2008
Rv0931c	Be et al. 2008; Be et al. 2012
Rv0986	Be et al. 2008; Jain et al. 2006
Rv0368c	Jain et al. 2006
Rv0573c	Jain et al. 2006
Rv0619	Jain et al. 2006
Rv0661c	Jain et al. 2006
Rv0662c	Jain et al. 2006
Rv0966c	Jain et al. 2006
Rv0967	Jain et al. 2006
Rv0968	Jain et al. 2006
Rv0970	Jain et al. 2006
Rv0971c	Jain et al. 2006
Rv0974c	Jain et al. 2006
Rv0975c	Jain et al. 2006
Rv0977	Jain et al. 2006
Rv0978c	Jain et al., 2006
Rv0980c	Jain et al. 2006
Rv0982	Jain et al. 2006
Rv0983	Jain et al. 2006
Rv0984	Jain et al. 2006
Rv0987	Jain et al. 2006
Rv0989c	Jain et al. 2006
Rv0990c	Jain et al. 2006
Rv0991c	Jain et al. 2006
Rv1726	Jain et al. 2006
Rv1801	Jain et al. 2006

Rv1966	Jain et al. 2006
Rv1968	Jain et al. 2006
Rv2318	Jain et al. 2006
Rv3021c	Jain et al. 2006
Rv3349c	Jain et al. 2006
Rv3351c	Jain et al. 2006
Rv3639c	Jain et al. 2006
Rv3833	Jain et al. 2006
Rv0079	Be et al. 2012
Rv0336	Be et al. 2012
Rv0727c	Be et al. 2012
Rv0755c	Be et al. 2012
Rv1273c	Be et al. 2012
Rv1673c	Be et al., 2012
Rv1914c	Be et al., 2012
Rv1932	Be et al., 2012
Rv2387	Be et al., 2012
Rv3094c	Be et al., 2012
Rv3159c	Be et al., 2012
Rv3224	Be et al., 2012
Rv3353c	Be et al., 2012
Rv1837c	Haldar et al., 2012
Rv0475	Pethe et al., 2001
Rv2947c	Tsenova et al., 2005
Rv2946c	Tsenova et al., 2005
Rv0015c	Av-Gay & Everett, 2000
Rv0014c	Av-Gay & Everett, 2000
Rv1743	Av-Gay & Everett, 2000
Rv1746	Av-Gay & Everett, 2000
Rv0410c	Av-Gay & Everett, 2000
Rv1266c	Av-Gay & Everett, 2000
Rv2914c	Av-Gay & Everett, 2000
Rv2088	Av-Gay & Everett, 2000
Rv3080c	Av-Gay & Everett, 2000
Rv2176	Av-Gay & Everett, 2000

Appendix 9

Table A9: The 141 proteins reported to be associated with *Streptococcus pneumonia* meningitis (Orihuela et al., 2004; Molzen et al., 2011; Mahdi et al., 2012).

Locus	Reference
SP_0019	Molzen et. al., 2011
SP_0029	Molzen et. al. , 2011
SP_0042	Molzen et. al. , 2011
SP_0058	Molzen et. al. , 2011
SP_0067	Molzen et. al. , 2011
SP_0079	Molzen et. al. , 2011
SP_0098	Molzen et. al. , 2011
SP_0099	Molzen et. al. , 2011
SP_0101	Molzen et. al. , 2011
SP_0149	Molzen et. al. , 2011
SP_0151	Molzen et. al. , 2011
SP_0152	Molzen et. al. , 2011
SP_0197	Molzen et. al. , 2011
SP_0198	Molzen et. al. , 2011
SP_0199	Molzen et. al. , 2011
SP_0276	Molzen et. al. , 2011
SP_0279	Molzen et. al. , 2011
SP_0282	Molzen et. al. , 2011
SP_0308	Molzen et. al. , 2011
SP_0342	Molzen et. al. , 2011

SP_0348	Molzen et. al. , 2011
SP_0349	Molzen et. al. , 2011
SP_0350	Molzen et. al. , 2011
SP_0351	Molzen et. al. , 2011
SP_0358	Molzen et. al. , 2011
SP_0416	Molzen et. al. , 2011
SP_0494	Molzen et. al. , 2011
SP_0552	Molzen et. al. , 2011
SP_0585	Molzen et. al. , 2011
SP_0665	Molzen et. al. , 2011
SP_0695	Molzen et. al. , 2011
SP_0731	Molzen et. al. , 2011
SP_0746	Molzen et. al. , 2011
SP_0748	Molzen et. al. , 2011
SP_0749	Molzen et. al. , 2011
SP_0751	Molzen et. al. , 2011
SP_0752	Molzen et. al. , 2011
SP_0753	Molzen et. al. , 2011
SP_0822	Molzen et. al. , 2011
SP_0823	Molzen et. al. , 2011
SP_0826	Molzen et. al. , 2011
SP_0881	Molzen et. al. , 2011
SP_0925	Molzen et. al. , 2011
SP_0931	Molzen et. al. , 2011
SP_1025	Molzen et. al. , 2011
SP_1059	Molzen et. al. , 2011
SP_1062	Molzen et. al. , 2011
SP_1063	Molzen et. al. , 2011
SP_1068	Molzen et. al. , 2011
SP_1069	Molzen et. al. , 2011
SP_1121	Molzen et. al. , 2011

SP_1296	Molzen et. al. , 2011
SP_1297	Molzen et. al. , 2011
SP_1298	Molzen et. al. , 2011
SP_1299	Molzen et. al. , 2011
SP_1330	Molzen et. al. , 2011
SP_1331	Molzen et. al. , 2011
SP_1336	Molzen et. al. , 2011
SP_1356	Molzen et. al. , 2011
SP_1376	Molzen et. al. , 2011
SP_1393	Molzen et. al. , 2011
SP_1462	Molzen et. al. , 2011
SP_1465	Molzen et. al. , 2011
SP_1466	Molzen et. al. , 2011
SP_1502	Molzen et. al. , 2011
SP_1507	Molzen et. al. , 2011
SP_1544	Molzen et. al. , 2011
SP_1563	Molzen et. al. , 2011
SP_1635	Molzen et. al. , 2011
SP_1645	Molzen et. al. , 2011
SP_1799	Molzen et. al. , 2011
SP_1931	Molzen et. al. , 2011
SP_1966	Molzen et. al. , 2011
SP_1995	Molzen et. al. , 2011
SP_2021	Molzen et. al. , 2011
SP_2098	Molzen et. al. , 2011
SP_2116	Molzen et. al. , 2011
SP_2198	Molzen et. al. , 2011
SP_2205	Molzen et. al. , 2011
SP_2206	Molzen et. al. , 2011
SP_2231	Molzen et. al. , 2011
SP_0088	Orihuela et al ., 2004

SP_0111	Orihuela et al., 2004
SP_0117	Orihuela et al., 2004
SP_0149	Orihuela et al., 2004
SP_0151	Orihuela et al., 2004
SP_0173	Orihuela et al., 2004
SP_0205	Orihuela et al., 2004
SP_0265	Orihuela et al., 2004
SP_0266	Orihuela et al., 2004
SP_0348	Orihuela et al., 2004
SP_0446	Orihuela et al., 2004
SP_0481	Orihuela et al., 2004
SP_0483	Orihuela et al., 2004
SP_0629	Orihuela et al., 2004
SP_0641	Orihuela et al., 2004
SP_0736	Orihuela et al., 2004
SP_0750	Orihuela et al., 2004
SP_0751	Orihuela et al., 2004
SP_0752	Orihuela et al., 2004
SP_0753	Orihuela et al., 2004
SP_0792	Orihuela et al., 2004
SP_0794	Orihuela et al., 2004
SP_0800	Orihuela et al., 2004
SP_0841	Orihuela et al., 2004
SP_1014	Orihuela et al., 2004
SP_1122	Orihuela et al., 2004
SP_1160	Orihuela et al., 2004
SP_1249	Orihuela et al., 2004
SP_1267	Orihuela et al., 2004
SP_1268	Orihuela et al., 2004
SP_1269	Orihuela et al., 2004
SP_1270	Orihuela et al., 2004

SP_1271	Orihuela et al., 2004
SP_1276	Orihuela et al., 2004
SP_1277	Orihuela et al., 2004
SP_1310	Orihuela et al., 2004
SP_1468	Orihuela et al., 2004
SP_1550	Orihuela et al., 2004
SP_1551	Orihuela et al., 2004
SP_1587	Orihuela et al., 2004
SP_1648	Orihuela et al., 2004
SP_1649	Orihuela et al., 2004
SP_1650	Orihuela et al., 2004
SP_1675	Orihuela et al., 2004
SP_1710	Orihuela et al., 2004
SP_1725	Orihuela et al., 2004
SP_1735	Orihuela et al., 2004
SP_1804	Orihuela et al., 2004
SP_1805	Orihuela et al., 2004
SP_1888	Orihuela et al., 2004
SP_1889	Orihuela et al., 2004
SP_1890	Orihuela et al., 2004
SP_1909	Orihuela et al., 2004
SP_1968	Orihuela et al., 2004
SP_2021	Orihuela et al., 2004
SP_2095	Orihuela et al., 2004
SP_2101	Orihuela et al., 2004
SP_2196	Orihuela et al., 2004
SP_2201	Orihuela et al., 2004
SP_2185	Mahdi et al., 2012

Appendix 10

Table A10: The reported 164 virulence genes derived from *Neisseria meningitis* (Hao et al., 2011).

ID	Reference
NMA0692	Hao et al.,2011
NMA0693	Hao et al.,2011
NMA0694	Hao et al.,2011
NMA1792	Hao et al.,2011
NMA1793	Hao et al.,2011
NMA1794	Hao et al.,2011
NMA1795	Hao et al.,2011
NMA1796	Hao et al.,2011
NMA1797	Hao et al.,2011
NMA1798	Hao et al.,2011
NMA1799	Hao et al.,2011
NMB0014	Hao et al.,2011
NMB0017	Hao et al.,2011
NMB0018	Hao et al.,2011
NMB0033	Hao et al.,2011
NMB0038	Hao et al.,2011
NMB0051	Hao et al.,2011
NMB0064	Hao et al.,2011
NMB0065	Hao et al.,2011
NMB0071	Hao et al.,2011

NMB0072	Hao et al.,2011
NMB0073	Hao et al.,2011
NMB0074	Hao et al.,2011
NMB0082	Hao et al.,2011
NMB0083	Hao et al.,2011
NMB0178	Hao et al.,2011
NMB0180	Hao et al.,2011
NMB0182	Hao et al.,2011
NMB0199	Hao et al.,2011
NMB0205	Hao et al.,2011
NMB0216	Hao et al.,2011
NMB0218	Hao et al.,2011
NMB0278	Hao et al.,2011
NMB0293	Hao et al.,2011
NMB0294	Hao et al.,2011
NMB0299	Hao et al.,2011
NMB0318	Hao et al.,2011
NMB0319	Hao et al.,2011
NMB0329	Hao et al.,2011
NMB0330	Hao et al.,2011
NMB0332	Hao et al.,2011
NMB0333	Hao et al.,2011
NMB0364	Hao et al.,2011
NMB0365	Hao et al.,2011
NMB0374	Hao et al.,2011
NMB0375	Hao et al.,2011
NMB0382	Hao et al.,2011
NMB0391	Hao et al.,2011
NMB0394	Hao et al.,2011
NMB0407	Hao et al.,2011

NMB0413	Hao et al.,2011
NMB0460	Hao et al.,2011
NMB0467	Hao et al.,2011
NMB0483	Hao et al.,2011
NMB0493	Hao et al.,2011
NMB0496	Hao et al.,2011
NMB0497	Hao et al.,2011
NMB0546	Hao et al.,2011
NMB0555	Hao et al.,2011
NMB0567	Hao et al.,2011
NMB0584	Hao et al.,2011
NMB0585	Hao et al.,2011
NMB0586	Hao et al.,2011
NMB0632	Hao et al.,2011
NMB0633	Hao et al.,2011
NMB0634	Hao et al.,2011
NMB0652	Hao et al.,2011
NMB0653	Hao et al.,2011
NMB0663	Hao et al.,2011
NMB0675	Hao et al.,2011
NMB0686	Hao et al.,2011
NMB0718	Hao et al.,2011
NMB0750	Hao et al.,2011
NMB0757	Hao et al.,2011
NMB0828	Hao et al.,2011
NMB0879	Hao et al.,2011
NMB0884	Hao et al.,2011
NMB0886	Hao et al.,2011
NMB0888	Hao et al.,2011
NMB0902	Hao et al.,2011

NMB0905	Hao et al.,2011
NMB0906	Hao et al.,2011
NMB0992	Hao et al.,2011
NMB0995	Hao et al.,2011
NMB1048	Hao et al.,2011
NMB1072	Hao et al.,2011
NMB1077	Hao et al.,2011
NMB1162	Hao et al.,2011
NMB1206	Hao et al.,2011
NMB1207	Hao et al.,2011
NMB1210	Hao et al.,2011
NMB1214	Hao et al.,2011
NMB1220	Hao et al.,2011
NMB1283	Hao et al.,2011
NMB1332	Hao et al.,2011
NMB1368	Hao et al.,2011
NMB1372	Hao et al.,2011
NMB1398	Hao et al.,2011
NMB1403	Hao et al.,2011
NMB1405	Hao et al.,2011
NMB1409	Hao et al.,2011
NMB1412	Hao et al.,2011
NMB1414	Hao et al.,2011
NMB1415	Hao et al.,2011
NMB1428	Hao et al.,2011
NMB1429	Hao et al.,2011
NMB1442	Hao et al.,2011
NMB1475	Hao et al.,2011
NMB1494	Hao et al.,2011
NMB1527	Hao et al.,2011

NMB1540	Hao et al.,2011
NMB1541	Hao et al.,2011
NMB1572	Hao et al.,2011
NMB1591	Hao et al.,2011
NMB1621	Hao et al.,2011
NMB1622	Hao et al.,2011
NMB1646	Hao et al.,2011
NMB1668	Hao et al.,2011
NMB1688	Hao et al.,2011
NMB1704	Hao et al.,2011
NMB1705	Hao et al.,2011
NMB1714	Hao et al.,2011
NMB1715	Hao et al.,2011
NMB1716	Hao et al.,2011
NMB1717	Hao et al.,2011
NMB1719	Hao et al.,2011
NMB1730	Hao et al.,2011
NMB1738	Hao et al.,2011
NMB1739	Hao et al.,2011
NMB1753	Hao et al.,2011
NMB1763	Hao et al.,2011
NMB1768	Hao et al.,2011
NMB1780	Hao et al.,2011
NMB1802	Hao et al.,2011
NMB1808	Hao et al.,2011
NMB1809	Hao et al.,2011
NMB1810	Hao et al.,2011
NMB1811	Hao et al.,2011
NMB1812	Hao et al.,2011
NMB1814	Hao et al.,2011

NMB1820	Hao et al.,2011
NMB1821	Hao et al.,2011
NMB1822	Hao et al.,2011
NMB1829	Hao et al.,2011
NMB1845	Hao et al.,2011
NMB1870	Hao et al.,2011
NMB1882	Hao et al.,2011
NMB1906	Hao et al.,2011
NMB1926	Hao et al.,2011
NMB1928	Hao et al.,2011
NMB1929	Hao et al.,2011
NMB1946	Hao et al.,2011
NMB1989	Hao et al.,2011
NMB1990	Hao et al.,2011
NMB1991	Hao et al.,2011
NMB2001	Hao et al.,2011
NMB2015	Hao et al.,2011
NMB2039	Hao et al.,2011
NMB2048	Hao et al.,2011
NMB2127	Hao et al.,2011
NMB2132	Hao et al.,2011
NMB2152	Hao et al.,2011
NMB2156	Hao et al.,2011
NMB2160	Hao et al.,2011

Appendix 11

Table A11: The neurotropic genes (73) reported to confer tissue tropism in *Neisseria meningitis* and *E. coli K*1 (Pouttu et al., 1999; Huang et al., 2001; Yao et al., 2006).

Gene	Reference
FimH	Pouttu et al.,1999
IbeA	Huang et al.,2001
CNF1	Yao et al., 2006
aslA	Yao et al., 2006
fhuA	Yao et al., 2006
fyuA	Yao et al., 2006
hlyABCD	Yao et al., 2006
hlyE	Yao et al., 2006
iha	Yao et al., 2006
iroN	Yao et al., 2006
iucABCD	Yao et al., 2006
iutA	Yao et al., 2006
malX	Yao et al., 2006
ompA	Yao et al., 2006
papA	Yao et al., 2006
flu	Yao et al., 2006
traJ	Yao et al., 2006
sitABCD	Yao et al., 2006
apbE	Yao et al., 2006

blc	Yao et al., 2006
cutF	Yao et al., 2006
lgt	Yao et al., 2006
lnt	Yao et al., 2006
lolA	Yao et al., 2006
lpp	Yao et al., 2006
nlpB	Yao et al., 2006
nlpC	Yao et al., 2006
nlpD	Yao et al., 2006
nlpI	Yao et al., 2006
osmB	Yao et al., 2006
rlpAB	Yao et al., 2006
slyB	Yao et al., 2006
spr	Yao et al., 2006
vacJ	Yao et al., 2006
yafL	Yao et al., 2006
yajG	Yao et al., 2006
yehR	Yao et al., 2006
yfiO	Yao et al., 2006
yqhH	Yao et al., 2006
yhiU	Yao et al., 2006
bglH	Yao et al., 2006
btuB	Yao et al., 2006
cirA	Yao et al., 2006
fadL	Yao et al., 2006
fecA	Yao et al., 2006
fepA	Yao et al., 2006
nfrA	Yao et al., 2006
nmpC	Yao et al., 2006
ompC	Yao et al., 2006

ompF	Yao et al., 2006
ompG	Yao et al., 2006
ompN	Yao et al., 2006
ompT	Yao et al., 2006
ompW	Yao et al., 2006
pldA	Yao et al., 2006
sfmD	Yao et al., 2006
slp	Yao et al., 2006
yaeT	Yao et al., 2006
yaiV	Yao et al., 2006
ybiL	Yao et al., 2006
ycbS	Yao et al., 2006
yehB	Yao et al., 2006
yejO	Yao et al., 2006
yfiB	Yao et al., 2006
yiaD	Yao et al., 2006
yiaT	Yao et al., 2006
yjcP	Yao et al., 2006
yjhA	Yao et al., 2006
yjiK	Yao et al., 2006
yncD	Yao et al., 2006
yohG	Yao et al., 2006
ypjA	Yao et al., 2006
ytfM	Yao et al., 2006

Appendix 12

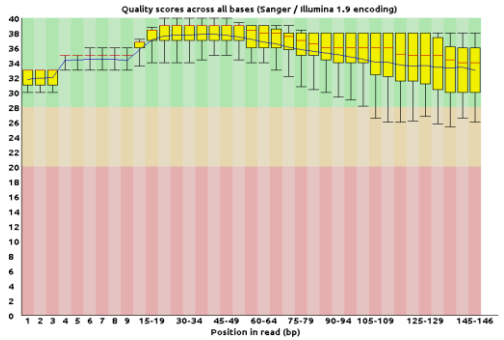
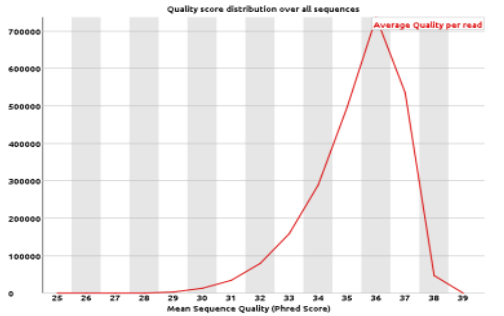
Table A12: Nucleotide sequences of the seven subtracted DNAs**.** These sequences have been filtered and trimmed from other 'contaminated' sequences such as the vector's sequences and background noises.

