

COLLEGE OF BASIC AND APPLIED SCIENCES

SCHOOL OF BIOLOGICAL SCIENCES

**COMPARATIVE GENOMIC ANALYSIS OF MYCOBACTERIUM
TUBERCULOSIS COMPLEX STRAINS FROM GHANA**

BY

ISAAC DARKO OTCHERE

(10170819)

**THIS THESIS IS SUBMITTED TO THE DEPARTMENT OF BIOCHEMISTRY, CELL
AND MOLECULAR BIOLOGY, UNIVERSITY OF GHANA, LEGON, IN PARTIAL
FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF DOCTOR OF
PHILOSOPHY DEGREE IN BIOCHEMISTRY.**

JULY, 2017

DECLARATION

I, Isaac Darko Otchere do hereby declare that except for references to other people's work for which I have duly acknowledged, this thesis is the product of my own research work carried out under the supervision of Professor Dorothy Yeboah-Manu and Professor Sebastien Gagneux. I do further declare that it has not been presented in its entirety or parts elsewhere to acquire another degree.

.....
ISAAC DARKO OTCHERE
(Candidate)

DATE.....

.....
PROFESSOR DOROTHY YEBOAH-MANU
(Supervisor)

DATE.....

.....
PROFESSOR SEBASTIEN GAGNEUX
(Co-Supervisor)

DATE.....

ABSTRACT

Background: Tuberculosis (TB), caused by a group of acid fast bacilli together called the *Mycobacterium tuberculosis* complex (MTBC) is one of the leading causes of adult mortality globally albeit treatable. TB in humans is mainly caused by the globally distributed *M. tuberculosis* sensu stricto (*Mtbss*) and West Africa restricted *M. africanum* (*Maf*), however, global efforts for TB control are *Mtbss*-biased to the neglect of *Maf*. However, *Maf* is an important pathogen in West Africa especially Ghana (harbouring the two lineages (L) 5 and 6 in substantial numbers) where it causes about 20% of TB. This thesis sought to determine the genomic diversity among the dominant MTBC lineages (L4, L5 and L6) in Ghana and the functional implications of the observed genomic variability.

Methodology: A total of 297 clinical MTBC genomes (L4, L5 and L6) from Ghana were whole genome sequenced to establish the population structure of *Maf* and to explore the genomes of the three lineages for differential fixation of mutations, selection pressures among conserved genes as well as comparative genomic erosions stratified by lineages with their potential phenotypic implications. In addition, this study used 1490 prospectively isolated and genotyped MTBC strains to determine the potential influence of the strain genetic background on emergence and fixation of drug resistance conferring mutations

Results: L6 was found to have accumulated significantly more fixed single nucleotide polymorphisms (SNPs) (median = 1037; $p < 0.05$) compared to L4 (median = 822) and L5 (median = 928) in addition to significantly higher pairwise nucleotide diversity (π).

The whole genome phylogeny of *Maf* in Ghana upon clustering showed 5 and 3 clusters of L5 and L6 respectively. Average nucleotide identity (ANI) between any two lineages (L4/L5, L4/L6 and L5/L6) was always greater than 99%. The π of the functional genes was all significantly higher among the L6 compared to L4 and L5 with prominence among T cell epitopes (0.000061, 0.000063 and 0.000149 for L4, L5 and L6). Apart from T cell epitopes and regulatory proteins, all the gene families were under purifying selection thus the ratio of change in nonsynonymous to change in synonymous mutations less than one ($dN/dS < 1.0$) irrespective of the lineage. T cell epitopes of L4 and L5 were under purifying selection; thus the ratio of change in nonsynonymous to change in synonymous mutations (dN/dS) less than one (L4: median $dN/dS = 0.57$ and median $dN/dS = 0.64$ for fL4 and L5 respectively) compared to positive selection in L6 (median $dN/dS = 1.53$). Most T cell epitopes were conserved with only 10.4% (128/1226) mutated with lineage-specific variations. Genomic erosion was predicted to be more pronounced in L6 (3992 genes present) compared to L5 (4035 genes present) with predicted lineage-specific impaired expression of proteins such ESAT-6, mpt64 and a number of mce3 proteins. This study found Ghana genotype of L4 to be associated with DR-TB in Ghana ($p < 0.05$) compared to the other dominant genotypes. INH resistant *Mtbss* and *Maf* were more likely *kat* G S315T and *inhApro* -15C/T mutants respectively. This study identified potential genotype-specific epistatic interaction DR-conferring and compensatory mutations.

Conclusion: *Maf* L6 and L5 are genomically two different biovars of the MTBC, so different from each other as they are from L4. The observed genomic difference among the lineages has potential implications for virulence and applicability of TB control tools.

DEDICATION

I dedicate this work to all TB patients in Ghana



ACKNOWLEDGEMENT

I thank The Holy Trinity One God, without whose providence my efforts toward this PhD would have been futile. I thank Him once more for even giving me this opportunity to express my deepest appreciation to some individuals, groups and agencies that supported this work.

The first person to be appreciated is my Principal Supervisor; Professor Dorothy Yeboah-Manu of the Noguchi Memorial Institute for Medical Research (NMIMR), University of Ghana. I call you Mum because you have always been more than just my boss and supervisor. The care you have shown since we first met in 2009 can only be likened to that of a mother. I do now and would forever appreciate your contribution towards my intellectual development.

Secondly, I would like to express my profound gratitude to Professor Sebastien Gagneux of the Swiss Tropical and Public Health Institute (Swiss TPH), University of Basel for your guidance and support throughout this work. Even before you accepted to Co-Supervise my PhD, you were very supportive in making me who I have become today. I say thank you Prof.

I am also grateful to the Yeboah-Manu Research Group of the Bacteriology Department, NMIMR from 2011 to date (Dr. Adwoa Asante-Poku, Mr. Prince Asare, Miss. Emelia Danso, Miss. Evelyn Owusu-Mireku, Miss. Zuleihatu Nakobu, Dr. Grace Kpeli, and Dr. Kobina Assan Ampah, Mr. Haruna Ganiyu Abdul, Mr. Stephen Osei-Wusu, Mr. Delali Mensah, Doris Aboagyewaa Boampong, Miss. Esther Sarpong, Miss. Diana Prah, Miss.

Diana Asandem, Miss. Pheona Badu, Mr. Selesie Ameke, Miss. Portia Morgan all National Service Personnel) for all the support and hard work.

My bioinformatics training started with the training I received from Dr. Sonia Borrell, Dr. David Stucki, Dr. Daniela Da Silva Brites and Dr. Mireia Coscolla all from the Sebastien Gagneux group of the Swiss TPH. I also do appreciate the support of Professor Julian Parkhill and Dr. Simon Harris of the Wellcome Trust Sanger Institute for the sequencing of the genomes and the advanced training I received from the pathogens genomics lab, in Hinxton, United Kingdom.

My deepest appreciation goes to the entire members of the Bacteriology Department and friends in the Virology Department of NMIMR. You all contributed to making this journey a memorable one. Furthermore, I appreciate the invaluable support I received from members, staff and colleagues of the Department of Biochemistry, Cell and Molecular Biology, University of Ghana.

I want to thank The Wellcome Trust Foundation, United Kingdom for funding this PhD work under the intermediate fellowship (097134/Z/11/Z) awarded to Dorothy Yeboah-Manu.

Finally, I want to thank my parents and siblings for their support and sacrifices since my kindergarten days. I cannot forget my wife Naomi and Son Melchizedek, for tolerating my continuous absence and coming home late throughout this study.

TABLE OF CONTENTS

DECLARATION	i
ABSTRACT.....	ii
DEDICATION	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	xi
LIST OF TABLES.....	xiii
LIST OF ABBREVIATIONS.....	xiv
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 RATIONALE AND JUSTIFICATION	5
1.2 HYPOTHESES , AIM AND SPECIFIC OBJECTIVES	9
1.2.1 Hypotheses.....	9
1.2.3 Specific Objectives	10
CHAPTER TWO	11
LITERATURE REVIEW.....	11
2.1 HISTORICAL FACTS	11
2.2 GLOBAL BURDEN OF TB.....	17
2.3 TUBERCULOSIS IN GHANA	18
2.4 THE CAUSATIVE AGENT OF TUBERCULOSIS.....	21
2.4.1 <i>The Mycobacterium tuberculosis Complex</i>	21
2.4.2 <i>Other Mycobacterial Species</i>	26
2.4.3 <i>The Mycobacterial Cell Wall</i>	27
2.5 PATHOGENESIS OF TB.....	30
2.6 CONTROL OF TB	32
2.6.1 Vaccination against TB.....	32
2.6.2 Diagnosis of TB	34

2.6.3	Treatment of TB and Emergence of Drug resistance.....	36
2.6.4	Molecular Basis of MDR-TB.....	37
2.6.4.1	<i>Isoniazid Resistance</i>	37
2.6.4.2	<i>Rifampicin Resistance</i>	43
2.7	THE GENOME OF THE MTBC	46
2.8	GENOMIC DIVERSITY WITHIN THE MTBC	49
2.8.1	Repetitive and Mobile Genetic Elements	50
2.8.2	Large Sequence Polymorphisms (LSPs).....	51
2.8.3	Single Nucleotide Polymorphisms (SNPs).....	52
2.9	PHENOTYPIC CONSEQUENCES OF GENOMIC DIVERSITY	54
CHAPTER THREE		60
MATERIALS AND METHODS		60
3.1	MATERIALS.....	60
3.1.1	Reagents and Consumables.....	60
3.1.2	Equipment.....	61
3.1.3	Software	62
3.2	METHODS	63
3.2.1	Preparation of Culture Media and Drug Stocks.....	63
3.2.1.1	<i>Lowenstein-Jensen (L-J) Media</i>	63
3.2.1.2	<i>Middlebrook 7H11 Agar Medium</i>	63
3.2.1.3	<i>Blood Agar preparation</i>	64
3.2.1.4	<i>Preparation of Drug Stocks</i>	64
3.2.2	<i>Mycobacterium tuberculosis</i> complex Isolates (200)	65
3.2.3	Purification of Archived MTBC Isolates by Sub-Culturing.....	65
3.2.4	Genotyping of Mycobacterial Isolates	68
3.2.4.1	Confirmation of <i>Mycobacterium tuberculosis</i> complex strains.....	68
3.2.5	Chromosomal DNA Extraction.....	69
3.2.6.	Whole Genome Sequencing.....	70

3.2.7	Analysis of Whole Genome Data	71
3.2.7.1	<i>Mapping of Reads</i>	71
3.2.7.2	<i>Phylogenetics and Clustering Analyses</i>	73
3.2.7.3	<i>Genetic Diversity and Mutation Analysis</i>	74
3.2.7.4	<i>De novo Assembly of Bacterial Genomes and Annotations</i>	78
3.2.7.5	<i>Comparative Pan-Genome Analysis</i>	79
3.2.8	Drug Resistance Profiling and Target Gene Mutational Analysis.....	79
3.2.8.1	<i>Drug Resistance Screening</i>	79
3.2.8.2	<i>PCR Amplification of Drug Resistance Associated Genes</i>	83
3.2.8.3	<i>Bioinformatics Analysis of Sequenced Genes for Mutations</i>	83
3.2.8.4	<i>Statistical Analysis of Drug Resistance Data</i>	85
CHAPTER FOUR.....		86
DESCRIPTION OF RESULTS		86
4.1	SNP DISTANCE AND DIVERSITY OF MTBC LINEAGES IN GHANA	86
4.2	PHYLOGENETICS, CLUSTERING AND NUCLEOTIDE IDENTITY	86
4.3	PAIRWISE NUCLEOTIDE DIVERSITY (<i>IT</i>) OF ANNOTATED GENES	89
4.4	PAIRWISE DN/DS ANALYSIS OF ANNOTATED GENES	91
4.5	LINEAGE-SPECIFIC ACCUMULATION OF MUTATIONS.....	94
4.6	CONSERVATION OF HUMAN T CELL EPITOPES IN <i>MAF</i>	103
4.7	COMPARATIVE GENOMIC ANALYSIS	109
4.8	DETECTION OF LINEAGE-SPECIFIC MUTATIONS.....	113
4.9	DRUG RESISTANT TB IN GHANA.....	120
4.9.1	Characteristics of Patients with Drug Resistant MTBC Isolates	120
4.9.3	Mutations in Genes Associated with INH and RIF Resistance	122
CHAPTER FIVE		126
DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS.....		126
5.1	GENERAL DISCUSSION	126
5.1.1	First Insight Into the Whole Genome Phylogeny	127

5.1.2	Lineage Specific Antigenic Variation:.....	129
5.1.3	Genomic Erosion and Accumulation of Mutations Among MTBC	132
5.1.4	Drug Resistance Mutations, Implication for Rapid Diagnostics	136
5.2	LIMITATIONS OF THE STUDY.....	142
5.3	CONCLUSIONS.....	143
5.3.1	Key Findings	143
5.3.2	Summary.....	145
5.4	RECOMMENDATIONS	146
BIBLIOGRAPHY		148
APPENDICES		176



LIST OF FIGURES

Figure 2. 1: The first tuberculosis sanatorium built in Germany.	16
Figure 2.2: Global estimates of TB incidence rate in 2015	19
Figure 2. 3: Countries in the three TB high-burden lists 2016–2020	22
Figure 2.4: Global Phylogeny of the MTBC rooted on <i>M. canettii</i>	24
Figure 2.5: Global phylogeography of the human adapted <i>M. tuberculosis</i> complex.....	25
Figure 2.6: Phylogeny mycobacteria rooted on <i>Rhodococcus spp</i> as an outlier.....	28
Figure 2.7: Schematic representation of the mycobacterial cell wall	29
Figure 2.8: Pathogenesis of tuberculosis.	33
Figure 2.9: Mechanism of action and mode of resistance of INH	40
Figure 2.10: Proposed mechanism of action of INH and ETH.....	42
Figure 2.11: Putative compensatory mutations in <i>rpoA</i> and <i>rpoC</i>	45
Figure 2.12: Circular map of the chromosome of <i>Mtbs</i> H37Rv.....	48
Figure 2.13: Global phylogeny of clinical and veterinary MTBC isolates.....	55
Figure 3.1: Maps of Ghana Showing the Regions.	67
Figure 3.2: Schematic diagram of illumina sequencing.....	72
Figure 3.3: Diagrammatic set up of the micro-plate alamar blue assay.....	82
Figure 3.4: High-confidence detection of SNPs with the Staden package.	85
Figure 4.1: SNP distance and nucleotide diversity of hMTBC lineages in Ghana.	87
Figure 4.2: Clustered maximum likelihood phylogeny of Ghanaian <i>Maf</i>	88
Figure 4.3: Average pairwise nucleotide diversity (π) of annotated genes.....	90
Figure 4.4: Pairwise dN/dS of annotated genes among hMTBC lineages.....	92
Figure 4.5: Pairwise dN/dS of T cell epitopes (A) and regulatory proteins (B)	93
Figure 4.6: Human T cell antigenic and T cell epitope variation among the <i>Maf</i>	96
Figure 4.7: <i>Maf</i> lineage-specific variation of T cell epitopes.	97
Figure 4.8: Pairwise dN/dS of T cell epitopes (A) and regulatory proteins (B)..	104
Figure 4.9: Pairwise dN/dS of T cell epitopes of L5 <i>Maf</i> lineages.....	105
Figure 4.10: T cell antigenic variation stratified by ethnicity of L5 TB patients	106

Figure 4.11: Uniform distribution of L5 isolated from ethnic Ewe TB patients	107
Figure 4.12: Pairwise dN/dS of T cell epitopes of L6 stratified by patient HIV status. .	108
Figure 4.13: Comparative pan-genomes of the MTBC lineages.	110
Figure 4.14: MTBC lineage-specific genome erosion.....	112
Figure 4.15: Distribution of lineage-specific mutations among genes.	118
Figure 4.16: Abundance of specific mutations among the MTBC lineages.	119



LIST OF TABLES

Table 3.1: Preparation of PCR mix for amplification of MTBC-specific IS6110.....	68
Table 3.2: Selected genes with reported MTBC lineage-specific mutations.....	77
Table 3.3: Primer pairs, expected band size and annealing temperatures used for PCR..	84
Table: 4.1: dN/dS ratio of MTBC lineages within functional genes	93
Table 4.2: Human T cell antigens and T cell epitopes with amino acid substitutions.....	98
Table 4.3: Lineage-specific deletion of genes	111
Table 4.4: Lineage-specific mutations.....	114
Table 4.5: Prevalence of DR-TB between the north and south of Ghana.....	121
Table 4.6: The prevalence of drug resistance among MTBC species.....	121
Table 4.7: Comparison of DR-TB	123
Table 4.8: Identified mutations among drug resistant MTBC	124
Table 4.9: Distinct drug target mutations and MTBC species.....	125



LIST OF ABBREVIATIONS

7H9-GC	7H9 broth supplemented with glycerol and casitone
7H9-GCOT	7H9-GC supplemented with OADC and Tween 80
ADC	Albumin-Dextrose-Catalase
AFB	Acid-Fast Bacilli
AMGs	Aminoglycosides
ANI	Average Nucleotide Identity
BCG	Bacillus Calmette-Guerin
CFR	Case Fatality Ratio
Cfz	Clofazime
CRISPRs	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyl trimethylammonium Bromide
<i>cwallproc</i>	cell wall and cell processes
dm ³	Decimeters Cube
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic Acid
dN/dS	Ratio of the change in nonsynonymous mutations to change in synonymous mutations
dNTP	Deoxy-Ribonucleotide Phosphate
DOTS	Directly Observed Treatment Short-course
DR	Drug Resistant
DS	Drug Susceptible
DST	Drug Susceptibility Test
EAI	East- African Indian
EDTA	Ethylenediaminetetraacetic Acid
EMB	Ethambutol
<i>epit</i>	T cell epitopes

ESAT-6	Early Secretory Antigen Target 6
<i>esmac</i>	essential for growth in macrophages
ETH	Ethionamide
FQs	Fluoroquinolones
HIV	Human Immuno-eficiency Virus
hMTBC	Human Adapted <i>Mycobacterium tuberculosis</i> complex
IGRa	Interferon Gamma Release Assay
INDELS	Insertions and /or Deletions
<i>infopath</i>	information pathways
INH	Isoniazid
<i>intmedres</i>	intermediate metabolism and respiration
IS	Insertion Sequence
Km	Kanamycin
L1	Lineage 1
L2	Lineage 2
L3	Lineage 3
L4	Lineage 4
L5	Lineage 5
L6	Lineage 6
L7	Lineage 7
LAM	Latin American
LED	Light Emitting Diode
<i>lipmet</i>	lipid metabolism
L-J	Lowentein-Jensen
LSP	Large Sequence Polymorphism
LTBI	Latent Tuberculosis Infection
MABA	Microplate Alamar Blue Assay

MAC	<i>Mycobacterium avium complex</i>
Maf	<i>Mycobacterium africanum</i>
mAGP	Mycolyl Arabinogalactan-Peptidoglycan
MDR	Multi-Drug Resistant
Mfx	Moxifloxacin
MIRU	Mycobacterium Interspersed Repetitive Units
mL	Milliliters
mol	Moles
MTBC	<i>Mycobacterium tuberculosis complex</i>
Mtbss	<i>Mycobacterium tuberculosis sensu stricto</i>
NAD	Nicotinamide Adenine Dinucleotide
NADH	Nicotinamide Adenine Dinucleotide (reduced)
NMIMR	Noguchi Memorial Institute for Medical Research
nsSNP	Non-synonymous Single Nucleotide Polymorphism
NTMs	Non-Tuberculous <i>Mycobacterium species</i>
OADC	Oleate-Albumin-Dextrose-Catalase
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PIMs	Phosphatidylinositol Mannosides
Pi (π)	Nucleotide Diversity
PNZ	Pyrazinamide
Pto	Prothionamide
RD	Region of Difference
<i>regprot</i>	regulatory proteins
RFLP	Restriction Fragment Length Polymorphism
RIF	Rifampicin

RIFr	Rifampicin Mono-resistant
SNP	Single Nucleotide Polymorphism
STBs	Smooth Tuberculosis Bacilli
Swiss TPH	Swiss Tropical and Public Health Institute
TB	Tuberculosis
TBE	Tris-Borate- Ethylenediaminetetraacetic Acid
uL	Microliters
UV	Ultra Violet
VNTR	Variable Numbered Tandem Repeats
WHO	World Health Organization
WTSI	Wellcome Trust Sanger Institute
XDR	Extremely Drug Resistant



CHAPTER ONE

INTRODUCTION

Tuberculosis (TB), albeit an ancient disease, remains an important public health problem today and has been declared a global health emergency by the World Health Organisation (WHO) in 1993, hence requesting for intensified efforts to fight the disease. Approximately 33% of the global human population is latently infected with the TB pathogen; however, only 5 to 10% of these individuals will develop active TB in their lifetime. Even though treatable, TB remains one of the leading causes of adult mortality globally with an estimated 10.4 million new cases (incidence) and 1.8 million deaths worldwide in the year 2015 (WHO, 2016). Cases among men and women were 5.9 million (56%) and 3.5 million (34%) respectively excluding the 1 million (10%) of the affected who were children under 15 years.

The high global burden of TB is driven by factors such as urbanization, insensitive diagnosis, ineffective vaccine, human immunodeficiency virus (HIV) co-infection and emergence of drug resistance. Urbanization has led to development population dense communities mostly with poor sanitation which promotes transmission of TB because the chances of getting into contact with TB patients in such situations are high. Secondly, the continuous reliance on sputum smear microscopy, a less sensitive (detects approximately 50% of TB cases) century-old diagnostic tool means that about 50% of TB patients (mostly early stage with low bacteria load) are missed and sent back home without treatment. These people continue to transmit the disease in their communities and might

later be diagnosed positive for TB and start treatment after their conditions have worsened which in effect negatively impacts treatment. In addition, the fact that there is no efficacious vaccine against TB capable of preventing or curing of TB in adults makes the use of BCG vaccine (an attenuated *M. bovis* strain), the only option. However, the BCG vaccine only prevents TB (especially cerebral TB) in children under 5 years and it is ineffective in either preventing infection or development of active TB in adults, hence contributing to the global burden (WHO, 2016).

Furthermore, the body's fight against TB is mostly cell mediated involving macrophages and T cells, therefore, co-infection with HIV which compromises cell mediated immune response by deregulation of CD4/CD8 T cell balance (through depletion of CD4 T cells) makes the body weak in fighting TB. More so, some anti-retroviral drugs interact unfavourably with some anti-tubercular agents requiring that both diseases are managed at different time points with its accompanying public health implications.

Lastly, the emergence of drug resistance, especially multidrug resistant TB (MDR-TB) and extremely drug resistant TB (XDR-TB) are threatening to make TB an untreatable disease. The WHO recommended treatment regimen for TB is directly observed treatment short course (DOTS) which involves an intensive phase administration of four drugs: isoniazid (INH), rifampicin (RIF), pyrazinamide (PNZ) and ethambutol (EMB) for two months followed by a continuous phase administration of INH and RIF for an additional four months. The term MDR means resistance to INH and RIF which are the two backbone drugs of the DOTS regimen, an occurrence which renders DOTS

ineffective and hence requires the use of second line drugs such as fluoroquinolones (FQs) and injectable aminoglycosides (AMGs). Even though these second line drugs are more expensive, more toxic and are administered for prolonged duration, the efficacy is far less than first-line DOTS. An MDR-TB that is also resistant to any FQ in addition to resistance to any injectable AMGs is what is termed an XDR-TB, a situation that is more difficult to manage than even MDR-TB. These factors fraught the fight against TB, nevertheless, it is the contribution by HIV co-infection and drug resistance which have been quantified.

For instance, TB patients living with HIV accounted for 1.2 million (11%) of the global TB incidence in 2015 of which 400 thousand died. Similarly, in the same year, 480 thousand TB cases were MDR-TB and additional 100 thousand cases were RIF-mono resistant. The most affected regions include Africa and Asia. Africa is home to approximately 13.5% of the global human population but carries about 26.2% of the global TB burden, 71.3% of the TB/HIV burden, 32.1% TB related deaths without HIV and 76.9% of TB/HIV deaths making Africa the most affected continent. With prevalence of 356 cases per 100 thousand population, case fatality ratio (CFR) of over 25% and over 10 thousand TB/HIV co-infected cases per year, Ghana is ranked among the 30 high TB/HIV burdened countries (WHO, 2016).

Tuberculosis in mammals is caused mainly by a group of closely related acid-fast bacilli, all together termed the *Mycobacterium tuberculosis* complex (MTBC). The MTBC include *M. tuberculosis sensu stricto* (*Mtbss*), *M. africanum* (*Maf*), *M. mungi*, *M. bovis*,

M. microti, *M. pinnipedii*, *M. bovis* and *M. bovis* BCG (Brosch *et al.*, 2002; Cousins *et al.*, 2003; Frota *et al.*, 2004; Garnier *et al.*, 2003). The MTBC has been regarded as a typical representation of genetically monomorphic pathogens because of the low DNA and amino acid sequence diversity among the members compared to other bacteria (Achtman, 2008). Recent advances in molecular methods coupled with the completion of whole TB genome sequencing in 1998 (Cole *et al.*, 1998; Comas *et al.*, 2013; Comas *et al.*, 2009; Frota *et al.*, 2004; Gagneux *et al.*, 2006; Hershberg *et al.*, 2008) however, revealed that even though the MTBC are genetically monomorphic compared to other bacteria, there exist some substantial genetic diversity that could have phenotypic implications for virulence, diagnostics, vaccines and treatment.

Other than occasional cross species infections, members of the MTBC are adapted to causing disease in specific mammalian hosts. Human tuberculosis is mostly caused by *Mtbss* and *Maf* here-in referred as human adapted MTBC (hMTBC). Genetic characterization of 850 clinical MTBC strains from 80 countries divided the hMTBC into 6 major lineages; *Mtbss* subdivided into Lineages 1 to 4 (L1-L4) and *Maf*, Lineages 5 and 6 (L5 and L6). Recently an additional *Mtbss* lineage (Lineage 7; L7) has also been described (Blouin *et al.*, 2012; Firdessa *et al.*, 2013). Mapping the human origin of the isolates showed that the hMTBC lineages exhibit a phylogeographic structure, implying that distinct human populations are more susceptible to specific lineages. *Maf* lineages are restricted to West Africa where they cause up to 50% of TB cases in some countries (de Jong *et al.*, 2010; S. Gagneux *et al.*, 2006). In Ghana, *Maf* has been shown to account for approximately 20% of TB cases with the dominant L5 causing about 13% of TB

(Yeboah-Manu *et al.*, 2016). In addition to the phylogeography of the hMTBC lineages, different sub-lineages within specific lineages have been shown to exhibit varying phenotypic traits. For instance, a sub-lineage of L2 (some strains of the Beijing sub-lineage) has been associated with outbreaks of MDR TB in parts of Asia, Europe and South Africa (Johnson *et al.*, 2006; Kato-Maeda *et al.*, 2010; Kato-Maeda *et al.*, 2012). In addition, the Latin American (LAM) sub-lineage of L4 has been shown to exhibit traits of a generalist pathogen (that is globally distributed) whereas the Cameroon and Ugandan sub-lineages of L4 are more specialist pathogens restricted to West and East Africa, respectively (Stucki *et al.*, 2016).

Even though MTBC L5 and L6 are classified as once species under *Maf* and are both restricted to West Africa, they exhibit disparate phenotypic traits. For instance, L5 is able to use glycerol in media as the only source of carbon whereas L6 requires pyruvate supplemented media. In addition, L6 has been clinically associated with HIV-coinfection independently in the Gambia and Ghana (Asante-Poku *et al.*, 2016; de Jong *et al.*, 2010; Homolka *et al.*, 2010). Therefore, understanding the genomic difference between these lineages would be useful for design of control tools for use in the West African sub-region.

1.1 RATIONALE AND JUSTIFICATION

1.1.1 **Rationale**

Global efforts to fight TB have largely been biased towards *Mtbss* to the neglect of *Maf*. This is partly due to the global distribution of *Mtbss* compared to the geographic

restriction of *Maf* to West Africa. There is also, the assumption that strain diversity among members of MTBC has no relevant consequence due to the limited DNA sequence diversity exhibited by the MTBC compared to other bacterial pathogens (Musser *et al.*, 2000; Sreevatsan *et al.*, 1997). However, this notion was dependent on findings of earlier studies that utilised less discriminatory genotyping tools and limited and often biased strain collections. Nevertheless, more data are accumulating indicating that there are polymorphisms within the genomes of MTBC that may have functional implications (Comas *et al.*, 2010; Daugelat *et al.*, 2003; Hershberg *et al.*, 2008).

The outcomes of TB infections in humans are extremely variable, ranging from lifelong latent infection to active disease with variable degrees of extra-pulmonary involvement. This variation was primarily attributed to host and environmental factors. However, phenotypic variations such as preferred carbon source in *in vitro* cultures, colonial morphological and other metabolic differences such as ability to reduce nitrate suggest that some of the variations in clinical outcome might be due to strain differences. Indeed, findings from *ex vivo* (using macrophages) and *in vivo* (using animal models) studies which have shown that hMTBC strains from different geographical regions differ in virulence and immunogenicity and that this disparity stems from the genetic background of the individual strains. For instance, *Mtbss* was shown to elicit delayed innate immune response, allowing it to progress faster to active disease compared to *Maf* which was earlier shown to be relatively less virulent *in vitro* compared to *Mtbss* (de Jong *et al.*, 2008; Portevin *et al.*, 2011). However, even among *Mtbss* strains, there are some genotypes that are associated with high transmissions and are more globally distributed

(generalists) compared to those mostly concentrated in certain geographical locations (specialists). This distribution is associated with differential conservation of T cell epitopes by the sub-lineages (Stucki *et al.*, 2016).

Conclusion from such clinical studies using limited collection of MTBC strains has been that *Maf* would be eventually eliminated by *Mtbss* (Bold *et al.*, 2012; Comas *et al.*, 2009), leading partly to reduced attention to *Maf* in promulgation of control measures. For example, anti-TB interventions currently in use, under evaluation and those under development for efficient control of TB such as the Interferon Gamma Release Assay (IGRa), Xpert MTB/RIF, MTBDR*plus*, MTBDR*sl* and TB Ag MPT64 diagnostics as well as the ESAT-6 based vaccines have all been done with little attention to *Maf*. The notion that *Maf* would eventually be eliminated by *Mtbss* has been supported by some studies reporting that *Maf* is on the decline in West Africa (Niobe-Eyangoh *et al.*, 2003). Nevertheless, current reports from Ghana, Nigeria, Gambia, and Benin show that *Maf* is still important in the sub-region and could be maintained by either an association with a native ethnicity or HIV co-infection (Asante-Poku *et al.*, 2015; Gehre *et al.*, 2013; Koro *et al.*, 2013; Lawson *et al.*, 2012; Yeboah-Manu *et al.*, 2016).