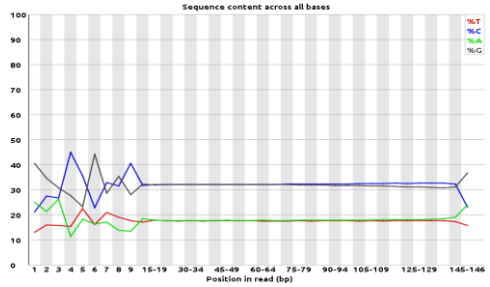
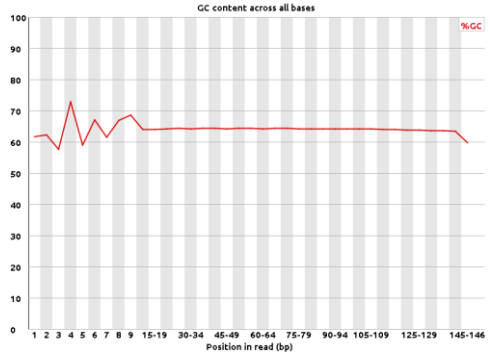
No	Positive Clones	Subtracted Sequences	Size (bp)
S1	pGT-F1-1-2	CTATTCGCCCTAGACTGCCGCAGTTCCGGGGGAAGTGAACCTATTGCGCCCGTG CATCACTGCACGGGTATGGGCTTTGGCGGTCGCTTCGCACCATCAACGCCGACA GTGCGGACAGCGCAAACCGACGGCGCACCCCTTGCACGGATGTGGGGTGTTTTT GAGATGGAGCGAAAGTAGGCGTGTCTTTTATTTTCACAACCCCCCAGGCATTGG ACAACGCGGCTAAGTCCGTGTCGGGGATTCACGATTTGTGGCGCAAAGGACGC TAAGGCATCGATCCCGGTGGTCAACGCTATTTGAGCCCCCCGCTTCCGACCCGG TGTCGAATAGGGATGAGGCCGCTCCTCCGCCAGCACATGAGGCAGTATCACCA GATCAGCTTTCCGGCCATAGAGCATCGTCACCGGGTTAGGCATGGTTTAGGCAG CGCTTAGCTGAGAACGCCGAGGCGTGTCGGCTCGCCGAGGCCCAAAACAGCAC AACCTTGCACTGATCTAGCTGAAGACCAAACCGGCACAGCAGACATTGCCATA CGCGACAACAGCCGTCATCAACCGAAAGGAGCAAAGAACAAACAGATGCATCC AATGATACCAGCGGAGTATATCTCCAACATAATATATGAAGGCCCGGGCGCTG ACTCATTGTTTTTCGCCTCCGGGCAATTGCGAGAATTGGCTTACTCAGTTGAAA CGACGGCTGAGTCGCTCGAGGACGAGCTCGACGAGCTGGATGAGAACTGGAAA GGTAGTTCGTCGGACTTGTTGGCCGACGCGGTTGAGCGGTATCTCCAATGGCTG TCTAACACTCCAGTCAGCTTAAGCATGCCGCCTGGGTGATCAACGGCCTCGCGA ACGCCTATAACGACACACGTCGGAAGGTG	886
S2	pGT-F1-4-24/30	AGGCGCTCGATAACCCGTTTCATCACCGAGCTGTTGTCGCCCAGCCAGATCGGC CCGCACAGCACCAGGATGTGCGCATCGAGGACACGCCGATACAGGGCGGGCCA TTCGTCGGTCGCCCAACCGTGTTCGGTCATGTCCGGCCATACGCCGGTCGCTAT GTCATGGTCAACTGCGCGCAGAGTGTCGACCTGGACGCCATGCTCACGCATGAT CCCCGAGCTGCGCTCAATGAGCCCGTCGGTATGGCTGAGCTCTGGCGAGCGCTT CAGTGTCGCGTTGATGAACAGCGCACGCAGCCCGTCGAATCGGGGTGGGGCCG CGGCGTTCTGGTCAGAGGTTGTGGTCATACGTCATACCCACCTGCCTGTCATCG TCGTGCCGGGTTGCCGCTGGGCGGCGGTGCTGGTGCCAAGAAATGACCGATCA GGCAGCAGCGTA	441

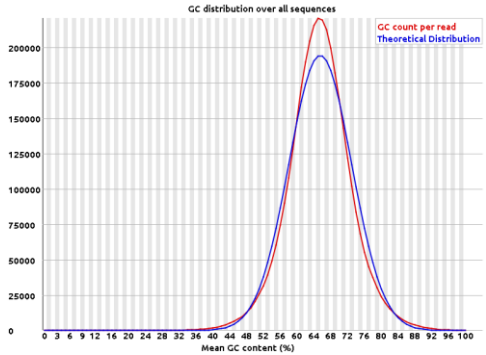
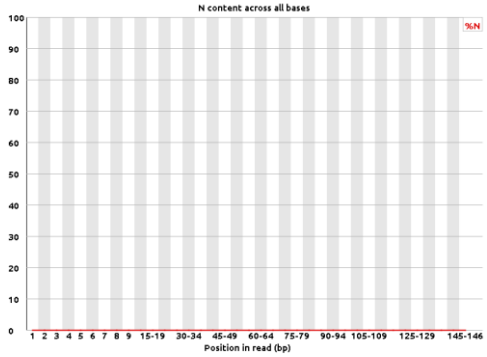
S3	pGT-F1-5-37/39	ATGGAAAATCACCGTGACGAGCTGGTGGTGTTTGGTGCGGGTTACGCCAAAGC GATGGAGAAAATGCTCGAGGTGAATCAGGGTCTGCGCCGGCGCTTTTCGACGG TGATCGAGTTCTTCAGCTACACCCCGCAGGAGCTGATCGCACTGACCCAGCTGA TGGGTCGGGAGAACGAAGACGTGATCACTGAGGAAGAGTCTCAAGTGTTGTTG CCGTCGTATACCAAGTTCTACATGGAGCAGAGCTACTCCGAGGACGGCGACCT GATCCGCGGGATCGATCTGTTGGGCAATGCCGGCTTTGTGCGCAACGTGGTGGA GAAGGCCCGCGACCACCGTAGTTTCCGTTTGGACGATGAGGATCTCGACGCCGT A	375
S4	pGT-F1-6-41	CAGCGCACGTGCTCCTCGAAGAGCTGGTGTTTCTCGGCCTCGGTCAAGCCTCCG CAGAAGTATTCGGCGACCTTGATGACAAGCATGAAAAACGTCGCATGCGCCCA GTAGAACGTATCTGGATTCAGCGCGTGATAGCGACGCCCCTCAGCGTCGACTCC CTTGATGGTTCGGTGGTAGCCCTTGATCTGCTGGCCGGTCTGGGCCGCTCGGTC ACCGTCATAGACCACACCCATGATCGGGTA	245
S5	pGT-B4-5-50/51	TTTCCGATCACCAAGCTTGAGGAGATGGTCAACGTCGGGACACTGTTCGCGTTC ATCCTCGTCTCGGCCGGAGTGGTCGTGCTGCGCCGGACCCGACCCGACCTTCAG CGGGGGTTCACAGCTCCGTGGGTGCCGTTACTTCCGATCGCCGCAGTGTGCGCG TGCCTGTGGCTGATGCTGAACCTCACCGCGTTGACTTGGATCCGGTTCGGGATC TGGCTGGTGGCCGGAACCGCGATTTATGTCGGCTACGGGCGCCGGCACTCGGC GCAGGGCCTTCGGCAAGCTCGAGAGAGCGCGACCCGGAGGTGTTGAACTAGCT TCGCCGCGTATTTACAAATTGCGTTATATGTCTAGACATAAGACGCAAACTGCT CTATTGTCAAATCCCAACGTGGTGTGGCTCATGAGGAGGTTTGGCATATGACCA CGCAAATCGGCTCTGGTGGACCCGAAGCCCCCCGGCCGCCCGAAGTCGAGGAG TGGCGGGACAAGAAGCGTTACCTGTGGCTTATGGGCCTCATCGCCCCGACGGCC TTGGTGGTGATGCTGCCGCTGATCTGGGGGATGAACCAGCTCGGCTGGCACGCC GCCGCGCAGGTGCCGCTGTGGATCGGACCGATCCTGCTCTACGTCTTGTTGCCG CTTCTTGACCTACGCTTCGGGCCCGACGGGCAGAACCCGCCCGACGAGGTGACC GACCGGCTGGAGAATGACAAGTA	722
S6	pGT-B4-7-56	ACCGGCGTGCTGTGCTCACGAGCAAACCAGATTGCCGAGCCCAGGTGGTCGAT GTGCGCGTGGGTGAGCACGATGGCGCGCACGTCACCCGGTGTGTAGCCCAGTTT GTTCAGCGAGGCCAGCACCTCCGCACGGTCGCCGGGATAGCCGGCGTCGATCA GCAGCACGCCGGTGTCGTCGGTGACTAGCACCCAGTTGACCGCGTGGCCGCGA GCGAGGTGAACCTTGTCGGTGATCTGAACAAGCTCCGCCATGCCCGCGAGTCTA GGAGCGAGCGCGAGCGCGGCAAGCCGGGTGCCGCGGGTCGCGACCATGGGAT ATGGAGCGATCGCGAGCGCGGCGAAGCCGGGCGTGGCGGGTCGCGTTTATGGC ATAGGAGTAGAAGGAACTGGTGGCTGAACTGAAGCTAGGTTACAAAGCATCGG CCGAACAATTCGCACCGCGCGAGCTCGTCGAACTAGCCGTCGCCGCCGAAGCC CACGGCATGGACAGCGCGACCGTCAGCGACCATTTTCAGCCTTGGCGCCACCA GGGCGGCCATGCCCCGTTCTCGCTGTCCTGGATGACCGCTGTCGGCGAAC	581
S7	pGT-B4-9-65/67	CACGGCAGCCTCCGGGAACGCTGCCAGCACAGCCTGCTCGCCGGCGAGTTTGCT GCGGGCATACACGCCCTGCGGCGCGGTTTCATCGGTGGGCTCGTAGGGCCGGG GCTCGGCGCCGCCGAAGTCGCCATCGAATACGTAGTCGGTGGAGACGTGGATT AACCGAGCACCCACACGAGCGCACGCACGGGCGAGGTGTTGCGGGCCAGTGGC ATTGACCGCATAGGCGACTGCCTCGTTGCTCTCGGCGCCGTCGACGTCGGTGTA GGCGGCGCAATTGATCACCACGTCACCGTGTCGGATGATCCGCTCGGCCGCAGC GGGGTCGGTGATATCCCACTGCGAGGAAGTCAGCGCCAGCATATCGCGG	370

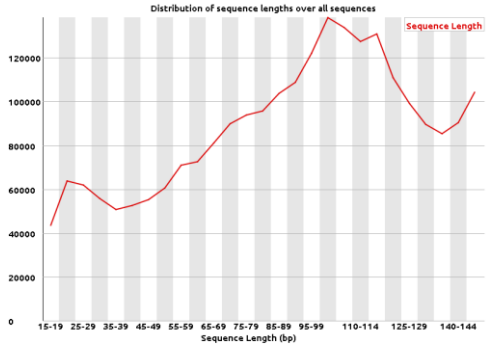
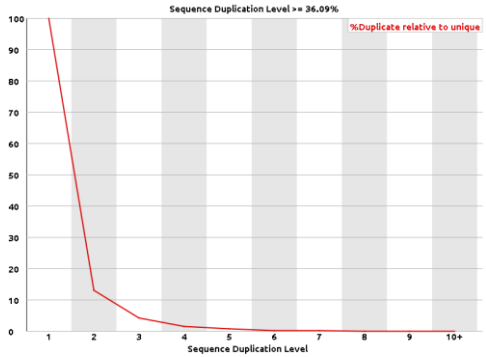
Appendix 13

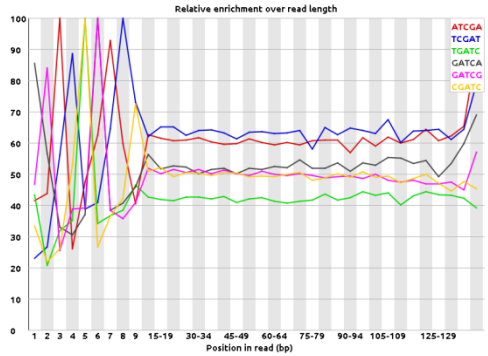
Table A13: The quality of raw sequences generated from MiSeq was checked using FastQC. Raw reads were trimmed at Phred probability score of 30 and were *de novo* assembled using CLC Genomic Workbench 5.1 (Qiagen Inc., Netherlands). FastQC is a free piece of software written by Babraham Bioinformatics group. One of the most important plots from FastQC is 'per sequence quality'. Phred score on the other hand is the base calling program for DNA sequence traces. Phred reads DNA sequence chromatogram files and analyses the peaks to call bases, assigning quality scores to each base call. This table shows the plots available for FastQC and the results obtained from UM-CSF01 as an example.

Summary	Description	Calculation	Graph	Result (UM-CSF01)
Per base sequence quality	The view shows an overview of the range of quality values across all bases at each position in the FastQ file.	FastQC attempts to automatically determine which encoding method is used and the title of the graph will describe the encoding FastQC.	BoxWhisker type plot shows: •Central red line: median value. •Yellow box: inter-quartile range (25-75%).	
Per sequence quality scores	The report allows determining if a subset of the sequences has universally low quality values.	A subset sequences will have universally poor quality because they are poorly imaged, however these should represent only a small percentage of the total sequences.	If a significant proportion of the sequences in a run have overall low quality then this could indicate some kind of systematic problem.	

Per base sequence content	The content plots out the proportion of each base position in a file for which each of the four normal DNA bases has been called.	The relative amount of each base should reflect the overall amount of these bases in the genome, but in any case they should not be hugely imbalanced from each other. If strong bias has been observed in different bases, it indicates an overrepresented sequence which is contaminating the library.	The line in this plot should run parallel with each other assuming that there would be little or no difference bases of a sequence run.	
Per base GC content	It plots out the GC content of each base position in a file.	The overall GC content should reflect the GC content of the underlying genome. GC bias (changes in different bases) indicates an overrepresented sequence which is contaminating the library. A bias which is consistent across all bases either indicates that the original library was sequence biased, or that there was a systematic problem during the sequencing of the library.	The line in this plot should run horizontally across the graph since there would be little or no difference between the different bases of a sequence run.	

Per sequence GC content	This module measures the GC content across the whole length of each sequence in a file and compares it to a modelled normal distribution of GC content.	A normal distribution of GC content will be seen in a normal random library, where the central peak corresponds to the overall GC content of the underlying genome. The modal GC content is calculated from the observed data and used to build a reference distribution.	A contaminated library or some other kinds of biased subset will be seen as an unusually shaped distribution. A normal distribution which is shifted indicates some systematic bias which is independent of base position.	
Per base N content	This module plots out the percentage of base calls at each position for which an N was called. It may identify cycles that are unreliable.	If proportion of Ns rises above a few % suggests that the analysis pipeline was unable to interpret the data well enough to make valid base calls.	A very low proportion of Ns will appear in a sequence especially nearer the end of a sequence.	

Sequence length distribution	This module shows the distribution of fragment sizes in the file which was analysed.	It generates the sequence fragments of uniform length or wildly varying lengths.	A simple graph showing a peak only at one size, but for variable length.	
Sequence duplication levels	This module counts the degree of duplication for every sequence in the set and creates a plot showing the relative number of sequences with different degrees of duplication. The first 200k sequences in each file are analysed to cut down on the memory requirements.	Low value indicates a very high level of coverage of the target sequence but high value indicates some kind of enrichment bias (PCR over amplification).	Each sequence is tracked to the end of the file to give a representative count of the overall duplication level. Sequences with more than 10 duplicates are placed into the 10 duplicates category.	

Kmer content	This module counts the enrichment of every 5-mer within the sequence library. Any k-mer showing more than a 3-fold overall enrichment or a 5-fold enrichment at any given base position will be reported. Only 20% of the whole library is analysed and the results are extrapolated to the rest of the library in order to allow the module to run in a reasonable time.	It calculates based on the base content of the library as a whole and uses the actual count to calculate an observed/expected ratio.	The top 6 hits which show the pattern of enrichment of that Kmer across the length of the reads. This will show if a general enrichment or a pattern of bias at different points over the read length occurs.	

Appendix 14

Table A14: The accession number and summary description of the whole genome shotgun project for the eight UM-CSF samples which have been deposited at DDBJ/ENA/GenBank. The assembly is at scaffold or contig level and the method is by CLC Genomics Workbench of different version. It is a *de-novo* guided assembly. The sequencing technology used was Illumina MiSeq

UM-CSF Strain	Accession Number	BioSample	Date	Assembly level	Assembly method	Genbank assembly accession	RefSeq assembly accession	Genome coverage	WGS project
CSF01	LLXF00000000	SAMN04159245	2015/12/15	Scaffold	V 5.1	GCA_001468445.1	GCF_001468445.1	58.0x	LLXF01
CSF04	LXGA00000000	SAMN04910091	2016/05/09	Contig	V 9.0	GCA_001640575.1	GCF_001640575.1	86.0x	LXGA01
CSF05	LLXG00000000	SAMN04159257	2015/12/15	Scaffold	V 5.1	GCA_001468545.1	GCF_001468545.1	84.0x	LLXG01

CSF06	LXGB00000000	SAMN04910094	2016/05/09	Contig	V 9.0	GCA_001640605.1	GCF_001640605.1	57.0x	LXGB01
CSF08	LXGC00000000	SAMN04910096	2016/05/09	Contig	V 9.0	GCA_001640615.1	GCF_001640615.1	83.0x	LXGC01
CSF09	LXGD00000000	SAMN04910097	2016/05/09	Scaffold	V 9.0	GCA_001640695.1	GCF_001640695.1	55.0x	LXGD01
CSF15	LXGE00000000	SAMN04910098	2016/05/09	Contig	V 9.0	GCA_001640705.1	GCF_001640705.1	83.0x	LXGE01
CSF17	LXGF00000000	SAMN04910113	2016/05/09	Scaffold	V 9.0	GCA_001640645.1	GCF_001640645.1	92.0x	LXGF01

Appendix 15

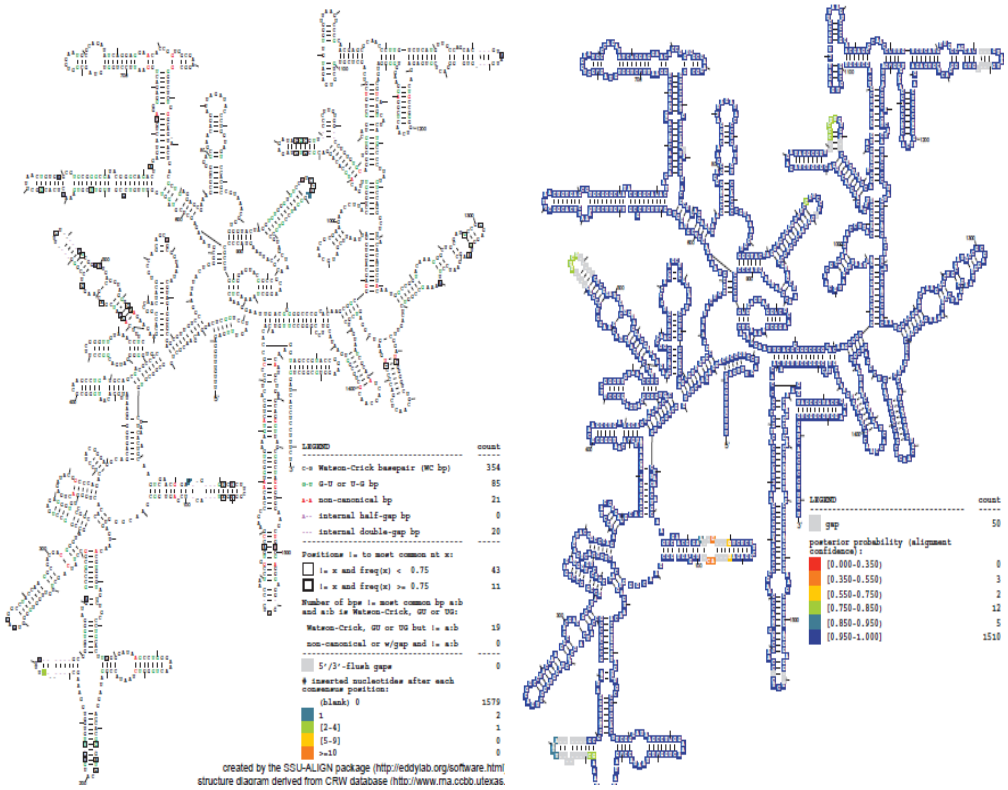


Figure A15: The structure of 16S rDNA generated from multiple sequence alignment of the gene from the study strains and H37Rv. All sequences are identical, thus the structure is identical. The alignment is created by the SSU-ALIGN package (http://eddylab.org/software.html) and the structure diagram is derived from CRW database (http://www.rna.ccbb.utexas.edu/).