1.1.2 Justification

There are only three publications, two original studies and one review on comparative whole genome sequence analysis involving *Maf* (Gehre, Otu, *et al.*, 2013; Winglee *et al.*, 2016). Even though these studies indicate a substantial genomic difference between *Maf*

and *Mtbss* and reveal genomic signatures reminiscent of the relatively less virulence of *Maf* reported earlier (Homolka *et al.*, 2010), the strain collections were either numerically limited or biased towards a single lineage. For instance, Gehre *et al.* used only 4 L6 genomes without any L5 whereas Winglee *et al.* used 22 L6 and 2 L5 strains. The limited and completely biased collection of *Maf* (mostly L6 which is dominant in Western West Africa) used in these studies undermines the generalization of the findings as the true depiction of the genomic difference between *Maf* (especially for L5 which dominates Eastern West Africa) and *Mtbss* (de Jong *et al.*, 2010). In addition, the limited number of genomes restricts the use of these genomes to construct a reliable phylogenetic population structure of *Maf* to study transmission dynamics, evolution of drug resistance and could contribute relevant information such as molecular markers towards control of TB in West Africa. This therefore called for whole genome sequencing of enough *Maf* genomes for comparative genomics studies.

Ghana is one of the few West African countries where all the MTBC lineages with the exception of the newly characterized *Mtbss* L7 (restricted to the horn of Africa) circulate in appreciable numbers and proportions. Therefore, Ghana serves as a unique place for comparative genomic studies of the human adapted MTBC. This study used MTBC isolated from TB patients reporting to health centers in the Greater Accra region and the Northern region of Ghana. The Greater Accra region, the capital of Ghana is the most cosmopolitan geographical setting in Ghana with TB patients of almost all the ethnic groups (Asante-Poku *et al.*, 2016) and represents a typical urban Ghanaian setting. The region reports the largest number of TB cases to the National Tuberculosis control Program (NTP) [Personal communication with the NTP program manager]. On the other

hand, the Northern region especially the Mamprusi East district is largely rural especially limited human ethnic diversity (Ganle, 2015). These two selected sites thence ideal representation of Ghana for this study.

1.2 HYPOTHESES , AIM AND SPECIFIC OBJECTIVES

Even though the MTBC is considered genetically a clone, DNA sequencing analyses of 89 gene concatenates from 108 (Hershberg *et al.*, 2008) and 97 (Comas *et al.*, 2009) global collection of strains revealed that *Maf* lineages emerged earlier during the evolution of the MTBC, thus, placing *Maf* lineages closer to the animal adapted strains relative to *Mtbss*. This disparate evolutionary paths of *Maf* and *Mtbss* strains has led to significant innate genomic differences that has implications for virulence, transmission, propensity to acquire drug resistance.

1.2.1 Hypotheses

I therefore hypothesized that;

1. There would be substantial genomic differences between the two *Maf* lineages and also between *Maf* and *Mtbss* L4 with implications for virulence and universal application of control tools.
2. The two *Maf* lineages exhibit different population structures among the global collection of MTBC strains.
3. The individual lineage background of the MTBC strains affects the propensity to acquire drug resistance and epistatic interaction between drug resistance

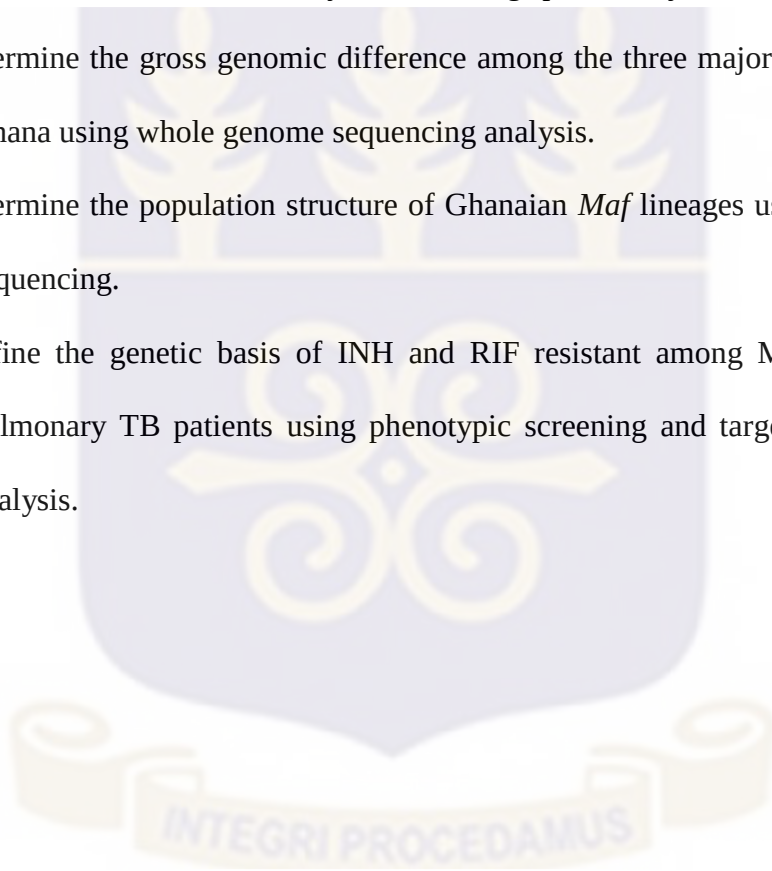
conferring and compensatory mutations.

1.2.2 Aim To understand the biology of the human adapted MTBC in Ghana using whole genome analysis

1.2.3 Specific Objectives

These hypotheses were addressed by the following specific objectives;

1. To determine the gross genomic difference among the three major MTBC lineages in Ghana using whole genome sequencing analysis.
2. To determine the population structure of Ghanaian *Maf* lineages using whole genome sequencing.
3. To define the genetic basis of INH and RIF resistant among MTBC strains from pulmonary TB patients using phenotypic screening and target gene sequencing analysis.



CHAPTER TWO

LITERATURE REVIEW

2.1 HISTORICAL FACTS

Tuberculosis is an ancient disease that has been called many names depending on the culture and the site of the disease. Names such as *phthisis* (Greek), *consumptione* (Latin), *yaksma* or *balasa* (India), *shizhu* (China) and *chaky oncay* (Incan), make reference to the "drying" or "consuming" effect of the illness that has made headlines throughout recorded history. In ancient Greece for instance, the name *phthisis* first appeared in writing when Hippocrates in 460 BC identified and described it as the cause of the most fatal illness in his time. Aristotle believed *phthisis* to be contagious but many of his contemporaries thought it was hereditary (Madkour *et al.*, 2004). A prominent Greek physician Galen defined *phthisis* as the ulceration of the lungs, thorax or throat, accompanied by cough, fever and consumption of the body by pus.

The first mention of tuberculosis in a non-European civilization was in the Vedas (large body of ancient Indian literature including Hindu scriptures) where TB is either called *yaksma* or *balasa* in 1500 BC (Zysk, 1998). In 600 BC India, breast milk, various meat product, alcohol and high altitude rest were recommended for treating TB (Ghose, 2003). In 200-250 CE ancient China, tuberculosis was mentioned as a contagious *shizhu* (translated as 'corpse' disease due to the high mortality) and was described as dynamic disease having many symptoms ranging from thirty-six to ninety-nine different kinds capable of wiping out a whole family (Liu, 1971). In the 19th century, the high mortality

rate among young and middle-aged adults and the glossy dying look of the infected caused many to refer to the disease as the "romantic disease" (Herzog, 1998).

The exact period when TB emerged seems not to be recorded. However, evidence of TB as human disease goes back as far as the beginning of human civilizations. For instance, evidence of TB have been discovered in human remains from the Neolithic era dating from 7000 BC in Atlit Yam (an ancient eastern Mediterranean civilization) and in human remains from a cemetery close to Heidelberg in the Neolithic era (Hershkovitz *et al.*, 2008; Madkour *et al.*, 2004). Signs of TB have also been found in Egyptian mummies dated between 3000 and 2400 BC (Madkour *et al.*, 2004; Zink *et al.*, 2003). Mummies of the priest Nesperehen, the priest Philoc, Akhenaten and his wife Nefertiti all have given indications that they suffered from TB (Madkour *et al.*, 2004).

Two independent studies have proposed that TB accompanied the out of Africa migrations of modern humans (*Homo sapiens*) some 40 thousand to 60 thousand years ago via land and the oceans to the rest of the world (Gutierrez *et al.*, 2005; Hershberg *et al.*, 2008). Another study using coalescent analysis of 259 representative *Mycobacterium tuberculosis* complex (MTBC) genomes predicted the pathogen emerged about 70,000 years ago in Africa and did accompanied the migrations of anatomically modern humans and expanded as a consequence of growing human population density during the Neolithic period (Comas *et al.*, 2013). However, report from another genome study in 2014 suggested that TB is significantly younger. Using genomes of the bacteria from 1,000-year-old human skeletons in southern Peru, the investigators concluded that the pathogen was less than 6,000 years old and theorized that seals may have been the mode of transmission from Africa to South America (Bos *et al.*, 2014).

Understanding the disease and identification of the causative agent of TB took many years. By the end of the 19th century, several major breakthroughs by various scientists gave hope that a cause and cure might be found. One of the most important physicians frequently remembered today for his major breakthrough in understanding the disease was Rene Laennec inventor of stethoscope in 1819 (Daniel, 2000, 2006). His clear description of the pathogenesis of TB using terms (such as murmurs, thrills, bronchial and vesicular breathing as well as pectoriloquy as signs of tubercular cavities) still relevant today paved the way for modern understanding of TB (Jay, 2000; Roguin, 2006). In Laennec's era, the disease extended rapidly across Europe in relation with industrialization. During that period, urbanization was galloping, and poverty, poor hygienic conditions and overcrowding became the order of the day - an environment conducive for the disease to thrive. Between 1851 and 1910, the death rate due to TB soared occurring mostly in the youth; in England and Wales alone, four million aged 20 to 24 died from tuberculosis (Bynum, 2012), giving the disease the name '*the robber of youth*' (Segen, 1992). Surrounded by so many deaths from one disease, medical practitioners and scientist sought to understand its aetiology.

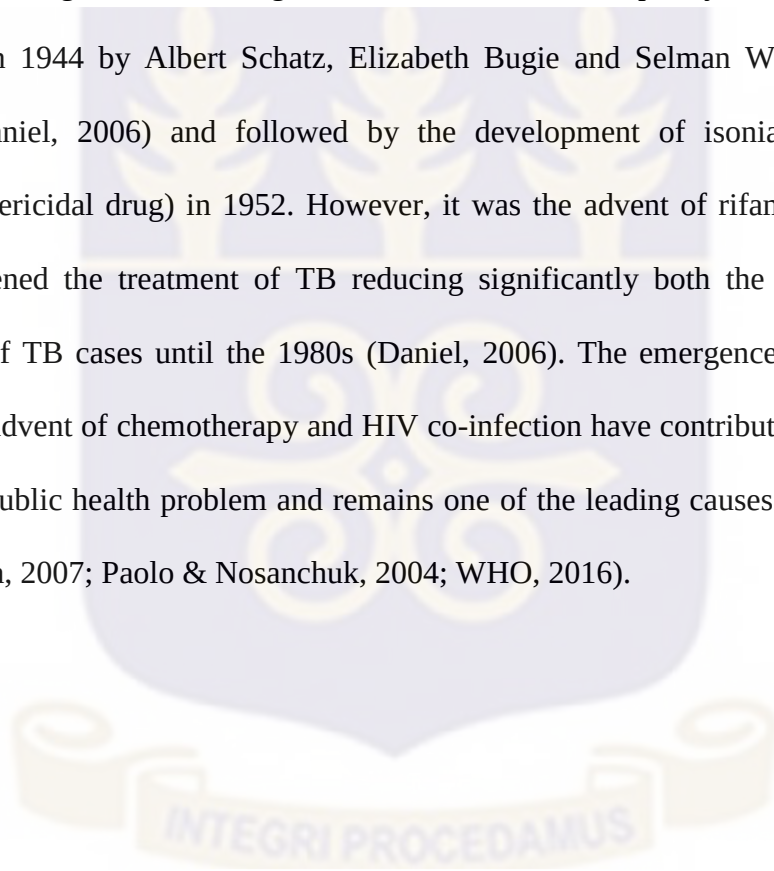
The history of TB was changed dramatically on March 24, 1882, with the famous Hermann Heinrich Robert Koch presentation, *Die Aetiologie der Tuberkulose*, to the Berlin Physiological Society where Dr. Koch demonstrated the infectious aetiology of the disease and presented *Mycobacterium tuberculosis* as the causative agent (Cambau & Drancourt, 2014) leading to the recognition of March 24th as the world TB day. With the identification of TB as an infectious disease and the recognition of the illness, a new era

of visualizing the disease as a public health problem began. Dr. Koch purified the tuberculin protein in 1890 intended to be used as a vaccine against TB but proved to be ineffective means of immunization against TB (Magner, 2002) but was demonstrated in 1908 as an effective intradermic test for diagnosing TB by Charles Mantoux (Madkour *et al.*, 2004). The first successful immunization against TB was the use an attenuated bovine-strain of TB bacilli called BCG (Bacille Calmette-Guerin) by Albert Calmette and Camille Guerin in 1906 (Bonah, 2005). The first human use was in 1921 in France and received global acceptance after the Second World War (Comstock, 1994; Macpherson, 1997).

The need for institutions to house and treat sufferers of TB became urgent even before understanding the aetiology of the disease. Even though this proposal met a lot of resistance, the first sanatorium (Fig 2.1) 650 meters above sea level in Germany in 1854 was built by Herman Brehmer who was convinced that TB was as a results of inability of the heart to irrigate the lungs and hence proposed housing of TB patients at high altitudes with less pressure to enable the heart to function better (Dubos and Dubos, 1990). The concept of sanatorium housing spread to other emerging economies including the North America, which were mostly patronized by the rich who could afford. However, sanatoriums were not the only treatment facilities. Specialized TB clinics and dispensaries also began in major metropolitan areas such as Edinburgh in what was termed “Edinburgh Anti-tuberculosis Scheme” geared towards the middle class the lower class (Edinburgh-Anti-Tuberculosis-Scheme, 1914). The surgical intervention

pneumothorax or *plombage*: (the collapsing of an infected lung to ‘rest’ it and allow the lesions to heal) were mostly used during these specialized TB clinics (Daniel, 2006).

The use of these interventions together with improved sanitation and nutrition reduced the TB mortality by 90% to 95% before the introduction of the first chemical agents for treatment (Colgrove, 2002; Deaton & Stone, 2013). The introduction of different anti tuberculosis agents which began with the isolation streptomycin from *Streptomyces griseus* in 1944 by Albert Schatz, Elizabeth Bugie and Selman Waksman (Colgrove, 2002; Daniel, 2006) and followed by the development of isoniazid (the first oral mycobactericidal drug) in 1952. However, it was the advent of rifampicin in the 1970s that hastened the treatment of TB reducing significantly both the recovery time and number of TB cases until the 1980s (Daniel, 2006). The emergence of drug resistance with the advent of chemotherapy and HIV co-infection have contributed to TB still being a major public health problem and remains one of the leading causes of deaths (Antoine & Watson, 2007; Paolo & Nosanchuk, 2004; WHO, 2016).





The Brehmer Sanatorium at Goerbersdorf.



Figure 2. 1: The first tuberculosis sanatorium built in Germany.

(source:<https://blog.wellcome.ac.uk/2016/03/18/image-of-the-week-tuberculosis-sanatorium/>)

2.2 GLOBAL BURDEN OF TB

Tuberculosis is the leading cause of adult mortality by a single infectious disease globally. Approximately a third of the 7 billion human population is latently infected with the TB bacilli, thus provide a large reservoir for future cases and active transmission that will last for decades (WHO, 2016). The 2016 WHO global tuberculosis report for the year 2015 estimated an alarming 10.4 million new cases (equivalent of 142 cases per 100 thousand population), out of which 56%, 34% and 10% were respectively among men, women and children under 15 years. People living with HIV accounted for 12.0% of all the new TB cases. There were a total of 1.4 million deaths among TB patients without HIV with additional 400 thousand deaths among TB/HIV patients (WHO, 2016).

Countries in the developing world carry the highest burden of TB (Fig 2.2). South-East Asia including China, India and Bangladesh accounted for 45.6% of the global TB incidence, 50.7% TB deaths without HIV, 19.4% global TB/HIV co-infections and 18.9% addition deaths due to TB/HIV even though home to over 50% of the global population. However, the African region which is home to 13.5% of the global population, accounted for 26.2% of the global TB burden, 71.3% of Global TB/HIV burden, 32.1% of TB related deaths without HIV and 76.9% of TB/HIV deaths (WHO, 2016). This makes the region a hot-spot for concern because the down-regulation of cell-mediated immune response in HIV reduces the ability of infected individuals to contain latent infectious as well as prevent new infections.

The emergence of drug resistance strains especially multi-drug resistance (MDR)

(Crofton & Mitchison, 1948) and the failure of the BCG vaccine to protect against TB other than cerebral TB in children less than 5 years (Perkins & Cunningham, 2007) also impedes TB control. In 2015 for instance, there were an estimated 480 thousand (representing 5% of the total incidence) new cases of MDR-TB in addition to 100 thousand RIF mono-resistant TB (RIF^r-TB) cases recorded (WHO, 2016). Individuals infected with resistant strains would have to take second line anti-TB drugs that are more toxic for a long time (at least 18 months), with side effect limiting compliance and treatment success rates.

2.3 TUBERCULOSIS IN GHANA

In 2015, 44 thousand new TB cases among which 9.9 thousand were TB/HIV and 1.5 thousand MDR/RIF^r-TB were estimated for Ghana. This translates into an annual TB prevalence of 356 per 100 thousand population (0.4%) which is more than twice the global rate of 142 cases per 100 thousand population (0.1%) making Ghana one of the most burdened countries in Africa. A total of 15 thousand deaths (34%) were estimated in 2015 stratified as 10 thousand among TB/HIV patients and 5 thousand among HIV negative TB patients even though the actual recorded deaths was 22% (WHO, 2016). The TB/HIV co-infected are 22.5% of the total TB cases, thus Ghana is among a group of 30 countries classified as high TB/HIV burdened (Fig 2.3).

Treatment of TB in Ghana begun in 1954 facilitated by the establishment of help groups, even though; treatment was mostly sporadic and poorly coordinated.

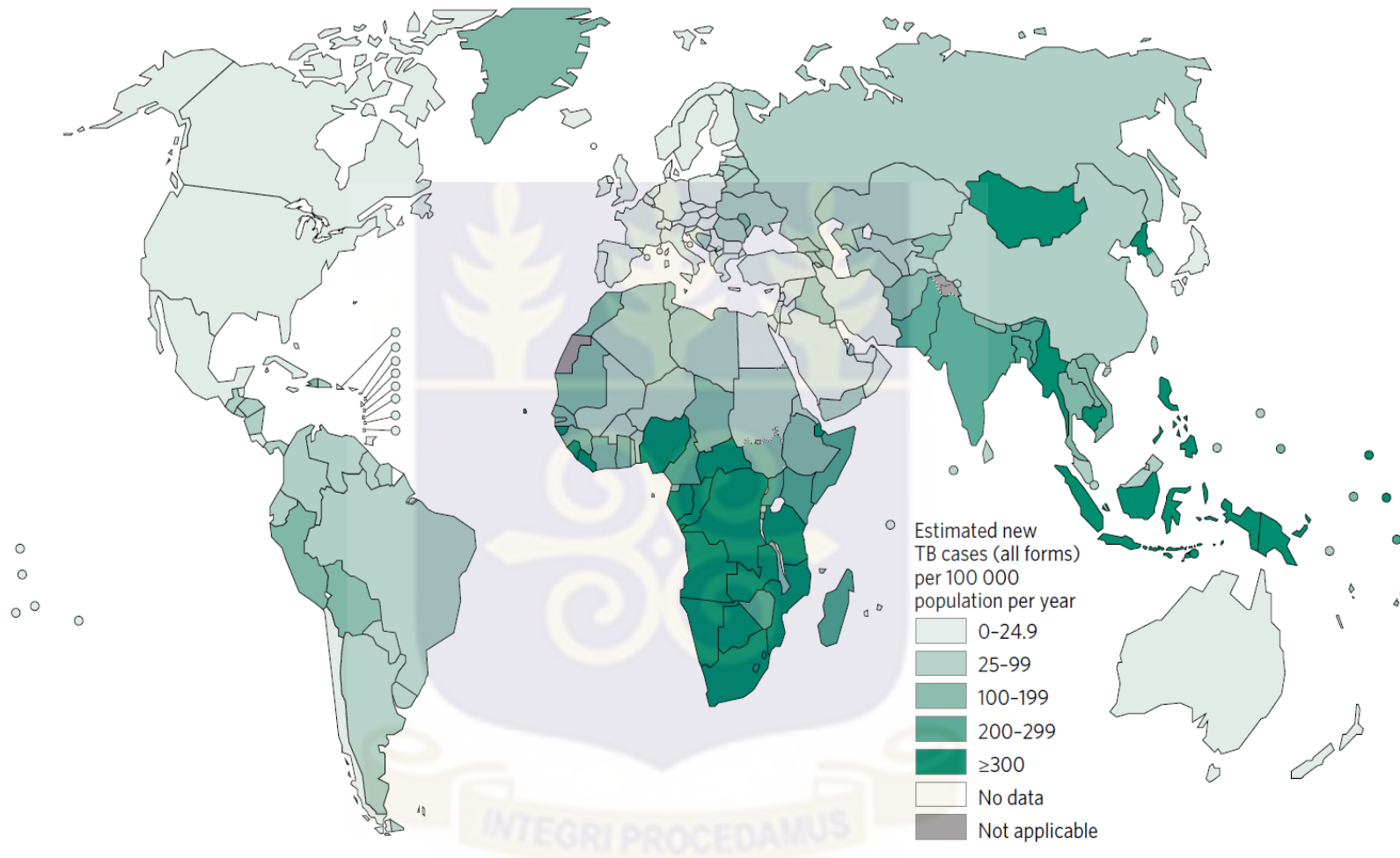


Figure 2.2: Global estimates of TB incidence rate in 2015. (Adapted from (WHO, 2016)).

Access to treatment was not free and hence restricted to some selected few who could afford the cost (Amo-adjei & Kumi-kyereme, 2013; Koch & Marolda, 1961). Treatment of TB became accessible and free for all following the adaptation of the Directly Observed Treatment Short-Course (DOTS) strategy in 1994 by the WHO. Currently, Ghana boasts of over 700 treatment centres strewn all over the country (Personal conversation with the control program manager). The annual cure rate as of 2015 stood at 76% of all TB patients stratified as 85% among all new TB cases, 76% among TB/HIV patients, 69% among MDR-TB patients and 81% among all patients with history of previous TB treatment (WHO, 2016).

The control of tuberculosis in Ghana has chalked some notable successes over the years and prominently among them are the rise of case notification from 7,425 in 1996 to 15,207 in 2012 (Personal communication with the Director of the National Tuberculosis Control Programme, Ghana), decline in death rate from 51 per 100 thousand population in 1990 to approximately 4 per 100 thousand population in 2013 and finally the rise in the rate of treatment success from 44.0% in 1997 to 76% (including drug resistance TB) in 2015 (WHO, 2016). The current WHO report indicates that the Greater Accra region reported the highest number of cases (averagely 70 cases) in a 100 thousand population whereas the Northern region had the least case report (averagely 50) in 100 thousand population. However, the national average case detection rate stands at an estimated 31%, which is far below the global target of 70%, meaning that about 69% of all TB cases in the country go undetected with the probability of continuous transmission within the populace (WHO, 2016).

2.4 THE CAUSATIVE AGENT OF TUBERCULOSIS

2.4.1 The *Mycobacterium tuberculosis* Complex

A group of closely related non-motile, aerobic, acid-fast bacilli collectively called *M. tuberculosis* complex (MTBC) are the main cause of TB in mammals (Brosch *et al.*, 2002; Cousins *et al.*, 2003; Frota *et al.*, 2004; Garnier *et al.*, 2003). These include *M. tuberculosis* sensu stricto (*Mtbss*), *M. africanum* (*Maf*), *M. mungi*, *M. orygis*, *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii* and *M. bovis* BCG. The MTBC belong to the phylum actinobacteria, in the order actinomycetales, suborder corynebacteriaceae and the genus *Mycobacterium* (Garrity, 2004)

In spite of their genetic relatedness, the MTBC seems to differ in host specificity, although there are occasional cross-species infections. *M. bovis* is a pathogen of cattle and sheep, *M. caprae* is goats, *M. microti* is rodents, *M. pinnipedii* is sea seals and sea lions, *M. mungi* is Mongoose whereas *M. orygis* infects antelope (Aranaz, 2003; Cavanagh *et al.*, 2002; Cousins *et al.*, 2003; Frota *et al.*, 2004; Van Ingen *et al.*, 2012; Wells, 1937). The main human pathogens are *Mtbss* and *Maf*, herein referred as human-adapted MTBC (Brosch *et al.*, 2002; Gagneux & Small, 2007; Garnier *et al.*, 2003; Yeboah-Manu *et al.*, 2016). *M. bovis* is reportedly the pathogen with the broadest host-range infecting all kinds of mammals including humans (Garnier *et al.*, 2003). Nevertheless, the contribution of *M. bovis* infection to the global burden of human TB is very low, below 1%.

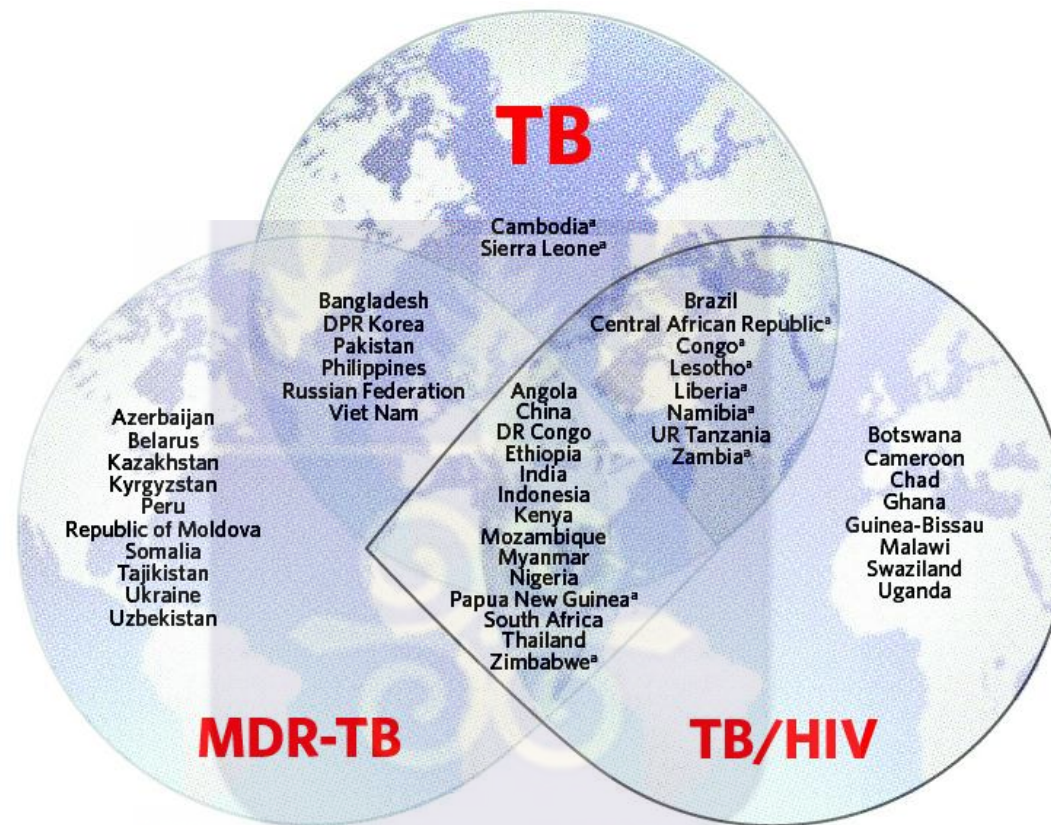


Figure 2. 3: Countries in the three TB high-burden lists 2016–2020. (Adapted from (WHO, 2016))

A mutations within the *PhoP/PhoR* regulation system (an important regulator of virulence factors involved in adaptation to survival in human macrophages) leads to down-regulation of the *PhoP* regulon and limits the ability of *M. bovis* to transmit within human populations (Smith *et al.*, 2006; Zhou & Xie, 2014).

Seven distinct sub-families of the human adapted MTBC have been identified to date designated as Gagneux lineages 1 to 7 defined by specific genomic signatures (Fig 2.4). Using a combination of comparative genome hybridization and multi-locus sequencing to analyse a collection of 875 human-adapted MTBC isolates from 80 countries, the first six lineages (L1 to L4 under *Mtbss* and L5 and L6 under *Maf*) were defined (Gagneux *et al.*, 2006). The additional *Mtbss* lineage designated L7 was later isolated from some TB patients at the horn of Africa (Blouin *et al.*, 2012; Firdessa *et al.*, 2013). Subsequent mapping of the Global collection of MTBC strains according to the country of origin revealed a strong bio-geographical structure (S. Gagneux, 2012; S. Gagneux, DeRiemer, *et al.*, 2006). Generally the *Mtbss* lineages with the exception of L7 which is restricted to the Horn of Africa are globally distributed whereas the two *Maf* lineages are restricted to West Africa (Fig 2.5).

These lineages are also colour coded and given names associated with specific geographical regions where there are most dominant. For instance, L1 is also called the pink or indo-oceanic, L2 is the blue or East-Asian, L3 is purple or East-African Indian (EAI), L4 is red or Euro-American, L5 is brown or West Africa 1 or *Maf*1, L6 is green or West Africa 2 or *Maf*2 whereas L7 is yellow or Ethiopian lineage (Blouin *et al.*, 2012; Firdessa *et al.*, 2013; Gagneux *et al.*, 2006). These lineages are further divisible into sub-

lineages with specific genotypic and phenotypic traits that can affect their ability to transmit and establish infections. For instance within L4, there are sub-lineages that are widely distributed globally (generalists) as well as others that are more restricted to specific geographical settings (specialist) (Stucki *et al.*, 2016).