Appendix 16

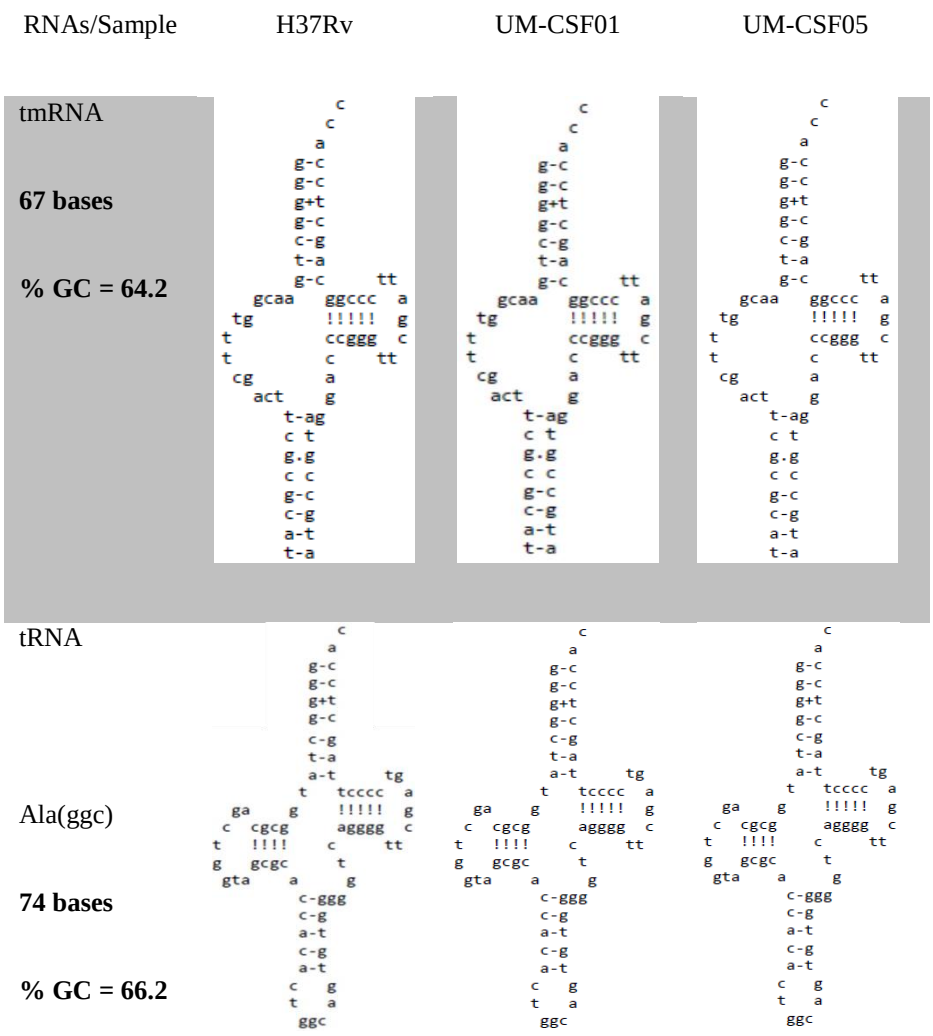


Figure A16: Structural similarities between tmRNA and tRNA of H37Rv, UM-CSF-01, and UM-CSF05. The illustrations were generated from http://mbio-serv2.mbioekol.lu.se/ARAGORN/.

Appendix 17

Genome rearrangement analysis was performed using Gepard and Mauve by comparing the UM-CSF strains with H37Rv and 29 respiratory *Mtb* genomes.

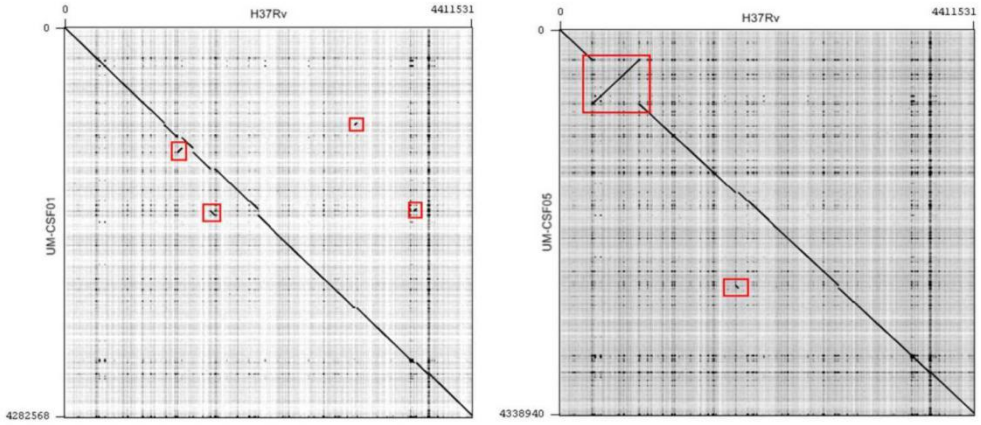


Figure A17.1: Dotplot presentation of the structural differences from H37Rv and UM-CSF01 (left); H37Rv and UM-CSF05 (right).

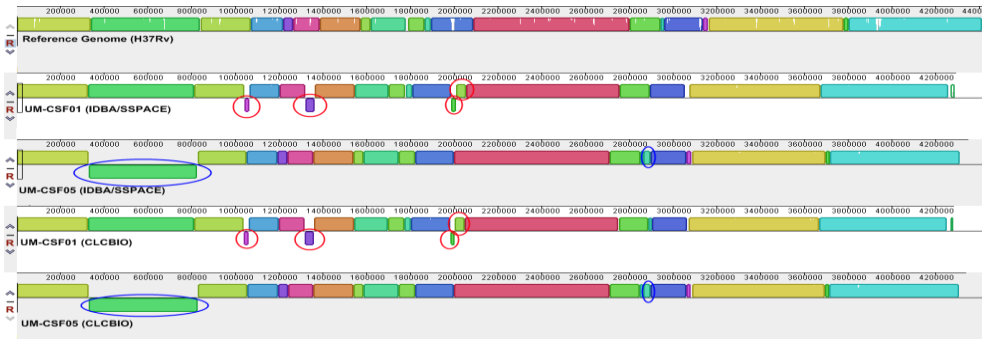


Figure A17.2: Mauve genome alignment of the reference genome H37Rv with UM-CSF01 and UM-CSF05, both assembled using IDBA/SSPACE and CLCBIO, respectively.

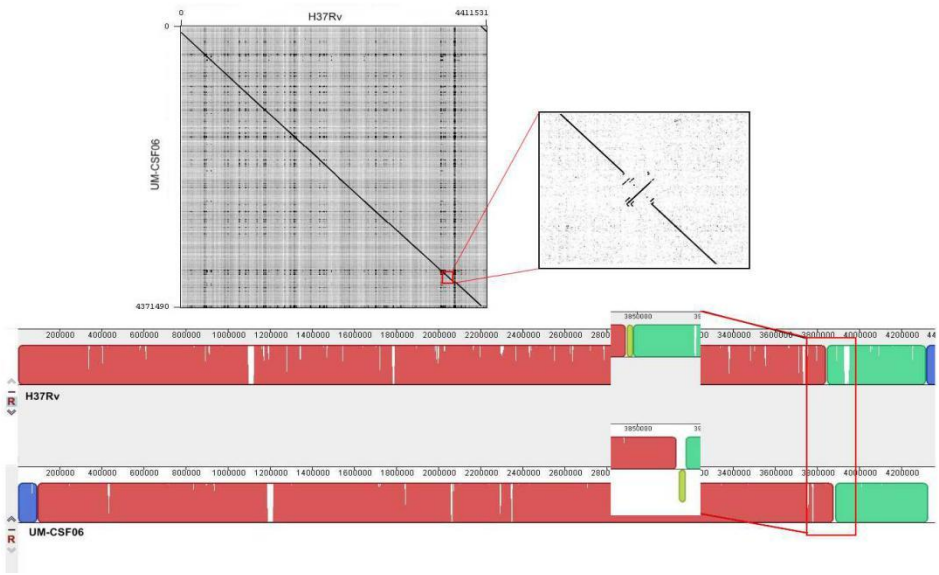


Figure A17.3: Structural differences between the reference genome H37Rv and UM-CSF06, shown in dotplot (top) and connected block (bottom).

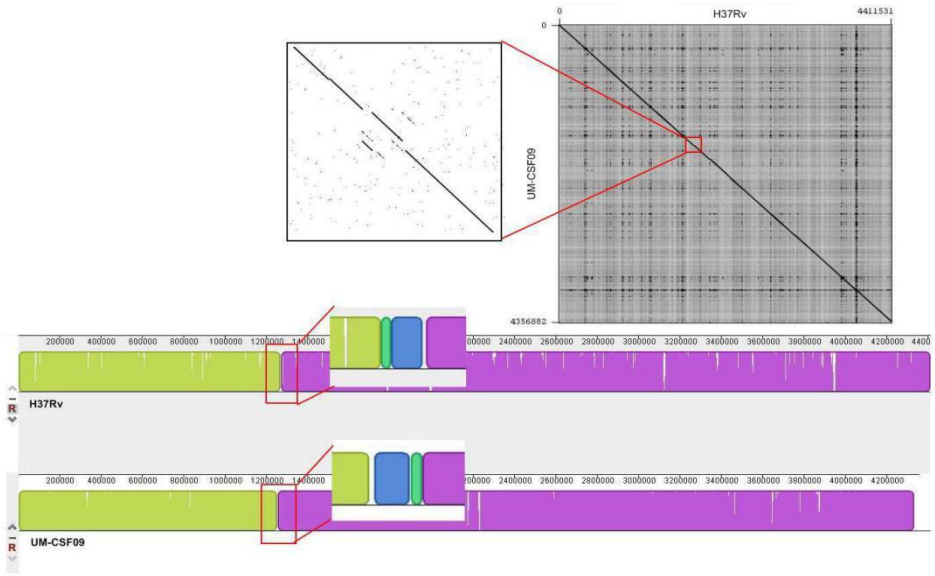


Figure A17.4: Structural differences between the reference genome H37Rv and UM-CSF09, shown in dotplot (top) and connected block (bottom).

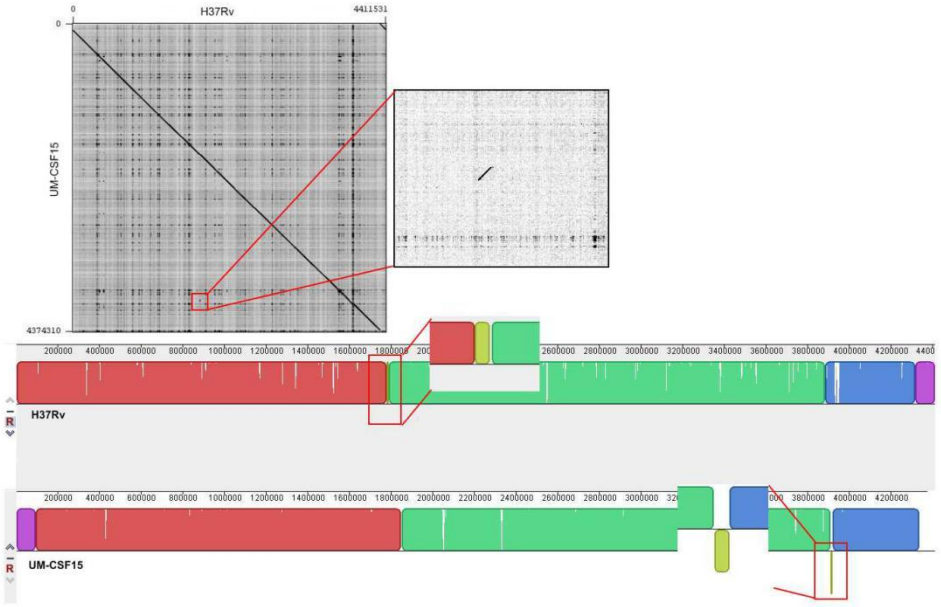


Figure A17.5: Structural differences between the reference genome H37Rv and UM-CSF015, shown in dotplot (top) and connected block (bottom).

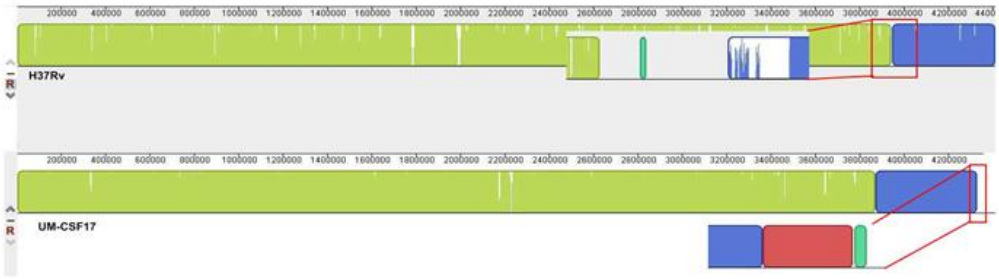


Figure A17.6: Structural differences between the reference genome H37Rv and UM-CSF017, shown in connected block (bottom).

Appendix 18

Table A18: Genome rearrangement events in six of the eight UM-CSF strains affected regions ranging from 223 to 500 211 bp and many genes affected belonged to PE/PPE/PGRS gene families, mammalian cell entry (mce) proteins, transcriptional factors, metabolic enzymes and toxin-antitoxin (TA) proteins.

Strand	Location	Gene	Gene Description
+	1211560..1213863	Rv1087	PE-PGRS family protein PE_PGRS21
+	1214040..1214360	Rv1087A	hypothetical protein
+	1214513..1214947	Rv1088	PE family protein PE9
+	<1214769..1215131	Rv1089	PE family protein PE10
+	1215517..1215621	Rv1089A	Probable cellulase CelA2a (endo-1,4-beta-glucanase) (endoglucanase) (carboxymethylcellulase)
+	1215599..1216054	Rv1090	Probable cellulase CelA2b (endo-1,4-beta-glucanase) (endoglucanase) (carboxymethylcellulase)
+	1216469..1219030	Rv1091	PE-PGRS family protein PE_PGRS22
-	1219248..1220186	Rv1092c	Probable pantothenate kinase CoaA (pantothenic acid kinase)
-	1220388..1220487	RVnc0034	Putative small regulatory RNA
+	1220574..1221854	Rv1093	Serine hydroxymethyltransferase 1 GlyA1
+	1221959..1222786	Rv1094	Possible acyl-[acyl-carrier protein] desaturase DesA2 (acyl-[ACP] desaturase) (stearoyl-ACP desaturase)
+	1222997..1224298	Rv1095	Probable PHOH-like protein PhoH2 (phosphate starvation-inducible protein PSIH)
+	1224385..1225260	Rv1096	Possible glycosyl hydrolase
-	1225263..1226144	Rv1097c	Probable membrane glycine and proline rich

			protein
-	1226141..1227565	Rv1098c	Probable fumarase Fum (fumarate hydratase)
-	1227596..1228684	Rv1099c	Fructose 1,6-bisphosphatase GlpX
+	1228683..1229384	Rv1100	hypothetical protein
-	1229391..1230548	Rv1101c	hypothetical protein
-	1230660..1230971	Rv1102c	Toxin MazF3
-	1230971..1231291	Rv1103c	Possible antitoxin MazE3
+	1231301..1231990	Rv1104	Possible para-nitrobenzyl esterase (fragment)
+	1232311..1232826	Rv1105	Possible para-nitrobenzyl esterase (fragment)
-	1232844..1233956	Rv1106c	3-beta-hydroxysteroid dehydrogenase
-	1233966..1234223	Rv1107c	Probable exodeoxyribonuclease VII (small subunit) XseB (exonuclease VII small subunit)
-	1234213..1235460	Rv1108c	Probable exodeoxyribonuclease VII (large subunit) XseA (exonuclease VII large subunit)
-	1235457..1236095	Rv1109c	hypothetical protein
+	1236185..1237192	Rv1110	Probable LYTB-related protein LytB2
-	1237209..1238192	Rv1111c	hypothetical protein
+	1238255..1239328	Rv1112	Probable GTP binding protein
+	1239416..1239613	Rv1113	Possible antitoxin VapB32
+	1239610..1239984	Rv1114	Possible toxin VapC32 contains PIN domain
+	1240187..1240885	Rv1115	Possible exported protein
+	1241003..1241188	Rv1116	Hypothetical protein
-	1241115..1241390	Rv1116A	Conserved hypothetical protein (fragment)
+	1241633..1241956	Rv1117	hypothetical protein
-	1241971..1242831	Rv1118c	hypothetical protein
-	1242864..1243013	Rv1119c	Hypothetical protein
-	1243010..1243504	Rv1120c	hypothetical protein
+	1243707..1245107	Rv1121	Probable glucose-6-phosphate 1-dehydrogenase Zwf1 (G6PD)
+	1245129..1246151	Rv1122	Probable 6-phosphogluconate dehydrogenase,decarboxylating Gnd2
-	1246144..1247052	Rv1123c	Possible peroxidase BpoB (non-haem peroxidase)
+	1247127..1248077	Rv1124	Probable epoxide hydrolase EphC (epoxide

			hydratase)
+	1248082..1249326	Rv1125	hypothetical protein
-	1249330..1249935	Rv1126c	hypothetical protein
-	1249932..1251404	Rv1127c	Probable pyruvate, phosphate dikinase PpdK
-	1251617..1252972	Rv1128c	hypothetical protein
-	1253074..1254534	Rv1129c	Probable transcriptional regulator protein
+	1254555..1256135	Rv1130	Possible methylcitrate dehydratase PrpD
+	1256132..1257313	Rv1131	Probable methylcitrate synthase PrpC
+	1257325..1259055	Rv1132	hypothetical protein
-	1259067..1261346	Rv1133c	Probable 5-methyltetrahydropteroyltriglutamate--homocysteinemethyltransferase MetE (methionine synthase, vitamin-B12independent isozyme)
+	1261922..1262158	Rv1134	Hypothetical protein
-	1262272..1264128	Rv1135c	PPE family protein PPE16
+	1264314..1264556	Rv1135A	Possible acetyl-CoA acetyltransferase (acetoacetyl-CoA thiolase)
+	1264606..1264947	Rv1136	Possible enoyl-CoA hydratase
+	1561464..1561772	Rv1386	PE family protein PE15
+	1561769..1563388	Rv1387	PPE family protein PPE20
+	1563694..1564266	Rv1388	Putative integration host factor MihF
+	1564401..1565027	Rv1389	Probable guanylate kinase Gmk
+	1565093..1565425	Rv1390	Probable DNA-directed RNA polymerase (omega chain) RpoZ (transcriptase omega chain) (RNA polymeraseomega subunit)
+	1565441..1566697	Rv1391	Probable DNA/pantothenate metabolism flavoprotein homolog Dfp
+	1566825..1568036	Rv1392	Probable S-adenosylmethionine synthetase MetK (mat) (AdoMet synthetase) (methionineadenosyltransferase)
-	1568109..1569587	Rv1393c	Probable monoxygenase
-	1569584..1570969	Rv1394c	Probable cytochrome P450 132 Cyp132
+	1571047..1572081	Rv1395	Transcriptional regulatory protein
-	1572127..1573857	Rv1396c	PE-PGRS family protein PE_PGRS25
-	1574112..1574513	Rv1397c	Possible toxin VapC10
-	1574510..1574767	Rv1398c	Possible antitoxin VapB10
-	1574850..1575809	Rv1399c	Probable non lipolytic carboxylesterase NlhH

-	1575834..1576796	Rv1400c	Probable lipase LipH
+	1576930..1577532	Rv1401	Possible membrane protein
+	1577613..1579580	Rv1402	Putative primosomal protein N' PriA (replication factor Y)
-	1579598..1580422	Rv1403c	Putative methyltransferase
+	1580591..1581073	Rv1404	Probable transcriptional regulatory protein
-	1581145..1581969	Rv1405c	Putative methyltransferase
+	1582166..1583104	Rv1406	Probable methionyl-tRNA formyltransferase Fmt
+	1583101..1584474	Rv1407	Probable Fmu protein (sun protein)
+	1584499..1585197	Rv1408	Probable ribulose-phosphate 3-epimerase Rpe (PPE) (R5P3E) (pentose-5-phosphate 3-epimerase)
+	1585194..1586213	Rv1409	Probable bifunctional riboflavin biosynthesis protein RibG : diaminohydroxyphosphoribosylaminopyrimidinedeaminase (riboflavin-specific deaminase) +5-amino-6-(5-phosphoribosylamino) uracil reductase (HTPreductase)
-	1586210..1587766	Rv1410c	Aminoglycosides/tetracycline-transport integral membrane protein
-	1587772..1588482	Rv1411c	Conserved lipoprotein LprG
+	1588567..1589172	Rv1412	Probable riboflavin synthase alpha chain RibC (RibE)
+	1589386..1589901	Rv1413	hypothetical protein
+	1589891..1590292	Rv1414	hypothetical protein
+	1590397..1591674	Rv1415	Probable riboflavin biosynthesis protein RibA2 : GTP cyclohydrolase II + 3,4-dihydroxy-2-butanone4-phosphate synthase (DHBP synthase)
+	1591671..1592153	Rv1416	Probable riboflavin synthase beta chain RibH (6,7-dimethyl-8-ribityllumazine synthase) (DMRL synthase)(lumazine synthase)
+	1592150..1592614	Rv1417	Possible conserved membrane protein
+	1592639..1593325	Rv1418	Probable lipoprotein LprH
+	1593505..1593978	Rv1419	hypothetical protein
+	1594042..1595982	Rv1420	Probable excinuclease ABC (subunit C-nuclease) UvrC

+	1595979..1596884	Rv1421	hypothetical protein
+	1596881..1597909	Rv1422	hypothetical protein
+	1597906..1598883	Rv1423	Probable transcriptional regulatory protein WhiA
-	1598893..1599654	Rv1424c	Possible membrane protein
+	1599658..1601037	Rv1425	Possible triacylglycerol synthase (diacylglycerol acyltransferase)
-	1601059..1602321	Rv1426c	Probable esterase LipO
-	1602321..1603928	Rv1427c	Possible long-chain-fatty-acid--CoA ligase FadD12 (fatty-acid-CoA synthetase) (fatty-acid-CoAsynthase)
-	1603932..1604759	Rv1428c	hypothetical protein
+	1604878..1606146	Rv1429	hypothetical protein
+	1606386..1607972	Rv1430	PE family protein PE16
+	1608083..1609852	Rv1431	hypothetical protein
+	1609849..1611270	Rv1432	Probable dehydrogenase
+	1611434..1612249	Rv1433	Possible conserved exported protein
+	1612256..1612393	Rv1434	Hypothetical protein
-	1612342..1612950	Rv1435c	Probable conserved proline, glycine, valine-rich secreted protein
+	1613307..1614326	Rv1436	Probable glyceraldehyde 3-phosphate dehydrogenase Gap (GAPDH)
+	1614329..1615567	Rv1437	Probable phosphoglycerate kinase Pgk
+	1615564..1616349	Rv1438	Probable triosephosphate isomerase Tpi (TIM)
-	1616961..1617386	Rv1439c	hypothetical protein
+	1617837..1618070	Rv1440	Probable protein-export membrane protein (translocase subunit) SecG
-	1618209..1619684	Rv1441c	PE-PGRS family protein PE_PGRS26
+	1619791..1622091	Rv1442	Probable biotin sulfoxide reductase BisC (BDS reductase) (BSO reductase)
-	1622207..1622692	Rv1443c	hypothetical protein
-	1623287..1623697	Rv1444c	hypothetical protein
-	1623714..1624457	Rv1445c	Probable 6-phosphogluconolactonase DevB (6PGL)
-	1624454..1625365	Rv1446c	Putative OXPP cycle protein OpcA
-	1625418..1626962	Rv1447c	Probable glucose-6-phosphate 1-dehydrogenase Zwf2 (G6PD)
-	1626959..1628080	Rv1448c	Probable transaldolase Tal
-	1628097..1630199	Rv1449c	Transketolase Tkt (TK)

-	1630638..1634627	Rv1450c	PE-PGRS family protein PE_PGRS27
-	3133709..3134593	Rv2826c	Hypothetical protein
-	3134596..3135483	Rv2827c	Hypothetical protein
-	3135788..3136333	Rv2828c	hypothetical protein
-	3136330..3136599	Rv2828A	hypothetical protein
-	3136620..3137012	Rv2829c	Possible toxin VapC22
-	3137009..3137224	Rv2830c	Possible antitoxin VapB22
+	3137271..3138020	Rv2831	Probable enoyl-CoA hydratase EchA16 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
-	3138099..3139181	Rv2832c	Probable Sn-glycerol-3-phosphate transport ATP-binding protein ABC transporter UgpC
-	3139174..3140484	Rv2833c	Probable Sn-glycerol-3-phosphate-binding lipoprotein UgpB
-	3140487..3141314	Rv2834c	Probable Sn-glycerol-3-phosphate transport integral membrane protein ABC transporter UgpE
-	3141311..3142222	Rv2835c	Probable Sn-glycerol-3-phosphate transport integral membrane protein ABC transporter UgpA
-	3142309..3143628	Rv2836c	Possible DNA-damage-inducible protein F DinF
-	3143635..3144645	Rv2837c	hypothetical protein
-	3144620..3145171	Rv2838c	Probable ribosome-binding factor a RbfA (P15B protein)
-	3145171..3147873	Rv2839c	Probable translation initiation factor if-2 InfB
-	3147959..3148258	Rv2840c	hypothetical protein
-	3148385..3149428	Rv2841c	Probable N utilization substance protein A NusA
-	3149425..3149976	Rv2842c	hypothetical protein
+	3150171..3150716	Rv2843	Probable conserved transmembrane alanine rich protein
+	3150713..3151201	Rv2844	Conserved alanine rich protein
-	3151202..3152950	Rv2845c	Probable prolyl-tRNA synthetase ProS (proline--tRNA ligase) (PRORS) (global RNA synthesisfactor) (proline translase)
-	3153039..3154631	Rv2846c	Possible integral membrane efflux protein EfpA

-	3154654..3155871	Rv2847c	Possible multifunctional enzyme siroheme synthase CysG: uroporphyrin-III C-methyltransferase(urogen III methylase) (SUMT) (uroporphyrinogen IIImethylase) (UROM) + precorrin-2 oxidase + ferrochelatase
-	3156148..3157521	Rv2848c	Probable cobyrinic acid A,C-diamide synthase CobB
-	3157521..3158144	Rv2849c	Probable cob(I)alamin adenosyltransferase CobO (corrinoid adenosyltransferase) (corrinoid adotransferaseactivity)
-	3158165..3160054	Rv2850c	Possible magnesium chelatase
-	3160051..3160521	Rv2851c	GCN5-related N-acetyltransferase
-	3160580..3162061	Rv2852c	Probable malate:quinone oxidoreductase Mqo (malate dehydrogenase [acceptor])
+	3162268..3164115	Rv2853	PE-PGRS family protein PE_PGRS48
+	3770773..3771048	Rv3357	Antitoxin RelJ
+	3771045..3771302	Rv3358	Toxin RelK
+	3771344..3772534	Rv3359	Possible oxidoreductase
+	3772651..3773019	Rv3360	hypothetical protein
-	3773016..3773567	Rv3361c	hypothetical protein
-	3773574..3774155	Rv3362c	Probable ATP/GTP-binding protein
-	3774136..3774504	Rv3363c	hypothetical protein
-	3774482..3774874	Rv3364c	hypothetical protein
-	3774871..3777501	Rv3365c	hypothetical protein
+	3777737..3778201	Rv3366	Probable tRNA/rRNA methylase SpoU (tRNA/rRNA methyltransferase)
+	3778568..3780334	Rv3367	PE-PGRS family protein PE_PGRS51
-	3780335..3780979	Rv3368c	Possible oxidoreductase
+	3780978..3781412	Rv3369	hypothetical protein
-	3781501..3784740	Rv3370c	Probable DNA polymerase III (alpha chain) DnaE2 (DNA nucleotidyltransferase)
+	3784932..3786272	Rv3371	Possible triacylglycerol synthase (diacylglycerol acyltransferase)
+	3786314..3787489	Rv3372	Trehalose 6-phosphate phosphatase OtsB2 (trehalose-phosphatase) (TPP)
+	3787726..3788367	Rv3373	Probable enoyl-CoA hydratase EchA18 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
+	3788368..3788616	Rv3374	Probable enoyl-CoA hydratase (fragment)

			EchA18.1 (enoyl hydrase) (unsaturated acyl-CoA hydratase)(crotonase)
+	3788621..3790048	Rv3375	Probable amidase AmiD (acylamidase) (acylase)
+	3790156..3790809	Rv3376	hypothetical protein
-	3790848..3792353	Rv3377c	Halimadienyl diphosphate synthase
-	3792358..3793248	Rv3378c	Diterpene synthase
-	3793257..3794867	Rv3379c	Probable 1-deoxy-D-xylulose 5-phosphate synthase Dxs2 (1-deoxyxylulose-5-phosphate synthase) (DXP synthase)(DXPS)
-	3795100..>3796086	Rv3380c	Probable transposase
-	3796035..3796361	Rv3381c	Probable transposase for insertion sequence element IS6110 (fragment)
-	3796448..3797437	Rv3382c	Probable LYTB-related protein LytB1
-	3797437..3798489	Rv3383c	Possible polyprenyl synthetase IdsB (polyprenyl transferase) (polyprenyl diphosphate synthase)
-	3799243..3799635	Rv3384c	Possible toxin VapC46 Contains PIN domain
-	3799635..3799943	Rv3385c	Possible antitoxin VapB46
+	3800092..3800796	Rv3386	Possible transposase
+	3800786..3801463	Rv3387	Possible transposase
+	3801653..3803848	Rv3388	PE-PGRS family protein PE_PGRS52
-	3803919..3804791	Rv3389c	Probable 3-hydroxyacyl-thioester dehydratase HtdY
+	3804865..3805575	Rv3390	Probable conserved lipoprotein LpqD
+	3805621..3807573	Rv3391	Possible multi-functional enzyme with acyl-CoA-reductase activity AcrA1
-	3807574..3808437	Rv3392c	Cyclopropane-fatty-acyl-phospholipid synthase 1 CmaA1 (cyclopropane fatty acid synthase) (CFA synthase)(cyclopropane mycolic acid synthase 1)
-	332708..333136	Rv0277c	Possible toxin VapC25 Contains PIN domain
-	333160..333417	Rv0277A	Possible antitoxin VapB25
-	333437..336310	Rv0278c	PE-PGRS family protein PE_PGRS3
-	336560..339073	Rv0279c	PE-PGRS family protein PE_PGRS4
+	339364..340974	Rv0280	PPE family protein PPE3
+	340998..341906	Rv0281	Possible S-adenosylmethionine-dependent

			methyltransferase
+	342130..344025	Rv0282	ESX conserved component EccA3 ESX-3 type VII secretion system protein
+	344022..345638	Rv0283	ESX conserved component EccB3 ESX-3 type VII secretion system protein Possible membrane protein
+	345635..349627	Rv0284	ESX conserved component EccC3 ESX-3 type VII secretion system protein Possible membrane protein
+	349624..349932	Rv0285	PE family protein PE5
+	349935..351476	Rv0286	PPE family protein PPE4
+	351525..351818	Rv0287	ESAT-6 like protein EsxG (conserved protein TB9.8)
+	351848..352138	Rv0288	Low molecular weight protein antigen 7 EsxH (10 kDa antigen) (CFP-7) (protein TB10.4)
+	352149..353036	Rv0289	ESX-3 secretion-associated protein EspG3
+	353083..354501	Rv0290	ESX conserved component EccD3 ESX-3 type VII secretion system protein Probable transmembrane protein
+	354498..355883	Rv0291	Probable membrane-anchored mycosin MycP3 (serine protease) (subtilisin-like protease) (subtilase-like)(mycosin-3)
+	355880..356875	Rv0292	ESX conserved component EccE3 ESX-3 type VII secretion system protein Probable transmembrane protein
-	356862..358064	Rv0293c	hypothetical protein
+	358171..358956	Rv0294	Probable trans-aconitate methyltransferase Tam
-	358945..359748	Rv0295c	hypothetical protein
-	359758..361155	Rv0296c	Probable sulfatase
+	361334..363109	Rv0297	PE-PGRS family protein PE_PGRS5
+	363252..363479	Rv0298	Hypothetical protein
+	363476..363778	Rv0299	Hypothetical protein
+	363826..364047	Rv0300	Possible antitoxin VapB2
+	364044..364469	Rv0301	Possible toxin VapC2
+	364605..365237	Rv0302	Probable transcriptional regulatory protein (probably TetR/AcrR-family)
+	365234..366142	Rv0303	Probable dehydrogenase/reductase
-	366150..372764	Rv0304c	PPE family protein PPE5

-	372820..375711	Rv0305c	PPE family protein PPE6
+	375914..376585	Rv0306	Putative oxidoreductase
-	376573..377055	Rv0307c	hypothetical protein
+	377113..377829	Rv0308	Probable conserved integral membrane protein
+	377931..378587	Rv0309	Possible conserved exported protein
-	378657..379148	Rv0310c	hypothetical protein
+	379172..380401	Rv0311	hypothetical protein
+	380556..382418	Rv0312	Conserved hypothetical proline and threonine rich protein
+	382490..382876	Rv0313	hypothetical protein
-	382879..383541	Rv0314c	Possible conserved membrane protein
+	383602..384486	Rv0315	Possible beta-1,3-glucanase precursor
+	384535..385149	Rv0316	Possible muconolactone isomerase
-	385173..385943	Rv0317c	Possible glycerophosphoryl diester phosphodiesterase GlpQ2 (glycerophosphodiesterphosphodiesterase)
-	386204..386274	Rvnt04	tRNA-Gly
-	386305..387099	Rv0318c	Probable conserved integral membrane protein
+	387148..387816	Rv0319	Probable pyrrolidone-carboxylate peptidase Pcp (5-oxoprolyl-peptidase) (pyroglutamyl-peptidase I) (PGP-I)(pyrase)
+	387888..388550	Rv0320	Possible conserved exported protein
+	388582..389154	Rv0321	Probable deoxycytidine triphosphate deaminase Dcd (dCTP deaminase)
+	389260..390591	Rv0322	Probable UDP-glucose 6-dehydrogenase UdgA (UDP-GLC dehydrogenase) (UDP-GLCDH) (UDPGDH)
-	390580..391251	Rv0323c	hypothetical protein
+	391352..392032	Rv0324	Possible transcriptional regulatory protein (possibly ArsR-family)
+	392039..392263	Rv0325	Hypothetical protein
+	392273..392728	Rv0326	Hypothetical protein
-	392696..394045	Rv0327c	Possible cytochrome P450 135A1 Cyp135A1
+	394111..394713	Rv0328	Possible transcriptional regulatory protein (possibly TetR/AcrR-family)
-	394694..395320	Rv0329c	hypothetical protein
-	395347..396087	Rv0330c	Hypothetical protein

+	396201..397367	Rv0331	Possible dehydrogenase/reductase
+	397442..398227	Rv0332	hypothetical protein
+	398254..398628	Rv0333	hypothetical protein
+	398658..399524	Rv0334	Alpha-D-glucose-1-phosphate thymidylyltransferase RmlA (dTDP-glucose synthase)(dTDP-glucose pyrophosphorylase)
-	399535..400050	Rv0335c	PE family protein PE6
+	400192..401703	Rv0336	Conserved 13E12 repeat family protein
-	401873..403162	Rv0337c	Probable aspartate aminotransferase AspC (transaminase A) (ASPAT)
-	403193..405841	Rv0338c	Probable iron-sulfur-binding reductase
-	405950..408448	Rv0339c	Possible transcriptional regulatory protein
+	408634..409173	Rv0340	hypothetical protein
+	409362..410801	Rv0341	Isoniazid inductible gene protein IniB
+	410838..412760	Rv0342	Isoniazid inductible gene protein IniA
+	412757..414238	Rv0343	Isoniazid inductible gene protein IniC
-	414381..414941	Rv0344c	Probable lipoprotein LpqJ
+	415050..415460	Rv0345	hypothetical protein
-	415502..416965	Rv0346c	Possible L-asparagine permease AnsP2 (L-asparagine transport protein)
+	417304..418290	Rv0347	Probable conserved membrane protein
+	418293..418946	Rv0348	Possible transcriptional regulatory protein
+	418949..419608	Rv0349	Hypothetical protein
+	419835..421712	Rv0350	Probable chaperone protein DnaK (heat shock protein 70) (heat shock 70 kDa protein) (HSP70)
+	421709..422416	Rv0351	Probable GrpE protein (HSP-70 cofactor)
+	422452..423639	Rv0352	Probable chaperone protein DnaJ1
+	423639..424019	Rv0353	Probable heat shock protein transcriptional repressor HspR (MerR family)
-	424269..424694	Rv0354c	PPE family protein PPE7
-	424777..434679	Rv0355c	PPE family protein PPE8
-	434830..435474	Rv0356c	hypothetical protein
-	435471..436769	Rv0357c	Probable adenylosuccinate synthetase PurA (imp--aspartate ligase) (ADSS) (ampsase)
+	436860..437507	Rv0358	hypothetical protein
+	437518..438297	Rv0359	Probable conserved integral membrane protein
-	438302..438739	Rv0360c	hypothetical protein

+	438822..439649	Rv0361	Probable conserved membrane protein
+	439871..441253	Rv0362	Possible Mg2+ transport transmembrane protein MgtE
-	441265..442299	Rv0363c	Probable fructose-bisphosphate aldolase Fba
+	442395..443078	Rv0364	Possible conserved transmembrane protein
-	443067..444197	Rv0365c	hypothetical protein
-	444222..444815	Rv0366c	hypothetical protein
-	444844..445233	Rv0367c	Hypothetical protein
-	445314..446525	Rv0368c	hypothetical protein
-	446531..447046	Rv0369c	Possible membrane oxidoreductase
-	447147..448043	Rv0370c	Possible oxidoreductase
-	448040..448633	Rv0371c	hypothetical protein
-	448630..449385	Rv0372c	hypothetical protein
-	449404..451803	Rv0373c	Probable carbon monoxyde dehydrogenase (large chain)
-	451800..452279	Rv0374c	Probable carbon monoxyde dehydrogenase (small chain)
-	452294..453154	Rv0375c	Probable carbon monoxyde dehydrogenase (medium chain)
-	453230..454372	Rv0376c	hypothetical protein
+	454421..455386	Rv0377	Probable transcriptional regulatory protein (probably LysR-family)
+	455637..455858	Rv0378	Conserved hypothetical glycine rich protein
+	455977..456192	Rv0379	Possible protein transport protein SecE2
-	456268..456819	Rv0380c	Possible RNA methyltransferase (RNA methylase)
-	456915..457823	Rv0381c	Hypothetical protein
-	457841..458380	Rv0382c	Probable orotate phosphoribosyltransferase PyrE (OPRT) (oprtase)
-	458461..459315	Rv0383c	Possible conserved secreted protein
-	459456..462002	Rv0384c	Probable endopeptidase ATP binding protein (chain B) ClpB (ClpB protein) (heat shock protein F84.1)
+	462135..463307	Rv0385	Probable monooxygenase
+	463411..466668	Rv0386	Probable transcriptional regulatory protein (probably LuxR/UhpA-family)
-	466672..467406	Rv0387c	hypothetical protein
-	467459..468001	Rv0388c	PPE family protein PPE9
+	468335..469594	Rv0389	Probable phosphoribosylglycinamide

			formyltransferase 2 PurT (GART 2) (gar transformylase 2)(5'-phosphoribosylglycinamide transformylase 2)(formate-dependent gar transformylase)
+	469591..470013	Rv0390	hypothetical protein
+	470010..471230	Rv0391	Probable O-succinylhomoserine sulfhydrylase MetZ (OSH sulfhydrylase)
-	471227..472639	Rv0392c	Probable membrane NADH dehydrogenase NdhA
+	472781..474106	Rv0393	Conserved 13E12 repeat family protein
-	474122..474841	Rv0394c	Possible secreted protein
+	474940..475344	Rv0395	Hypothetical protein
+	475350..475742	Rv0396	Hypothetical protein
+	475816..476184	Rv0397	Conserved 13E12 repeat family protein
+	476394..476642	Rv0397A	hypothetical protein
-	476679..477320	Rv0398c	Possible secreted protein
-	477327..478556	Rv0399c	Possible conserved lipoprotein LpqK
-	478566..479753	Rv0400c	Acyl-CoA dehydrogenase FadE7
+	479789..480160	Rv0401	Probable conserved transmembrane protein
-	480355..483231	Rv0402c	Probable conserved transmembrane transport protein MmpL1
-	483228..483656	Rv0403c	Probable conserved membrane protein MmpS1
+	483977..485734	Rv0404	Fatty-acid-AMP ligase FadD30 (fatty-acid-AMP synthetase) (fatty-acid-AMP synthase)
+	485731..489939	Rv0405	Probable membrane bound polyketide synthase Pks6
-	489887..490705	Rv0406c	Beta lactamase like protein
+	490783..491793	Rv0407	F420-dependent glucose-6-phosphate dehydrogenase Fgd1
+	491786..493858	Rv0408	Probable phosphate acetyltransferase Pta (phosphotransacetylase)
+	493851..495008	Rv0409	Probable acetate kinase AckA (acetokinase)
-	495062..497314	Rv0410c	Serine/threonine-protein kinase PknG (protein kinase G) (STPK G)
-	497314..498300	Rv0411c	Probable glutamine-binding lipoprotein GlnH (GLNBP)
-	498300..499619	Rv0412c	Possible conserved membrane protein
+	499713..500366	Rv0413	Possible mutator protein MutT3 (7,8-

			dihydro-8-oxoguanine-triphosphatase) (8-oxo-dGTPase)(dGTP pyrophosphohydrolase)
-	500350..501018	Rv0414c	Thiamine-phosphate pyrophosphorylase ThiE (TMP pyrophosphorylase) (TMP-PPASE) (thiamine-phosphatesynthase)
+	501148..502170	Rv0415	Possible thiamine biosynthesis oxidoreductase ThiO
+	502167..502373	Rv0416	Possible protein ThiS
+	502366..503124	Rv0417	Probable thiamin biosynthesis protein ThiG (thiazole biosynthesis protein)
+	503496..504998	Rv0418	Probable lipoprotein aminopeptidase LpqL
+	505086..506582	Rv0419	Possible lipoprotein peptidase LpqM
-	506561..506971	Rv0420c	Possible transmembrane protein
-	507132..507761	Rv0421c	hypothetical protein
-	507758..508555	Rv0422c	Probable phosphomethylpyrimidine kinase ThiD (HMP-phosphate kinase) (HMP-P kinase)
-	508582..510225	Rv0423c	Probable thiamine biosynthesis protein ThiC
-	510377..510652	Rv0424c	Hypothetical protein
-	510702..515321	Rv0425c	Possible metal cation transporting P-type ATPase CtpH
-	515373..515816	Rv0426c	Possible transmembrane protein
-	516017..516892	Rv0427c	Probable exodeoxyribonuclease III protein XthA (exonuclease III) (EXO III) (AP endonuclease VI)
-	516895..517803	Rv0428c	GCN5-related N-acetyltransferase
-	517803..518396	Rv0429c	Probable polypeptide deformylase Def (PDF) (formylmethionine deformylase)
+	518733..519041	Rv0430	hypothetical protein
+	519073..519567	Rv0431	Putative tuberculin related peptide
+	519600..520322	Rv0432	Periplasmic superoxide dismutase [Cu-Zn] SodC
+	520324..521454	Rv0433	hypothetical protein
+	521514..522167	Rv0434	hypothetical protein
-	522347..524533	Rv0435c	Putative conserved ATPase
-	524530..525390	Rv0436c	Probable CDP-diacylglycerol--serine O-phosphatidyltransferase PssA (PS synthase)(phosphatidylserine synthase)