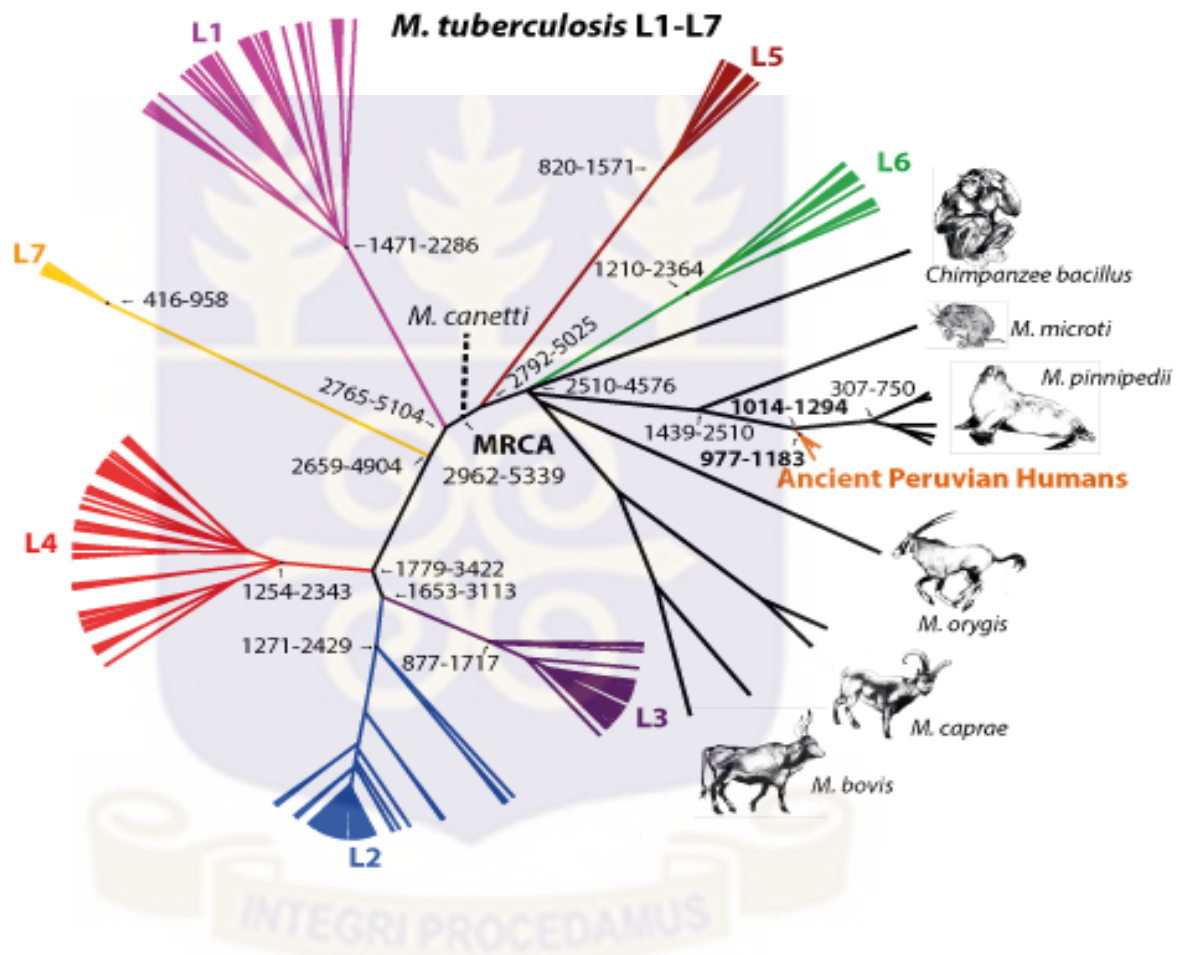


Figure 2.4: Global Phylogeny of the MTBC rooted on *M. canettii*

Each clade represents an MTBC lineage and the clade size corresponds to the number of individual strain genomes. the clades are coloured according to the lineages with pink, blue, purple, red, brown, green and yellow representing the human adapted MTBC lineages 1 to 7 respectively whereas the animal strains are in black. Showing in gold in addition is ancient Peruvian human strains reconstructed from archeological remains (Adapted from Bos *et al.*, 2014).



Figure 2.5: Global phylogeography of the human adapted *M. tuberculosis* complex.

*The seven identified lineages of the human adapted *M. tuberculosis* complex are geographically structured. Each dot corresponds to one of the countries represented in the global strain collection. The colours of the dots relate to the lineages and indicate the dominant lineage(s) in the respective countries. Lineages 1, 2, 3, 4, 5, 6 and 7 are represented by the colours pink, blue, purple, red, brown, green and yellow respectively. (Adapted and modified from Gagneux et al., 2006).*

2.4.2 Other Mycobacterial Species

In addition to the MTBC, mycobacteria such as the smooth tubercle bacilli (STBs) and non-tuberculous mycobacteria species (NTMs) have also been found to occasionally cause disease symptoms similar to TB caused by the MTBC (Gutierrez *et al.*, 2005). The STBs are considered outliers from the MTBC albeit the closest relatives of the MTBC by phylogenetics (Blouin *et al.*, 2012; Supply *et al.*, 2013). Chief among the STBs is *M. canettii*, named after Georges Canetti, a pioneer of STBs research (D. van Soolingen *et al.*, 1997). *M. canettii* was previously thought to belong to the MTBC (Brisse *et al.*, 2006). However, the large number of SNPs among selected *M. canettii* strains and between *M. canettii* and MTBC strains (Blouin *et al.*, 2014) together with the observed ongoing horizontal gene transfer among *M. canettii* (Boritsch *et al.*, 2016; Gutierrez *et al.*, 2005) compared to the highly clonal genomes of the MTBC were evidence enough to dismember *M. canettii* from the MTBC. Phylogenetic analysis places *M. canettii* at the deep root of MTBC phylogenies (Fig 2.3, Fig 2.5) making it an MTBC progenitor-like pathogen (Blouin *et al.*, 2014; Derbyshire & Gray, 2014; Mortimer & Pepperell, 2014; Supply *et al.*, 2013).

The NTMs, thought to have emerged millions of years before members of the MTBC, (Fig 2.6) are increasingly becoming a major public health problem (Galagan, 2014). This is mainly due to the emergence of HIV/AIDS coupled with other immune suppression conditions such as diabetes and cancer chemotherapy (Kasperbauer & De Groote, 2015; Nunes-Costa *et al.*, 2016; Philley & Griffith, 2015). The most commonly isolated from TB cases are the *M. avium* complex (MAC) including *M. intracellulare*, *M. avium*

subspecies *paratuberculosis* and *M. colombiense*, which are mostly opportunistic infections (de Mello *et al.*, 2013; Nunes-Costa *et al.*, 2016).

2.4.3 The Mycobacterial Cell Wall

A unique feature of the genus *Mycobacterium* is the cell wall, which is thicker than other bacteria and makes the species essentially adapted for intracellular survival and persistence (Bhamidi *et al.*, 2011). The mycobacterial cell wall (Fig 2.7), which is composed of upper and lower segments, is made up of over 60% lipids. Prominent among these lipids are mycolic acids (branched, long-chain fatty acids), cord factor and wax-D. The lower segment directly above the plasma membrane, is composed of peptidoglycan that is covalently attached to arabinogalactan esterified to the mycolic acids, thence forming the core of the cell wall also called the mycolyl arabinogalactan–peptidoglycan (mAGP) complex. The upper segment is composed of free lipids with long and short chain fatty acids to complement the shorter α -chains and the longer chains respectively.

In addition to the upper segment are interspersed cell-wall proteins (including porins) and free lipids such as, phenolic glycolipids, phthiocerol dimycocerosates, dimycolyltrehalose or cord factor, sulpholipids and phosphatidylinositol mannosides (PIMs), lipomannan and lipoarabinomannan intercalated within the mycolic acids giving the cell wall its rigid hydrophobic property (Abdallah *et al.*, 2007; Beukes *et al.*, 2010; Brennan, 2003). These lipids, proteins, and lipoglycans other than being beneficial to the bacteria by making them relatively impermeable to hydrophilic agents including disinfectants and drugs, also serve as the signaling effector molecules in the disease process.

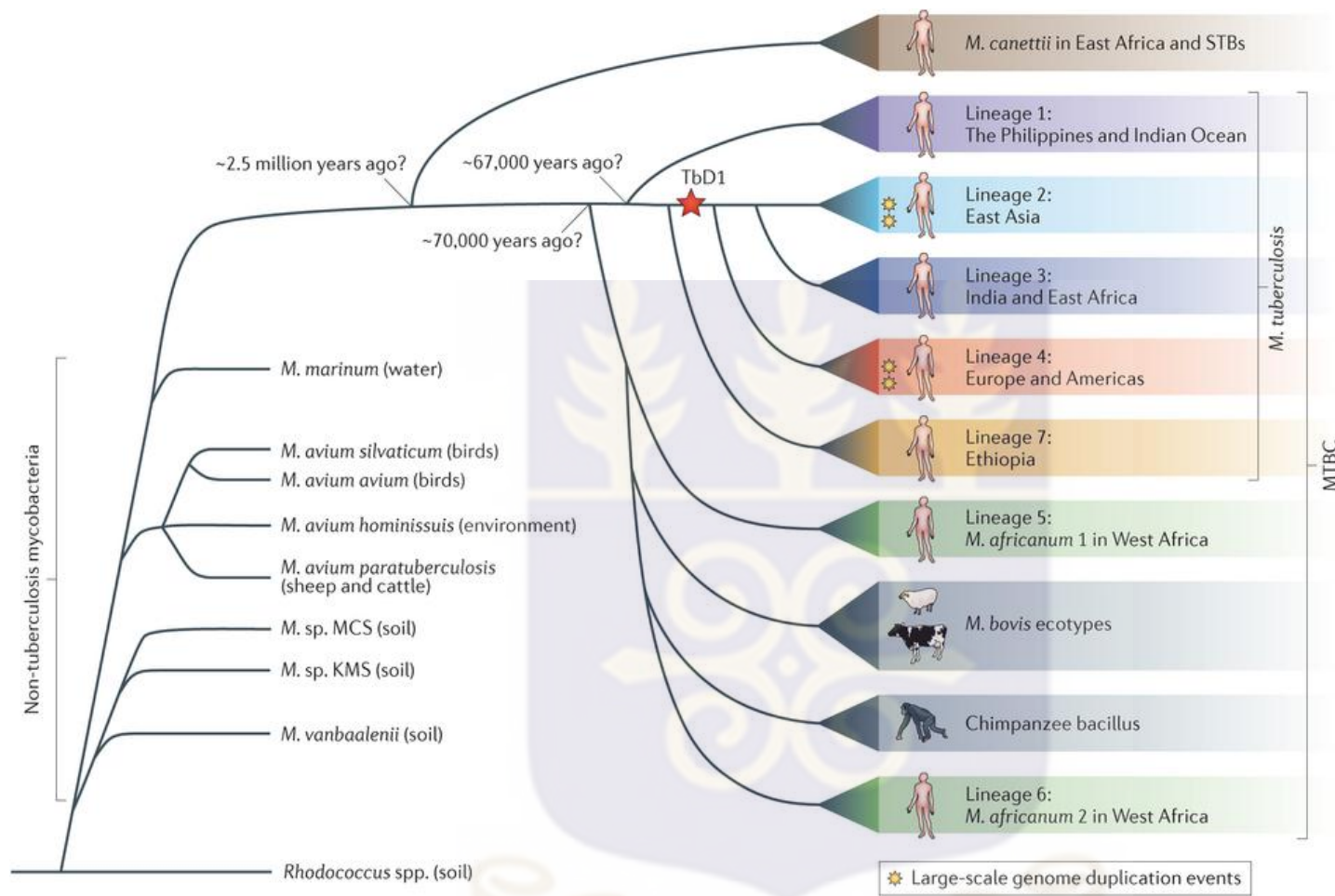


Figure 2.6: Phylogeny mycobacteria rooted on *Rhodococcus* spp as an outlier. (Adopted from (Galagan, 2014))

Even though the inherent biosynthetic pathways of these unique mycobacterial cell wall components potentially render them critical targets for selective drugs development, it is the core of the cell wall which is most essential for cell viability and hence best target for development of new drugs (Brennan, 2003).

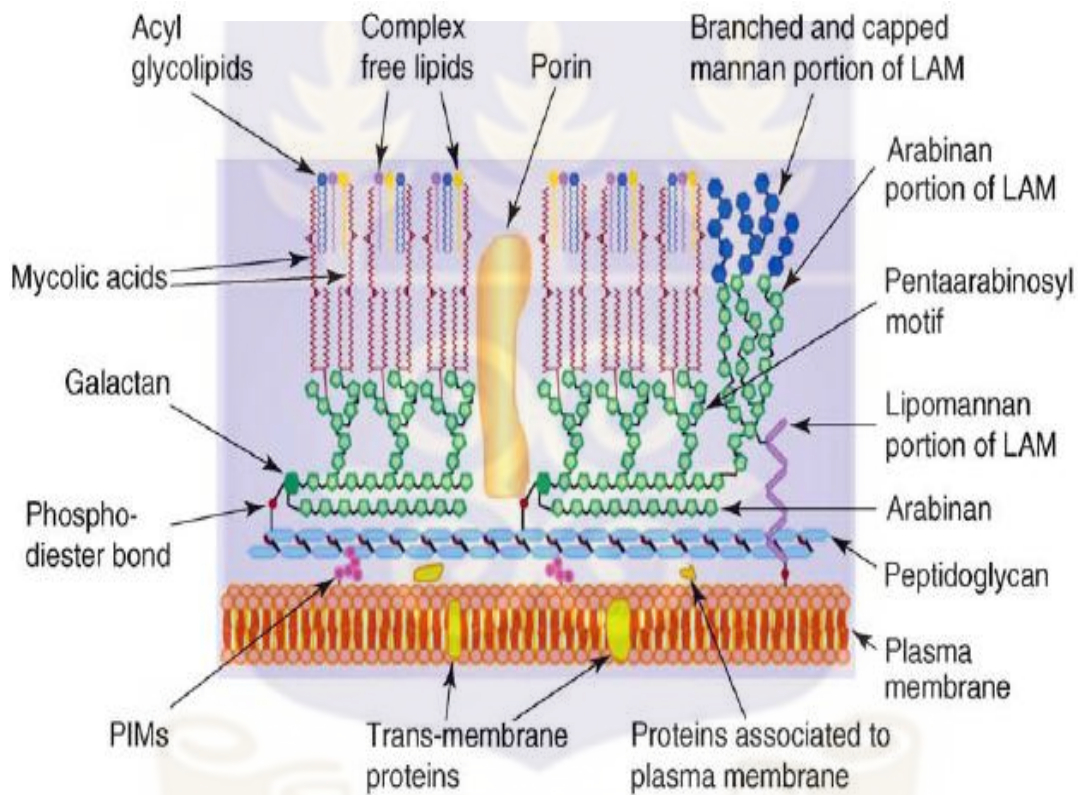


Figure 2.7: Schematic representation of the mycobacterial cell wall (adapted from Medjahed *et al.*, 2010).

2.5 PATHOGENESIS OF TB

The primary mode of transmission is via inhalation of aerosol droplets expelled by infected people, containing viable bacilli. When MTBC strain enters the host mostly via the airways into the lungs, it is engulfed by alveolar macrophages, which may subsequently lead to rapid elimination of the bacillus, containment of the infection or development of active tuberculosis disease (Fig 2.8 I). The infectious dose is 1-200 bacilli, however an aerosol droplet can contain up to 400 bacilli, therefore it is nearly impossible to remain uninfected upon inhalation of a bacilli laden droplet. The understanding of the exact processes required for establishment of TB infection is limited; however, animal model experiments have given important clues of the early events of primary infection (Orme *et al.*, 2015) [Fig 2.8 II].

The MTBC bacilli first encounter alveolar macrophages (also called resident alveolar macrophages which are the most dominant immune cells involved with TB infection) and/or alveolar type II pneumocytes in the lower respiratory tract after its inhalation. The bacilli are engulfed by these phagocytic cells via receptor-mediated phagocytosis involving mannose and complement as well as surfactant protein A (a glycoprotein present on alveolar surfaces shown to enhance binding and uptake of the bacilli by up-regulation of mannose receptor activity (Schlesinger, 1993). After internalization, the bacterium has been shown to block the subsequent fusion of the resulting phagosome with lysosome which would lead to its destruction (Russell, 2011). Inhibition of Ca^{2+} signalling has been implicated for this failed phagosome-lysosome fusion (Malik *et al.*, 2000). Through the action of ESX-1 secretion system, the bacilli raptures the phagosomal

membrane then releasing bacterial products such as mycobacteria DNA, cell debris and even whole bacterial cells into the cytosol of the alveolar macrophage within days after initial infection (Houben *et al.*, 2012; van der Wel *et al.*, 2007). This possibly activates the cytosolic surveillance system culminating with induction of interferon response which inadvertently promotes intracellular proliferation of the bacilli (Manzanillo *et al.*, 2012; Pandey *et al.*, 2009).

The TB bacilli then translocates into the lung interstitial cells after alveolar macrophage infection to continue the process of infection. Infection of the parenchyma cells of the epithelium leads to increase roaming of immune cells (including dendritic cells, extra-pulmonary macrophages and monocytes) to the site of infection which lead to generation of multicellular host response called granuloma (a growing collection of infected phagocytic cells). T cell priming is then initiated after presentation of the bacilli to lymph nodes by infected dendritic cells (Chackerian *et al.*, 2002; Samstein *et al.*, 2013; Wolf *et al.*, 2008).

Depending on the bacilli burden of the granuloma and host immunity, it may contain the infection in what is termed latent TB infection (LTBI) or degenerate into active TB situation. However, the exact processes involved with progression from LTBI to active TB characterized by overwhelming replication of the bacilli, necrotizing of the granuloma with release of the bacilli into the sputum are not well understood (Comas *et al.*, 2010; Pai *et al.*, 2016). Initial necrosis of the granuloma is followed by the host attempt to contain the spread of the infection by killing of tissue around the ruptured

granuloma and deploying T cells to the site which results in the inflammation and some other symptoms associated with TB. The infection then spreads by evading the immune response via infection of the very cells aimed at containing the bacilli. This leads to growing areas of the lung infected with the bacilli called tubercles. These tubercles may liquefy, thence leading to bacteria rich debris lining the walls of the lungs ready to be expelled during coughing (Lázár-Molnár *et al.*, 2010; Mayer-Barber *et al.*, 2011).

2.6 CONTROL OF TB

2.6.1 Vaccination against TB

Currently, the only licensed vaccine to protect against development of active TB diseases, BCG has been in use since 1921 (Mangtani *et al.*, 2014; Roy *et al.*, 2014). It has been evaluated under countless interventional trials and observational studies with variable reports of efficacy ranging from 0% to 80% under different geographical locations (Mangtani *et al.*, 2014). The most reliable and reproducible efficacy of BCG is the protection against pulmonary and severe forms of extra-pulmonary TB disease in children under 5 years (Roy *et al.*, 2014; Trunz *et al.*, 2006). Since the morbidity and mortality associated with neonate TB can be so devastating, BCG vaccination of newborns is an invaluable activity to prevent active TB among children of this age group. Over 90% of newly born babies are annually vaccinated with BCG (Zwerling *et al.*, 2011). However, majority of active transmission of pulmonary TB occurs among adolescents and adults for whom the efficacy of BCG is unreliable (Barreto *et al.*, 2011; Tuberculosis research centre, 1999)

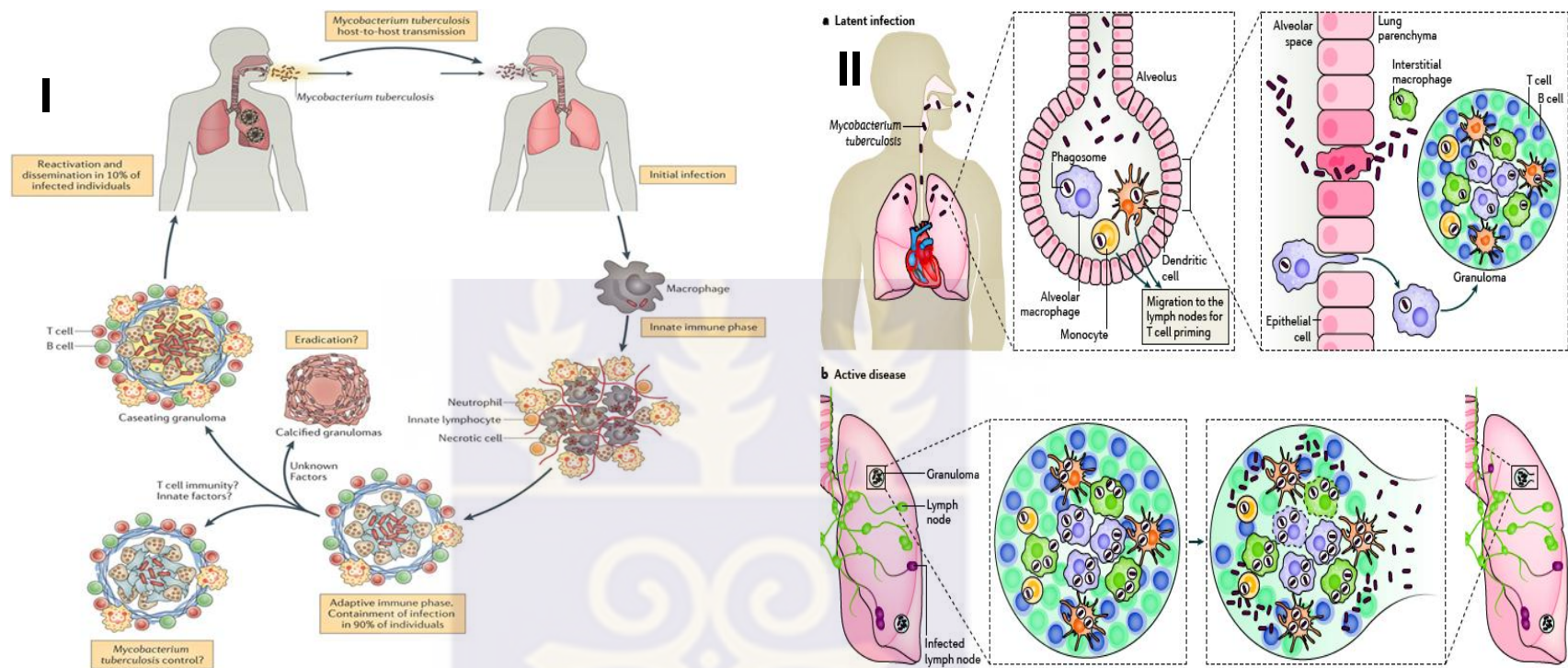


Figure 2.8: Pathogenesis of tuberculosis.

I: Following inhalation of MTBC contaminated droplet, there could be either total elimination of the bacilli if inoculum size is small enough or establishment of infection. Most infected individuals will remain in a 'latent' state of infection with no clinical symptoms. Up to 10% of the infected eventually progress to active disease with its accompanying clinical manifestations. **II:** [a: Infection begins with inhalation of the MTBC and the encounter with resident alveolar macrophages. Failure of this first line defence to eliminate the bacteria leads to the bacteria invading the lung interstitial tissue, either by the bacteria directly infecting the alveolar epithelium or the infected alveolar macrophages migrating to the lung parenchyma. Subsequently, either dendritic cells or inflammatory monocytes transport the bacilli to pulmonary lymph nodes for T cell priming. This event leads to the recruitment of immune cells, including T cells and B cells, to the lung parenchyma to form a granuloma]. [b The bacteria replicate within the growing granuloma. If the bacterial load becomes too great, the granuloma will fail to contain the infection and bacteria could disseminate to other parts of the lungs and eventually to other organs.]. (Adapted from (Nunes-Alves et al., 2014; Pai et al., 2016))

This makes the immediate search for an active vaccine imperative until which control of TB would continue to rely on early detection and treatment with appropriate anti-tubercular agents.

2.6.2 Diagnosis of TB

People reporting to hospitals/clinics with productive cough lasting over 2 weeks and sometimes streaked with blood with or without unexplained fever, chest pain and weight loss are clinically considered presumptive TB cases may be asked to undergo chest X-ray (Cudahy & Haven, 2016). The definitive method for diagnosis of TB is bacteriological confirmation and the gold standard diagnostic assay for confirmation of pulmonary TB is culture. However, due to the biosafety issues, the required expertise, the cost of equipment and the long time to culture positivity (at least 4 weeks), culture is mainly reserved for research purposes and monitoring of treatment. Therefore, at point of care settings, sputum smear microscopy and molecular based rapid assays are the commonly used diagnostic tools. However, there are few instances where microscopy or molecular diagnostics give negative results (especially in children and HIV positive patients due to either the inability to produce quality sputum or paucibacilliary) but are deemed confirmed pulmonary TB and hence managed as such based only on suggestive chest X-ray (Piccini *et al.*, 2014; Walters *et al.*, 2008).

The cheapest, simplest and most commonly used diagnostic assay for pulmonary TB is direct smear detection of acid-fast bacilli (AFB) by Ziel-Neelsen staining and light

microscopy which is available in most point of care centres (especially resource limited settings) (TB CARE I, 2014; WHO, 2011). However, it is confounded by its relatively less sensitivity (20% to 80% averagely 50%) and less specificity for detection of members of the MTBC (Davis, 2014). Conventional fluorescence microscopy is relatively more sensitive (10% more sensitive) and requires less time for detection than Ziel-Neelsen staining cum light microscopy, however, it is limited by high cost of equipment and reagents and specificity (Steingart *et al.*, 2006). The WHO has recently recommended the use of light emitting diode (LED) based microscopy for clinical diagnosis of TB due its qualitative and operational, and cost advantage over Ziel-Neelsen and fluorescence microscopy. However, it has not yet been implemented in most point of care centres around the globe (WHO, 2010). The major limitations of the microscopy-based diagnostics are the inability to differentiate MTBC from NTMs and inability to tell the sensitivity of the bacilli to the available anti-TB drugs to advice an appropriate treatment regimen, an essential requirement for the fight against transmission of DR-TB. However, it remains the most popular diagnostic assay in most public health labs across the globe because it is rapid, cheap and requires no sophisticated machinery.

Molecular-based methods for diagnosis of TB such as the DNA-based line probe MTBDR*plus* and Xpert MTB/RIF assays are increasingly becoming popular as they are very rapid and have high specificity, sensitivity and positive predictive values. An added advantage of the DNA-based assays is the ability to differentiate MTBC from NTMs, ability to confirm the presence of MTBC in 50% - 80% of microscopy negative sputum samples, ability to diagnose sensitivity of infecting pathogen to available ant-TB drugs

such as RIF (Xpert MTB/RIF) and INH and RIF (MTBDR_{plus}) to advice proper management of TB (Nhu *et al.*, 2014). However, these assays rely on the detection of the presence of DNA and hence cannot differentiate between viable and non-viable bacteria. Therefore, they are not recommended for monitoring patients who are on anti-tubercular chemotherapy whom culture remains the only reliable option.

2.6.3 Treatment of TB and Emergence of Drug resistance

The control of TB relies on detection of active cases followed by chemotherapy using the directly observed treatment short-course (DOTS) strategy, which aims to cure the affected and also cut the source of infection, as man is the major host of the causative agent. The first line treatment involves 2 months administration of isoniazid (INH), rifampicin (RIF), pyrazinamide (PNZ) and ethambutol (EMB) called the intensive phase followed by 4 months of INH and RIF administration called the continuous phase. Even though TB is treatable, its control is being threatened by the emergence of drug resistant (DR) bacilli especially multi-drug resistant (MDR) TB, defined as resistance to at least rifampicin (RIF) and isoniazid (INH), the two most important first line drugs (Kirimuhuzya, 2007). Treatment of MDR-TB requires the use of a combination of high concentration first line drugs (such INH and PNZ) and second line drugs (including moxifloxacin (Mfx), prothionamide (Pto), clofazime (Cfz) and kanamycin (Km)) that are more toxic (with severe side effects) for prolonged duration even though has lower success rate as compared to drug sensitive TB.

Until very recently, MDR patients required chemotherapy lasting for not less than 18 months. However, this regimen had marked reduced success rate compared to DOTS. The recurrent WHO recommended short course treatment for MDR or rifampicin resistant TB is 4-6 months intensive phase administration of Km-Mfx-Pto-Cfz-PNZ_{high-dose}-EMB followed by 5 months continuous phase of Mfx-Cfz-PNZ-EMB (World Health Organization, 2016). Even though treatment of MDR-TB is more expensive (averagely USD 1100) compared to drug susceptible (DS) TB (averagely USD 500), the success rate is highly reduced (52%) compared to 83% for DS-TB (Davies, 2001; World Health Organization, 2016). A more severe form of drug resistant TB called extensively drug resistant tuberculosis (XDR-TB) is TB caused by MDR bacilli that is also resistant to fluoroquinolones (FQs) and any of the second line aminoglycoside injectable drugs (Kirimuhuzya, 2007). This form of TB is treated with more toxic drugs for longer periods with much lower success rates (28%). If nothing is done to curb the transmission of DR-TB, the once treatable disease would eventually become incurable. Therefore, more stringent efforts are needed for early detection and characterization of TB cases to effectively control the disease (Barry *et al.*, 2009) if the new global effort (the End TB Strategy) is to succeed (WHO, 2015).

2.6.4 Molecular Basis of MDR-TB

2.6.4.1 Isoniazid Resistance

Isoniazid (generic name for isonicotinic acid hydrazide) is a pro-drug which is bactericidal in its active form against actively growing MTBC and one of the most potent first-line drugs for the treatment of TB. Resistance to INH is a complex process and has

been associated with mutations in several mycobacterial genetic elements key among which are *katG*, *inhApro*, *ahpC*, *inhA* and *ndh* (Fig 2.9) in order of importance (Heym, Saint-Joanis, & Cole, 1999; S Ramaswamy & Musser, 1998). The pro-drug INH requires activation by the catalase/oxidase enzyme encoded by *katG* of the target bacterium. Activated isoniazid (isonicotinic acid) interferes with the biosynthesis of essential mycolic acids by inhibiting NADH-dependent *enoyl-ACP reductase* encoded by the *inhA*. Genetic alterations associated with the *katG* are the most common mutations responsible for high isoniazid resistance (Ando *et al.*, 2011).

Over a hundred mutations including; mis-sense and nonsense mutations, insertions, deletions, truncation and, more rarely, deletion of *katG* have been associated with blockage of pro-drug activation (Wei *et al.*, 2003). The most common INH resistance conferring mutation is S315T resulting from the nucleotide substitution G944C of the *katG*, which results in a *katG* product with an altered conformation. The abnormal *katG* product is unable to activate INH a requirement for the formation of INH-NAD adducts responsible for INH antimicrobial activity. Nevertheless, the mutant *katG* product is still able to protect the bacteria from the reactive oxygen and nitrogen species in the cytosol of the infected macrophages (Vilchèze & Jacobs, 2007).

Mutations associated with *inhA* and its promoter region cause not only resistance to isoniazid, but also structurally related second-line drug Ethionamide (Banerjee *et al.*, 1994). Isoniazid resistance associated *inhApro* mutations result in the up-regulation of the *inhA* gene causing the overproduction of the *inhA* gene product (NADH-dependent *enoyl-ACP reductase*) the target of INH, hence amounting to INH resistance (Baker *et al.*,

2005). The most common *inhA* mutations are those that occur at the promoter region -15C/T and -8T/C which frequently identified among INH resistant MTBC strains mostly associated with low level INH resistance among INH mono-resistant strains (Homolka *et al.*, 2010; Leung *et al.*, 2006). Other mutations in the *inhApro* (-16A/G, -17G/C and -24G/T) have been found in several INH resistant isolates however, their direct involvement with resistance is not clearly understood (Flandrois, Lina, & Dumitrescu, 2014; Homolka *et al.*, 2010). In Ghana, mutations (-47G/C and -102G/C) in addition to the afore mentioned have been detected in *Maf* isolates that are resistant to only INH with either *katG* S315T or *inhApro* -15C/T and hence believed to be phylogenetic markers specific for *Maf* strains (Homolka *et al.*, 2010).

In MTBC, the gene *ahpC* encoding an *alkyl hydroperoxidase reductase* is associated with detoxification of reactive oxygen and nitrogen species or intermediates. Mutations in the promoter of *ahpC* (the enzyme is co-expressed with *oxyR* for sharing the same promoter region) have been used as proxy markers for the detection of isoniazid resistance (Rinder *et al.*, 1998). Nevertheless, other studies had found the up-regulation of the *ahpC* as a more compensatory mechanism for the loss of *katG* activity rather than the basis for INH resistance because mutations accounting for the up-regulation of the expression of *ahpC* always accompany either the *katG* S315T or deletion of *katG* (Costa *et al.*, 2009; Sherman *et al.*, 1996). Even though the *alkyl hydroperoxidase reductase* is less efficient with detoxification compared to the *catalase peroxidase* encoded by the *katG*, the over production ensures that the affected bacterium is inundated with enough for detoxification even though *katG* activity may be absent or defective (Costa *et al.*, 2009).

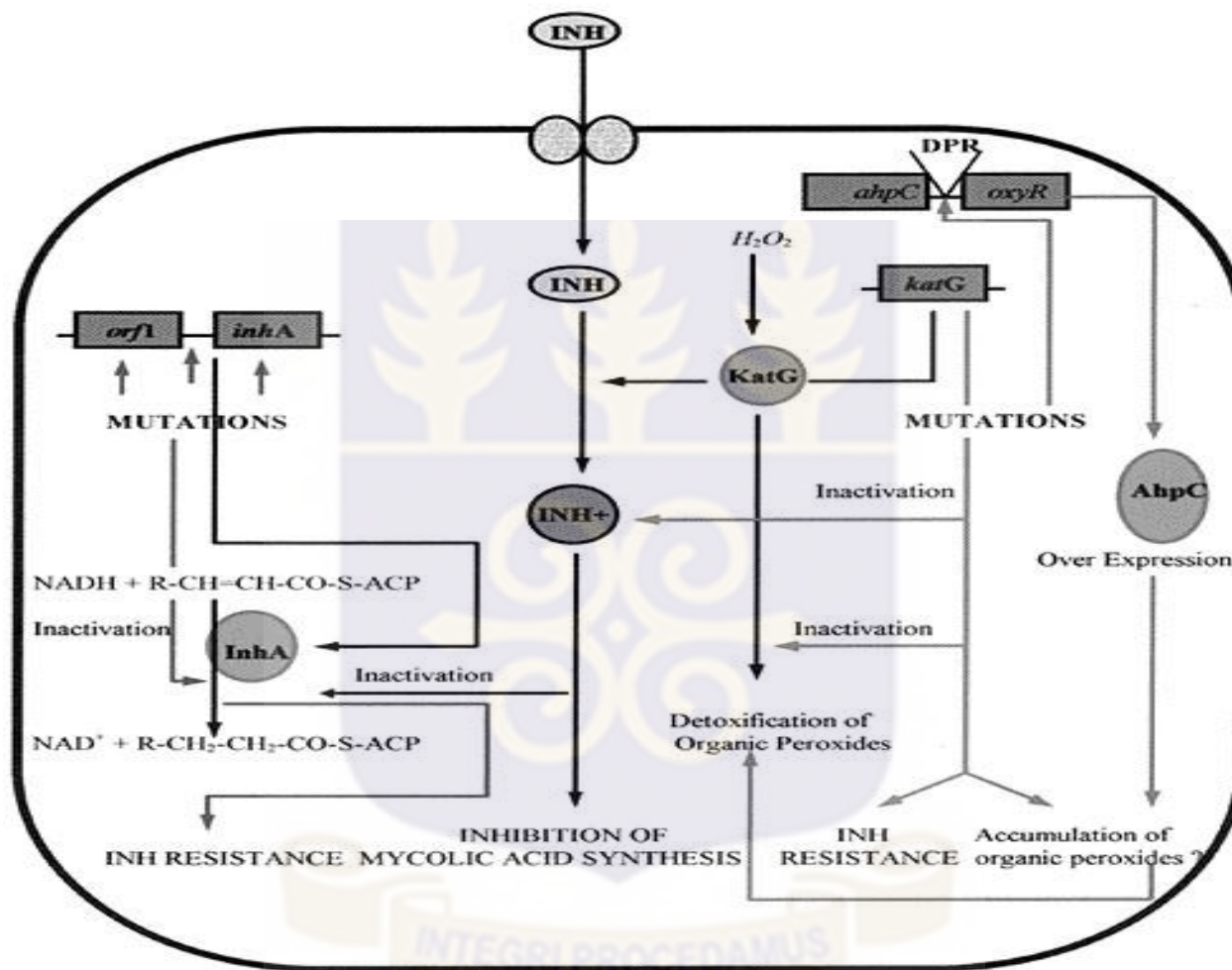


Figure 2.9: Mechanism of action and mode of resistance of INH. (Adapted and modified from Rattan *et al.*, 1998).

Mutations such as -48G/A, -51G/A, -54C/T, -74G/A and -81C/T have been identified in the promoter region of *ahpC* which have been found to confer over-expression of *alkyl hydroperoxidase reductase* in some INH resistant isolates (Santos, 2012).

The last among the genetic elements associated with INH resistance is *ndh* gene encoding an *NADH dehydrogenase* that is bound to the active site of NADH-dependent *enoyl-ACP reductase* to form the ternary complex with activated INH has been implicated with resistance to INH in some MTBC species (Lee *et al.*, 2001; Santos, 2012). In one study using *Mycobacterium bovis* and *Mycobacterium smegmatis*, it was found that about 9.5% of INH resistant isolates showed point mutations in the *ndh* gene at codons 110 and 168 and these mutations were detectable in only INH resistant isolates (Lee *et al.*, 2001). In those species, it was found that mutations in the *ndh* gene led to substantial increase in the NADH/NAD ratio. The rise in NADH/NAD ratio was shown to protect the *NADH-dependent enoyl-ACP reductase* from inhibition by INH-NAD adduct (Vilchèze *et al.*, 2005). Thus, mutation(s) with the *ndh* leads to defects in the NADH dehydrogenase which subsequently results in reduction of the rate of oxidation of NADH to NAD and hence NADH accumulation and NAD depletion (Lee *et al.*, 2001). The occurrence appears to protect the bacteria from INH-mediated toxicity by inhibiting the binding of the INH-NAD adduct to the active site of the NADH-dependent *enoyl-ACP reductase* (Hazbon *et al.*, 2006; Lee *et al.*, 2001; Vilchèze *et al.*, 2005). An interesting observation is that *ndh* mutations give cross resistance to ethionamide (ETH) a structural analogue of INH (Fig 2.10), which is also used as anti-tuberculosis drug as part of combination therapy for treating MDR-TB.

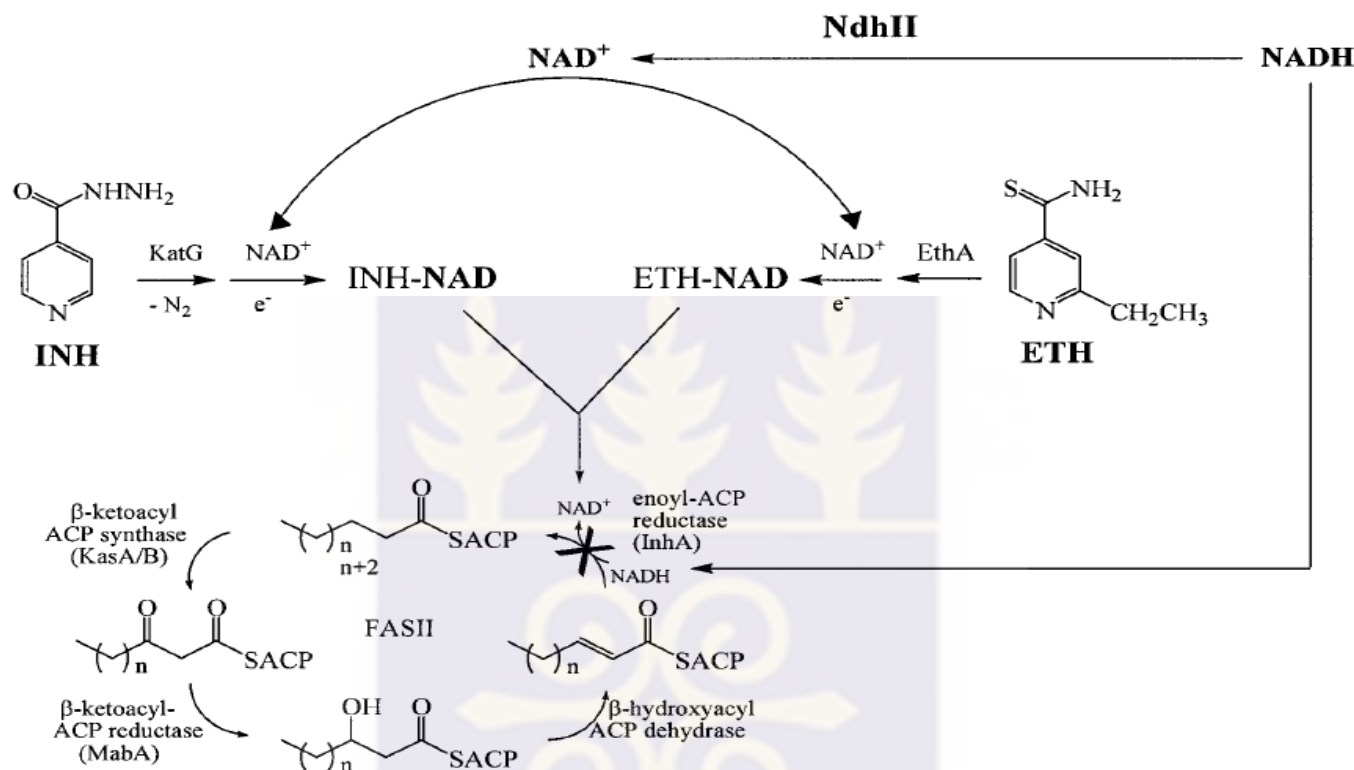


Figure 2.10: Proposed mechanism of action of INH and ETH.

INH and ETH are both pro-drugs that are activated by the catalase-peroxidase *katG* or the monooxygenase *ethA*, respectively. Activated INH or ETH reacts with NAD to form an INH-NAD or ETH-NAD adduct which inhibits the common target *InhA* required for NADH-dependent enoyl-ACP reductase of the fatty acid synthase type II system. This inhibition thence halts mycolic acid biosynthesis and results in cell lysis. A novel mechanism of co-resistance to INH and ETH is by mutations in *ndh*, which increase the intracellular concentration of NADH and cause resistance by competitively inhibiting the binding of the INH-NAD or ETH-NAD adduct to *InhA*). (Adapted from Vilchèze *et al.*, 2005)

2.6.4.2 Rifampicin Resistance

Rifampicin is a complex lipophilic macro-cyclic broad-spectrum bactericidal antibiotic to both actively dividing and dormant bacilli. It inhibits prokaryotic DNA-dependent RNA polymerase by binding to the β -subunit of the polymerase and hence prevents subsequent protein synthesis. It is one of the widely used anti-tubercular agents and together with INH form the backbone of the current first line treatment regimen (Blanchard, 1996; Rattan *et al.*, 1998). Resistance of clinical *MTBC* isolates to RIF is mostly due to mutations in the *rpoB* gene that encodes the β -subunit of the prokaryotic DNA-dependent RNA polymerase which leads to conformational changes in the resulting protein that reduce the affinity of the protein for binding to the drug (Telenti *et al.*, 1993).

Mutations within an 81bp region of the β -subunit the RNA polymerase called rifampicin resistance determining region (RRDR) are mostly associated with RIF resistance (Homolka *et al.*, 2010; Telenti *et al.*, 1993). The RRDR region spans codons 426-452 corresponding to 507-533 when referring to the *E. coli* reference. The most common mutations are those found in codons 445 and 450 corresponding to *E. coli* loci 526 and 531 respectively and are considered high resistance RIF mutations (minimal inhibitory concentration; MIC > 32 ug/mL) (Pang *et al.*, 2013). However, not all mutations within this region lead to such high level RIF resistance. Mutations in codons 430, 435, 437, and 441 respectively corresponding to the *E. coli* equivalent loci 511, 516, 518 and 522 lead to low level RIF resistance (MIC < 30 ug/mL) (Moghazeh *et al.*, 1996; Ohno *et al.*, 1996). Over 96% of all RIF resistance are attributed to these genetic alterations in the RRDR (Somoskovi *et al.*, 2001) hence other mutations in codons outside the

RRDR of the β -subunit of the prokaryotic DNA-dependent RNA polymerase or other genes may also account for some of the RIF resistant TB cases which require further research. There is evidence implicating the genetic background of the infecting bacilli and/or the geographical source of the bacilli on the kind of *rpoB* mutations (Comas *et al.*, 2011; Santos, 2012).

Remarkably, low level RIF resistance is known to be an innate property of *M. smegmatis* and other fast growing *Mycobacterium species* which do not harbour any *rpoB* mutations (Hetherington *et al.*, 1995). In these bacteria, resistance of RIF has been attributed to ADP-ribosylation (Quan *et al.*, 1997). The ADP-ribosyl transferase is encoded by the *arr-ms* gene which catalyses the transfer of an ADP-ribose unit to a susceptible amino acid residue on the target protein with the loss of a nicotinamide molecule (Baysarowich *et al.*, 2008). Interestingly, this gene is absent in the genome of the MTBC hence probably accounting for the general RIF susceptibility among the members of the complex. This observation shows that other genetic elements may contribute somehow to drug resistance TB even though this contribution may be relatively insignificant in comparison to *rpoB* mutations.

In addition to the *rpoB* gene, nsSNPs within the *rpoA* and the *rpoC* genes encoding the α and β' subunits of the RNA polymerase respectively have also been associated with RIF resistant strains of the MTBC. Non-synonymous SNPs within these genes affect the conformation of the respective subunits hence affecting the 3D conformation of the entire polymerase. However, almost all the strains with these mutations also harbour at least one *rpoB* mutation known to confer RIF resistance. Comas *et al* therefore proposed that non-synonymous *rpoA* and *rpoC*

mutations occurring in only RIF resistant MTBC may be mutations that restore the fitness cost associated with the acquisition of RIF resistance conferring *rpoB* mutations hence are compensatory and would not on their own confer drug resistance (Comas *et al.*, 2011). Indeed, it has been shown that non-synonymous SNPs in these genes actually localize at the interface between the subunits (Fig 2.11) indicating the importance of these mutations. Codons 183 and 187 of *rpoA* and 483, 484, 485, 491, 517 and 698 of *rpoC* have been predicted to be hot-spots for compensatory RIF resistance mutations (Comas *et al.*, 2011).

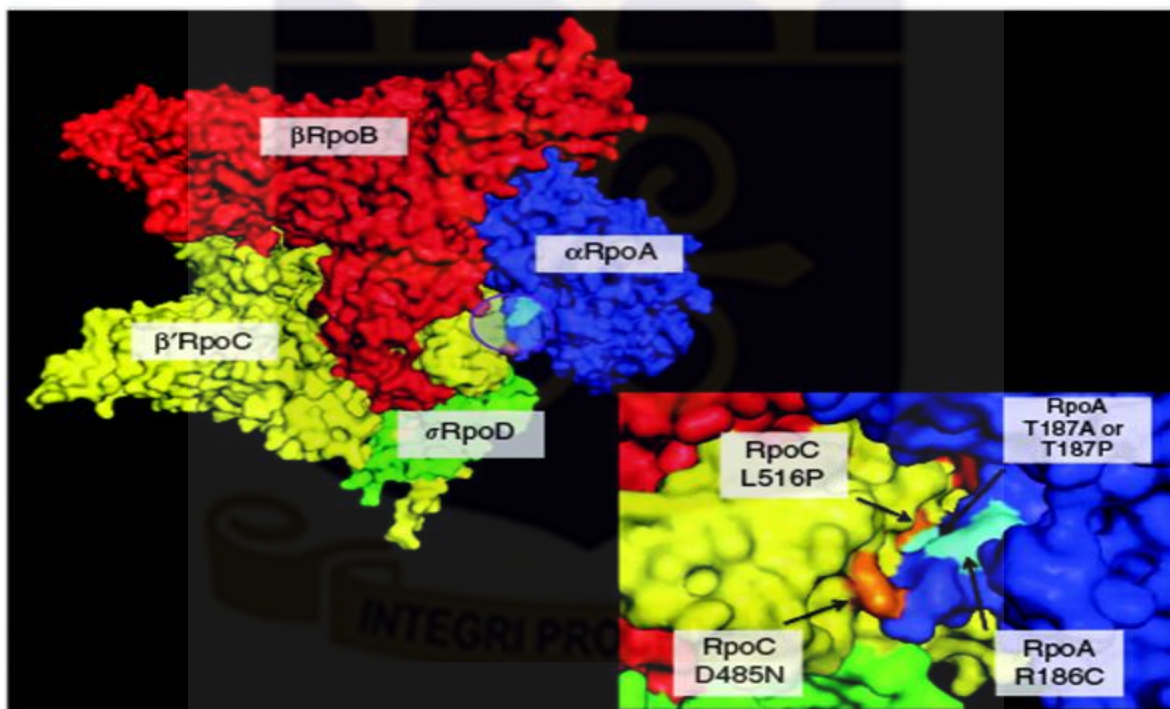


Figure 2.11: Putative compensatory mutations in *rpoA* and *rpoC*.

The compensatory mutations fall in regions encoding the interface of the RNA polymerase subunits. The alterations are localized to residues of RpoA (light blue) and RpoC (orange) that are predicted to have roles in RNA polymerase subunit interaction. Residue numbers are indicated according to M. tuberculosis coordinates. RpoA (α subunit), blue; RpoB (β subunit), red; RpoC (β' subunit), yellow; RpoD (σ subunit), green) (Adapted from Comas et al., 2011).

2.7 THE GENOME OF THE MTBC

Prior to the late 1990s, very little was known about the genetics of the MTBC. However, the successful completion and publishing of the first whole genome of a candidate *Mtbc* strain H37Rv (Cole *et al.*, 1998) has led to the deciphering of the intricacies of the genome of the MTBC. Annotation of the sequenced H37Rv genome reveals genome size of 4.4 Mbp containing approximately 4,000 genes (Fig 2.12). Among these genes were those encoding proteins involved with lipid metabolism (225; 5.7%), information pathways (207; 5.2%), intermediate metabolism and respiration (877; 22.0%), cell wall and cell processes (517; 13.0%), regulation (188; 4.7%), virulence, detoxification and adaptation (91; 2.3%), supposedly conserved hypothetical functions (911; 22.9%) and unknown functions (607; 15.3%). In addition genes encoding stable RNAs (50; 1.3%), insertion elements and remnants of bacteriophages (137; 3.4%) and those associated with PE (Pro-Glu) and PPE (Pro-Pro-Glu) metabolism and respiration (877; 22.0%) were also identified.

One unique feature gleaned from this genome is the amount of genes involved with fatty acid metabolism (200 genes) especially those involved with β -oxidation of fatty acids (over 100 genes) compared to 50 genes found to perform the same function in *E. coli*. This large number of the MTBC enzymes dedicated to fatty acid catabolism enhances its ability to grow in the tissues of the infected host, where fatty acids are the major source of carbon (Cole *et al.*, 1998). Secondly, the presence of glycine rich proteins of the PE (104) and PPE (68) families is very unique and similar levels of these acidic proteins have been found in other members of the MTBC. Even though the actual function of these PE/PPE genes are not clearly understood, the

association of similar genes in *M. marinum* with virulence and reported antigenicity of some localized PE subfamily of proteins called PGRS (polymorphic GC-rich repetitive sequence) with conserved PE domain followed by Gly-Gly-Ala or Gly-Gly-Asn call for a critical look at these genes. Perhaps one unique feature of all the genomes of the MTBC sequenced to date is the evidence of no horizontal gene transfer since the pathogens evolved for intracellular adaptation, which has contributed to the clonal nature of members of the complex (Ioerger *et al.*, 2010).

However, there is still enough diversity within the complex that could account for the different host adaptations, disease phenotypes and interventions (Coscolla & Gagneux, 2014; Coscolla & Gagneux, 2010). For instance, analysis of the first *Maf* whole genome sequence (L6 strain GM041182) revealed the presence of a unique sequence RD900 encoding a protein involved with trans-membrane transportation of macromolecules. This sequence is also present in all so-called ‘ancient’ lineages of the MTBC however; it is independently lost in all ‘modern’ lineages of *Mtbss* and *M. bovis*. In addition, The *Maf* genome however shares a number of uniquely lost genes (including genes for biosynthesis of some vitamins) due to generation of pseudogenes, a property it shares with *M. bovis* but not *Mtbss*. In addition, the *Maf* L6 genome has an intact copy of a gene capable of increasing its susceptibility to antibiotics that are not active against *Mtbss*. Comparative genomics analysis of the available genomic data of MTBC strains shows events of extensive pseudogene, an indication of on-going genome erosion linked to intracellular lifestyle (Bentley *et al.*, 2012). In that same line, all classical human MTBC strains (*Mtbss* and *Maf*) have a conserved *mpt40* gene which is missing from the genomes of all classical animal adapted MTBC such as *M. bovis* an indication of specific host adaptation (Koivula *et al.*, 2002) a

discovery that has been incorporated into a rapid nested assay for differential diagnosis of TB (Sinha *et al.*, 2017).

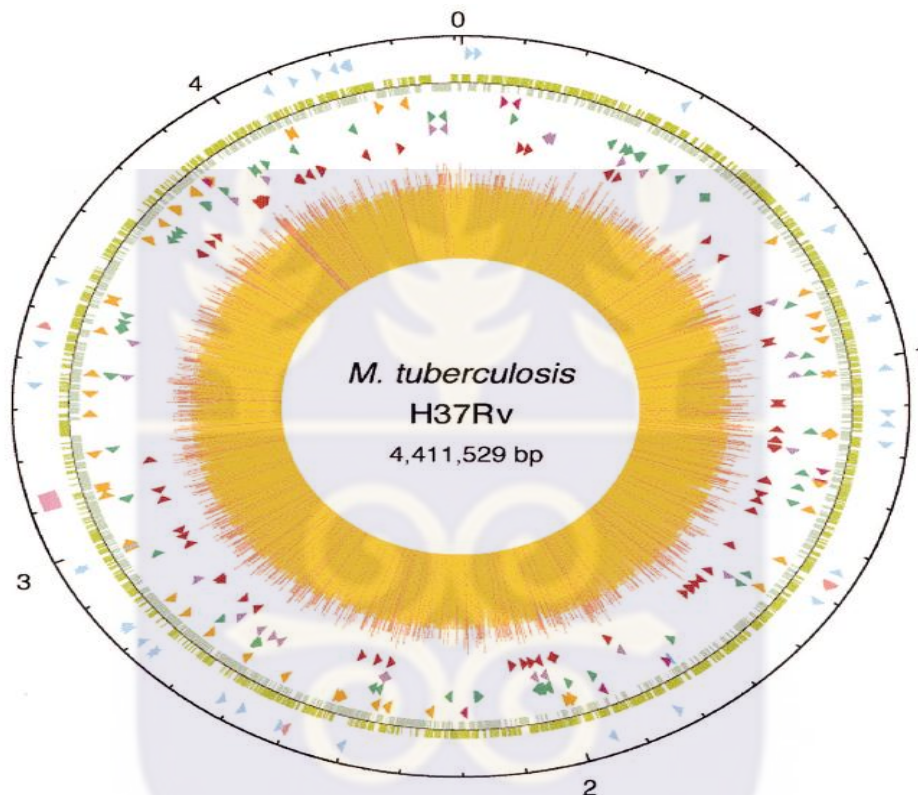


Figure 2.12: Circular map of the chromosome of *Mtbs* H37Rv.

The outer circle shows the scale in megabases. 0 represents the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes including tRNAs (blue) and others (pink). There is also the direct-repeat region (pink cube). The second ring shows the coding sequence by strand (clockwise, dark green; anticlockwise, light green). The third ring depicts repetitive DNA (insertion sequences, orange; 13E12 REP family, dark pink; prophage, blue). The fourth ring shows the positions of the PPE family members (green) whereas the fifth ring shows the positions of the PE family members (purple, excluding PGRS). The sixth ring shows the positions of the PGRS sequences (dark red). The histogram (center) represents the GC content, < 65% GC in yellow and >65% in brown. (Modified from Cole *et al.* 1998)

2.8 GENOMIC DIVERSITY WITHIN THE MTBC

The MTBC had over the years been described as genetically monomorphic bacteria with a high level of genomic sequence similarity over 99.95% and a clonal population structure (Comas *et al.*, 2013, 2009; Sreevatsan *et al.*, 1997). Therefore, little attention was paid to strain to strain variations albeit there were disparities in virulence of different collection of the bacteria (Homolka & Niemann, 2016; Mihret *et al.*, 2012; Portevin *et al.*, 2011). It is widely established that there is no/low recombination (horizontal gene transfer) within the MTBC unlike its closely related smooth tubercle bacilli, *M. canettii* (Blouin *et al.*, 2014; Comas *et al.*, 2009; Coscolla & Gagneux, 2014; Supply *et al.*, 2013). The variable outcome of TB infections was thence attributed to host and environmental factors to the almost total neglect of strain to strain diversity even though there were phenotypic differences (such as macroscopic morphology, niacin production, inhibition by PNZ and the ability to grow on glycerol containing media) between different identified lineages of the MTBC (Collins *et al.*, 1982; Niemann *et al.*, 2004).

Therefore, global interventions put across for prevention and control of TB were largely based on limited and biased collection of isolates which may not have been representative enough for the global MTBC burden. However, this notion of genetic homogeneity relied on the findings from early DNA sequencing studies that were conducted using limited and/or biased strain collections and less discriminatory genotyping tools (Sreevatsan *et al.*, 1997; van Embden *et al.*, 1993). With the advancement in molecular biology, more data are accumulating indicating and already point to substantial strain diversity among the different members of MTBC at the genomic level which could have functional

implications (Gehre *et al.*, 2013; Krishnan *et al.*, 2011; Sarkar *et al.*, 2012)

Genomic variability within the MTBC have been mostly associated with genetic events such as polymorphisms within insertion sequences 6110 (IS6110), substitution of nucleotides (single nucleotide polymorphisms; SNPs) insertion and/or deletion of single nucleotides (INDELS) or large sequence deletions (leading to large sequence polymorphism; LSPs) and duplication of genetic elements in tandem or interspersed (variable number of tandem repeats; VNTRs) (Groenheit *et al.*, 2011; Mahairas *et al.*, 1996; Ozcaglar *et al.*, 2011).

2.8.1 Repetitive and Mobile Genetic Elements

Repetitive elements such as clustered regularly interspaced short palindromic repeats (CRISPRs), variable number of tandem repeats (VNTRs) and mobile genetic elements such as insertion sequences (ISs) especially IS6110 are major source of genetic variability within the MTBC. They are respectively used as molecular markers for MTBC strain genotyping in assays such as spoligotyping (spacer oligonucleotide typing which detects the presence or absence of 43 different spacers within the direct repeat region of the MTBC genome), *Mycobacterium* interspersed repetitive units (MIRU) VNTR typing (detects variations in the number of tandem repeats) (Mokrousov *et al.*, 2004) and restriction fragment length polymorphisms within IS6110 (IS6110-RFLP). These elements are an important source of MTBC genomic polymorphisms and this variability is used in identifying individual strains (spoligotyping) and tracing outbreaks (MIRU-

VNTR typing). For instance, a typical spoligotype of an *Mtbss* strain would have most of the spacers between 33 and 36 deleted compared to a *Maf* strain which would have most of them present if not all. However, MIRU-VNTR and spoligotyping are limited by homoplasy. *IS6110*-RFLP is more discriminatory with variable copy numbers, mostly greater 10 in *Mtbss* and less than 10 in *Maf* and sometimes complete absence in some strains) make it unreliable and not reproducible (Niemann *et al.*, 2002; Viana-niero *et al.*, 2001).

2.8.2 Large Sequence Polymorphisms (LSPs)

Large genomic deletions which results in LSPs are integral source of genomic diversity within the MTBC and are the results of genomic deletions resulting from transposition of mobile genetic elements, insertion of insertion sequences and/or homologous recombination between adjacent insertion sequences leading to the loss of intervening DNA segment (Behr *et al.*, 1999). Studies such as microarray comparative hybridizations, subtractive genomic hybridization were used to identify a number of genomic deletions specific to certain MTBC species (Behr *et al.*, 1999; Gordon *et al.*, 1999; Mahairas *et al.*, 1996) such as RD1 deleted in *M. bovis* BCG, RD9 (deleted in *M. bovis* and *Maf*). In addition to these species-specific deletions, there are also sub-species level deletions such as deletion of RD711 in *Maf* L5, deletion of RD702 in *Maf* L6, deletions of RD105, RD181 and RD142 in Beijing strains of *Mtbss* (Gagneux *et al.*, 2006; Yeboah-manu *et al.*, 2011). LSPs are less prone to homoplasy and are more reliable for identification of individual members of the MTBC.