-	525387..526082	Rv0437c	Possible phosphatidylserine decarboxylase Psd (PS decarboxylase)
-	526143..527360	Rv0438c	Probable molybdopterin biosynthesis protein MoeA2
-	527379..528314	Rv0439c	Probable dehydrogenase/reductase
+	528608..530230	Rv0440	60 kDa chaperonin 2 GroEL2 (protein CPN60-2) (GroEL protein 2) (65 kDa antigen) (heat shock protein 65)(cell wall protein A) (antigen A)
-	530296..530724	Rv0441c	Hypothetical protein
-	530751..532214	Rv0442c	PPE family protein PPE10
+	532396..532911	Rv0443	hypothetical protein
-	533091..533789	Rv0444c	Anti-sigma factor RskA (regulator of sigma K)
-	533833..534396	Rv0445c	Alternative RNA polymerase sigma factor SigK
-	534445..535215	Rv0446c	Possible conserved transmembrane protein
-	535224..536507	Rv0447c	Probable cyclopropane-fatty-acyl-phospholipid synthase UfaA1 (cyclopropane fatty acid synthase) (CFAsynthase)
-	536504..537169	Rv0448c	hypothetical protein
-	537229..538548	Rv0449c	hypothetical protein
-	538588..541491	Rv0450c	Probable conserved transmembrane transport protein MmpL4
-	541488..541910	Rv0451c	Probable conserved membrane protein MmpS4
+	542142..542852	Rv0452	Possible transcriptional regulatory protein
+	543174..544730	Rv0453	PPE family protein PPE11
+	544835..545185	Rv0454	hypothetical protein
-	545375..545821	Rv0455c	hypothetical protein
-	545889..546803	Rv0456c	enoyl-CoA hydratase EchA2 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
-	547076..547357	Rv0456A	Possible toxin MazF1
-	547344..547517	Rv0456B	Possible antitoxin MazE1
-	547586..549607	Rv0457c	Probable peptidase
+	549675..551198	Rv0458	Probable aldehyde dehydrogenase
+	551198..551689	Rv0459	hypothetical protein
+	551749..551988	Rv0460	Conserved hydrophobic protein

+	552026..552550	Rv0461	Probable transmembrane protein
+	552614..554008	Rv0462	Dihydrolipoamide dehydrogenase LpdC (lipoamide reductase (NADH)) (lipoyl dehydrogenase) (dihydrolipoyldehydrogenase) (diaphorase)
+	554016..554309	Rv0463	Probable conserved membrane protein
-	554313..554885	Rv0464c	hypothetical protein
-	554882..556306	Rv0465c	Probable transcriptional regulatory protein
+	556458..557252	Rv0466	hypothetical protein
+	557527..558813	Rv0467	Isocitrate lyase Icl (isocitrase)
+	558895..559755	Rv0468	3-hydroxybutyryl-CoA dehydrogenase FadB2 (beta-hydroxybutyryl-CoA dehydrogenase) (BHBD)
+	559888..560748	Rv0469	Possible mycolic acid synthase UmaA
-	560848..561711	Rv0470c	Mycolic acid synthase PcaA (cyclopropane synthase)
-	561854..562294	Rv0470A	Hypothetical protein
-	562225..562713	Rv0471c	Hypothetical protein
-	562723..563427	Rv0472c	Probable transcriptional regulatory protein (possibly TetR-family)
+	563564..564934	Rv0473	Possible conserved transmembrane protein
+	565021..565443	Rv0474	Probable transcriptional regulatory protein
+	565797..566396	Rv0475	Iron-regulated heparin binding haemagglutinin HbhA (adhesin)
+	566508..566771	Rv0476	Possible conserved transmembrane protein
+	566776..567222	Rv0477	Possible conserved secreted protein
+	567222..567896	Rv0478	Probable deoxyribose-phosphate aldolase DeoC (phosphodeoxyriboaldolase) (deoxyriboaldolase)
-	567921..568967	Rv0479c	Probable conserved membrane protein
-	568964..569806	Rv0480c	Possible amidohydrolase
-	569988..570512	Rv0481c	Hypothetical protein
+	570539..571648	Rv0482	Probable UDP-N-acetylenolpyruvoylglucosamine reductase MurB (UDP-N-acetylmuramate dehydrogenase)
+	571710..573065	Rv0483	Probable conserved lipoprotein LprQ
-	573046..573801	Rv0484c	Probable short-chain type oxidoreductase
+	573984..575300	Rv0485	Possible transcriptional regulatory protein
+	575033..575069	RVnc0019	Fragment of putative small regulatory RNA

+	575348..576790	Rv0486	Glycosyltransferase MshA
+	576787..577338	Rv0487	hypothetical protein
+	577664..578269	Rv0488	Probable conserved integral membrane protein
+	578426..579175	Rv0489	Probable phosphoglycerate mutase 1 Gpm1 (phosphoglyceromutase) (PGAM) (BPG-dependent PGAM)
+	579349..580581	Rv0490	Putative two component sensor histidine kinase SenX3
+	580809..581492	Rv0491	Two component sensory transduction protein RegX3 (transcriptional regulatory protein) (probablyLuxR-family)
-	581489..583378	Rv0492c	Probable oxidoreductase GMC-type
-	583375..583704	Rv0492A	Hypothetical protein
-	583701..584690	Rv0493c	hypothetical protein
+	584695..585423	Rv0494	Probable transcriptional regulatory protein (probably GntR-family)
-	585424..586314	Rv0495c	hypothetical protein
+	586394..587380	Rv0496	hypothetical protein
+	587377..588309	Rv0497	Probable conserved transmembrane protein
+	588325..589167	Rv0498	hypothetical protein
+	589183..590058	Rv0499	hypothetical protein
+	590083..590970	Rv0500	Probable pyrroline-5-carboxylate reductase ProC (P5CR) (P5C reductase)
+	591111..591347	Rv0500A	hypothetical protein
+	591475..591576	Rv0500B	hypothetical protein
+	591654..592784	Rv0501	Possible UDP-glucose 4-epimerase GalE2 (galactowaldenase) (UDP-galactose 4-epimerase) (uridinediphosphate galactose 4-epimerase) (uridinediphospho-galactose 4-epimerase)
+	592791..593867	Rv0502	hypothetical protein
-	593871..594779	Rv0503c	Cyclopropane-fatty-acyl-phospholipid synthase 2 CmaA2 (cyclopropane fatty acid synthase) (CFA synthase)(cyclopropane mycolic acid synthase 2) (mycolic acidtrans-cyclopropane synthetase)
-	594802..595302	Rv0504c	hypothetical protein
-	595464..596585	Rv0505c	Possible phosphoserine phosphatase SerB1 (PSP) (O-phosphoserine phosphohydrolase)

			(pspase)
+	596759..597202	Rv0506	Probable conserved membrane protein MmpS2
+	597199..600105	Rv0507	Probable conserved transmembrane transport protein MmpL2
+	600098..600391	Rv0508	hypothetical protein
+	600441..601847	Rv0509	Probable glutamyl-tRNA reductase HemA (GLUTR)
+	601857..602786	Rv0510	Probable porphobilinogen deaminase HemC (PBG) (hydroxymethylbilane synthase) (HMBS)(pre-uroporphyrinogen synthase)
+	602819..604516	Rv0511	Probable uroporphyrin-III C-methyltransferase HemD (uroporphyrinogen III methylase) (urogen IIImethylase) (SUMT) (urogen III methylase) (UROM)
+	604602..605591	Rv0512	Probable delta-aminolevulinic acid dehydratase HemB (porphobilinogen synthase) (ALAD) (ALADH)
+	605604..606152	Rv0513	Possible conserved transmembrane protein
+	606149..606448	Rv0514	Possible transmembrane protein
+	606551..608062	Rv0515	Conserved 13E12 repeat family protein
-	608059..608535	Rv0516c	Possible anti-anti-sigma factor
+	608746..610056	Rv0517	Possible membrane acyltransferase
+	610188..610883	Rv0518	Possible exported protein
-	611172..612074	Rv0519c	Possible conserved membrane protein
+	612255..612605	Rv0520	Possible methyltransferase/methylase (fragment)
+	612598..612903	Rv0521	Possible methyltransferase/methylase (fragment)
+	613038..614342	Rv0522	Probable GABA permease GabP (4-amino butyrate transport carrier) (GAMA-aminobutyrate permease)
-	614326..614721	Rv0523c	hypothetical protein
+	614835..616223	Rv0524	Probable glutamate-1-semialdehyde 2,1-aminomutase HemL (GSA) (glutamate-1-semialdehydeaminotransferase) (GSA-at)
+	616223..616831	Rv0525	hypothetical protein
+	616846..617496	Rv0526	Possible thioredoxin protein (thiol-disulfide interchange protein)
+	617493..618272	Rv0527	Possible cytochrome C-type biogenesis

			protein CcdA
+	618305..619894	Rv0528	Probable conserved transmembrane protein
+	619891..620865	Rv0529	Possible cytochrome C-type biogenesis protein CcsA
+	620907..622124	Rv0530	hypothetical protein
-	622121..622282	Rv0530A	hypothetical protein
+	622329..622646	Rv0531	Possible conserved membrane protein
+	622793..624577	Rv0532	PE-PGRS family protein PE_PGRS6
-	624473..625480	Rv0533c	3-oxoacyl-[acyl-carrier-protein] synthase III FabH (beta-ketoacyl-ACP synthase III) (KAS III)
-	625562..626440	Rv0534c	1,4-dihydroxy-2-naphthoate octaprenyltransferase MenA (DHNA-octaprenyltransferase)
+	626457..627251	Rv0535	Probable 5'-methylthioadenosine phosphorylase Pnp (MTA phosphorylase)
+	627248..628288	Rv0536	Probable UDP-glucose 4-epimerase GalE3 (galactowaldenase) (UDP-galactose 4-epimerase) (uridinediphosphate galactose 4-epimerase) (uridinediphospho-galactose 4-epimerase)
-	628298..629731	Rv0537c	Probable integral membrane protein
+	630040..631686	Rv0538	Possible conserved membrane protein
+	631743..632375	Rv0539	Probable dolichyl-phosphate sugar synthase (dolichol-phosphate sugar synthetase) (dolichol-phosphatesugar transferase) (sugar phosphoryldolichol synthase)
+	632372..633034	Rv0540	hypothetical protein
-	633055..634404	Rv0541c	Probable conserved integral membrane protein
-	634416..635504	Rv0542c	Possible O-succinylbenzoic acid--CoA ligase MenE (OSB-CoA synthetase) (O-succinylbenzoate-CoA synthase)
-	635573..635875	Rv0543c	hypothetical protein
-	635935..636213	Rv0544c	Possible conserved transmembrane protein
-	636210..637463	Rv0545c	Probable low-affinity inorganic phosphate transporter integral membrane protein PitA
-	637583..637969	Rv0546c	hypothetical protein
-	638032..638916	Rv0547c	Possible oxidoreductase
-	639012..639956	Rv0548c	Naphthoate synthase MenB

			(dihydroxynaphthoic acid synthetase) (DHNA synthetase)
-	640228..640641	Rv0549c	Possible toxin VapC3
-	640638..640904	Rv0550c	Possible antitoxin VapB3
-	641096..642811	Rv0551c	Probable fatty-acid-CoA ligase FadD8 (fatty-acid-CoA synthetase) (fatty-acid-CoA synthase)
+	642889..644493	Rv0552	hypothetical protein
+	644490..645470	Rv0553	Probable muconate cycloisomerase MenC (cis,cis-muconate lactonizing enzyme) (MLE)
+	645467..646255	Rv0554	Possible peroxidase BpoC (non-haem peroxidase)
+	646298..647962	Rv0555	Probable bifunctional menaquinone biosynthesis protein MenD :2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylatesynthase (SHCHC synthase) + 2-oxoglutarate decarboxylase(alpha-ketoglutarate decarboxylase) (KDC)
+	647959..648474	Rv0556	Probable conserved transmembrane protein
+	648536..649672	Rv0557	Mannosyltransferase MgtA
+	649689..650393	Rv0558	Probable ubiquinone/menaquinone biosynthesis methyltransferase MenH (2-heptaprenyl-1,4-naphthoquinonemethyltransferase)
-	650407..650745	Rv0559c	Possible conserved secreted protein
-	650779..651504	Rv0560c	Possible benzoquinone methyltransferase (methylase)
-	651529..652755	Rv0561c	Possible oxidoreductase
+	652771..653778	Rv0562	Probable polyprenyl-diphosphate synthase GrcC1 (polyprenyl pyrophosphate synthetase)
+	653879..654739	Rv0563	Probable protease transmembrane protein heat shock protein HtpX
-	654924..655949	Rv0564c	Probable glycerol-3-phosphate dehydrogenase [NAD(P)+] GpdA1 (NAD(P)H-dependent glycerol-3-phosphatedehydrogenase) (NAD(P)H-dependentdihydroxyacetone-phosphate reductase)

-	656010..657470	Rv0565c	Probable monooxygenase
-	657548..658039	Rv0566c	hypothetical protein
+	658109..658189	Rvnt05	tRNA-Tyr
+	658321..659340	Rv0567	Probable methyltransferase/methylase
+	659450..660868	Rv0568	Possible cytochrome P450 135B1 Cyp135B1
+	661003..661269	Rv0569	hypothetical protein
+	661295..663373	Rv0570	Probable ribonucleoside-diphosphate reductase (large subunit) NrdZ (ribonucleotide reductase)
-	663487..664818	Rv0571c	hypothetical protein
-	665042..665383	Rv0572c	Hypothetical protein
-	665851..667242	Rv0573c	Nicotinic acid phosphoribosyltransferase PncB2
-	667252..668394	Rv0574c	hypothetical protein
-	668579..669745	Rv0575c	Possible oxidoreductase
+	669848..671152	Rv0576	Probable transcriptional regulatory protein (possibly ArsR-family)
+	671166..671951	Rv0577	Conserved protein TB27.3
-	671996..675916	Rv0578c	PE-PGRS family protein PE_PGRS7
+	676238..676996	Rv0579	hypothetical protein
-	677125..677616	Rv0580c	hypothetical protein
+	677710..677925	Rv0581	Possible antitoxin VapB26
+	677922..678329	Rv0582	Possible toxin VapC26 Contains PIN domain
-	678389..679075	Rv0583c	Probable conserved lipoprotein LpqN
+	679229..681862	Rv0584	Possible conserved exported protein
-	681885..684272	Rv0585c	Probable conserved integral membrane protein
+	684410..685132	Rv0586	Probable transcriptional regulatory protein Mce2R (GntR-family)
+	685129..685926	Rv0587	Conserved hypothetical integral membrane protein YrbE2A
+	685928..686815	Rv0588	Conserved hypothetical integral membrane protein YrbE2B
+	686821..688035	Rv0589	Mce-family protein Mce2A
+	688032..688859	Rv0590	Mce-family protein Mce2B
+	688808..689062	Rv0590A	Mce-family related protein
+	689059..690504	Rv0591	Mce-family protein Mce2C
+	690501..692027	Rv0592	Mce-family protein Mce2D

+	692024..693232	Rv0593	Possible Mce-family lipoprotein LprL (Mce-family lipoprotein Mce2E)
+	693237..694787	Rv0594	Mce-family protein Mce2F
-	694839..695231	Rv0595c	Possible toxin VapC4
-	695228..695485	Rv0596c	Possible antitoxin VapB4
-	695668..696903	Rv0597c	hypothetical protein
-	697154..697567	Rv0598c	Possible toxin VapC27 Contains PIN domain
-	697564..697800	Rv0599c	Possible antitoxin VapB27
-	697904..698410	Rv0600c	Two component sensor kinase [second part]
-	698524..698994	Rv0601c	Two component sensor kinase [first part]
-	699038..699799	Rv0602c	Two component DNA binding transcriptional regulatory protein TcrA
+	699856..700167	Rv0603	Possible exported protein
+	700239..701189	Rv0604	Probable conserved lipoprotein LpqO
+	701406..702014	Rv0605	Possible resolvase
+	702016..702759	Rv0606	Possible transposase (fragment)
+	702813..703199	Rv0607	Hypothetical protein
+	703244..703489	Rv0608	Possible antitoxin VapB28
+	703486..703887	Rv0609	Possible toxin VapC28 Contains PIN domain
+	703830..704057	Rv0609A	hypothetical protein
+	704187..704247	RVnc0005	Putative small regulatory RNA
-	704752..705909	Rv0610c	Hypothetical protein
-	705961..706344	Rv0611c	Hypothetical protein
+	706324..706929	Rv0612	hypothetical protein
-	706948..709515	Rv0613c	hypothetical protein
+	709356..710348	Rv0614	hypothetical protein
+	710345..710587	Rv0615	Probable integral membrane protein
-	710584..710850	Rv0616c	Hypothetical protein
+	710782..711009	Rv0616A	Possible antitoxin VapB29
+	711006..711407	Rv0617	Possible toxin VapC29 Contains PIN domain
+	711536..712231	Rv0618	Probable galactose-1-phosphate uridylyltransferase GalTa [first part]
+	<712174..712719	Rv0619	Probable galactose-1-phosphate uridylyltransferase GalTb [second part]
+	712716..713807	Rv0620	Probable galactokinase GalK (galactose kinase)
+	714202..715266	Rv0621	Possible membrane protein