Using a combination of comparative genome hybridization and multi-locus sequence analyses, human-adapted MTBC isolates from 80 countries, were divided into six main lineages; 4 classified under *Mycobacterium tuberculosis* sensu stricto (*Mtbss*) and 2 under *M. africanum* (*Maf*). Subsequent mapping of origins of these strains explicitly revealed a strong bio-geographical structure [Fig 2.5] (Gagneux, 2012; Gagneux, DeRiemer, *et al.*, 2006). Another *Mtbss* lineage designated Lineage 7 (the yellow lineage) has recently been isolated from humans. However, this lineage unlike the ubiquitous distribution of the other *Mtbss* lineages is geographically restricted to the horn of Africa (Blouin *et al.*, 2012; Firdessa *et al.*, 2013). Analysis of these human adapted MTBC shows genetically distinct groups with evidence of different independent evolutionary paths from each other as from the animal-adapted forms of MTBC (Comas *et al.*, 2013).

2.8.3 Single Nucleotide Polymorphisms (SNPs)

Hershberg and colleagues in 2008, working with 89 genes in 108 strains, showed that an average pair of MTBC strains would have approximately 300 functional differences which further bolster the argument that the genetic diversity in the MTBC had been underestimated over the years (Hershberg *et al.*, 2008). However, the maximum genetic distances between any two human adapted MTBC and between human adapted and any animal strain have been estimated to be respectively 1,800 and 2,400 SNPs from a current study using WGS analysis of approximately 260 global collection of diversified MTBC strains (Comas *et al.*, 2013).

This diversity compared to the 9000 SNPs identified among a group of *M. canettii* strains (Blouin *et al.*, 2014; Supply *et al.*, 2013) further emphasizes the highly clonal nature of MTBC. Nevertheless, unique branches with exclusive phenotypes and biological associations do exist within this complex of highly mono-clonal bacteria as shown in the phylogeny of global collection of 283 MTBC strains (Fig 2.13). From this phylogeny, the more ancestral strains (L1, L5, L6 and L7) are either limited to Africa (L5 and L6 to West Africa and L7 to East Africa) or concentrated in Africa (L1 in East Africa) compared to Asian (L2 and L3) and global (L4) distribution of the designated modern lineages (Fig 2.13). Even though L4 is the most ubiquitously distributed among the MTBC, it exhibits geographical restrictions at the sub-lineage level (Fig 2.13) evidenced by the restriction of sub-lineages L4.1 (Ghana sub-lineage) and L4.6.2 (Cameroon sub-lineage) to West Africa and sub-lineage L4.1.2 (sub-lineage Haarlem) in Europe (Asante-Poku *et al.*, 2015, 2016; Brudey *et al.*, 2006; Homolka *et al.*, 2012; Niobe-Eyangoh *et al.*, 2003). This distribution of the MTBC is in favour of the African origin of MTBC and possible coevolution of the pathogen with the host. This theory is further bolstered by the sympatric associations between genetically distinct strains and hosts which are mostly distorted with immune-compromised hosts leading to allopatric associations (Fenner *et al.*, 2013; Gagneux *et al.*, 2006). Notwithstanding the observed diversity within the MTBC, the classification of Lineages 1 to 4 and 7 as *Mtbss* and Lineages 5 and 6 as *Maf* species may have to change because current evidence from whole genome analysis have shown that the average nucleotide identity (ANI) between any two lineages including the animal strains is always greater than 96% which is the cut-off ANI value for different species (Winglee *et al.*, 2016).

These genomic revelations are in line with earlier observations made by taxonomist using adansonian classification, immuno-diffusion analysis and DNA homology (Baess, 1979; Bradley, 1972; David *et al.*, 1978; Stanford *et al.*, 1974; Tsukamura, 1976).

2.9 PHENOTYPIC CONSEQUENCES OF GENOMIC DIVERSITY

The consequence of genomic diversity of bacterial pathogens such as *Escherichia coli*, *Neisseria meningitidis*, *Haemophilus influenzae*, *Bordetella* and *Streptococcus* species is well documented (Bogaert *et al.*, 2004; Buboltz *et al.*, 2009; Lipsitch, 1999; Madoff *et al.*, 1996; Schouls *et al.*, 2008; Spratt *et al.*, 2000). In these species, some strains are more likely to cause invasive disease than others because such virulent phenotypes pose distinct virulent markers such as genes associated with the production of toxins (Buboltz *et al.*, 2009). However, no such genetic marker is identified for the MTBC. Though, infection with the MTBC results in a range of clinical phenotypes ranging from asymptomatic infection through localized diseased lungs to different forms of disseminated disease.

It is likely that diverse virulence factors may be presented by distinct lineages/strains of MTBC, which then influence the immune response that will be elicited by the host, and subsequently the outcome of the host-pathogen interaction (Forrellad *et al.*, 2013; Nunes-Alves *et al.*, 2014). This assertion has been supported by results from *in vivo* animal model studies which involving different MTBC lineages.

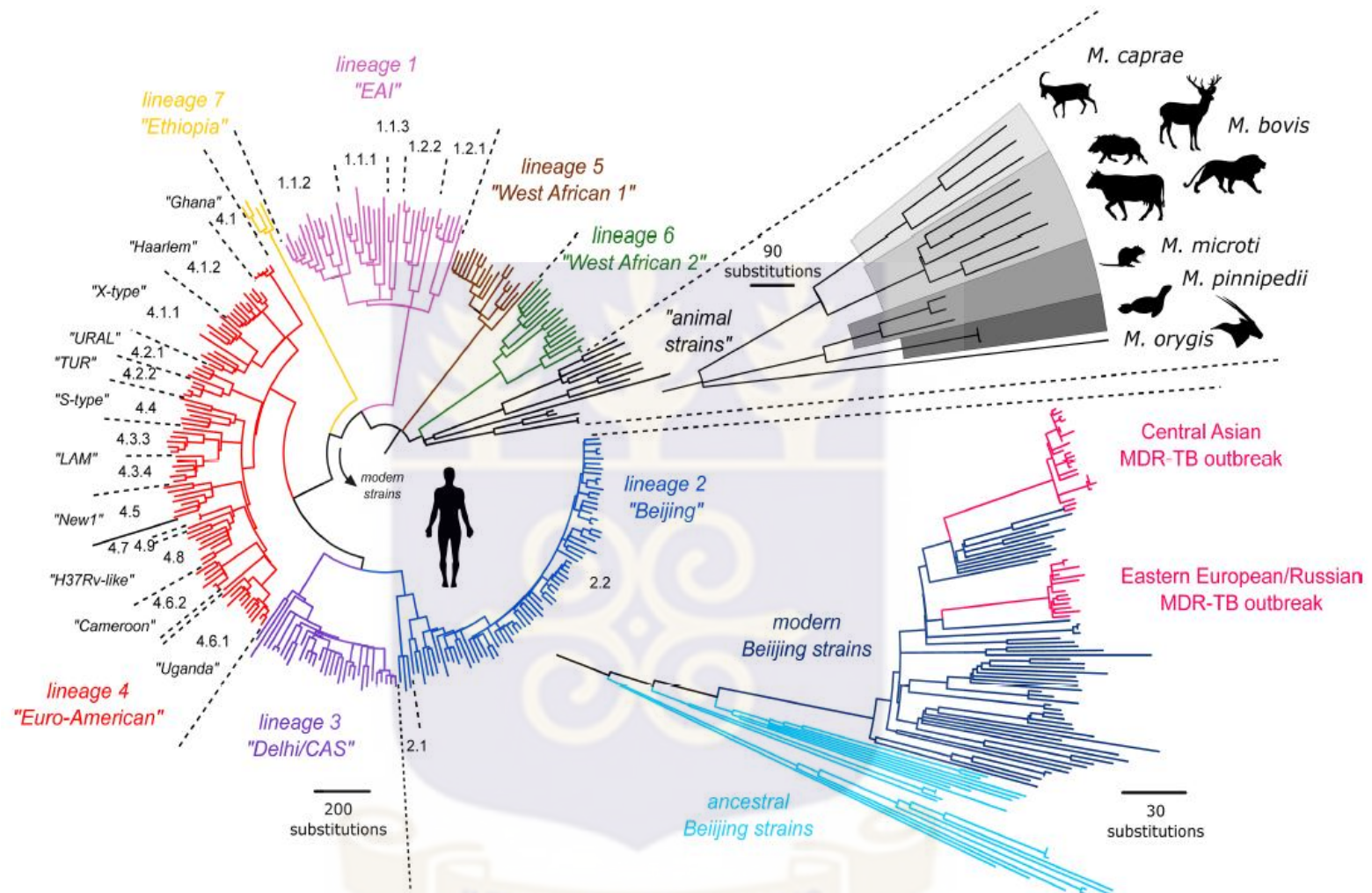


Figure 2.13: Global phylogeny of clinical and veterinary MTBC isolates.

The phylogeny illustrates the higher diversity at the whole genome level. (Adapted from (Niemann et al., 2016)).

For example it has been found that some strains of TB bacilli of the L2 progress faster to active disease compared to those of L6 (de Jong *et al.*, 2008). In addition, experiments involving candidates of L1 and L4 found L4 to progress faster compared to L1 (Portevin *et al.*, 2011).

Additional support for strain virulence variation is hyper-virulence of the Beijing family of *Mtbss* L2 and the attenuated nature of the *M. bovis* BCG due to the deletion of RD1 (Frota *et al.*, 2004; López *et al.*, 2003). Even within the Beijing family, there is indication that there exist phenotypic differences among strains of the same lineage as evidence by the variable transmissibility of sub-lineages of the Beijing strains (Kato-Maeda *et al.*, 2010; Kato-Maeda *et al.*, 2012; Niemann *et al.*, 2010). This suggests that competitive advantage of the MTBC might not necessarily be fixed at the lineage level but rather strain level. In spite of these indications, how MTBC genomic diversity influences disease presentation remains an open question.

Clinical studies comparing *Mtbss* and *Maf* (mostly Lineage 6), have shown that infection of *Mtbss* is associated with early progression to active pulmonary TB as compared to *Maf* hence suggesting that *Mtbss* is more virulent compared to *Maf* (Bold *et al.*, 2012; de Jong *et al.*, 2010; Gehre *et al.*, 2013; Meyer *et al.*, 2008). Lineage and strain-specific mutations in genes of functional categories such as intermediate metabolism and respiration, cell wall and cell processes, lipid metabolism, regulatory proteins, information pathways and virulence, detoxification and adaptation could be responsible for the differential presentation of infection by different Lineages. Interestingly, comparative target gene

sequence analysis of candidate MTBC strains have shown that approximately two-thirds of all SNPs in coding genes are non-synonymous SNPs (nsSNPs) thus emphasizing the potential implications of even the limited genomic diversity therein compared to other bacteria (Fleischmann *et al.*, 2002; Garnier *et al.*, 2003; Hershberg *et al.*, 2008; Sreevatsan *et al.*, 1997). Other than SNPs, the reductive evolution of the MTBC involving the genomic deletion of blocks of specific genomic regions including phylogenetic markers can involve blocks of diverse functional genes involved in either immunogenicity or host evasion (Gagneux & Small, 2007; Mostowy *et al.*, 2004; Tsolaki *et al.*, 2004). For instance, even though *M. microti* is virulent in voles, it is virtually harmless in humans an observation that has been attributed to the deletion of RD1 an even similar to the generation of the widely used BCG vaccine from continuous *in vitro* passage of *M. bovis*. RD1 encodes the *esx1* locus which is responsible for the type VII secretory system driving the secretion of germane T cell antigens ESAT6 and CFP-10 required for virulence (Guinn *et al.*, 2004; Pym *et al.*, 2002; Stanley *et al.*, 2003).

A very large proportion of the MTBC genome (250 genes) as compared to *E. coli* (50 genes) is devoted to lipid metabolism albeit having approximately equal genome size (Neyrolles & Guilhot, 2011). The cell wall lipids play essential role in pathogenesis making the encoding genes potential virulence elements that may be exploited in drug/vaccine design and development (Forrellad *et al.*, 2013). Experiments with mutant *Mtbss* strains defective in one of the 2 β -keto acyl synthetases of the FASII system encoded by the *kasB* gene though not too essential (unlike *casA* gene) for the survival of the bacilli, has shown that the bacilli produced mycolates with relatively shorter lengths

compared to the wild-type. In addition, the bacilli persisted in infected experimental mice for up to 600 days without causing disease or mortality as compared to the wild-type strains (Bhatt *et al.*, 2008; Bhatt *et al.*, 2007). Additional studies targeting other genes such as (*mmaA4*, *pcaA*, *fbpA*, *Rv2869c* and *mymA*) involved in the synthesis of mycolic acids have also implicated the mycobacterial cell wall in virulence (Barkan *et al.*, 2012; Belisle *et al.*, 1997; Glickman *et al.*, 2000; Makinoshima & Glickman, 2005; Manganeli *et al.*, 2002; Rao *et al.*, 2005).

The MTBC uses cholesterol as source of carbon for energy and biosynthesis of the virulence-associated lipid PDIM (Griffin *et al.*, 2012; Pandey & Sassetti, 2008). With the avalanche of cholesterol in the host cells, the MTBC has unlimited source of carbon. The very survival of the bacilli in the host is shown to depend on the ability to catabolize cholesterol (Miner *et al.*, 2009; Pandey & Sassetti, 2008). Mutations in some of the genes (*choD*, *hsaC* and *mce4* operon) involved with cholesterol catabolism have been shown to attenuate the bacilli in experimental animal models (Yang *et al.*, 2011) and can even be lethal due to accumulation of catabolic intermediates (Brzostek *et al.*, 2007; Yang *et al.*, 2011).

Different lineage and sub-lineages of MTBC differ qualitatively and/or quantitatively in the modulation of the host immune response. For example, the hyper-virulence of *Mtbss* strain HN878 of L2 has been linked to the production of phenolic glycolipids (PGL) with immune-modulatory effects. Studies in human monocytes and mice have also shown that unlike other strains which up-regulate expression of pro-inflammatory cytokines including TNF- α , IL-1 β , IL12, and IFN- γ , strain HN878 down-regulate expression of

such pro-inflammatory cytokines and rather induce an elevated production of macrophage deactivating cytokines such as IL-11 and IL-13 (Sinsimer *et al.*, 2008). In fact, the production of PGL was shown to vary among different MTBC lineages which could have dire consequence on the disease presentation.



CHAPTER THREE

MATERIALS AND METHODS

3.1 MATERIALS

3.1.1 Reagents and Consumables

96-well micro-plates (Nunc International, Rochester, New York, USA), glycerol [Sigma-Aldrich, Steinheim, Germany], casitone [Difco, Detroit, Michigan USA], Tween 80 [Sigma-Aldrich, Steinheim, Germany], alamar blue reagent [Trek Diagnostic Systems, Ohio, USA], OADC (oleate-albumin-dextrose-catalase) [Difco, Detroit, Michigan USA], Middlebrook 7H9 [Difco, Detroit, Michigan USA], 50 mL falcon tubes [greiner bio-one, Frickenhausen, Germany], transfer pipettes [alpha laboratories, Hampshire, UK], Media tubes [Corning Life Sciences, New York, USA], 2 mL eppendorf tubes [Sarstedt AG & Co, Nümbrecht, Germany], nuclease-free water [Sigma-Aldrich, Steinheim, Germany], proteinase K [Qiagen, Hilden, Germany], lysozyme [Sigma-Aldrich, St Louis USA], Isoniazid [Sigma-Aldrich, St Louis, USA], Rifampicin [Sigma-Aldrich, St Louis, USA], Dimethylsulphoxide (DMSO) [Sigma-Aldrich, St Louis, USA], Molecular grade absolute ethanol [Sigma-Aldrich, St Louis USA], molecular grade isopropanol [Sigma-Aldrich, St Louis USA], chloroform isoamyl alcohol [Sigma-Aldrich, St Louis, USA], Cetyl trimethylammonium bromide (CTAB) [Sigma-Aldrich, St Louis, USA], SDS [Sigma-Aldrich, St Louis USA], Ethylenediaminetetraacetic acid (EDTA) [Sigma-Aldrich, St Louis USA], Trisma base [Sigma-Aldrich, St Louis USA], ethidium bromide [Sigma-Aldrich, St Louis USA], DNA ladder [Promega corporation, Madison, USA,], Boric acid [Sigma-Aldrich, St Louis USA], Phosphate buffered saline (PBS) [Invitrogen,

Massachusetts, USA], micro pipettes [Thermo Scientific, Vantaa, Finland], pipette tips [Biotix Inc., San Diego, USA], McFarland standards [BD and company, CA, USA], PCR tubes [greiner bio-one, Kremsmünster, Austria], 10X PCR buffer with coryl dye [Qiagen, Hilden, Germany], 5X Q solution [Qiagen, Hilden, Germany], 2.5 mg/mL Monopotassium phosphate [Sigma-Aldrich, Steinheim, Germany], Magnesium chloride [Qiagen, Hilden, Germany], Magnesium sulphate heptahydrate [Qiagen, Hilden, Germany] molecular grade nuclease free water [Qiagen, Hilden, Germany], dNTPs [Qiagen, Hilden, Germany], Hot-start Taq polymerase [Qiagen, Hilden, Germany], Agarose [Sigma Aldrich, Steinheim, Germany], PCR primers [Macrogen, Amsterdam, Netherlands], inoculating loops [medical wire and equipment Co. Ltd., Wiltshire, UK], Oxalic acid [Sigma-Aldrich, St Louis, USA], Phenol [Sigma-Aldrich, St Louis USA], Sodium chloride [Sigma-Aldrich, Steinheim, Germany], Malachite green [Sigma-Aldrich, Steinheim, Germany], Magnesium citrate [Sigma-Aldrich, Steinheim, Germany], Sodium pyruvate [Sigma-Aldrich, Steinheim, Germany], Asparagine [Sigma-Aldrich, Steinheim, Germany], powder-free nitrile gloves [BM Polyco Ltd., Enfield, UK], disposable lab coats [Medline, Oxon, UK], N95 particulate nose mask [Safety company, CA, USA], Powdered gloves [TG Medical (USA) Inc., CA, USA, UK].

3.1.2 Equipment

Class IIA biosafety cabinet [Medical air technology, Manchester, UK], PCR work station [Airclean systems, North Carolina, USA], thermocycler [Applied Biosystems, Singapore city, Singapore], Gel electrophoresis tank set up [BIO-RAD, Paris, France], UV trans-

illuminator [UVP BioDoc, Cambridge, UK], Fridge-freezer [LG, London, UK], Nano drop [Thermo Scientific, Delaware, USA], pH meter [Thermo Scientific, Massachusetts, USA], incubator [Thermo Scientific, Massachusetts, USA], eppendorf centrifuge [Eppendorf AG, Hamburg, Germany], refrigerated centrifuge [Eppendorf AG, Hamburg, Germany], vortex [scientific industries Inc, New York, USA], heating block [Haep Labor Consult, Göttingen, Germany].

3.1.3 Software

R statistical package version 3.2.3 (R Core Team, 2015), Rstudio interface (<https://www.rstudio.com/products/RStudio/>), *ape* package of R (Popescu *et al.*, 2012), *seqinr* package of R (Charif, 2007), *ggplot2* package of R (Wickham, 2009), *reshape2* package of R (Wickham, 2007), BWAtools (Li & Durbin, 2009), Samtools (Li *et al.*, 2009), RAxML (Stamatakis, 2014), figtree (<http://tree.bio.ed.ac.uk/software/figtree/>), phandango (Hadfield *et al.*, 2017), ARIBA (Hunt *et al.*, 2017), In-house customized R and python scripts, roary pan-genome pipeline (Page *et al.*, 2015), Prokka (Seemann, 2014), Prodigal (Hyatt *et al.*, 2010), RNAmmer (Lagesen *et al.*, 2007), Aragorn (Laslett & Canback, 2004), hierBAPS (Cheng *et al.*, 2013), SignalP (Petersen *et al.*, 2011) and Infernal (Nawrocki & Eddy, 2013), ANI Calculator (<http://enve-omics.ce.gatech.edu/ani/>).

3.2 METHODS

3.2.1 Preparation of Culture Media and Drug Stocks

3.2.1.1 Lowenstein-Jensen (L-J) Media

The L-J media containing 0.0012 mole/dm³ magnesium sulphate heptahydrate (MgSO₄·7H₂O), 0.0109 mol/dm³ mono-potassium phosphate (KH₂PO₄), 0.0017 mol/dm³ magnesium citrate (C₆H₆MgO₇), 0.0169 mol/dm³ asparagine (C₄H₈N₂O₃) and supplemented with either 0.1017 mol/dm³ glycerol (C₃H₈O₃) or 0.0549 mol/dm³ pyruvate (C₃H₃NaO₃) were prepared by weighing the corresponding salts into 100 mL of sterile distilled water, adjusted to pH 7.0 and autoclaved at 121°C and 15 lbs of pressure for 15 minutes and allowed to cool room temperature. Homogenized eggs (167 mL) and filter-sterilized malachite green (0.01% v/v) (for inhibition of bacteria other than mycobacteria) were added to the sterilized salts solution, mixed homogenously by gentle swirling to avoid foaming. Five milliliters of the resulting suspension were aliquot into screwed cup glass tubes, slanted and solidified at 85 °C for 55 min in an inspissation machine. The freshly prepared media incubated at 37 °C overnight to check for sterility and stored in a fridge at 4 °C until use.

3.2.1.2 Middlebrook 7H11 Agar Medium

Middlebrook 7H11 agar was prepared by following the manufacturers instruction. Briefly, 7H11 agar base powder (20.5 g) was weighed and 900 mL distilled water plus 5 mL glycerol added, heated to dissolve the agar base completely before sterilization by autoclaving at 121 °C and 15 lbs of pressure for 15 minutes. The sterile medium was

allowed to cool to about 50 °C and 100 mL of Middlebrook OADC growth supplement (containing Oleic, albumin, dextrose and casitone) aseptically added, mixed thoroughly before dispensing 20 mL into sterile petri dishes and allowed to solidify. Sterility check was done by incubation of the media at 37 °C overnight and stored at 4 °C until use.

3.2.1.3 Blood Agar preparation

Following the manufacturers instruction, the blood agar medium was prepared by weighing 50 g of blood agar base powder into a flat bottom flask followed by the addition of distilled water (1 L) and autoclaving at 121 °C and 15 lbs of pressure for 15 min. the liquefied mixture was then transferred to water pre-warmed to 50 °C in a water bath. When the melted agar had acclimatized to the 50 °C temperature of the water, 50 mL of fresh sheep blood pre-warmed to 50 °C was aseptically added and mixed thoroughly by gentle swirling to avoid foaming. Aliquots of the molten blood agar (15 mL) were poured into sterile petri dishes and allowed to set at room temperature. Sterility checked was done by overnight incubation of the blood agar plates at 37 °C and stored at 8 °C in sterile zip-locked bags until use.

3.2.1.4 Preparation of Drug Stocks

Drug stocks; 2 mg/mL each of isoniazid (INH) and rifampicin (RIF) were prepared by weighing 100 mg each of powdered INH and RIF and dissolving with 50 mL of double

deionized water and DMSO respectively. Aliquots of the drugs (1 mL) were put into 2 mL cryovials and stored at -20 °C until use.

3.2.2 *Mycobacterium tuberculosis* complex Isolates (200)

Isolates used in this study (19 L4, 143 L5 and 38 L6) were obtained from archived MTBC strains isolated from pulmonary TB patients reporting to selected hospitals in three regions of Ghana namely; Greater Accra, Central and Northern regions (Fig 3.1) from 2007 to 2014 (Asante-Poku *et al.*, 2015, 2016; Yeboah-manu *et al.*, 2011). The isolates were from pulmonary TB patients routinely reporting to the selected health centres. Isolates were consecutively selected based on the order of their isolation and genotyping for this study. Ethical approval for this study was provided by the Institutional Review Board of the Noguchi Memorial Institute for Medical Research, University of Ghana with Federal-wide Assurance number FWA00001824 after reviewing this study and its protocols.

3.2.3 *Purification of Archived MTBC Isolates by Sub-Culturing*

Selected frozen MTBC strains were first allowed to thaw then checked for purity by inoculation of 100 µL of suspension on blood agar and incubation at 37 °C for 24 hrs. Vials from which no macroscopic growth was observed after the purity check were inoculated unto Middlebrook 7H11 agar and incubated at 37 °C until confluent growth was observed. Those from which macroscopic growth was observed on the blood agar

were purified with 5% oxalic acid decontamination method (Yeboah-manu *et al.*, 2004) before inoculation. Briefly, equal volume (1X mL) of 5%(w/v) of oxalic acid was added to thawed bacterial suspension in 50 mL falcon tubes and vortexed intermittently at room temperature in a class II biosafety cabinet in a biosafety level 3 lab for 30 min.

One normal PBS solution was added to the suspension up to the 45 mL mark to neutralize it before centrifugation at 4000 rpm to pellet mycobacteria in the suspension. The supernatant was decanted carefully leaving just about 2 mL of it to re-suspend the pellet for inoculation. A 100 µL aliquots of decontaminated suspension were inoculated unto two pairs of LJ-media slants supplemented with either glycerol (to enhance growth of *Mtbss*) or pyruvate (to enhance growth of *Maf*) because *Maf* known to have a defective pyruvate kinase due to a nsSNP in the encoding gene (Keating *et al.*, 2005). The inoculated media slants were incubated at 37 °C until confluent growth was observed. Isolates were then transferred with sterile inoculating loops into either Middlebrook 7H9 broth supplemented with Middlebrook ADC and glycerol for viable cells or nuclease-free water for heat inactivation at 98 °C for 60 min for further DNA-based molecular works.



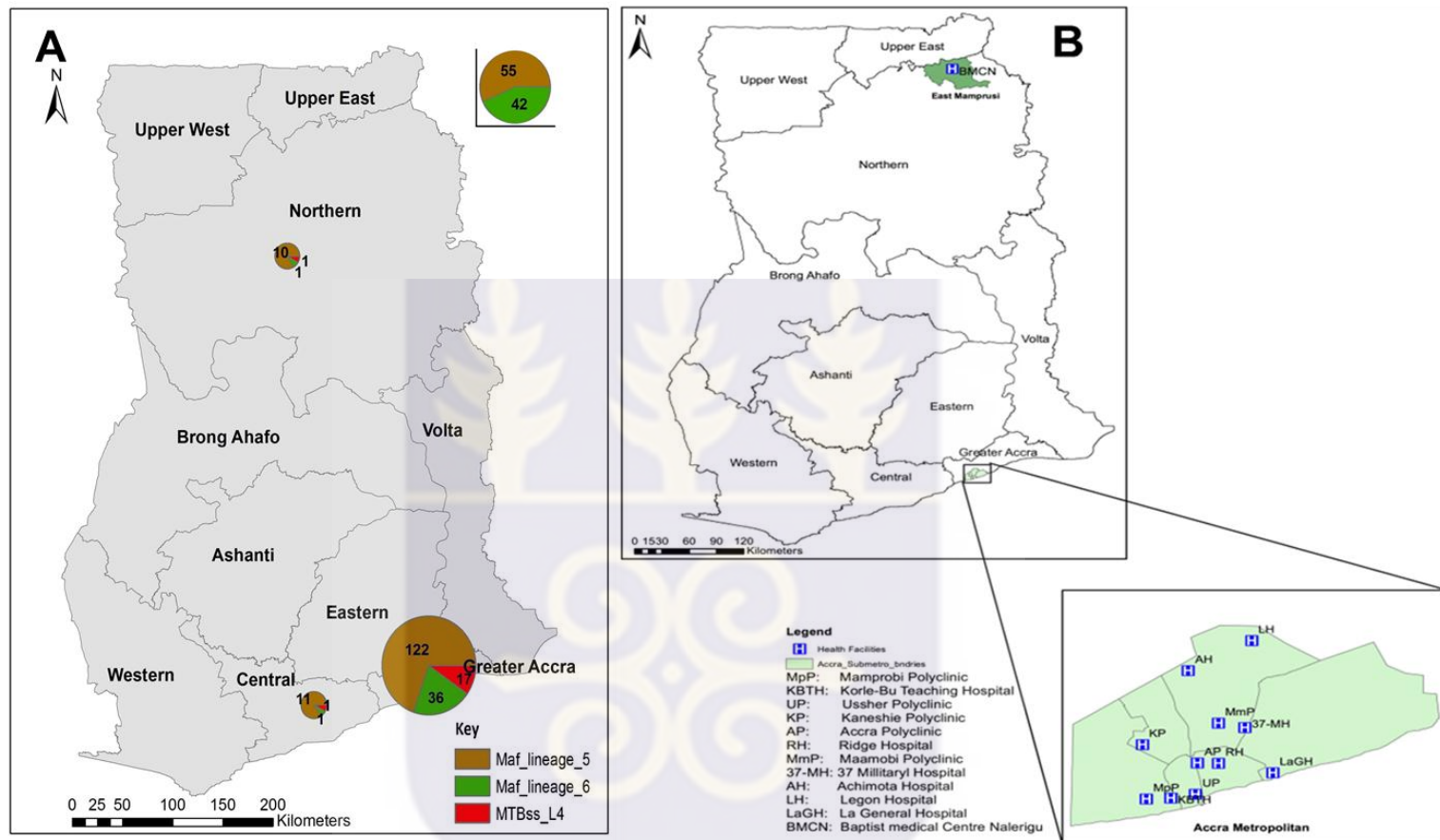


Figure 3.1: Maps of Ghana Showing the Regions from which TB Patients were recruited.

A: Isolates for whole genome sequencing analysis for which (there is no regional data on 92 MTBC genomes sequenced at Borstel). B: Isolates for the follow-up drug resistance and mutation analysis). (Fig 3.1B is adapted from (Otchere et al., 2016))

3.2.4 Genotyping of Mycobacterial Isolates

3.2.4.1 Confirmation of *Mycobacterium tuberculosis* complex strains

In order to confirm that the archived isolates were indeed members of the MTBC and belonged to the assigned lineages, PCR amplification of the MTBC-specific IS6110 (1361bp) was amplified with the primer pairs (Forward: 5'-TTC GGA CCA CCA GCA CCT AAC C -3', Primer Reverse: 5'-CCT TCT TGT TGG CGG GTC CAG -3') (Mashta *et al.*, 2011) using the reagents and quantities shown in Table 3.1. Five microliters of bacterial suspension after heat-inactivation (described above 3.2.3) was used as template DNA.

Table 3.1: Preparation of PCR mix for amplification of MTBC-specific IS6110

REAGENT	QUANTITY FOR ONE REACTION MIX (μL)	QUANTITY FOR (N)REACTION MIX (μL)
nf (H ₂ O)	16.2	16.2×N
10x Buffer	5	5×N
Q-Solution (5x)	10	10×N
MgCl ₂ (25mM)	2.5	2.5×N
Primer_F (10pmol/ μl)	2.5	2.5×N
Primer_R (10pmol/ μl)	2.5	2.5×N
dNTPs (10mM)	1	1×N
10x Coral dye	5	5×N
HotStarTaq Pol (5U/ μl)	0.3	0.3×N
Total Reaction Mix	45	45xN
Template	5	-

The thermal cycling conditions used were initial heat activation of the polymerase (Hot Start Taq polymerase) at 96 °C for 15 min followed by thirty five cycles of denaturation at 95 °C for 1 min, annealing at 65 °C for 1 min and primer extension at 72 °C for 1 min then final amplicon extension at 72 °C for 4 min. Gel electrophoresis analysis of the PCR

amplicons was then carried out by loading 10 μ L of the amplicons unto labelled wells of 2% agarose gel prepared with 1X tris-borate EDTA buffer (TBE) containing 0.03% ethidium bromide and run with 1X TBE at 100 volts for an hour. After the electrophoresis, the resolved DNA on the gel was observed under UV trans-illumination. The amplicons were sized using 100bp ladder from bacteriophage lambda DNA.

3.2.5 Chromosomal DNA Extraction

Heat-inactivated bacterial suspensions (approximately 1800 μ L) of PCR confirmed MTBC strains were spun at 14, 000 rpm for 5 min to pellet cells. Chromosomal DNA was extracted using a combined protocol of enzyme assisted lysis (Käser *et al.*, 2009) and CTAB purification (Soolingen *et al.*, 1991) methods. Briefly, the mycobacterial cell wall was disrupted by adding 50 μ L of 10 mg/mL lysozyme inverted briefly to mix and incubated overnight at 37 °C without shaking. To solubilize the plasma membrane, 75 μ L of 10% SDS was added and mixed gently by inverting the tube 5 times, after which the proteins were denatured with 10 μ L proteinase K (20 mg/mL), mixed and incubated at 65 °C for 15 min followed by addition of 100 μ L of 5 M NaCl then 100 μ L CTAB pre-warmed at 65 °C to preserve the integrity of the chromosomal DNA. After mixing by inverting the tube 5 times, the extracted DNA was purified by the addition of 900 mL of chloroform/isoamyl alcohol (24:1 v/v) and mixed thoroughly by inverting 30 times followed by centrifuging at 14 000 rpm for 10 min. The upper phase (aqueous phase) containing the DNA was pipetted out into a new 2 mL sterile tubes for purification. 700 mL of ice-cold isopropanol was added to the aqueous phase in the new tubes and

incubated at -20 °C for 3 hrs followed by centrifugation at 14, 000 rpm for 10 min to pellet the DNA. The supernatant was decanted and the precipitate washed with 500 mL of ice-cold 70% ethanol followed by centrifugation at 14000 rpm and decanting the supernatant. The DNA pellet was air dried at room temperature and suspended in 100 µL of nuclease free water and stored at -20 °C until use.

3.2.6. Whole Genome Sequencing

An aliquot of chromosomal DNA (60 µL) of each selected isolate was sequenced by the Illumina platform at the Wellcome Trust Sanger Institute, Genome campus, Hinxton United Kingdom, following the manufacturers protocols for library preparation and sequencing (Metzker, 2010). The bridge amplification process was used after sonication to fragment the ~5 ug of DNA on a glass plate called flow-cell (Fig 3.2). Briefly, the DNA was fragmented, denatured and adaptors (oligo nucleotides of known sequences) ligated to both ends of the single stranded DNA fragments. The single stranded fragments were then attached both ends to the surface of the flow-cells (Fig 3.2: A1) with pre-coated complementary sequences of the adapter molecules on its surface (Fig 3.2: A2) to form bridges with 3' overhangs of the probes on the flow cell. PCR reaction was then carried out which extended the overhung across the bridge (Fig 3.2: A3, A4). The cycle ended with denaturation of the bridge and separation into two distinct single stranded DNA sequences ready for the next bridge formation and amplification. Through the bridge amplification, multiple single strand copies of the DNA fragments were made creating a dense cluster on the flow-cell (Fig 3.2: B). The flow-cell with dense cluster of

DNA library (100-200 million molecules) was then transferred into the sequencer (Fig 3.2: C1) after addition of 3'dye-labeled dNTPs and sequencing primers complementary to the adapters. The libraries were sequenced by synthesis leading to the cleavage of the fluorescence dye at the 3' of the dNTPs. The released dye was then excited by laser irradiation leading to fluorescence which is was detected and recorded according the wavelength of the light as either green, blue, red or black corresponding to adenine, cytosine, thymine and guanine respectively (Fig 3.2: C2).

3.2.7 Analysis of Whole Genome Data

3.2.7.1 Mapping of Reads

The sequence reads in fastq format were added to an unpublished *Maf* genomes from Ghana (97 strains from the lab of Stefan Niemann in Borstel, Germany isolated from Ghanaian TB patients in Ghana from 2001 to 2004 (Fig 3.1A) and the global collection of 21 MTBC genomes (Comas *et al.*, 2010) and mapped unto the reference hypothetical MTBC ancestral genome (Comas *et al.*, 2010) using combinations of customized python and bash algorithms communicating to BWA and SAMtools interpreter softwares. The consensus sequence of each genome was extracted from the mapping of each genome with insertion of "N" into the individual nucleotides within regions encoding repetitive elements such as PE-PGRS, PPE, IS elements and saved in a fasta format for further analyses. All genomes with less than 10% coverage were excluded from further analysis.

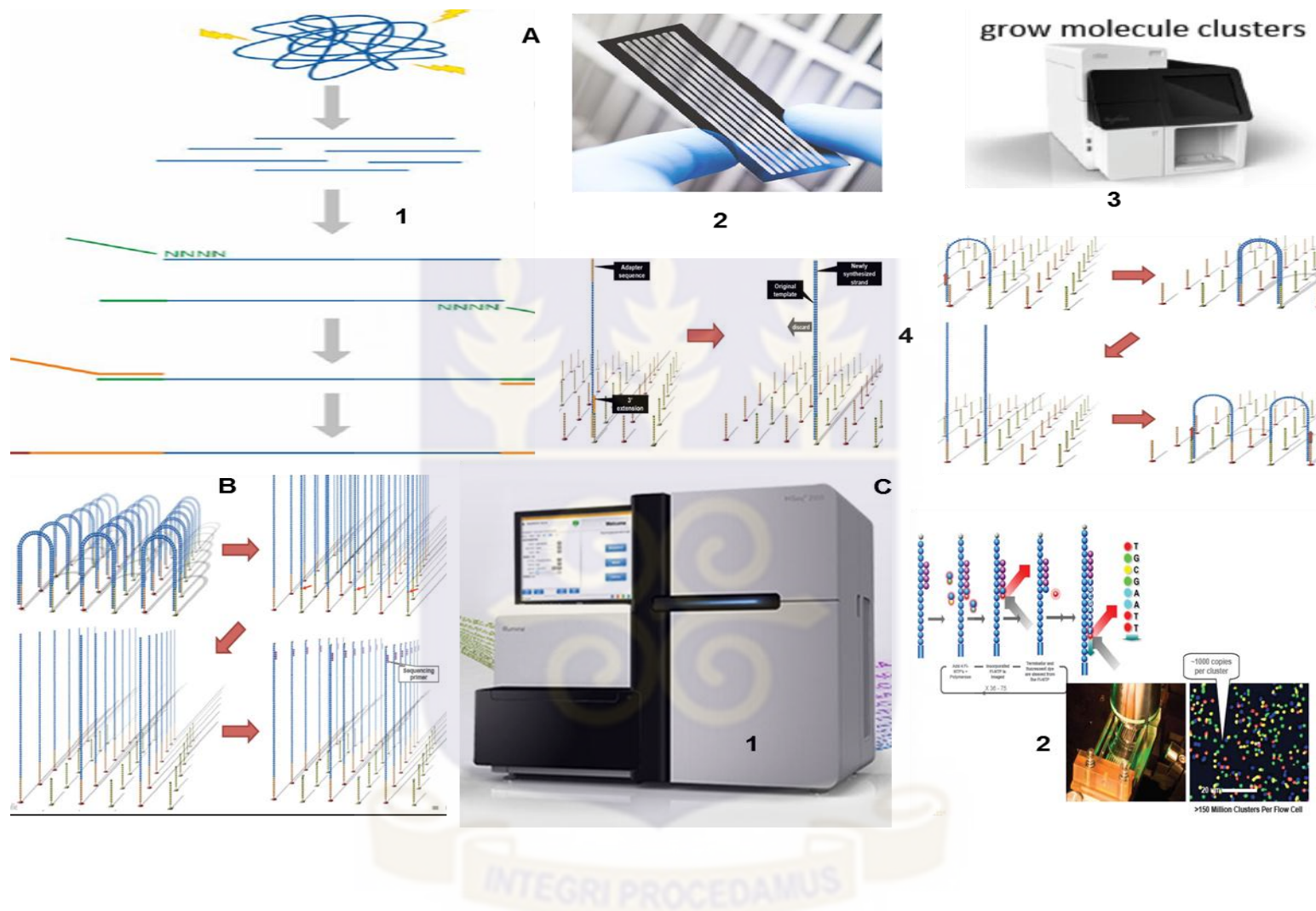


Figure 3.2: Schematic diagram of illumina sequencing.

A: library preparation, B: complete library by cycle amplification, C: sequencing of libraries.

(modified from <http://bitesizebio.com/13546/sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/>).

In addition, all genomes from mixed infections (more than a single MTBC lineage per genome or same lineage but different strains) indicated by lineage-specific SNP calls and/or more non-synonymous SNPs (heterozygous SNPs) than synonymous SNPs (homozygous SNPs) were also filtered out leaving only those for distinct strains for further analysis.

The total number of homozygous SNPs between each genome and the MTBC ancestor were compared stratified by lineage and the total number of pairwise homozygous SNPs within each lineage and plotted with *ggplot* function of the *ggplot2* package in R version 3.2.3. A statistical significant difference between the lineages was tested using the non-parametric Wilcoxon rank-sum test in R version 3.2.3.

3.2.7.2 Phylogenetics and Clustering Analyses

Gross genetic variance between each of the Ghanaian MTBC genomes as well as the global collection of 21 MTBC genomes and the reference ancestor genome excluding variance within genes associated with drug resistance were extracted from the fasta files with customized python algorithm and aligned into a multi-fasta file for phylogenetic analyses. All the “N” inserted into regions encoding repetitive elements such as PE-PGRS, IS and PE were replaced with “-“ in the multifasta file. Maximum likelihood phylogenetic tree with 1000 bootstraps and rooted on *M. canettii* was generated with RaxML and annotated with figtree. Clustering analysis of the resulting phylogeny was carried out using hierBAPS with 3 clustering iterations and 10 expected number of

clusters ($L = 3$, $K = 10$). In addition, the average nucleotide identities (ANI) between any two lineages (complete genomes in fasta format) were estimated with the ANI calculator from Kostas lab (<http://enve-omics.ce.gatech.edu/ani/>).

3.2.7.3 Genetic Diversity and Mutation Analysis

Firstly, whole genome pairwise nucleotide diversity of the lineages was determined using raw counts of mutations and pairwise deletions for sites with gaps (variable positions) of each genome compared to the MTBC ancestor reference making use of the *dna.dist* function of the *ape* package with 1000 bootstraps of size 50 with replacement in R using a customized R script and plotted with *ggplot* function of *ggplot2* package in R version 3.2.3. For the comparison of total pairwise number of SNP distributions, a mean SNP distance to all isolates in each lineage was calculated and a distribution of the mean pairwise distances plotted. The non-parametric Wilcoxon rank-sum was used to test for differences between lineages because the data was assumed as not normally distributed. A p-value less than 0.05 at 95% confidence level was accepted for statistical significance. Average pairwise nucleotide diversity (π) was calculated per non-synonymous site as the mean number of pairwise mismatches among each lineage (group) divided by the total length of base pairs in the sequences under consideration (Hartl & Clark, 2007). The lower and upper confidence intervals were obtained by calculating the 2.5th and 97.5th quantiles for π using 1000 bootstrap replicates with replacement of the polymorphic sites of the alignment using the *sample* and plotted with *ggplot2* in R. Non-overlapping confidence intervals of π were taken as evidence of statistically significant difference.

In addition, annotated MTBC genes according to the H37Rv reference genome and grouped into 7 functional categories namely those encoding proteins essential for growth in macrophages (*esmac*), involved with cell wall and cell processes (*cwallproc*), involved with lipid metabolism (*limpet*), involved with intermediate metabolism and respiration (*intmedres*), encoding regulatory proteins (*regprot*), proteins involved with virulence, detoxification and adaptation (*virDETAD*), involved with information pathways (*infopath*) as well as all the currently identified 1226 human T cell epitopes (*epit*) spanning 304 T cell antigens (Stucki *et al.*, 2016) were also analysed for diversity as well as mutations. All genes encoded by the forward strand were extracted from the individual genomes and concatenated together directly. However, those encoded by the reverse strand were first, concatenated and then used as template to generate the complement of the concatenate which was then concatenated to the forward strand concatenate into a multi-fasta file for the analyses. All unknown nucleotides (“Xs” and “Ns”) resulting from non-specific assigning of reads to repetitive regions such as PE-PGRS and PPE were replaced with a “-” before the analysing for pairwise nucleotide diversity using the *nuc.div* function of the *ape* package in R version 3.2.3 using customized R scripts. The lower and upper confidence intervals were calculated as already described above and plotted with *ggplot* function of *ggplot2* package in R version 3.2.3.

In addition, the multifasta concatenates of the genes of the 7 functional categories as well as human T cell epitopes were analysed for pairwise ratio of the change in nonsynonymous mutations (dN) to the change in synonymous mutations (dS) simplified

as dN/dS within each lineage using the *kaks* function of *seqinr* package and plotted with *ggplot* function of *ggplot2* package in R using a customized R algorithms. The dN/dS measures were estimated using the polymorphic sites within each lineage to reflect both lineage-specific substitutions and transient strain-specific polymorphisms. The mean pairwise dN of each isolate was divided by its mean pairwise dS with relative to the other isolates within each lineage to obviate the possibility of underestimating the dN/dS. Statistical differences between the estimates for each gene category within each lineage was accessed using the non-parametric Wilcoxon rank-sum tests with continuity correction implemented in R version 3.2.3. A p-value less than 0.05 at 95% confidence level was accepted for statistical significance

In addition, pairwise nucleotide diversity and dN/dS of T cell epitopes and regulatory proteins of *Maf* L5 and L6 were estimated and plotted as described above stratified by patient ethnicity and HIV status of the patients, respectively, due to the association of L5 with ethnicity and L6 with a positive HIV status in Ghana (Asante-Poku *et al.*, 2015, 2016). In furtherance, using customized bash algorithms, the number of mutated T cell epitopes (epitopes with synonymous and non-synonymous SNPs) was compared between L5 and L6 as well as the lineages stratified by either patient ethnicity (L5) or HIV status (L6) of the TB patients. The number of variable sites (non-redundant SNPs) within the T cell epitopes was also compared. Lastly, some selected genes reported to harbour mutations among some MTBC strains with phenotypic implications (Table 3.2) were analyzed and compared among the three MTBC lineages in Ghana relative to orthologous genes of H37Rv using ARIBA for mutations.

Table 3.2: Selected genes with reported MTBC lineage-specific mutations

Family of Genes	List of Genes
Sulfur transporters	Rv2400c,Rv2398c,Rv2397c,Rv2399c,Rv1739c,Rv1707
Phosphorus transporters	Rv0545c,Rv2281,Rv0932c,Rv0931c,Rv0934,Rv0933, Rv0828c,Rv0930,Rv0929,Rv0820
Nitrogen transporters	Rv2920c,Rv0267,Rv0261c,Rv1737c,Rv2329c
Ion transporters	Rv1348,Rv1349,Rv2691,Rv2692,Rv1857,Rv1858,Rv1859, Rv0924c,Rv1239c,Rv2856,Rv1607,Rv2287,Rv3236c, Rv2877c,Rv0092,Rv0103c,Rv3270,Rv1469,Rv3200c, Rv3679,Rv3680,Rv3682,Rv0908,Rv1997,Rv1992c, Rv0425c,Rv0107c,Rv3743c,Rv0969,Rv1029,Rv1030, Rv1031,Rv1028A,Rv0143c,Rv0265c,Rv2025c,Rv3041c, Rv1811,Rv0362,Rv3044
Carbohydrate transporters	Rv3331,Rv2833c,Rv2835c,Rv2832c,Rv2834c,Rv1236, Rv1237,Rv1238,Rv1235,Rv2037c,Rv2038c,Rv2039c, Rv2040c,Rv2041c,Rv2316,Rv2317,Rv2318,Rv1200,Rv2456c
Lipid transporters	Rv0402c,Rv0507,Rv0206c,Rv0450c,Rv0676c,Rv1557, Rv2942,Rv3823c,Rv2339,Rv1183,Rv0202c,Rv1522c,Rv1145, Rv1146
Peptide transporters	Rv2585c,Rv1281c,Rv1280c,Rv1283c,Rv1282c,Rv3666c, Rv3665c,Rv3664c,Rv3663c
Amino acid transporters	Rv3759c,Rv3758c,Rv3757c,Rv3756c,Rv0072,Rv0073, Rv0048c,Rv1496,Rv1747,Rv1979c,Rv1986,Rv1999c, Rv2563, Rv3253c,Rv0522,Rv0899,Rv1704c,Rv0917, Rv2320c, Rv2564,Rv2127,Rv0346c
Cell invasion operon	Rv0169,Rv0170,Rv0171,Rv0172,Rv0174,Rv0173,Rv0175, Rv0176,Rv0177,Rv0178
Lipid catabolism operon	Rv3545c,Rv3544c,Rv3543c,Rv3542c,Rv3541c,Rv3540c
Cholesterol catabolism operon	Rv3570c,Rv3569c,Rv3568c,Rv3567c
ESAT-6 secretory system 1	Rv3864,Rv3865,Rv3866,Rv3867,Rv3868,Rv3869,Rv3870, Rv3871
ESAT-6 secretory system 2	Rv3874,Rv3875,Rv3880c,Rv3877,Rv3878
Growth attenuation genes	Rv0557,Rv0469,Rv0930,Rv0862c,Rv1096,Rv1125,Rv1178, Rv1328,Rv1592c,Rv2112c,Rv2241,Rv2383c,Rv2737c, Rv3282,Rv3696c,Rv3409c,Rv0642c,Rv0470c,Rv2869c, Rv0186A,Rv2524c,Rv3499c,Rv3082c,Rv1242,Rv1246c, Rv1241,Rv1323,Rv1345,Rv1347c,Rv1350
Polyketide synthase genes	Rv0405,Rv1013,Rv1180,Rv1181,Rv1527c,Rv1660,Rv1661, Rv1662,Rv1663,Rv1664,Rv1665,Rv2048c,MTB000057, Rv2946c,Rv2947c,Rv3800c,Rv3825c
Virulence gene	Rv3414c

3.2.7.4 *De novo Assembly of Bacterial Genomes and Annotations*

Before the *de novo* assembly of the sequence reads, paired-end reads were obtained by concatenating the forward and reverse read files of each genome into one using a customized *perl* script. The assembly was then carried out on the concatenated reads with Velvet (Zerbino & Birney, 2008) using default settings and range of k-mers between 16 and 31 bp and estimated genome size of 4.4Mbp long by execution of the code: “*VelvetOptimiser.pl -s 16 -e 31 -f “-short -fastq.gz reads.fa” -g 4.4*” on the command prompt (terminal of the Unix system) in the directory containing the sequence reads. This instruction calls into action two programs *velveth* and *velvetg* which carry out the pre-processing of the reads into single files and *de novo* assembly of short reads based on de Bruijn graphs respectively (Zerbino & Birney, 2008). The assembled genomes were then annotated using *Prokka*. *Prokka* is a rapid command line prokaryotic genome annotation *perl* algorithm that relies on external feature prediction tools namely *Prodigal*, *RNAmmmer*, *Aragorn*, *SignalP* and *Infernal* to identify the coordinates of genomic features within contigs of assembled genomes and annotate protein coding genes in a hierarchical order. Identified open reading frames (ORFs) were translated and compared to existing protein databases one after the other starting with the one with relatively smaller trustworthy databases, progressing to medium-sized domain-specific databases and ending with curated models of protein families and assigns an e-value threshold of 10^{-6} to signify confidence of annotation.

3.2.7.5 Comparative Pan-Genome Analysis

Comparative pan-genome analysis was carried out with the *roary* pan-genome pipeline with default settings. Briefly, the pipeline extracted all annotated genes from the annotated de novo assembled genomes and concatenated together to create a core gene alignment per genome followed by creation of gene presence/absences file and summary of core and accessory genes. The summary files were exported into excel for comparative analysis and summary into Tables and Figures.

3.2.8 Drug Resistance Profiling and Target Gene Mutational Analysis

3.2.8.1 *Drug Resistance Screening*

To test for the potential influence of bacterial genetics on acquisition and fixation of DR-associated mutations, a prospective MTBC were isolated from consecutive TB patients reporting to selected health centers in the Greater Accra and the Northern regions of Ghana for DR analyses. Micro-plate alamar blue assay (MABA) for drug susceptibility testing (DST) was performed in clear-bottomed 96 well plates. Aliquots of drugs stocks (2 μ L/mL of INH and 4 μ L/mL of RIF) prepared with deionized water and DMSO respectively) were stored at -20 °C until use. The DST was set up by following the method described by Franzblau *et al.* (Franzblau *et al.*, 1998) with slight modification. In summary, a loop-full of bacteria growing at the log-phase was emulsified with sterile Middlebrook 7H9 broth supplemented with 0.2% glycerol and 0.001% casitone (7H9-GC). The obtained suspension was then adjusted to a McFarland no 1 standard with 7H9-GC and a 1:20 dilution used as the inoculum for the assay. As shown in the Figure below (Fig 3.3), all the outer wells (A1 to A12, H1 to H12, then B1, C1, D1, E1, F1 and G1 as

well as B12, C12, D12, E12, F12 and G12) of the clear bottomed, 96 well micro-plates were filled with 200 μ L of sterile water to prevent evaporation of contents of the test wells. All the remaining wells but those on column 11 (which were left blank and filled with sterile water) were filled with 100 μ L of 7H9-GC supplemented with OADC (10%) and Tween 80 (0.01%) thence called 7H9-GCOT media.

The drug stocks were thawed at room temperature and 100 μ L aliquots pipetted into the first labelled wells on the row B as indicated in the Figure (Fig 3.3), mixed by gentle pipetting to avoid foaming and serially diluted by pipetting 100 μ L from those wells into the wells below on row C. This procedure was repeated until row G after which the excess 100 μ L was discarded after mixing leaving 100 μ L of media-drug mixture in each well. Bacterial inoculum (100 μ L) was pipetted into isolate-specific labelled drug-containing wells as well as drug-free positive control wells whereas 100 μ L of 7H9-GC was added to the negative control wells. The final concentrations of drugs after inoculation of 100 μ L of bacterial suspension were 0.03125-1.0 μ L/ml and 0.0625-2.0 μ L/ml for INH and RIF respectively. The inoculated plates were incubated for 7 days' at 37 °C, and 50 μ L of the indicator (1:1 v/v Alamar Blue and sterile 10% Tween 80) added to a drug-free growth-positive control and drug free un-inoculated negative control wells. Plates were re-incubated for 24 h. When the positive growth control wells turned pink without the negative growth control changing (indication of no contamination), the indicator was subsequently added to all other wells. After a further 24 h of incubation, the colours of all wells were recorded. Wells remaining blue were scored as no growth. The critical concentration of drugs were $<0.25 \mu\text{L/mL}$ for both drugs. The minimum inhibitory concentration of the drugs for each isolate was then recorded.



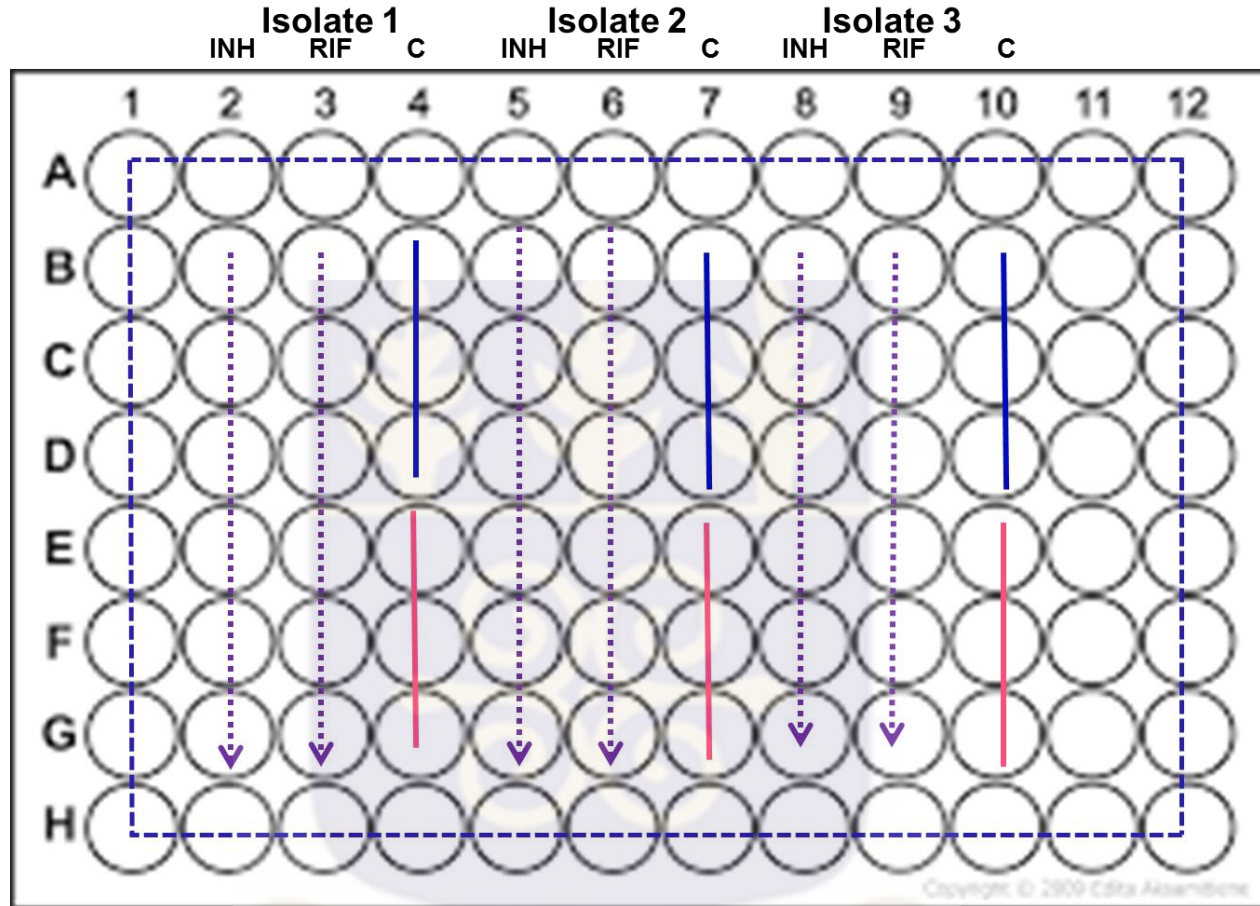


Figure 3.3: Diagrammatic set up of the micro-plate alamar blue assay.

The solid blue lines mark negative control wells whereas the solid red lines mark positive control wells. The broken blue lines demarcates outer wells filled with sterile water to prevent evaporation of the contents of the inner wells

3.2.8.2 PCR Amplification of Drug Resistance Associated Genes and DNA Sequencing

The following genomic targets, the *rpoB*, *rpoA* and the *rpoC* genes were analysed for RIF resistant isolates and the *katG*, *inhA*, and *ndh* genes as well as the promoter region of the *inhA* (*inhA_{pro}*) and *oxyR-ahpC* for INH resistant isolates for mutations. Specific regions within these targets (known to be hot-spots of mutations associated with drug resistance) were amplified by PCR with specific primers (Table 3.3). The cycling conditions and the size of the amplicons were confirmed by gel electrophoresis as already described in section 3.3.1 above. The PCR amplicons were then sequenced by outsourcing to Macrogen Europe (Macrogen, Netherlands) using the traditional Sanger chain termination method on both directions using the same respective primers used for the PCR amplification of each target.

3.2.8.3 Bioinformatics Analysis of Sequenced Genes for Mutations

The sequences were analyzed for specific mutations by comparing them with reference sequences of the *Mtbss* strain H37Rv downloaded from the Tuberculist database (Lew *et al.*, 2011) using the Staden package for sequence assembly (Staden *et al.*, 2000). The software was used in its default form settings. However, all vector sequence remnants were clipped off and mis-matches up to 50% were allowed for the analysis. SNPs were called only when the said SNP is supported by a clear chromatogram without background noise (Fig 3.4).

Table 3.3: Primer pairs, expected band size and annealing temperatures used for PCR

Drug	Primer Name	Primer Sequence (5'-3')	Amplicon Size	T _m	Reference
INH	<i>katG_F</i>	CCAGCGGCCCAAGGTATC	850 bp	65 °C	(Malla, 2013)
	<i>katG_R</i>	GCTGTGGCCGGTCAAGAAGAAGTA			
	<i>inhA1_F</i>	CTACATCGACACCGATATGAC	600 bp	57 °C	(Ho <i>et al.</i> , 2009)
	<i>inhA1_R</i>	GACCGTCATCCAGTTGTAG			
	<i>inhA2_F</i>	GCATCAACCCGTTCTTCGAC	550 bp	57 °C	(Ho <i>et al.</i> , 2009)
	<i>inhA2_R</i>	TAATGCCATTGATCGGTGATAC			
	<i>inhApro_F</i>	GGCACGTACACGTCTTTATGTA	478 bp	65 °C	(Gagneux <i>et al.</i> , 2006)
	<i>inhApro_R</i>	GGTGCTCTTCTACCGCCGTGAA			
	<i>oxyR-ahpC_F</i>	ACCACTGCTTTGCCGCCACC	236 bp	70 °C	(Homolka <i>et al.</i> , 2010)
	<i>oxyR-ahpC_R</i>	CCGATGAGAGCGGTGAGCTG			
	<i>ndh_F</i>	ATCACCACCGCCGCTGAAGC	589 bp	64 °C	(Ramaswamy <i>et al.</i> , 2003)
	<i>ndh_R</i>	GTTCGGGTACCCGGAATG			
RIF	<i>rpoA1_F</i>	GCATTCCAGTCGATTCCATC	560 bp	57 °C	(de Vos <i>et al.</i> , 2013)
	<i>rpoA1_R</i>	CCAAGATCGCCTTCTGATGT			
	<i>rpoA2_F</i>	GGACGTCGAAAGGAAGAAGA	542 bp	57 °C	(de Vos <i>et al.</i> , 2013)
	<i>rpoA2_R</i>	GTCTCCACGTCCAGGATCAG			
	<i>rpoB_F</i>	GTAGTCCACGCCGTAAACGG	601 bp	65 °C	(Gutierrez <i>et al.</i> , 2005)
	<i>rpoB_R</i>	ACGTCCATGTAGTCCACCTCAG			
	<i>rpoC_F</i>	CGAAAACCTCTACCGCGAAC	650 bp	64 °C	(de Vos <i>et al.</i> , 2013)
	<i>rpoC_R</i>	GCGACAGGATGTTGTTGGAG			

3.2.8.4 Statistical Analysis of Drug Resistance Data

The mutations identified were compared between *Mtbss* and *Maf* for statistical significance. Fisher's exact test was used to test for statistical association between the proportion of individual MTBC lineages and either drug resistance (INH-mono resistance [INH^r], RIF-mono resistance [RIF^r] or MDR) or specific DR-associated mutations using the R version 3.2.3. A p-value less than 0.05 at 95% confidence level was accepted for statistical significance



Figure 3.4: High-confidence detection of SNPs with the Staden package.