+	715370..716317	Rv0622	Possible membrane protein
+	716410..716664	Rv0623	Possible antitoxin VapB30
+	716664..717059	Rv0624	Possible toxin VapC30 Contains PIN domain
-	717153..717893	Rv0625c	Probable conserved transmembrane protein
+	718025..718285	Rv0626	Possible antitoxin VapB5
+	718282..718689	Rv0627	Possible toxin VapC5
-	718761..719912	Rv0628c	hypothetical protein
-	720005..721732	Rv0629c	Probable exonuclease V (alpha chain) RecD (exodeoxyribonuclease V alpha chain) (exodeoxyribonucleaseV polypeptide)
-	721729..725013	Rv0630c	Probable exonuclease V (beta chain) RecB (exodeoxyribonuclease V beta chain)(exodeoxyribonuclease Vpolypeptide) (chi-specific endonuclease)
-	725013..728306	Rv0631c	Probable exonuclease V (gamma chain) RecC (exodeoxyribonuclease V gamma chain)(exodeoxyribonucleaseV polypeptide)
-	728583..729278	Rv0632c	Probable enoyl-CoA hydratase EchA3 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
-	729327..730166	Rv0633c	Possible exported protein
-	730320..731033	Rv0634c	Possible glyoxalase II (hydroxyacylglutathione hydrolase) (GLX II)
+	731113..731364	Rv0634A	hypothetical protein
+	731494..731566	Rvnt06	tRNA-Thr
+	731603..731676	Rvnt07	tRNA-Met
+	731712..731879	Rv0634B	50S ribosomal protein L33 RpmG2
+	731930..732406	Rv0635	(3R)-hydroxyacyl-ACP dehydratase subunit HadA
+	732393..732821	Rv0636	(3R)-hydroxyacyl-ACP dehydratase subunit HadB
+	732825..733325	Rv0637	(3R)-hydroxyacyl-ACP dehydratase subunit HadC
+	733524..733596	Rvnt08	tRNA-Trp
+	733737..734222	Rv0638	Probable preprotein translocase SecE1
+	734254..734970	Rv0639	Probable transcription antitermination protein NusG
+	735022..735450	Rv0640	50S ribosomal protein L11 RplK

+	735517..736224	Rv0641	50S ribosomal protein L1 RplA
-	736298..737203	Rv0642c	Methoxy mycolic acid synthase 4 MmaA4 (methyl mycolic acid synthase 4) (MMA4) (hydroxy mycolic acidsynthase)
-	737268..738149	Rv0643c	Methoxy mycolic acid synthase 3 MmaA3 (methyl mycolic acid synthase 3) (MMA3) (hydroxy mycolic acidsynthase)
-	738297..739160	Rv0644c	Methoxy mycolic acid synthase 2 MmaA2 (methyl mycolic acid synthase 2) (MMA2) (hydroxy mycolic acidsynthase)
-	739327..740187	Rv0645c	Methoxy mycolic acid synthase 1 MmaA1 (methyl mycolic acid synthase 1) (MMA1) (hydroxy mycolic acidsynthase)
-	740234..741139	Rv0646c	Probable lipase/esterase LipG
-	741151..742617	Rv0647c	hypothetical protein
+	742719..746366	Rv0648	Alpha-mannosidase
+	746363..747037	Rv0649	Possible malonyl CoA-acyl carrier protein transacylase FabD2 (MCT)
+	747037..747945	Rv0650	Possible sugar kinase
+	748276..748812	Rv0651	50S ribosomal protein L10 RplJ
+	748849..749241	Rv0652	50S ribosomal protein L7/L12 RplL (SA1)
-	749234..749929	Rv0653c	Possible transcriptional regulatory protein (probably TetR-family)
+	750000..751505	Rv0654	Probable dioxygenase
+	751517..752596	Rv0655	Possible ribonucleotide-transport ATP-binding protein ABC transporter Mkl
-	752984..753367	Rv0656c	Possible toxin VapC6
-	753462..753617	Rv0657c	Possible antitoxin VapB6
-	753693..754409	Rv0658c	Probable conserved integral membrane protein
-	754685..754993	Rv0659c	Toxin MazF2
-	754980..755225	Rv0660c	Possible antitoxin MazE2
-	755335..755772	Rv0661c	Possible toxin VapC7
-	755769..756023	Rv0662c	Possible antitoxin VapB7
+	756137..758500	Rv0663	Possible arylsulfatase AtsD (aryl-sulfate sulphohydrolase) (arylsulphatase)
+	758532..758804	Rv0664	Possible antitoxin VapB8
+	758801..759139	Rv0665	Possible toxin VapC8
+	759136..759309	Rv0666	Possible membrane protein
+	759807..763325	Rv0667	DNA-directed RNA polymerase (beta

			chain) RpoB (transcriptase beta chain) (RNA polymerase beta subunit)
+	763370..767320	Rv0668	DNA-directed RNA polymerase (beta' chain) RpoC (transcriptase beta' chain) (RNA polymerase beta'subunit)
-	767684..769597	Rv0669c	Possible hydrolase
+	769792..770550	Rv0670	Probable endonuclease IV End (endodeoxyribonuclease IV) (apurinase)
+	770582..771424	Rv0671	Possible conserved lipoprotein LpqP
+	771484..773112	Rv0672	Probable acyl-CoA dehydrogenase FadE8
+	773123..774061	Rv0673	Possible enoyl-CoA hydratase EchA4 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
+	774064..774786	Rv0674	hypothetical protein
+	774783..775574	Rv0675	Probable enoyl-CoA hydratase EchA5 (enoyl hydrase) (unsaturated acyl-CoA hydratase) (crotonase)
-	775586..778480	Rv0676c	Probable conserved transmembrane transport protein MmpL5
-	778477..778905	Rv0677c	Possible conserved membrane protein MmpS5
+	778990..779487	Rv0678	hypothetical protein
-	779543..780040	Rv0679c	Conserved threonine rich protein
-	780042..780416	Rv0680c	Probable conserved transmembrane protein
+	780721..781311	Rv0681	Probable transcriptional regulatory protein (possibly TetR-family)
+	781560..781934	Rv0682	30S ribosomal protein S12 RpsL
+	781934..782404	Rv0683	30S ribosomal protein S7 RpsG
+	782485..784590	Rv0684	Probable elongation factor G FusA1 (EF-G)
+	784821..786011	Rv0685	Probable iron-regulated elongation factor TU Tuf (EF-TU)
+	786149..786946	Rv0686	Probable membrane protein
+	787099..787926	Rv0687	Probable short-chain type dehydrogenase/reductase
+	787940..789160	Rv0688	Putative ferredoxin reductase
-	789157..789411	Rv0689c	Hypothetical protein
-	790024..791073	Rv0690c	hypothetical protein
-	791070..791666	Rv0691c	Probable transcriptional regulatory protein
+	791658..791846	Rv0691A	Mycofactocin precursor protein
+	791831..792160	Rv0692	hypothetical protein

+	792157..793332	Rv0693	Probable coenzyme PQQ synthesis protein E PqqE (coenzyme PQQ synthesis protein III)
+	793335..794525	Rv0694	Possible L-lactate dehydrogenase (cytochrome) LldD1
+	794715..795470	Rv0695	hypothetical protein
+	795519..796931	Rv0696	Probable membrane sugar transferase
+	796933..798372	Rv0697	Probable dehydrogenase
+	798833..799444	Rv0698	hypothetical protein
+	799629..799850	Rv0699	Hypothetical protein
+	800487..800792	Rv0700	30S ribosomal protein S10 RpsJ (transcription antitermination factor NusE)
+	800809..801462	Rv0701	50S ribosomal protein L3 RplC
+	801462..802133	Rv0702	50S ribosomal protein L4 RplD
+	802133..802435	Rv0703	50S ribosomal protein L23 RplW
+	802528..803370	Rv0704	50S ribosomal protein L2 RplB
+	803411..803692	Rv0705	30S ribosomal protein S19 RpsS
+	803689..804282	Rv0706	50S ribosomal protein L22 RplV
+	804282..805106	Rv0707	30S ribosomal protein S3 RpsC
+	805110..805526	Rv0708	50S ribosomal protein L16 RplP
+	805526..805759	Rv0709	50S ribosomal protein L29 RpmC
+	805756..806166	Rv0710	30S ribosomal protein S17 RpsQ
+	806335..808698	Rv0711	Possible arylsulfatase AtsA (aryl-sulfate sulphohydrolase) (arylsulphatase)
+	808746..809645	Rv0712	hypothetical protein
+	809946..810887	Rv0713	Probable conserved transmembrane protein
+	811373..811741	Rv0714	50S ribosomal protein L14 RplN
+	811742..812059	Rv0715	50S ribosomal protein L24 RplX
+	812059..812622	Rv0716	50S ribosomal protein L5 RplE
+	812627..812812	Rv0717	30S ribosomal protein S14 RpsN1
+	812976..813374	Rv0718	30S ribosomal protein S8 RpsH
+	813398..813937	Rv0719	50S ribosomal protein L6 RplF
+	813940..814308	Rv0720	50S ribosomal protein L18 RplR
+	814328..814990	Rv0721	30S ribosomal protein S5 RpsE
+	814993..815190	Rv0722	50S ribosomal protein L30 RpmD
+	815190..815630	Rv0723	50S ribosomal protein L15 RplO
+	815663..817534	Rv0724	Possible protease IV SppA (endopeptidase IV) (signal peptide peptidase)
-	817531..>817866	Rv0724A	hypothetical protein
-	817539..818444	Rv0725c	hypothetical protein

-	818537..819640	Rv0726c	Possible S-adenosylmethionine-dependent methyltransferase
-	819843..820499	Rv0727c	Possible L-fuculose phosphate aldolase FucA (L-fuculose-1-phosphate aldolase)
-	820496..821476	Rv0728c	Possible D-3-phosphoglycerate dehydrogenase SerA2 (phosphoglycerate dehydrogenase) (PGDH)
+	821507..822853	Rv0729	Possible D-xylulose kinase XylB (xylulokinase) (xylulose kinase)
+	822866..823594	Rv0730	GCN5-related N-acetyltransferase
-	823683..824639	Rv0731c	Possible S-adenosylmethionine-dependent methyltransferase
+	824800..826125	Rv0732	Probable preprotein translocase SecY
+	826122..826667	Rv0733	Adenylate kinase Adk (ATP-AMP transphosphorylase)
+	826670..827470	Rv0734	Methionine aminopeptidase MapA (map) (peptidase M) (MetAP)
+	827543..828076	Rv0735	Probable alternative RNA polymerase sigma factor SigL
+	828140..828892	Rv0736	Anti-sigma factor RslA
+	829207..829704	Rv0737	Possible transcriptional regulatory protein
+	830062..830610	Rv0738	hypothetical protein
+	830855..831661	Rv0739	hypothetical protein
+	831776..832303	Rv0740	hypothetical protein
+	832534..832848	Rv0741	Probable transposase (fragment)
+	832981..833508	Rv0742	PE-PGRS family protein PE_PGRS8
-	833886..834443	Rv0743c	Hypothetical protein
-	834440..834946	Rv0744c	Possible transcriptional regulatory protein
+	835154..835681	Rv0745	hypothetical protein
+	835701..838052	Rv0746	PE-PGRS family protein PE_PGRS9
+	838451..840856	Rv0747	PE-PGRS family protein PE_PGRS10
+	840947..841204	Rv0748	Possible antitoxin VapB31
+	841228..841656	Rv0749	Possible toxin VapC31 Contains PIN domain
+	1855764..1856696	Rv1646	PE family protein PE17
+	1856774..1857724	Rv1647	Adenylate cyclase (ATP pyrophosphate-lyase) (adenylyl cyclase)
+	1857731..1858537	Rv1648	Probable transmembrane protein
+	1858733..1859758	Rv1649	Probable phenylalanyl-tRNA synthetase, alpha chain PheS

+	1859758..1862253	Rv1650	Probable phenylalanyl-tRNA synthetase, beta chain PheT
-	1862347..1865382	Rv1651c	PE-PGRS family protein PE_PGRS30
+	1865576..1866634	Rv1652	Probable N-acetyl-gamma-glutamyl-phoshate reductase ArgC
+	1866631..1867845	Rv1653	Probable glutamate N-acetyltransferase ArgJ
+	1867842..1868726	Rv1654	Probable acetylglutamate kinase ArgB
+	1868723..1869925	Rv1655	Probable acetylornithine aminotransferase ArgD
+	1869922..1870845	Rv1656	Probable ornithine carbamoyltransferase, anabolic ArgF
+	1870842..1871354	Rv1657	Probable arginine repressor ArgR (AHRC)
+	1871363..1872559	Rv1658	Probable argininosuccinate synthase ArgG
+	1872639..1874051	Rv1659	Probable argininosuccinate lyase ArgH
+	1874160..1875221	Rv1660	Chalcone synthase Pks10
+	1875304..1881684	Rv1661	Probable polyketide synthase Pks7
+	1881704..1886512	Rv1662	Probable polyketide synthase Pks8
+	1886512..1888020	Rv1663	Probable polyketide synthase Pks17
+	1888026..1891079	Rv1664	Probable polyketide synthase Pks9
+	1891226..1892287	Rv1665	Chalcone synthase Pks11
-	1892270..1893562	Rv1666c	Probable cytochrome P450 139 Cyp139
-	1893577..1894230	Rv1667c	Probable second part of macrolide-transport ATP-binding protein ABC transporter
-	1894224..1895342	Rv1668c	Probable first part of macrolide-transport ATP-binding protein ABC transporter
+	1895725..1896087	Rv1669	Hypothetical protein
+	1896120..1896467	Rv1670	hypothetical protein
+	1896475..1896867	Rv1671	Probable membrane protein
-	3839691..3840197	Rv3422c	tRNA threonylcarbamoyladenosine biosynthesis protein
-	3840194..3841420	Rv3423c	alanine racemase
-	3841714..3842076	Rv3424c	hypothetical protein
+	3842239..3842769	Rv3425	PPE family protein PPE57
+	3843036..3843734	Rv3426	PPE family protein PPE58
-	3843885..3844640	Rv3427c	transposase
-	3844738..3845970	Rv3428c	transposase
+	3847165..3847701	Rv3429	PPE family protein PPE59
-	3847642..3848805	Rv3430c	transposase
-	3848844..3848999	Rv3430a	hypothetical protein

-	3850372..3851754	Rv3432c	glutamate decarboxylase GadB
-	1266485..1266985	Rv1139c	hypothetical protein
+	1267347..1268195	Rv1140	integral membrane protein
-	1268203..1269009	Rv1141c	enoyl-CoA hydratase EchA11
-	1269152..1269958	Rv1142c	enoyl-CoA hydratase EchA10
+	1270062..1271144	Rv1143	alpha-methylacyl-CoA racemase
+	1271156..1271908	Rv1144	oxidoreductase
+	1272423..1273334	Rv1145	transmembrane transport protein
+	1273355..1274767	Rv1146	transmembrane transport protein
+	1777859..1778539	Rv1570	ATP-dependent dethiobiotin synthetase BioD
+	1778539..1779048	Rv1571	hypothetical protein
+	1779314..1779724	Rv1573	phage protein
+	1779930..1780241	Rv1574	phage protein
+	1780199..1780699	Rv1575	phage protein
-	1780643..1782064	Rv1576c	phage capsid protein
-	1782072..1782584	Rv1577c	phage prohead protease
-	1782758..1783228	Rv1578c	phage protein
-	1783309..1783623	Rv1579c	phage protein
-	1783620..1783892	Rv1580c	phage protein
-	1783906..1784301	Rv1581c	phage protein
-	1784497..1785912	Rv1582c	phage protein
-	1785912..1786310	Rv1583c	phage protein
-	1786307..1786528	Rv1584c	phage protein
-	1786584..1787099	Rv1585c	phage protein
-	1787096..1788505	Rv1586c	phage integrase
-	1788162..1789163	Rv1587c	hypothetical protein
-	1789168..1789836	Rv1588c	hypothetical protein
+	1790284..1791333	Rv1589	biotin synthetase
+	1791334..1791573	Rv1590	hypothetical protein
+	1791570..1792235	Rv1591	transmembrane protein
+	3939617..3941761	Rv3511	PE-PGRS family protein PE_PGRS55
+	3943812..3944963	Rv3512	PE-PGRS family protein PE_PGRS56
-	3945092..3945748	Rv3513c	fatty-acid--CoA ligase FadD18
+	3945794..3950263	Rv3514	PE-PGRS family protein PE_PGRS57
-	3950824..3952470	Rv3515c	long-chain-fatty-acid--CoA ligase FadD19
+	3952544..3953335	Rv3516	enoyl-CoA hydratase EchA19

Appendix 19

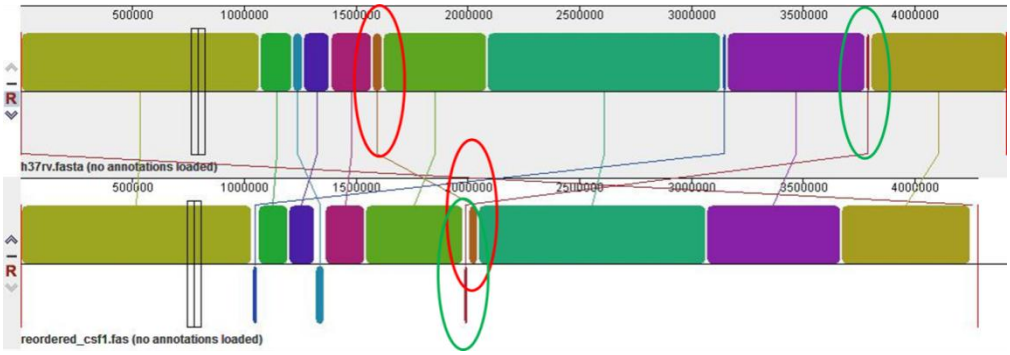


Figure A19.1: MAUVE alignment of UM-CSF01 and H37Rv shows the recombination sites (circled) for translocations, which were not located within genes. Hence, the rearrangements did not affect adjacent gene sequences.

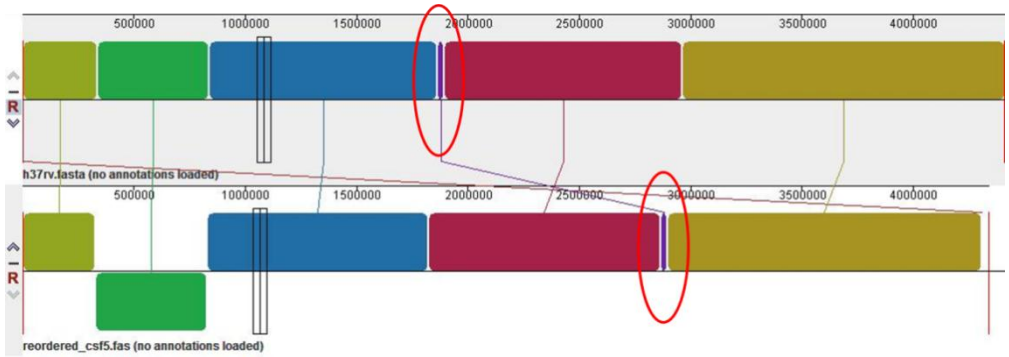
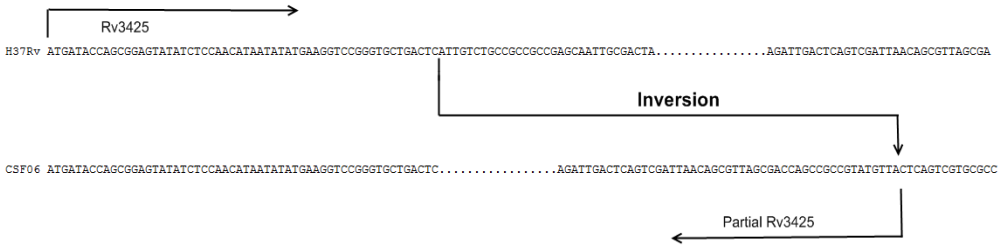


Figure A19.2: MAUVE alignment of UM-CSF05 and H37Rv shows the recombination sites (circled) for translocations, which were not located within genes. Hence, the rearrangements did not affect adjacent gene sequences.

Appendix 20

The information below show the genome rearrangements found in four affected UM-CSF strains. The rearrangements caused deletions resulting in gene truncations.