The red circle captures a polymorphic site. The wild-type nucleotide on reference sequence (in lower-case) is a “g” compared to the mutant nucleotide “C” detected in four out of five sequences shown.

CHAPTER FOUR

DESCRIPTION OF RESULTS

4.1 SNP DISTANCE AND DIVERSITY OF MTBC LINEAGES IN GHANA

Among the final 270 MTBC genomes (excluding 27 genomes not meeting the minimum requirement for mapping and variance calling), the number of fixed SNPs in a genome relative to the MTBC ancestor (Comas *et al.*, 2010) was significantly higher ($p < 0.0001$) in L5 (median = 928) and L6 (median = 1037) compared to L4 (median = 822). The number of fixed in SNPs in L6 was also higher ($p < 0.0001$) compared to L5 (Fig 4.1A). At the whole genome level, L6 had higher pairwise nucleotide diversity also called π (π) compared to L4 and L5 (Fig 4.1B). In addition, even though the π of L4 was relatively higher than L5, the difference was not statistically significant with overlapping error bars (Fig 4.1B).

4.2 PHYLOGENETICS, CLUSTERING AND NUCLEOTIDE IDENTITY ANALYSIS OF THE MTBC IN GHANA

Two phylogenetic trees were generated from this data set. The first tree (Appendix A) was constructed from 16, 976 variable nucleotide positions extracted from mapping of the 270 Ghanaian genomes (17 L4, 175 L5 and 78 L6) together with the global reference collection of MTBC genomes (Comas *et al.*, 2010) unto the MTBC ancestor reference. The second phylogeny (Fig 4.2) was generated from 11,027 variable nucleotide positions extracted from mapping of the Ghanaian *Maf* genomes (175 L5 and 78 L6) unto the ancestor reference, rooted on *M. canettii* and clustered using *hierBAPS* with 3 iterations. Five clusters of L5 (L5.C1 to L5.C5) and 3

clusters of L6 (L6.C1 to L6.C3) were observed (Fig 4.2). The average nucleotide identity (ANI) calculated with the ANI calculator between the two lineages was greater than 99% (Appendix B).

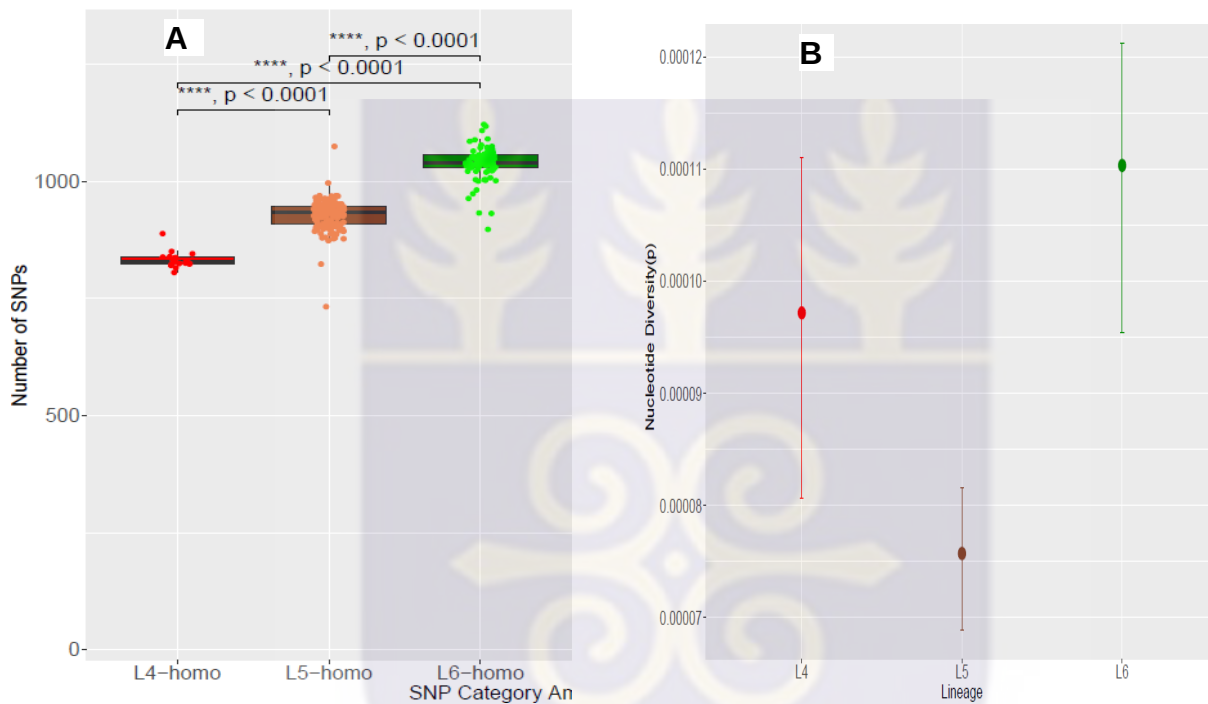


Figure 4.1: SNP distance and nucleotide diversity of hMTBC lineages in Ghana.

A: Number of fixed SNPs between MTBC lineages and the hypothetical MTBC ancestor (the median fixed SNPs of L4 (833) and L5 (934) were significantly lower ($p < 0.0001$) compared to L6 (1,039) with L4 having the least number of fixed SNPs. **B:** Whole genome average pairwise nucleotide diversity (π) among L4 (0.000097), L5 (0.000076) and L6 (0.000110). Error bars indicate 95% confidence intervals overlapping of which denotes no significant difference.

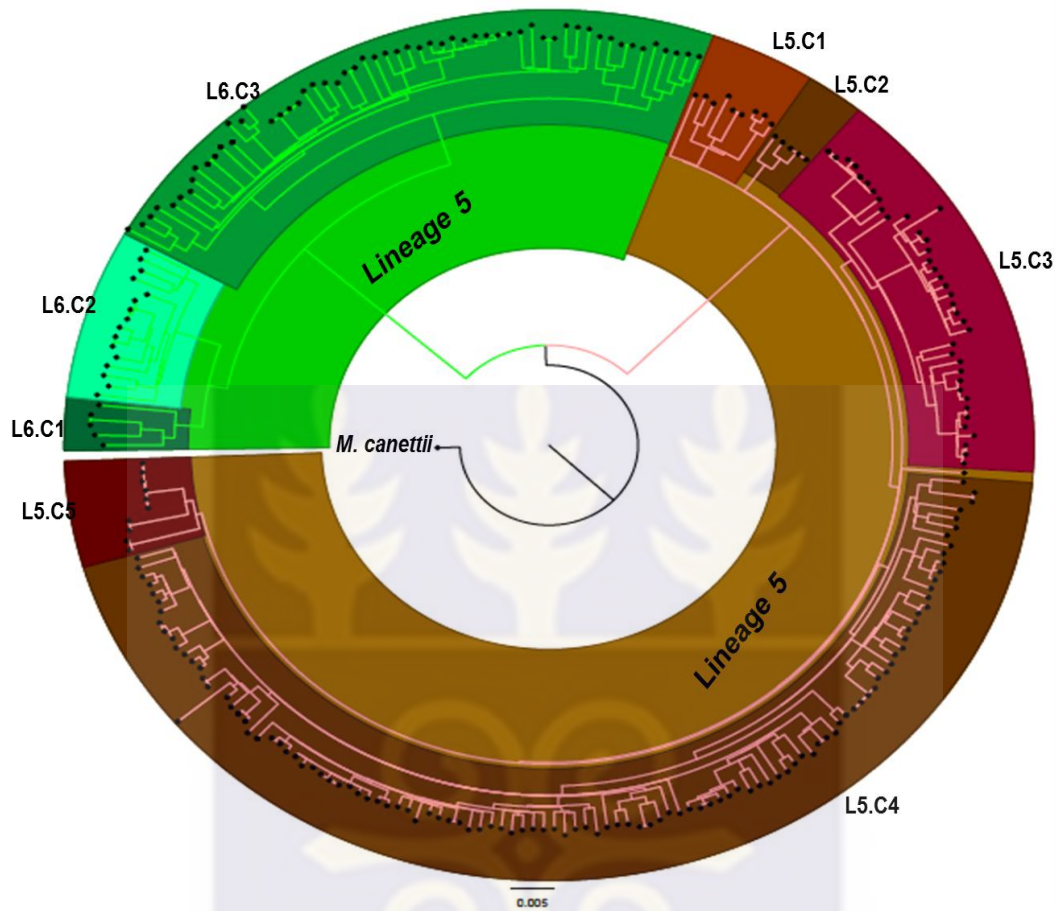


Figure 4.2: Clustered maximum likelihood phylogeny of Ghanaian *Maf* lineages rooted on *M. canettii*.

The L5 clade is coloured with different shades of brown whereas L6 is coloured with different shades of green. Among 175 L5 and 78 L6 genomes, 5 clusters of L5 (L5.C1 –L5.C5) were predicted compared to the three predicted clusters of L6 (L6.C1-L6.C3)

4.3 PAIRWISE NUCLEOTIDE DIVERSITY (π) OF ANNOTATED GENES

The mean, the 2.5%/ 97.5% quantiles of the estimated π were calculated per lineage and compared among T cell epitopes and genes of the other 7 functional categories such as essential for growth in macrophages, intermediate metabolism and respiration, lipid metabolism, virulence, detoxification and adaptation, cell wall and cell processes, regulatory proteins and information pathways (Fig. 4.3). The estimated π of L4 compared to L6 was only significant within T cell epitopes and information pathways genes as well as regulatory proteins. However, the π estimated for L5 was significantly lower compared to L6 across all the 8 functional groups of sequences and/or genes. Whereas the π for L5 across all functional categories were all below 0.0001, π for L6 were all above 0.0001. The most prominent difference in π was within T cell epitopes for which the mean π for L4 and L5 were 0.000061 and 0.000063, respectively, compared to 0.000149 estimated for L6 (more than two-fold increase in diversity in each instance). The π of L4 compared to L5 was comparable across all the gene families under study with the exception of genes for lipid metabolism and genes for involved with cell wall and cell processes (Fig 4.3).

Within L4, there was a statistically significant difference between the π for T cell epitopes and genes for lipid metabolism, genes for virulence, detoxification and adaption, and genes for involved with cell wall and cell processes (supported by Kruskal-Wallis test, $p < 0.0001$ in each instance). In a similar way, there was a statistically significant difference (Kruskal-Wallis test, $p < 0.0001$) between the estimated π for genes for lipid metabolism and genes for virulence,

detoxification and adaption as well as regulatory proteins. In addition, there was a statistically significant difference between the estimated π for genes involved with intermediate metabolism and respiration and genes for virulence, detoxification and adaption (Kruskal-Wallis test, $p < 0.0001$). However, within L5 there was no difference between the estimated π for T cell epitopes and any of the functionally categorized genes (Fig 4.3).

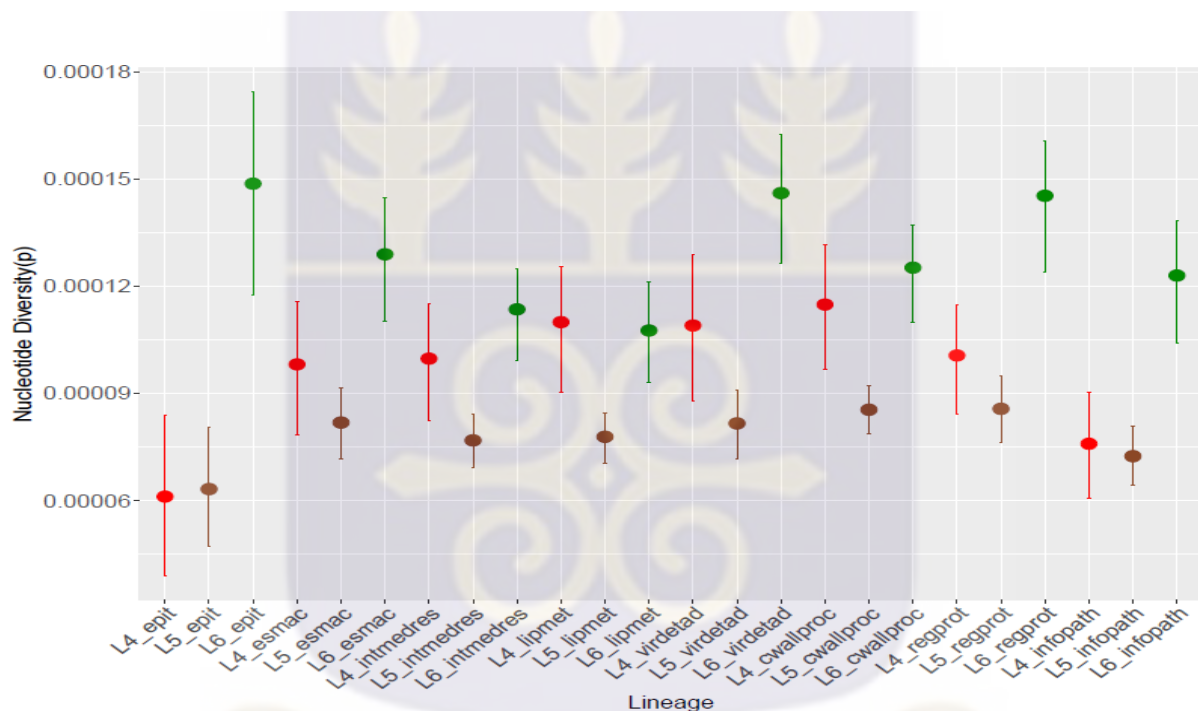


Figure 4.3: Average pairwise nucleotide diversity (π) of annotated genes among hMTBC lineages.

- *epit* – T cell epitopes,
- *esmac* – genes essential for growth in macrophages,
- *intmedres* – genes for intermediate metabolism and respiration,
- *lipmet* – genes for lipid metabolism,
- *virdetad* – genes for virulence, detoxification and adaptation,
- *cwallproc* – genes for cell wall and cell processes,
- *regprot* – genes for regulatory proteins
- *infopath* – genes for information pathways.

4.4 PAIRWISE DN/DS ANALYSIS OF ANNOTATED GENES OF THE MTBC IN GHANA

Average pairwise dN/dS of T cell epitopes as well as genes of the 7 different functional categories were calculated and compared among the lineages (Fig 4.4). Apart from T cell epitopes and regulatory proteins, for which the median of the average pairwise dN/dS ratios for at least one lineage was larger than 1.0 (Fig 4.4), none of the genes of the different functional categories had dN/dS ratio larger than 1.0 among the lineages. As shown in Table 4.1, T cell epitopes of L4 and L5 (median pairwise dN/dS of 0.57 and 0.64, respectively) were significantly more conserved (Wilcoxon rank-sum test, $W = 166$, $p < 0.0001$ and $W = 2265$, $p < 0.0001$, respectively) compared to L6 which exhibited higher diversity (median pairwise dN/dS = 1.53). In the case of regulatory proteins, it was L5 (with median pairwise dN/dS = 1.03) that was more diverse (Wilcoxon rank-sum test, $W = 6303$, $p = 0.0010$) compared to L6 (with median pairwise dN/dS = 0.85). In addition, L5 was more diverse ($p < 0.0001$) compared to L4 (with median pairwise dN/dS = 0.56). A similar trend was observed when three random samples of L5 of equal sample size as L6 were compared (Fig 4.5). dN/dS within T cell epitopes of all the 3 L5 samples were significantly lower than L6 ($p < 0.0001$ across the board) whereas dN/dS within regulatory proteins of the 3 L5 samples were significantly higher than L6 ($p < 0.0001$ for both sample 1 and sample 2 and $p = 0.0008$ was recorded for the comparison with sample 2).

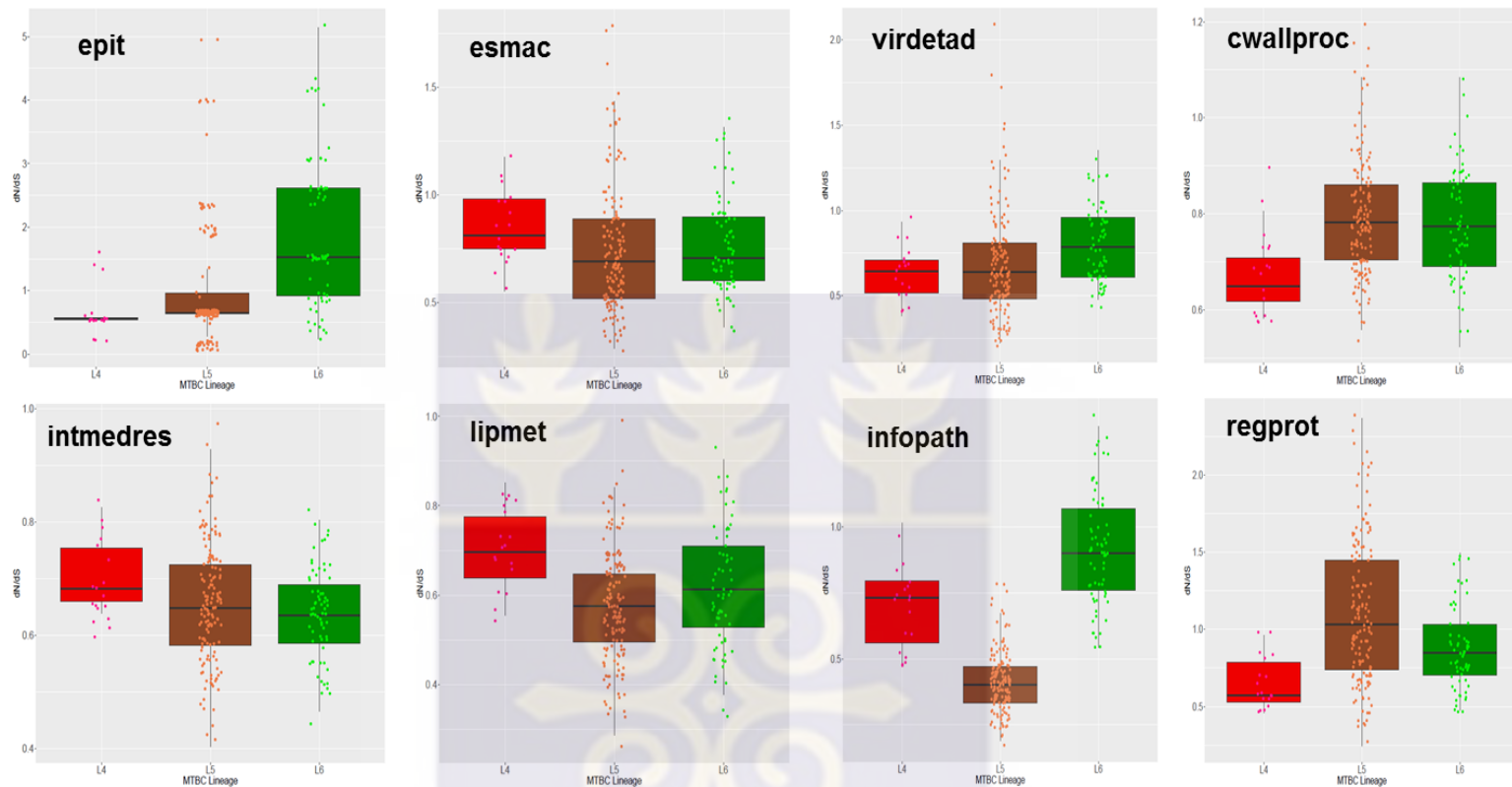
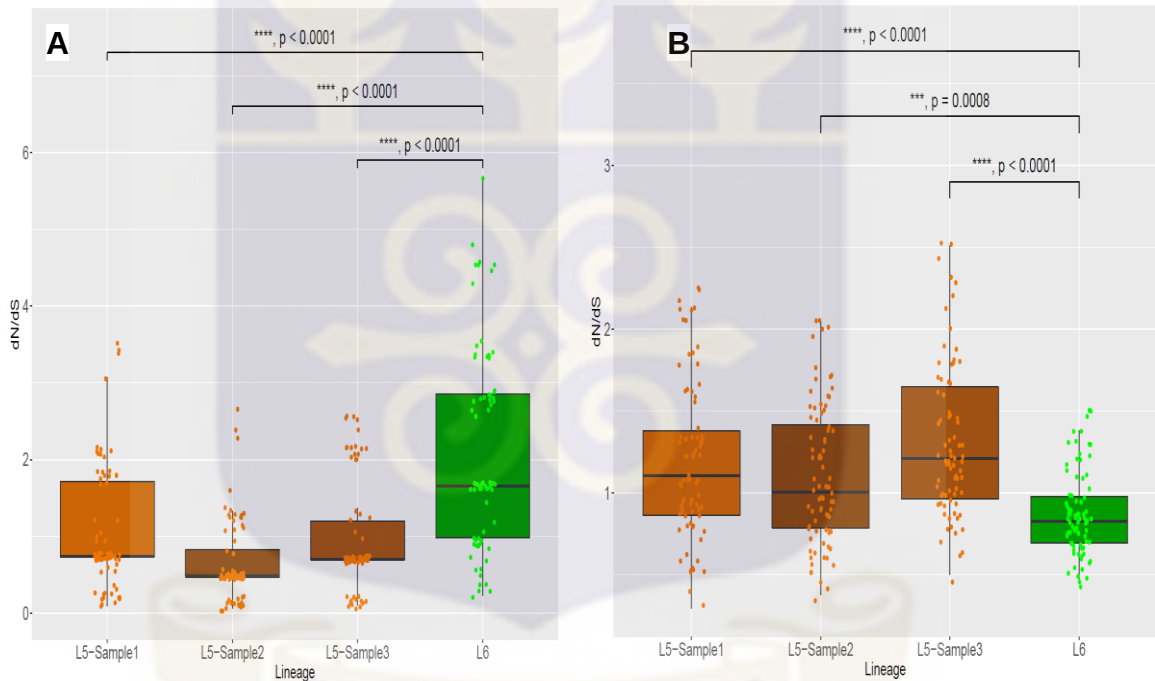


Figure 4.4: Pairwise dN/dS of annotated genes among hMTBC lineages

- *epit* – T cell epitopes,
- *esmac* – genes essential for growth in macrophages,
- *intmedres* – genes for intermediate metabolism and respiration,
- *lipmet* – genes for lipid metabolism,
- *virdetad* – genes for virulence, detoxification and adaptation,
- *cwallproc* – genes for cell wall and cell processes,
- *regprot* – genes for regulatory proteins
- *infopath* – genes for information pathways..

Table: 4.1: dN/dS ratio of MTBC lineages within functional genes

Functional category of genes	Median dN/dS			Wilcoxon test (p-value)		
	L4	L5	L6	L4-L5	L4-L6	L5-L6
T cell epitopes	0.57	0.64	1.53	0.01221	<0.0001	<0.0001
Essential for growth in macrophages	0.73	0.69	0.70	0.1242	0.3383	0.1703
Virulence, detoxification and adaptation	0.65	0.65	0.78	0.5209	0.00188	0.0128
Cell wall and cell processes	0.66	0.78	0.75	<0.0001	0.0007	0.5297
Regulatory proteins	0.56	1.03	0.85	<0.0001	0.0002	0.0010
Intermediate metabolism	0.72	0.66	0.62	0.0060	<0.0001	0.0009
Lipid metabolism	0.69	0.57	0.62	<0.0001	0.0157	0.0461
Information pathways	0.72	0.39	0.88	<0.0001	0.0061	<0.0001

**Figure 4.5: Pairwise dN/dS of T cell epitopes (A) and regulatory proteins (B) between *Maf* lineages.**

A: Comparative pairwise dN/dS of T cell epitopes of 3 random samples of L5 of size 67 with replacement against the 67 L6 genomes. **B:** Comparative pairwise dN/dS of regulatory protein genes using 3 random samples of L5 of size 67 with replacement against 67 L6 genomes.

4.5 LINEAGE-SPECIFIC ACCUMULATION OF MUTATIONS WITHIN HUMAN T CELL EPITOPES

Out of the total 1,226 characterized T cell epitopes of the MTBC, 1,068 of L5 versus 1137 of L6 were conserved without any mutation compared to 142 versus 74 with single mutations, respectively (Fig 4.6A). In addition, 16 T cell epitopes in L5 versus 15 in L6 had double mutations with no T cell epitopes with more than 2 mutations. Comparing the number of conserved T cell epitopes between the *Maf* lineages, revealed that T cell epitopes of the 147 L5 genomes were less conserved compared to the 67 L6 genomes ($X^2 = 20.82$, $p < 0.0001$). However, there was no significant difference between the number of T cell epitopes with mutations ($p > 0.05$) in L5 compared to L6 when 10 bootstrap samples of L5 of equal size as L6 were compared to L6 (Fig 4.6B). In addition, comparing the number of polymorphic sites (non-redundant mutations) between the 2 *Maf* lineages (Fig 4.7) revealed no difference between L5 and L6 even though L6 had numerically more mutations.

Human T cell antigens and T cell epitopes with mutations between the two *Maf* lineages were compared (Fig 4.7; Table 4.2). Seventy-two T cell epitopes were found to be specifically mutated in L5 compared to 54 uniquely mutated in L6. Only 2 T cell epitopes with Immune Epitope Database (IEDB) numbers 178644 and 178609 were mutated in both lineages but at different loci with different amino acid substitutions, respectively, as A183G and G278D in L5 compared to A177V and D277N in L6 (Fig 4.7A). In addition to the T cell epitopes, there were 28 T cell antigens among the 304 experimentally characterized human T cell antigens specifically mutated in L5 compared to 19 in L6 (Fig

4.8B). These include Rv0288 (encoding low molecular weight antigen 7 EsxH involved with cell wall and cell processes), Rv0934 (encoding periplasmic phosphate-binding lipoprotein PstS1 involved with cell wall and cell processes), Rv2029c (encoding 6-phosphofructokinase PfkB involved with intermediate metabolism and respiration), Rv2627c (encoding a conserved hypothetical protein), Rv3003c (encoding the large subunit of acetolactate synthase involved with valine and isoleucine biosynthesis), Rv3024c (encoding a probable tRNA involved with information pathways), Rv3763 (encoding a 19 kDa lipoprotein antigen precursor LpqH involved with cell wall and cell processes), Rv3804c (encoding the secreted antigen 85-a FbpA involved with lipid metabolism) and Rv3823c (encoding conserved integral membrane transport protein MmpL8 involved with cell wall and cell processes). The remaining antigens were Rv3825c (encoding polyketide synthase Pks2 involved with genes for lipid metabolism), Rv3879c (encoding ESX-1 secretion-associated protein EspK involved with genes for cell wall and cell processes) and Rv3883c (encoding membrane-anchored myosin MycP1 involved with genes for intermediate metabolism and respiration) that were independently mutated at unrelated T cell epitopes regions in both lineages (Table 4.2).



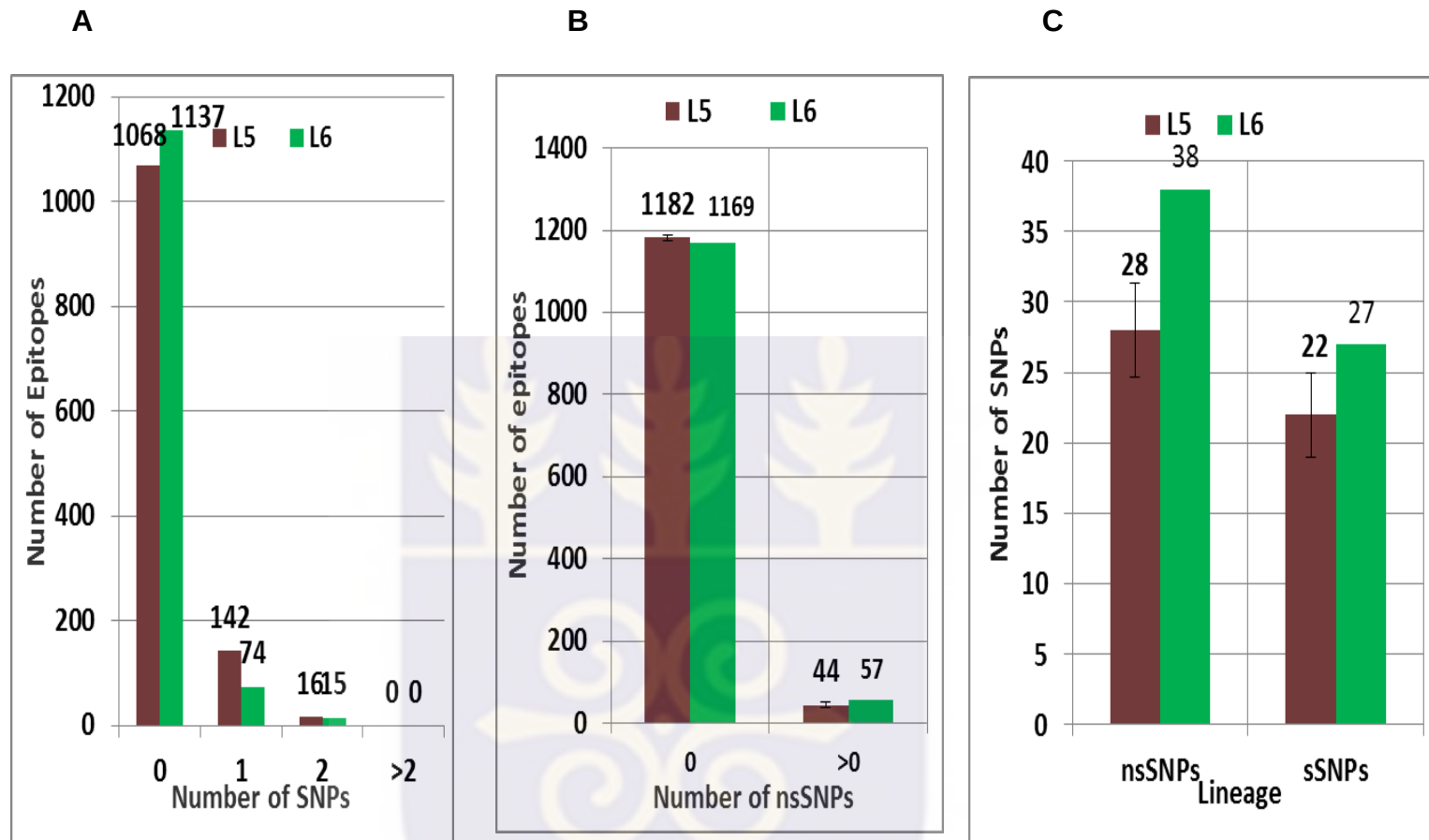


Figure 4.6: Human T cell antigenic and T cell epitope variation among the *Maf* lineages.