A20.1 CSF 06




```
H37Rv_Rv3425        ATGATACCAGCGGAGTATATCTCCAACATAATATATGAAGGCCCGGGCGCTGACTCATTG
CSF6_gene3775       ------------------------------------------------------------TTG
                                                                                ***

H37Rv_Rv3425        TTTTTCGCCTCCGGGCAATTGCGAGAATTGGCTTACTCAGTTGAAACGACGGCTGAGTCG
CSF6_gene3775       TATGCCGCCGACCAGCGATTGCGACAATTAGCTGACTCAGTTAGAACGACTGCCGAGTCG
                    * *   ****   *   ** ******* **** *** ********   ****** ** ******

H37Rv_Rv3425        CTCGAGGACGAGCTCGACGAGCTGGATGAGAACTGGAAAGGTAGTTCGTCGGACTTGTTG
CSF6_gene3775       CTCAACACCACGCTCGACGAGCTGCACGAGAACTGGAAAGGTAGTTCATCGGAATGGATG
                    *** *    *   ************** * ********************* ***** * * **

H37Rv_Rv3425        GCCGACGCGGTTGAGCGGTATCTCCAATGGCTGTCTAAACACTCCAGTCAGCTTAAGCAT
CSF6_gene3775       GCCGACGCGGCTTTGCGGTATCTCGACTGGCTGTCTAAACACTCCCGTCAGATTTTGCGA
                    ********** *   ********** * ****************** ***** **   **

H37Rv_Rv3425        GCCGCCTGGGTGATCAACGGCCTCGCGAACGCCTATAACGACACACGTCGGAAGGTGGTA
CSF6_gene3775       ACCGCCCGCGTGATCGAATCCCTCGTAATGGCCTATGAGGAGACACTTCTGAGGGTGGTA
                      ***** * ****** *    *****   *   ****** * ** **** ** **  *******

H37Rv_Rv3425        CCCCCGGAGGAGATCGCCGCCAACCGCGAGGAGAGGCGCAGGCTGATCGCGAGCAACGTG
CSF6_gene3775       CCCCCGGCGACTATCGCCAACAACCGCGAGGAGGTGCGCAGGCTGATCGCGAGCAACGTG
                    ******* *    ******  *************  *************************

H37Rv_Rv3425        GCCGGGGTAAACACTCCAGCAATCGCAGACCTCGATGCACAATACGACCAGTACCGGGCC
CSF6_gene3775       GCCGGGGTAAACACTCCAGCAATCGCAGACCTCGAGGCACAATACGAGCAGTACCGGGCC
                    ************************************ ********** ***********

H37Rv_Rv3425        CGCAATGTCGCTGTAATGAACGCCTATGTAAGTTGGACCCGATCTGCGCTATCGGATCTG
CSF6_gene3775       GAAAATATCCAAGCAATGGACCGCTATCTAAGTTGGACCCGATTTGCGCTATCGAAGCTG
                       *** **    * **** **   **** ************** ********** * ***

H37Rv_Rv3425        CCCCGGTGGCGGGAACCGCCGCAGATCTACAGGGGCGGGTAG
CSF6_gene3775       CCCCGATGGCGGGAGCCGCCGCAGATCCACAGGAGCGGGTAG
                    ***** ******** ************ ***** ********
```

Descriptions:

The affected regions consist of four genes, which are Rv3425 – Rv3428c. However, the affected truncation is only Rv3425, other three genes remain intact. The figures are illustrating the Rv3425 gene.

- The arrow labelled with Rv3425 in H37Rv marks the starting point of the gene.
- The arrow labelled with “Inversion” marks the site where the inversion occurred.

- In CSF06, the 5’ of the Rv3425 is present. However, the nucleotide starting from position 57th has been found at the position as marked with the arrow label “Partial 3425”.
- The alignment shown is the resulting annotation.

A20.2 CSF09

```
H37Rv_Rv1141c     ATGCTGCACATGCCAGATTCCGGGATTGCCGCATTAACGCCGGTCACAGGCCTCAACGTC
CSF9_gene3339     ATGCTGCACATGCCAGATTCCGGGATTGCCGCATTAACGCCGGTCACAGGCCTCAACGTC
                  ************************************************************

H37Rv_Rv1141c     ACCCTGACCGACAGAGTGTTGTCGGTGCGCATCAACCGCCCTAGCAGTCTCAACTCGCTG
CSF9_gene3339     ACCCTGACCGACAGAGTGTTGTCGGTGCGCATCAACCGCCCTAGCAGTCTCAACTCGCTG
                  ************************************************************

H37Rv_Rv1141c     ACCGTGCCAATCCTGACGGGGATCGCCGACACGCTGGAGCGCGCGGCGGCCGATCCCGTG
CSF9_gene3339     ACCGTGCCAATCCTGACGGGGATCGCCGACACGCTGGAGCGCGCGGCGGCCGATCCCGTG
                  ************************************************************

H37Rv_Rv1141c     GTCAAGGTGGTGCGCCTAGGCGGGGTGGGCCGCGGTTTCAGCTCCGGAGTGTCTATGTCT
CSF9_gene3339     GTCAAGGTGGTGCGCCTAGGCGGGGTGGGCCGCGGTTTCAGCTCCGGAGTGTCTATGTCT
                  ************************************************************

H37Rv_Rv1141c     GTGGACGATGTGTGGGGCGGAGGGCCGCCGACCGCCATCGTCGAAGAGGCCAACCGCGCA
CSF9_gene3339     GTGGACGATGTGTGGGGCGGAGGGCCGCCGACCGCCATCGTCGAAGAGGCCAACCGCGCA
                  ************************************************************

H37Rv_Rv1141c     GTACGCGCCGTGGCCGCGCTACCGCACCCGGTTGTAGCTGTCGTTCAAGGACCAGCGGTC
CSF9_gene3339     GTACGCGCCGTGGCCGCGCTACCGCACCCGGTTGTAGCTGTCGTTCAAGGACCAGCGGTC
                  ************************************************************

H37Rv_Rv1141c     GGCGTCGCTGTCTCGCTAGCGCTGGCGTGTGACTTCATATTGGCTTCTGATAGTGCATTT
CSF9_gene3339     GGCGTCGCTGTCTCGCTAGCGCTGGCGTGTGACTTCATATTGGCTTCTGATAGTGCATTT
                  ************************************************************

H37Rv_Rv1141c     TTCATGCTCGCCAACACCAAGGTAGCGTTGATGCCCGACGGCGGCGCATCGGCGTTAGTC
CSF9_gene3339     TTCATGCTCGCCAACACCAAGGTAGCGTTGATGCCCGACGGCGGC---------------
                  *******************************************

H37Rv_Rv1141c     GCCGCGGCCACCGGCCGGATCCGGGCGATGCGGCTGGCGCTGCTGGCCGAGCAACTGCCG
CSF9_gene3339     ------------------------------------------------------------

H37Rv_Rv1141c     GCCCGCGAGGCACTGGCCTGGGGCCTGATCAGCGCGGTATATCCGGACAGCGACTTCGAG
CSF9_gene3339     ------------------------------------------------------------

H37Rv_Rv1141c     GCCGAGGTGGACAAGGTGATTTCACGGTTGCTGGCCGGCCCGGCGCTGGCGTTCGCCCAG
CSF9_gene3339     ------------------------------------------------------------

H37Rv_Rv1141c     GCCAAAAACGCCATCAATGCAGCCGCCCTCACCGAATTGGAACCCACGTTCGCGCGCGAA
CSF9_gene3339     ------------------------------------------------------------

H37Rv_Rv1141c     TTGGATGGACAGGAAGTCCTGCTGCGAACACACGACTTCGCCGAGGGCGCAGCGGCGTTC
CSF9_gene3339     ------------------------------------------------------------

H37Rv_Rv1141c     CTGCAACGCCGCACCCCCAACTTCACCGGTTCCTGA
CSF9_gene3339     -----------------------------------
```

Description:

Rv1141c is truncated from nucleotide number 466.

A20.3 CSF15

```
H37Rv_Rv1587c       GACGGCATGTCACGGCTAAGTGGCTACCTGACCCCCCAAGCGCGGGCCACCTTTGAAGCC
CSF15_gene4288      GACGGCATGTCACGGCTAAGTGGCTACCTGACCCCCCAAGCGCGGGCCACCTTTGAAGCC
                    ************************************************************

H37Rv_Rv1587c       GTGCTAGCCAAACTGGCCGCCCCCGGCGCGACCAACCCCGACGACCACACCCCGGTCATC
CSF15_gene4288      GTGCTAGCCAAACTGGCCGCCCCCGGCGCGACCAACCCCGACGACCACACCCCGGTCATC
                    ************************************************************

H37Rv_Rv1587c       GACACCACCCCCGATGCGGCCGCCATCGACCGCGACACCCGCAGCCAAGCCCAACGCAAC
CSF15_gene4288      GACACCACCCCCGATGCGGCCGCCATCGACCGCGACACCCGCAGCCAAGCCCAACGCAAC
                    ************************************************************

H37Rv_Rv1587c       CACGACGGGCTGCTGGCCGGGCTGCGCGCGCTGATCGCCTCCGGGAAACTGGGCCAACAC
CSF15_gene4288      CACGACGGGCTGCTGGCCGGGCTGCGCGCGCTGATCGCCTCCGGGAAACTGGGCCAACAC
                    ************************************************************

H37Rv_Rv1587c       AACGGTCTTCCCGTCTCGATCGTGGTCACCACCACCCTGACCGACCTGCAAACCGGCGCC
CSF15_gene4288      AACGGTCTTCCCGTCTCGATCGTGGTCACCACCACCCTGACCGACCTGCAAACCGGCGCC
                    ************************************************************

H37Rv_Rv1587c       GGCAAGGGCTTCACCGGCGGCGGCACCCTGCTACCCATGGCCGATGTGATCCGCATGACC
CSF15_gene4288      GGCAAGGGCTTCACCGGCGGCGGCACCCTGCTACCCATGGCCGATGTGATCCGCATGACC
                    ************************************************************

H37Rv_Rv1587c       AGCCACGCCCACCACTACTCCCCCGCAAGCGGGAGGTACCCCCAGGCGATCTTCGACCAC
CSF15_gene4288      AGCCACGCCCACCACTACTCCCCCGCAAGCGGGAGGTACCCCCAGGCGATCTTCGACCAC
                    ************************************************************

H37Rv_Rv1587c       GGCACACCCCTGGCGCTGTATCACACCAAACGCCTAGCCTCCCCGGCCCAGCGGATCATG
CSF15_gene4288      GGCACACCCCTGGCGCTGTATCACACCAAACGCCTAGCCTCCCCGGCCCAGCGGATCATG
                    ************************************************************

H37Rv_Rv1587c       CTGTTCGCCAACGACCGCGGCTGCACCAAACCCGGCTGTGACGCACCGGCCTACCACAGC
CSF15_gene4288      CTGTTCGCCAACGACCGCGGCTGCACCAAACCCGGCTGTGACGCACCGGCCTACCACAGC
                    ************************************************************

H37Rv_Rv1587c       CAAGCCCACCACGTCACCGCCTGGACCAGCACCGGACGCACCGACATCACCGAGCTGACC
CSF15_gene4288      CAAGCCCACCACGTCACCGCCTGGACCAGCACCGGACGCACCGACATCACC---------
                    *************************************************

H37Rv_Rv1587c       CTGGCCTGCGGCCCCGACAACCGACTCGCCGAAAAAGGCTGGACCACCCACAACAACACC
CSF15_gene4288      ------------------------------------------------------------

H37Rv_Rv1587c       CACGGCCACACCGAATGGCTACCACCACCCCACCTCGACCACGGCCAACCGTGGACCTGT
CSF15_gene4288      ------------------------------------------------------------

H37Rv_Rv1587c       GAGATACACTACACCTGTGCGTGCTGCTGTCTACCTCCGAATCTCAGAAGACCGCTCCGG
CSF15_gene4288      ------------------------------------------------------------

H37Rv_Rv1587c       CGAACAGCTCGGCGTGGCCCGCCAACGCGAGGACTGCCTAAAGCTGTGCGGGCAGCGAAA
CSF15_gene4288      ------------------------------------------------------------
```

Description:

Rv1587c is truncated.

A20.4 CSF17

Genome analysis had showed the event which caused deletions in Rv3513c. But the involved contig is too short for further clarification.

Appendix 21

Table A21: Enrichment analysis using DAVID on proteins which showed globularity differences between UM-CSF strains and H37Rv. The single largest group of the enriched proteins involved transcription factors and transcription regulators.

Category	Term	p-Value
COG_ONTOLOGY	Transcription	3.66E-04
GOTERM_BP_FAT	GO:0006355~regulation of transcription, DNA-dependent	7.73E-07
GOTERM_BP_FAT	GO:0051252~regulation of RNA metabolic process	8.90E-07
GOTERM_BP_FAT	GO:0045449~regulation of transcription	4.39E-06
GOTERM_BP_FAT	GO:0006350~transcription	2.81E-05
GOTERM_BP_FAT	GO:0009187~cyclic nucleotide metabolic process	0.009860284
GOTERM_BP_FAT	GO:0009190~cyclic nucleotide biosynthetic process	0.009860284
GOTERM_BP_FAT	GO:0007242~intracellular signalling cascade	0.028798523
GOTERM_CC_FAT	GO:0016021~integral to membrane	4.42E-06
GOTERM_CC_FAT	GO:0031224~intrinsic to membrane	6.36E-05
GOTERM_CC_FAT	GO:0005886~plasma membrane	0.032338255
GOTERM_MF_FAT	GO:0003700~transcription factor activity	7.77E-07
GOTERM_MF_FAT	GO:0030528~transcription regulator	4.63E-06

	activity	
GOTERM_MF_FAT	GO:0003677~DNA binding	5.14E-05
GOTERM_MF_FAT	GO:0016849~phosphorus-oxygen lyase activity	0.009764838
INTERPRO	IPR001647:Transcriptional regulator, TetR-like, DNA-binding, bacterial/archaeal	1.41E-05
INTERPRO	IPR012287:Homeodomain-related	1.59E-04
INTERPRO	IPR001279:Beta-lactamase-like	0.001356161
INTERPRO	IPR013216:Methyltransferase type 11	0.009800718
INTERPRO	IPR002641:Patatin	0.032394203
INTERPRO	IPR001054:Adenylyl cyclase class-3/4/guanylyl cyclase	0.045571002
KEGG_PATHWAY	mtu00350:Tyrosine metabolism	3.55E-10
KEGG_PATHWAY	mtu00150:Androgen and oestrogen metabolism	2.19E-07
KEGG_PATHWAY	mtu00450:Selenoamino acid metabolism	1.70E-06
KEGG_PATHWAY	mtu00340:Histidine metabolism	2.48E-06
KEGG_PATHWAY	mtu00626:Naphthalene and anthracene degradation	1.84E-05
KEGG_PATHWAY	mtu00642:Ethylbenzene degradation	6.92E-04
KEGG_PATHWAY	mtu00360:Phenylalanine metabolism	0.00369836
KEGG_PATHWAY	mtu00860:Porphyrin and chlorophyll metabolism	0.014895679
KEGG_PATHWAY	mtu00780:Biotin metabolism	0.019713574
KEGG_PATHWAY	mtu03430:Mismatch repair	0.044554197
KEGG_PATHWAY	mtu00310:Lysine degradation	0.044572173
SWISS-PROT	transcription regulation	0.008901379
SWISS-PROT	DNA-binding	0.009996586
SWISS-PROT	Transcription	0.010891026

Appendix 22

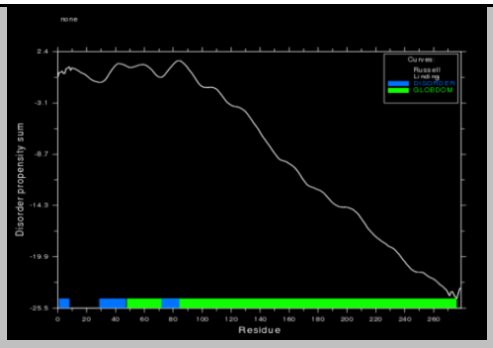
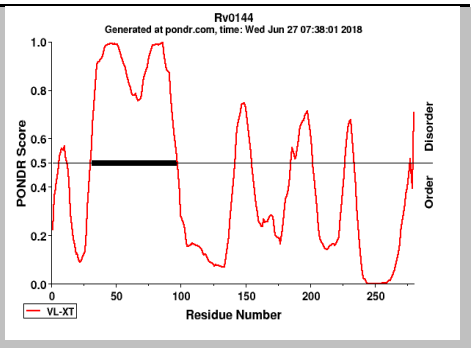
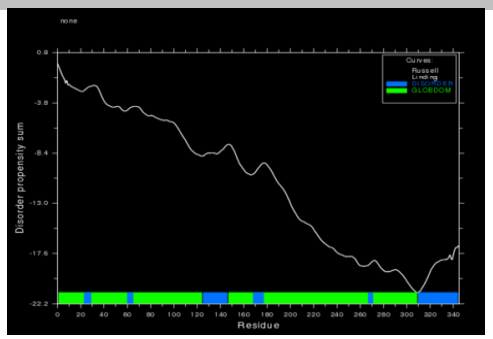
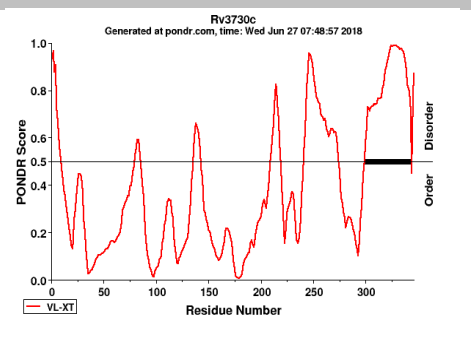
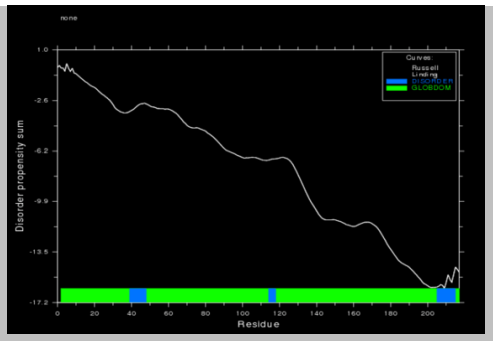
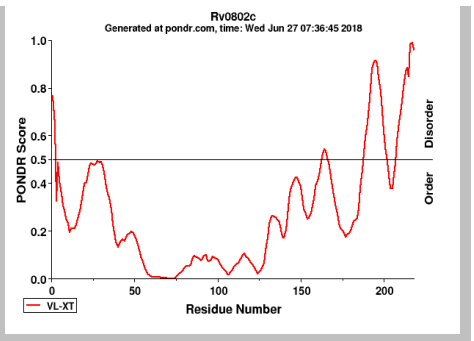
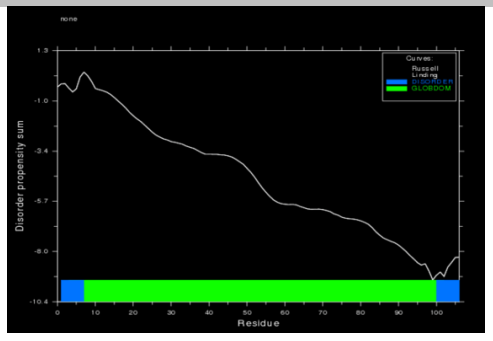
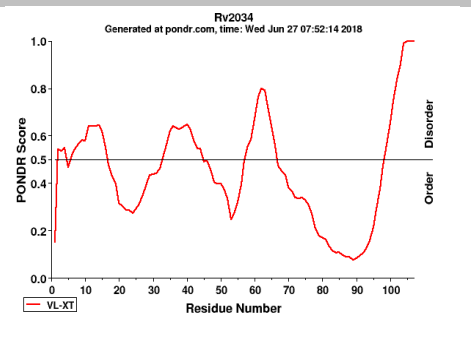
Protein	GlobPlot	PONDR Score
Rv0144		
Rv3730c		
Rv0802c		
Rv2034		

Figure A22: Four of the enriched proteins found in three different CSF strains: Rv0144 (probable transcriptional regulatory protein with Leu to Asp substitution), Rv3730c (hypothetical protein with Thr to Asp substitution), Rv0802c (possible succinyltransferase with Ser to Pro substitution) and Rv2034 (transcriptional regulatory protein with Ala to Thr substitution). GlobPlot (http://globplot.embl.de/cgiDict.py) on the left shows regions of globularity (green) and disorder (blue) within protein sequences. Protein disorder is due to the lack of regular secondary structure and a high degree of flexibility in the polypeptide chain, and thus a decrease in stability (Wright and Dyson, 1999). Ordered regions are termed globular and typically contain regular secondary structures packed into a compact globule. On the right panel, PONDR score (http://www.pondr.com/) was used to predict disorder score for the proteins (Garner et al., 1999). All four showed higher propensity to disorder (overall percentage disorder for Rv0144, Rv3730c, Rv0802c and Rv2034 are 40.0 %, 32.4 %, 14.2 % and 41.1 %, respectively). All shows a decrease in stability except for Rv0802c.

Appendix 23

Protein	Peptides	Hydrophobicity/ hydrophilicity
Rv0144 (280 aa)	VPHSWTPTSVMTPPLVVAAFRPVGHYRLATDRAGGPCSPPATGA KLTSSVASRPTVGTKPQWWHTLVMSMSLTAGRGPGRPPAAKAD ETRKRILHAARQVFSERGYDGATFQEIAVRADLTRPAINHYFANK RVLYQEVVEQTHELVIVAGIERARREPTLMGRLAVVVDFAMEAD AQYPASTAFLATTVLESQRHPELSRTENDAVRATREFLVWAVND AIERGELAADVDVSSLAETLLVVLCGVGFYIGFVGSYQRMATITD SFQQLLAGTLWRPPT	Hydrophobic: **48.21%** Acidic: 9.64% Basic: 12.86% Neutral: 29.29%
Rv3730c (346 aa)	MAAAAEELDVDGIAVRLTSPDRMYFPKLGSHGTKRRLVEYYFAV AGGPMLTALRDRPTHLQRFPDGVDGEQIYQKRIPRHRPDYLQTCR VTFPSGRMADALKVTHPAAIVWAAQMGTITLHPWQVRCPDTEHP DELRIDLDPQPGTGFVEARTVAVDVLRSVLDDLGLVGYPKTSGG RGIHVFLRIATDWDFVEVRRAGIALAREVERRAPDAVTTSWWKE ERGARIFIDFNQNARDRTMASAYSVRPTPIATVSMPLTWEELAGA DPDDYTMTTVPELVKIRDDPWAGMDDVAQSIAPLLDLAAADEER GLGDMPYPPNYPKMPGEPKRVQPSRDTDLKGGNTSK	Hydrophobic: **45.66%** Acidic: 14.45% Basic: 14.45% Neutral: 25.43%
Rv0802c (218 aa)	MSRHWPLFDLRITTPRLQLQLPTEELCDQLIDTILEGVHDPDRMPF SVPWTRASREDLPFNTLSHLWQQLAGFKRDDWSLPLAVLVDGR AVGVQALSSKDFPITRQVDSGSWLGLRYQGHGYGTEMRAAVLY FAFAELEAQVATSRSFVDNPASIAVSRRNGYRDNGLDRVAREGA MAEALLFRLTRDDWQRHRTVEVRVDGFDRCRPLFGPLEPPRY	Hydrophobic: **44.95%** Acidic: 13.30% Basic: 14.68% Neutral: 27.06%
Rv2034 (107 aa)	VSTYRSPDRAWQALADGTRRAIVERLAHGPLAVGELARDLPVSR PAVSQHLKVLKTARLVCDRPAGTRRVYQLDPTGLAALRTDLDRF WTRALTGYAQLIDSEGDDT	Hydrophobic: **43.93%** Acidic: 12.15% Basic: 16.82% Neutral: 27.10%

Figure A23: Hydrophobicity amino acids form less than 50.0 % of the amino acid sequence in these disordered proteins of Rv0144, Rv3730c, Rv0802c, and Rv2034. Green letters shows hydrophobic uncharged residues, blue letters shows basic residues and black letters shows other residues. The above hydrophobicity was calculated using the program Peptide 2.0 (https://www.peptide2.com/).

Appendix 24

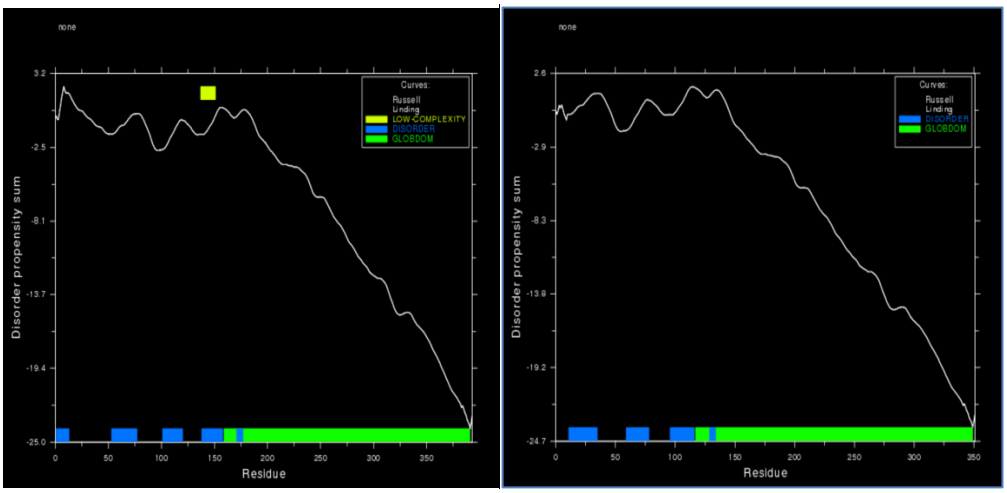
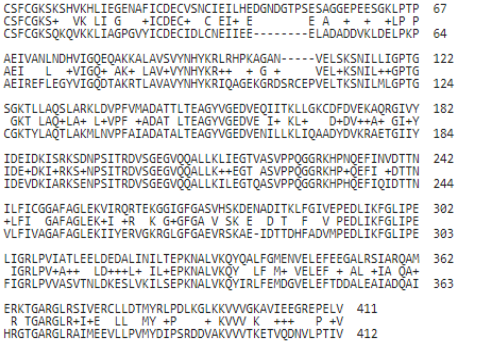
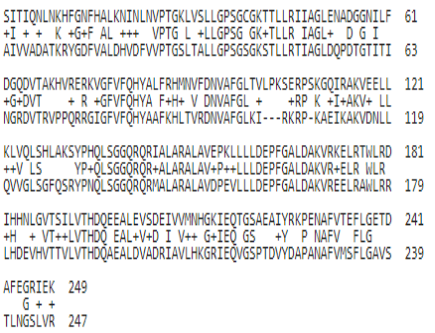
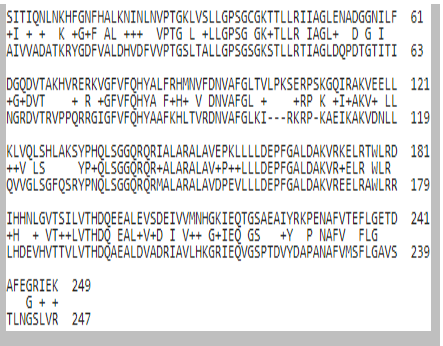
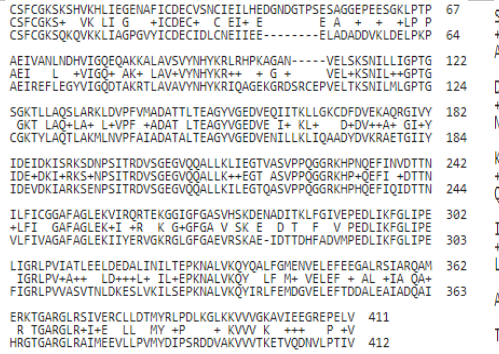
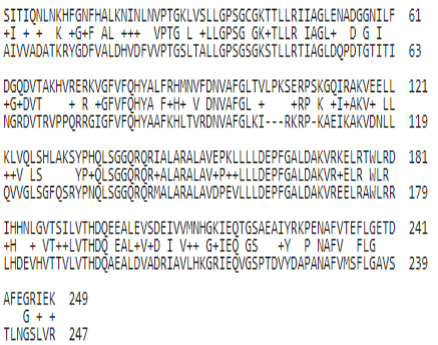


Figure A24: GlobPlot prediction for Rv0619 in H37Rv (left) and UM-CSF strains (right), which had a substitution of Ala (GCC) for Thr (ACC) at position 174. The disorder propensity versus residue graph shows an increase in the globular segment in UM-CSF strains with a change from polarity to hydrophobicity. The graph was generated from http://globplot.embl.de/.

Appendix 25

Table A25: Alignment of meningitis-associated protein (Rv2457c and Rv2397c) from *N. meningitidis*, H37Rv, UM-CSF01 and UM-CSF05 strains, using protein BLAST (BLASTP).

Protein from *N. meningitidis*	Rv2457c	Rv2397c
	60.0 % Identical	**56.0 % Identical**
H37Rv		
UM-CSF01		
UM-CSF05		

Appendix 26

Table A26: The summary of BLASTN results of meningitis-associated gene Rv1699, Rv2606c, Rv0357c from *S. pneumoniae* and Rv2457c and Rv2397c from *N. meningitidis* in 26 *Mtb* strains from NCBI, isolated from different sources and different times (National Center for Biotechnology Information, 2018).

No	Country	Strain	Sputum	CSF	Rv1699	Rv2606c	Rv0357c	Rv2457c	Rv2397c
1	India	LJ338	√		-	+	+	+	+
2	India	CAS		√	-	+	+	+	+
3	Papua New Guinea	Beijing2014PNGD	√		-	+	-	+	+
4	Hong Kong	WC078	√		-	+	+	+	+
5	India	LJ319		√	-	+	-	-	+

6	Guatemala	GG-229-10	√		**+**	**+**	**+**	**+**	**+**
7	Peru	TBV4952	√		**-**	**+**	**+**	**-**	**+**
8	Peru	TBDM2717	√		**+**	**+**	**+**	**-**	**+**
9	Peru	MDRMA2441	√		**+**	**+**	**+**	**+**	**+**
10	Peru	LN2900	√		**+**	**+**	**+**	**+**	**+**
12	Peru	CSV5769	√		**+**	**+**	**+**	**+**	**+**
13	Peru	SLM100	√		**+**	**+**	**+**	**-**	**+**
14	Peru	TBV4766	√		**+**	**+**	**+**	**+**	**+**
15	Peru	TBDM2489	√		**+**	**+**	**+**	**+**	**+**
16	Peru	ME1473	√		**+**	**+**	**+**	**+**	**+**
17	Peru	MDRDM1098	√		**+**	**+**	**+**	**+**	**+**
18	Peru	LE492	√		**+**	**+**	**+**	**+**	**+**
19	India	C3		√	**-**	**+**	**-**	**+**	**+**
20	India	S3	√		**-**	**+**	**+**	**+**	**+**
21	Seoul	MTB2	√		+	**+**	+	+	+

22	Hanoi	HN-506	√		-	-	+	+	-
23	India	TBMENG-03	√		-	+	+	+	+
24	Hong Kong	H112		√	-	+	+	+	+
25	Russia	RUS B0	√		-	+	+	+	+
26	USA	TRS2		√	+	-	-	-	-

"+" sign indicates the gene is present in the *Mtb* strain.

"-" sign indicates the gene is absent in the *Mtb* strain.

Appendix 27

Table A27: A summary of the genes and putative genetic features that might be associated with neurotropism in *Mtb*. A total of 35 genes were identified of which six (in bold) were selected for verification using PCR and Sanger sequencing.

Type of Analysis	Program	Study samples	Gene affected	Consequences
Genome rearrangement	Gepard	H37Rv	Rv1587c	Deletion-gene truncation
	Mauve	13 *Mtb* local sputum isolates	**Rv3425**	
	Mugsy	16 *Mtb* downloaded genome	**Rv1141c**	
		8 UM-CSF isolates	Rv3513c	
Micro-variants	Burrow-Wheeler Aligner	H37Rv	Rv1267c	Amino acid changes
	Mpileup of samtools	13 *Mtb* local sputum isolates	Rv1899c	
	snpEff	56 *Mtb* respiratory SRA	Rv0747	

		8 UM-CSF isolates	Rv1067c	
			Rv1087	
			Rv3344c	
			Rv3426	
			Rv0278c	
			Rv3159c	
			Rv1753c	
Protein globularity changes	GlobPlot	H37Rv	Rv0144	Higher propensity to disorder
	DAVID	8 UM-CSF isolates	Rv3730c	A decrease in stability
			Rv0802c	
			Rv2034	
Amino acid comparisons	GeneMarkS	8 UM-CSF isolates	**Rv0311**	Sequence homology detected
	ProteinOrthoProgram	63 proteins reported for *Mtb*	Rv0619	
		5 other mycobacteria *spp*	Rv0014c	
			Rv1837c	
			Rv2176	

			Rv0984 Rv0983 Rv0966 Rv0805 Rv1273c Rv2318 Rv2947c	
Amino acid comparisons	GeneMarkS ProteinOrthoProgram	8 UM-CSF isolates 141 *S. pneumonia* meningitis genes 164 *N. meningitidis* virulence genes	Rv1699 **Rv2606c** Rv0357c Rv2457c **Rv2397c**	51 – 68 % sequence similarity with their homologues

Appendix 28

Table A28: The characteristics of the 35 genes obtained from *in silico* study.

Rv No	Gene name	Size (bp)	Protein size (aa)	Function
Rv1587c	Rv1587c	1002	333	Unknown
Rv3425	PPE57	531	176	PPE family protein
Rv1141c	echA11	807	268	Probable enoyl-CoA hydratase EchA11 (enoyl hydrase) (unsaturated acyl-CoA hydratase)
Rv3513c	fadD18	657	218	Involved in lipid degradation
Rv1267c	embR	1167	388	Involved in transcriptional mechanism
Rv1899c	1ppD	1032	343	Possible lipoprotein LppD
Rv0747	PE_PGRS10	2406	801	PE-PGRS family protein
Rv1067c	PE_PGRS19	2004	667	PE-PGRS family protein
Rv1087	PE_PGRS21	2304	767	PE-PGRS family protein
Rv3344c	PE_PGRS49	1455	484	PE-PGRS family protein
Rv3426	PPE58	699	232	PPE family protein
Rv0278c	PE_PGRS3	2874	957	PE-PGRS family protein
Rv3159c	PPE53	1773	590	PPE family protein
Rv1753c	PPE24	3162	1053	PPE family protein
Rv0144	Rv0144	843	280	Probable transcriptional regulatory protein (possibly TetR-family)
Rv3730c	Rv3730c	1041	346	Conserved hypothetical protein
Rv0802c	Rv0802c	657	218	Possible succinyltransferase in the GCN5-related N-

				acetyltransferase family
Rv2034	Rv2034	324	107	Involved in transcriptional regulation. ArsR repressor protein
Rv0311	Rv0311	1230	409	Unknown
Rv0619	galTb	546	181	Probable galactose-1-phosphate uridylyltransferase GalTb
Rv0014c	pknB	1881	626	Involved in signal transduction (via phosphorylation) and regulates morphological changes associated with cell division. Transmembrane serine/threonine-protein kinase B PknB (protein kinase B) (STPK B)
Rv1837c	glcB	2226	741	Involved in glyoxylate bypass (second step), an alternative to the tricarboxylic acid cycle [catalytic activity: L-malate + CoA = acetyl-CoA + H(2)O + glyoxylate]
Rv2176	pknL	1200	399	Involved in signal transduction (via phosphorylation). Probable transmembrane serine/threonine-protein kinase L
Rv0984	moaB2	546	181	Thought to be involved in molybdopterin biosynthesis.
Rv0983	pepD	1395	464	Possibly hydrolyses peptides and/or proteins
Rv0966	Rv0966c	603	200	Function unknown
Rv0805	Rv0805	957	318	Hydrolyses cyclic nucleotide monophosphate to nucleotide monophosphate

Rv1273c	Rv1273c	1749	582	Involved in active transport of drugs across the membrane (export): multidrug resistance by an export mechanism. Responsible for energy coupling to the transport system and for the translocation of the substrate across the membrane
Rv2318	uspC	1323	440	Thought to be involved in active transport of sugar across the membrane (import). Probable periplasmic sugar-binding lipoprotein UspC
Rv2947c	pks15	1491	496	Polyketide synthase possibly involved in lipid synthesis
Rv1699	pyrG	1761	586	Pyrimidine biosynthesis. Probable CTP synthase PyrG
Rv2606c	snzP	900	299	Possibly involved in the biosynthesis of pyridoxine/pyridoxal 5-phosphate biosynthesis
Rv0357c	purA	1299	432	Involved in AMP biosynthesis (first committed step). Plays an important role in the de novo pathway of purine nucleotide biosynthesis
Rv2457c	clpX	1281	426	ATP-dependent specificity component of the CLP protease
Rv2397c	cysA1	1056	351	Involved in the active transport across the membrane of multiple sulphur-containing compounds, including sulfate and thiosulfate (import).

Appendix 29

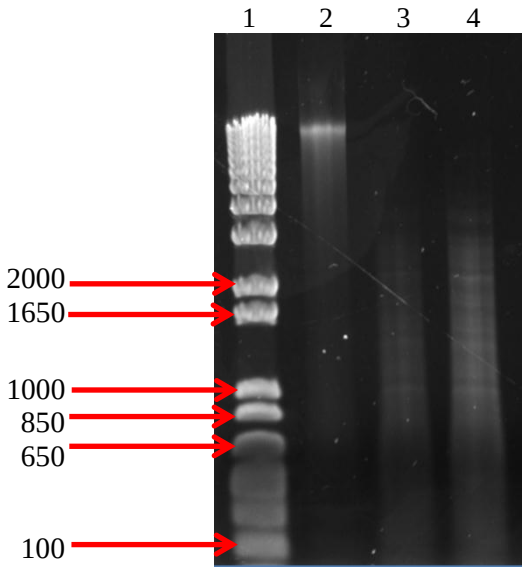


Figure A29: Agarose gel electrophoresis (0.8 %) of *Rsa*I digested genomic DNA from sputum and CSF *Mtb* isolates; Lane 1: 1 kb Plus DNA Ladder; Lane 2: Undigested *Mtb* genomic DNA; Lane 3: UM-CSF01; Lane 4: UM-Sputum samples. The DNA size distribution has reduced to less than 2 kb, as compared to the undigested genomic DNA and qualitatively the largest average size of fragments are at approximately 600 bp.

Appendix 30

Table A30: The homology studies (% of identity) between the 7 subtracted sequences and the genes obtained from our *in silico* studies. The red underlined letters indicated the most significant BLASTN results. The blue circled letters indicated the genes selected for verification using PCR.

No	Rv number	Identity/Query Cover (%)						
		S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**
1	Rv1587c	88.0/1.0	100/5.0	NA	**87.0/8.0**	NA	NA	100/6.0
2	**Rv3425**	**99.0/34.0**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
3	**Rv1141c**	**NA**	**NA**	**NA**	**84.0/7.0**	**NA**	**NA**	**NA**
4	Rv3513c	NA	NA	NA	NA	94.0/2.0	NA	NA
5	Rv1267c	NA	NA	NA	100/ 4.0	NA	NA	NA
6	Rv1899c	NA	100/ 5.0	NA	NA	NA	89.0/3.0	100/3.0
7	Rv0747	NA	100/2.0	100/3.0	NA	NA	NA	91.0/11.0
8	Rv1067c	NA	90.0/4.0	NA	NA	NA	100/2.0	**86.0/22.0**
9	Rv1087	NA	100/8.0	NA	NA	100/1.0	89.0/3.0	82.0/12.0
10	**Rv3344c**	**NA**	**NA**	**NA**	**NA**	**NA**	**100/ 3.0**	**87.0/6.0**
11	Rv3426	**80.0/34.0**	NA	NA	NA	NA	NA	NA
12	Rv0278c	NA	NA	100/3.0	NA	**93.0/6.0**	100/4.0	94.0/8.0
13	Rv3159c	NA	100/ 2.0	89.0/4.0	100/ 4.0	NA	NA	83.0/6.0
14	Rv1753c	100/1.0	NA	100/ 2.0	NA	88.0/2.0	NA	100/3.0
15	Rv0144	NA	NA	93.0/4.0	100/ 4.0	NA	NA	NA
16	Rv3730c	NA	NA	NA	NA	NA	NA	100/8.0
17	Rv0802c	NA	NA	NA	NA	NA	NA	NA

18	Rv2034	NA	NA	NA	NA	NA	NA	NA
19	**Rv0311**	**NA**	**83.0/5.0**	**NA**	**NA**	**NA**	**NA**	**NA**
20	Rv0619	NA	NA	100/ 2.0	NA	89.0/2.0	NA	NA
21	Rv0014c	NA	100/6.0	NA	NA	90.0/2.0	NA	100/4.0
22	Rv1837c	100/1.0	100/ 2.0	**94.0/7.0**	93.0/6.0	NA	100/2.0	89.0/15.0
23	Rv2176	NA	100/5.0	NA	NA	NA	NA	NA
24	Rv0984	NA	100/ 2.0	NA	NA	NA	NA	NA
25	Rv0983	NA	NA	NA	NA	NA	NA	83.0/6.0
26	Rv0966	NA	90.0/4.0	NA	NA	NA	NA	NA
27	Rv0805	NA	**81.0/7.0**	NA	NA	NA	NA	NA
28	Rv1273c	86.0/2.0	100/ 2.0	100/2.0	100/ 4.0	93.0/2.0	NA	NA
29	Rv2318	NA	NA	NA	100/ 4.0	NA	**100/4.0**	NA
30	Rv2947c	100/1.0	93.0/3.0	100/3.0	**100/ 9.0**	100/1.0	NA	NA
31	Rv1699	100/1.0	NA	NA	NA	NA	NA	NA
32	**Rv2606c**	**NA**	**NA**	**NA**	**NA**	**NA**	**82.0/3.0**	**NA**
33	Rv0357c	100/1.0	86.0/4.0	100/5.0	NA	NA	NA	100/2.0
34	Rv2457c	100/1.0	NA	NA	100/4.0	100/1.0	NA	NA
35	**Rv2397c**	**NA**	**NA**	**NA**	**100/ 4.0**	**NA**	**NA**	**NA**