A: Total number of conserved and mutated T cell epitopes by lineage. B: Number of conserved and mutated T cell epitopes by random sampling. C: Number of non-redundant SNPs within T cell epitopes of Maf lineages.

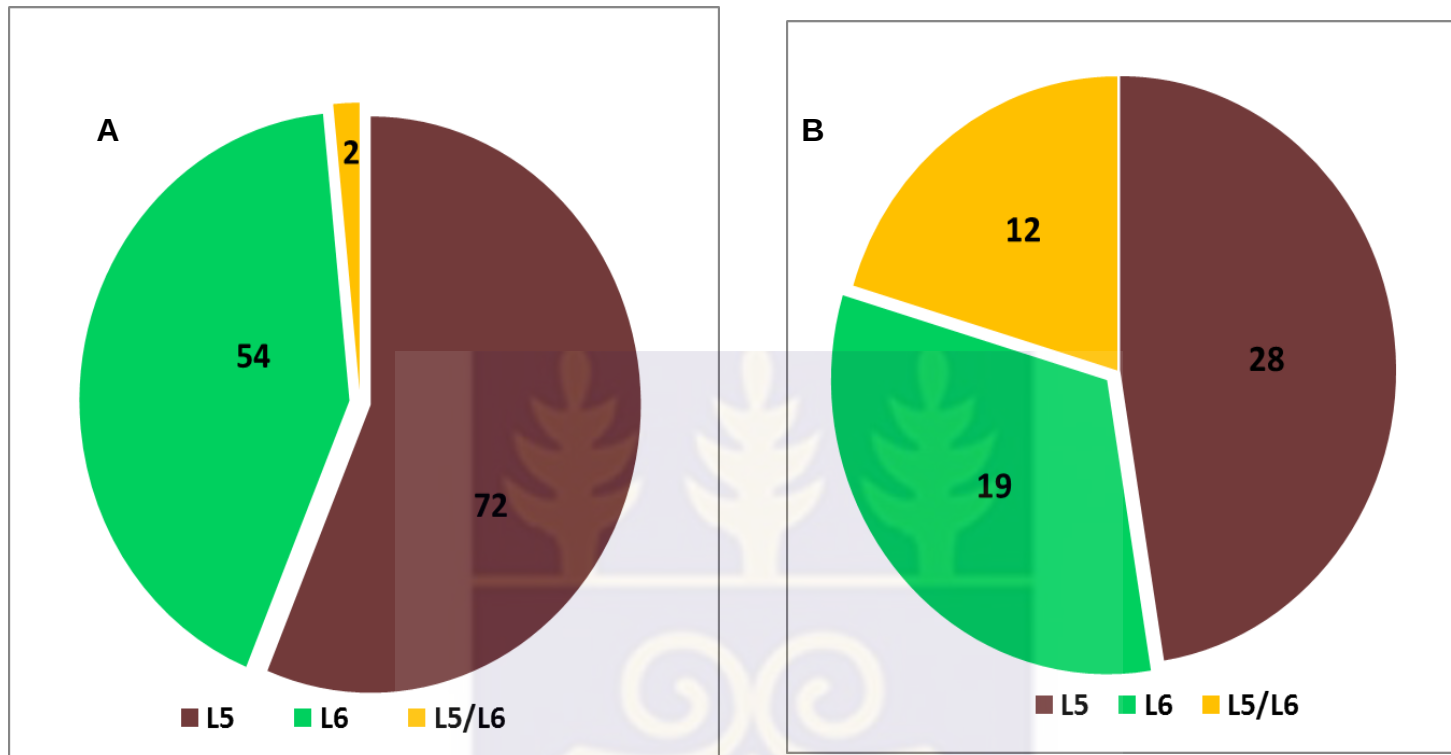


Figure 4.7: Maf lineage-specific variation of T cell epitopes. Number of human T cell epitopes (A) and human T cell antigens (B) with amino acid substitutions stratified by Maf lineage. Green represents L6-specific mutant antigens or T cell epitopes. Brown represents L5-specific mutant antigens or T cell epitopes. Yellow represents antigens or T cell epitopes mutated in both L5 and L6 but at different loci with different amino acid substitutions.

Table 4.2: Human T cell antigens and T cell epitopes with amino acid substitutions

MTBC Lineage	Gene (Antigen)	IEDB ID of T cell Epitope	Mutation	Gene product	Functional category of Gene
L6	<i>Rv0206c</i>	144970	L336V	Possible conserved transmembrane transport protein MmpL3	cell wall and cell processes
L5	<i>Rv0222</i>	65198	S241C	enoyl-CoA hydratase EchA1	lipid metabolism
L5	<i>Rv0288</i>	3196	M81V	Low molecular weight protein antigen 7 EsxH	cell wall and cell processes
L6	<i>Rv0288</i>	140717	P9L		
L6	<i>Rv0288</i>	12402	V34M		
L5	<i>Rv0289</i>	178694	P91Q	ESX-3 secretion-associated protein EspG3	cell wall and cell processes
L6	<i>Rv0290</i>	178433	S266R	ESX conserved component EccD3	cell wall and cell processes
L6	<i>Rv0290</i>	178957	R370C		
L6	<i>Rv0290</i>	179031	I456T		
L5	<i>Rv0291</i>	178530	V443A	Probable membrane-anchored mycosin MycP3	intermediary metabolism and respiration
L6	<i>Rv0292</i>	178499	A75V	ESX conserved component EccE3	cell wall and cell processes
L5	<i>Rv0294</i>	179284	P4S	Probable trans-aconitate methyltransferase Tam	intermediary metabolism and respiration
L6	<i>Rv0299</i>	178736	H15Y	Hypothetical protein	conserved hypotheticals
L5	<i>Rv0522</i>	178563	V231D	GAMA-aminobutyrate permease	cell wall and cell processes
L5	<i>Rv0529</i>	178614	P238S	Possible cytochrome C-type biogenesis protein CcsA	intermediary metabolism and respiration
L6	<i>Rv0853c</i>	179057	Q219R	Probable pyruvate or indole-3-pyruvate decarboxylase Pdc	intermediary metabolism and respiration

L5	<i>Rv0934</i>	9923	P356L	Periplasmic phosphate-binding lipoprotein PstS1	cell wall and cell processes
MTBC Lineage	Gene (Antigen)	IEDB ID of T cell Epitope	Mutation	Gene product	Functional category of Gene
L6	<i>Rv0934</i>	44020	A152D		
L6	<i>Rv0934</i>	3505	P310L		
L6	<i>Rv0993</i>	178909	V70A	UTP-glucose-1-phosphate uridylyltransferase GalU	intermediary metabolism and respiration
L5	<i>Rv1028c</i>	178770	F432L	Probable sensor protein KdpD	regulatory proteins
L6	<i>Rv1045</i>	179449	H93Y	Hypothetical protein	conserved hypotheticals
L5	<i>Rv1174c</i>	226572	A84T	Low molecular weight T cell antigen TB8.4	cell wall and cell processes
L5	<i>Rv1242</i>	16879	S53R	Possible toxin VapC33	virulence, detoxification, adaptation
L5	<i>Rv1255c</i>	55530	R152H	Probable transcriptional regulatory protein	regulatory proteins
L5	<i>Rv1503c</i>	178753	E54Q	Conserved hypothetical protein	conserved hypotheticals
L6	<i>Rv1617</i>	144897	G161E	Probable pyruvate kinase PykA	intermediary metabolism and respiration
L5	<i>Rv1866</i>	120392	C168S	Conserved protein	lipid metabolism
L5	<i>Rv1871c</i>	178713	R81L	Conserved protein	conserved hypotheticals
L5	<i>Rv1885c</i>	178607	G27S	Chorismate mutase	intermediary metabolism and respiration
L5	<i>Rv1886c</i>	76455	Y306C	Secreted antigen 85-B FbpB (85B)	lipid metabolism
L5	<i>Rv1886c</i>	181733	G18R		
L6	<i>Rv1908c</i>	125317	W328L	Catalase-peroxidase-peroxynitritase T KatG	virulence, detoxification, adaptation

L6	<i>Rv1985c</i>	179874	R256P	Probable transcriptional regulatory protein	regulatory proteins
L6	<i>Rv1985c</i>	179851	A188V		
L5	<i>Rv1987</i>	38245	T13P	Possible chitinase	cell wall and cell processes
MTBC Lineage	Gene (Antigen)	IEDB ID of T cell Epitope	Mutation	Gene product	Functional category of Gene
L5	<i>Rv2029c</i>	154044	L297R	6-phosphofructokinase PfkB	intermediary metabolism and respiration
L5	<i>Rv2029c</i>	154733	H59Y		
L6	<i>Rv2029c</i>	155082	S77I		
L6	<i>Rv2094c</i>	179580	P61L	Sec-independent protein translocase membrane-bound protein TatA	cell wall and cell processes
L5	<i>Rv2576c</i>	178470	H91P	Possible conserved membrane protein	cell wall and cell processes
L5	<i>Rv2627c</i>	144882	R188G	Conserved protein	conserved hypotheticals
L6	<i>Rv2627c</i>	179452	T144M		
L5	<i>Rv2628</i>	106659	R5Q & P6L	Hypothetical protein	conserved hypotheticals
L5	<i>Rv2628</i>	106713	P6L		
L5	<i>Rv2628</i>	106537	A35V		
L5	<i>Rv2628</i>	106364	P103S		
L5	<i>Rv2819c</i>	179215	E122A	Hypothetical protein	conserved hypotheticals
L6	<i>Rv2855</i>	179616	D290Y	NADPH-dependent mycothiol reductase Mtr	intermediary metabolism and respiration
L6	<i>Rv2865</i>	178778	P58T	Antitoxin RelF	virulence, detoxification, adaptation
L6	<i>Rv2873</i>	153271	V197A	Cell surface lipoprotein Mpt83	cell wall and cell processes
L6	<i>Rv2874</i>	179117	I130T	Possible integral membrane C-type cytochrome biogenesis protein DipZ	intermediary metabolism and respiration

L6	<i>Rv2874</i>	179424	S436P		
L5	<i>Rv2875</i>	20744	D32G	Major secreted immunogenic protein Mpt70	cell wall and cell processes
L5	<i>Rv2958c</i>	189566	L292S	Possible glycosyl transferase	intermediary metabolism and respiration
MTBC Lineage	Gene (Antigen)	IEDB ID of T cell Epitope	Mutation	Gene product	Functional category of Gene
L6	<i>Rv2994</i>	120878	M65I & G76S	Probable conserved integral membrane protein	cell wall and cell processes
L5	<i>Rv3003c</i>	178644	A183G	Acetolactate synthase (large subunit)	Involved in valine and isoleucine biosynthesis
L6	<i>Rv3003c</i>	178644	A177V		
L5	<i>Rv3010c</i>	178438	Y304H	Probable 6-phosphofructokinase PfkA	intermediary metabolism and respiration
L6	<i>Rv3020c</i>	178429	A56V	ESAT-6 like protein EsxS	cell wall and cell processes
L5	<i>Rv3024c</i>	178477	D80A	Probable tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase TrmU	information pathways
L6	<i>Rv3024c</i>	178496	E305G		
L5	<i>Rv3296</i>	142233	A32G	Probable ATP-dependent helicase Lhr	information pathways
L5	<i>Rv3333c</i>	70013	T103A	Hypothetical proline rich protein	conserved hypotheticals
L5	<i>Rv3617</i>	178550	L31R	Probable epoxide hydrolase EphA	virulence, detoxification, adaptation
L5	<i>Rv3634c</i>	178744	T159I	UDP-glucose 4-epimerase GalE1	intermediary metabolism and respiration
L5	<i>Rv3763</i>	105967	S20P	19 kDa lipoprotein antigen precursor LpqH	cell wall and cell processes

L6	<i>Rv3763</i>	18914	T40A		
L5	<i>Rv3775</i>	179234	W192S	Probable lipase LipE	intermediary metabolism and respiration
L5	<i>Rv3804c</i>	21482	N266D	Secreted antigen 85-a FbpA	lipid metabolism
L6	<i>Rv3804c</i>	195727	T145A		
L5	<i>Rv3823c</i>	178473	N946H	Conserved integral membrane transport protein MmpL8	cell wall and cell processes
MTBC Lineage	Gene (Antigen)	IEDB ID of T cell Epitope	Mutation	Gene product	Functional category of Gene
L5	<i>Rv3823c</i>	144909	P81S		
L6	<i>Rv3823c</i>	179015	P454A		
L5	<i>Rv3825c</i>	178609	G278D	Polyketide synthase Pks2	lipid metabolism
L6	<i>Rv3825c</i>	178609	D277N		
L5	<i>Rv3859c</i>	178759	A1177G	Probable ferredoxin-dependent glutamate synthase	intermediary metabolism and respiration
L6	<i>Rv3874</i>	34928	Q28P	10 kDa culture filtrate antigen EsxB (LHP) (CFP10)	cell wall and cell processes
L6	<i>Rv3878</i>	20015	G185A	ESX-1 secretion-associated protein EspJ	cell wall and cell processes
L5	<i>Rv3879c</i>	144864	A685T & A684D	ESX-1 secretion-associated protein EspK	cell wall and cell processes
L6	<i>Rv3879c</i>	120976	T143S		
L5	<i>Rv3883c</i>	144921	A12V	Membrane-anchored mycosin MycP1	intermediary metabolism and respiration
L6	<i>Rv3883c</i>	120428	L375P		

4.6 CONSERVATION OF HUMAN T CELL EPITOPES IN *MAF* IS INDEPENDENT OF PATIENT ETHNICITY

Pairwise dN/dS was estimated for T cell epitopes and regulatory proteins of L5 and compared by patient ethnicity as Ewe and Non-Ewe (Fig 4.8) using 10 bootstrap samples of L5 (size = 15, with replacement) from non-Ewe TB patients compared to the 15 L5 genomes from Ewe TB patients. The dN/dS of T cell epitopes were below 1.0 irrespective of patient ethnicity (Fig 4.8A). Similarly, the dN/dS of regulatory proteins were above 1.0 irrespective of patient ethnicity (Fig 4.8B). There was no difference between the estimated dN/dS of T cell epitopes stratified by patient ethnicity ($p = 0.1146$). Furthermore, dN/dS of regulatory proteins was also similar ($p = 0.9470$) between L5 from Ewe and non-Ewe TB patients. To confirm that the L5 strains isolated from patients of specific ethnicities making up the “other” as a group are not confounding the analysis with T cell epitopes, L5 genomes from patients of each major ethnicity (with more than 5 genomes) were independently compared to those from patients of the Ewe ethnicity as well as using the total number of genomes in each category (Ewe and non-Ewe ethnicities) with L6 as a control (Fig 4.9). Again we found no difference between dN/dS of T cell epitopes of L5 from Ewe and non-Ewe TB patients (Fig 4.9A).

In addition, there was no difference in the pairwise dN/dS of T cell epitopes between L5 genomes from TB patients of Ewe and Akan, Ewe and Ga and Ewe and other minor ethnic descents (Fig 4.9B). Furthermore, there was no difference in neither the number of T cell epitopes with non-synonymous SNPs amino acid substitutions (Fig 4.10A) nor the number of non-redundant SNPs (Fig 4.10B between L5 strains from patients of Ewe ethnicity and those of

other ethnicities using 10 bootstrap samples (size = 15, with replacement) of L5 from non-Ewe TB patients against the 15 L5 from Ewe TB patients. Furthermore, when patient ethnicity was mapped unto the *Maf* phylogenetic tree, there was no Ewe ethnicity specific cluster of *Maf* seen as the L5 from Ewe TB patients were uniformly distributed among the L5 clade (Fig 4.11).

Lastly, dN/dS of T cell epitopes was analyzed and compared between L6 from HIV negative and HIV positive TB patients using L5 as control (Fig 4.12). Both groups of L6 genomes had dN/dS greater than 1.0. Even though, the estimate for L6 among HIV positive patients was relatively higher compared to that for L6 from HIV negative patients, the difference was not statistically significant but rather suggestive ($p = 0.0546$).

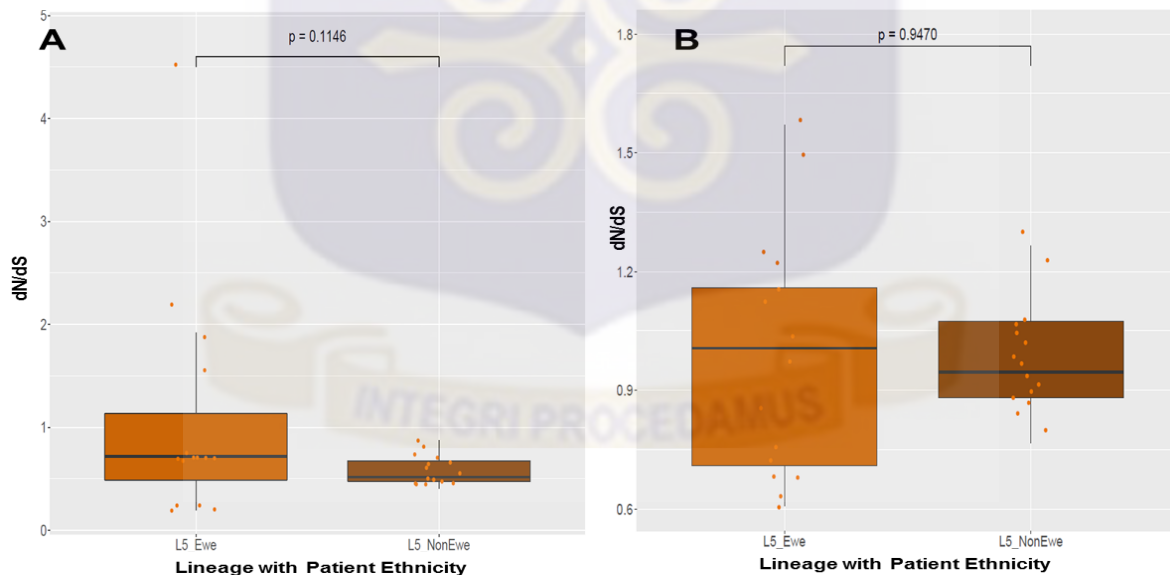


Figure 4.8: Pairwise dN/dS T cell epitopes (A) and regulatory proteins (B) of L5 by patient ethnicity. *L5 genomes from strains isolated from patients of Ewe ethnicity (15 genomes) against, average values of 10 random samples of size 15 of L5 genomes of isolates from non-Ewe patients.*

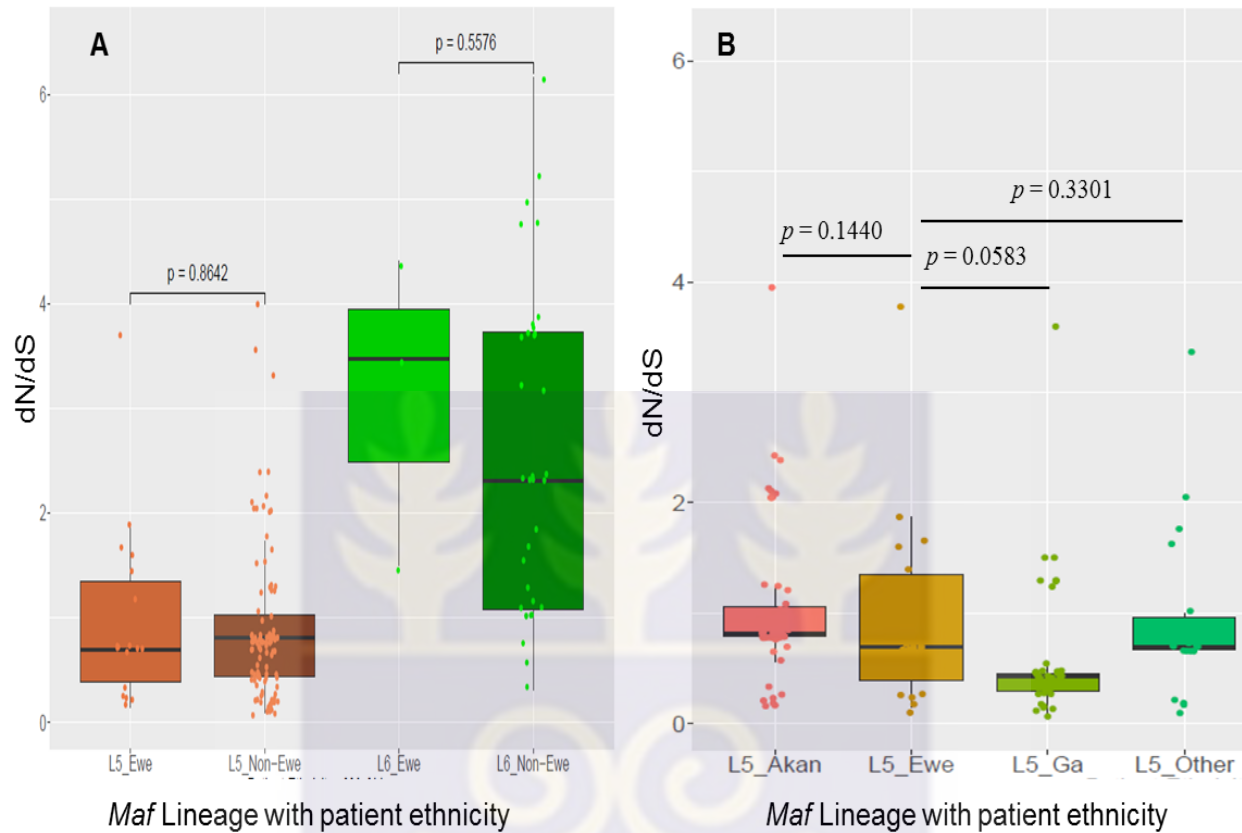


Figure 4.9: Pairwise dN/dS of T cell epitopes of L5 *Maf* lineages stratified by specific patient ethnicity.

A: Controlled analysis using the total number of L5 genomes stratified by Ewe and Non-Ewe ethnicities.

B: dN/dS analysis of L5 genomes stratified by the specific individual patient ethnicities.

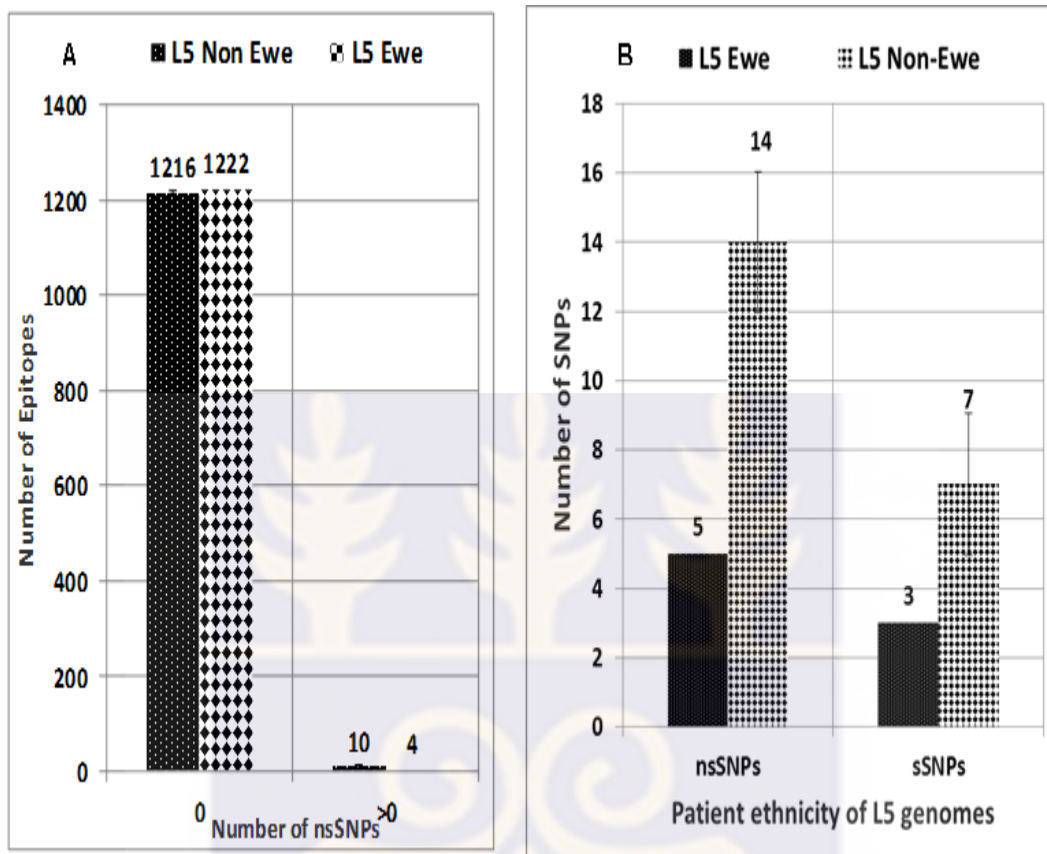


Figure 4.10: T cell antigenic variation stratified by ethnicity of L5 TB patients

Number of T cell epitopes with and without nsSNPs stratified by Lineage 5 with patient ethnicity (A). Number of non-redundant SNPs in T cell epitopes stratified by ethnicity of L5 infected patients. The values for the non-Ewe associated L5 are mean estimates of 10 random samples of size 15 with replacement.

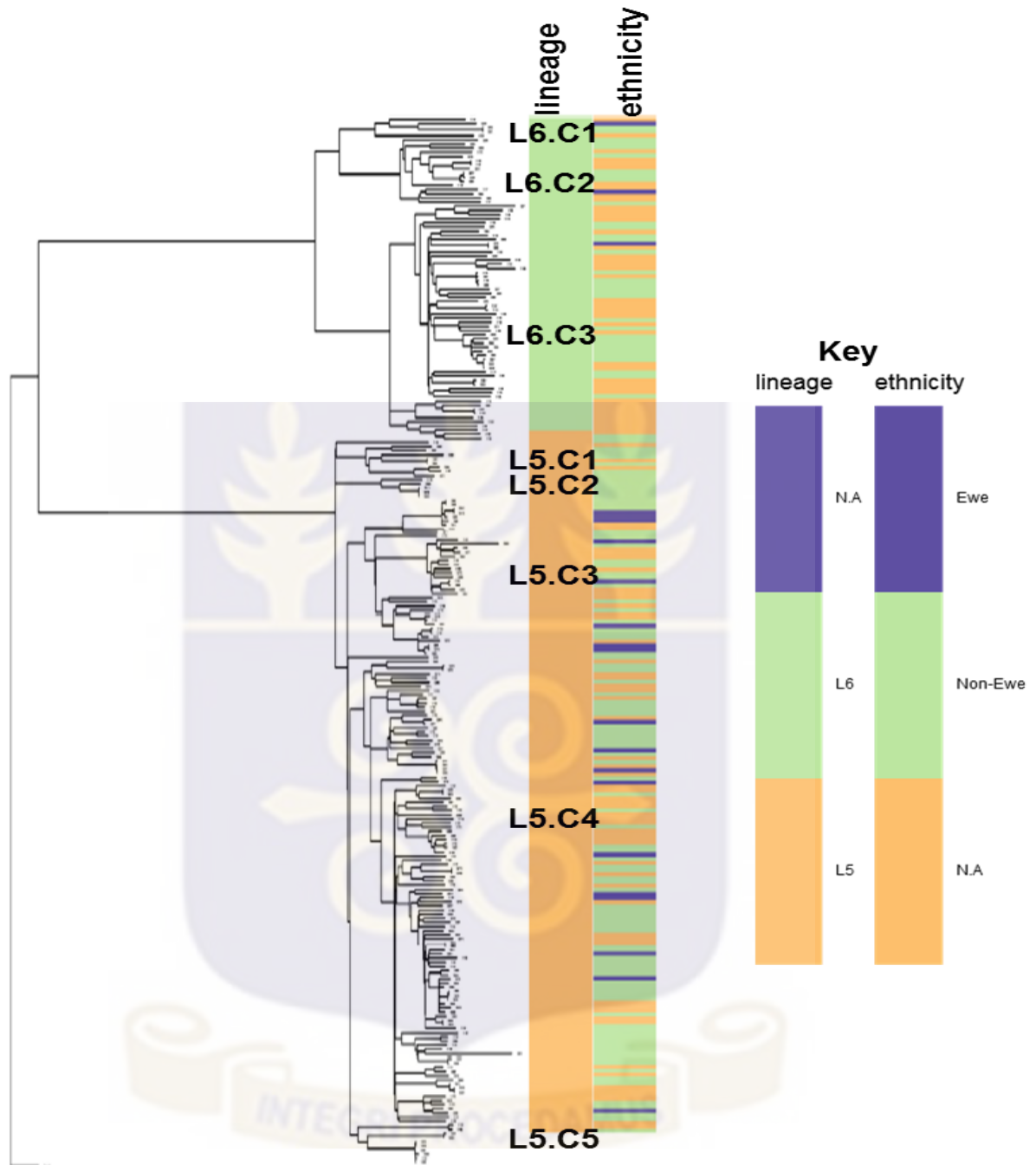


Figure 4.11: Uniform distribution of L5 isolated from ethnic Ewe TB patients among the L5 clade. No Ewe ethnicity specific clade of L5 was observed as L5 from TB patients of the Ewe ethnicity (coloured blue) were sporadically distributed along the phylogeny.

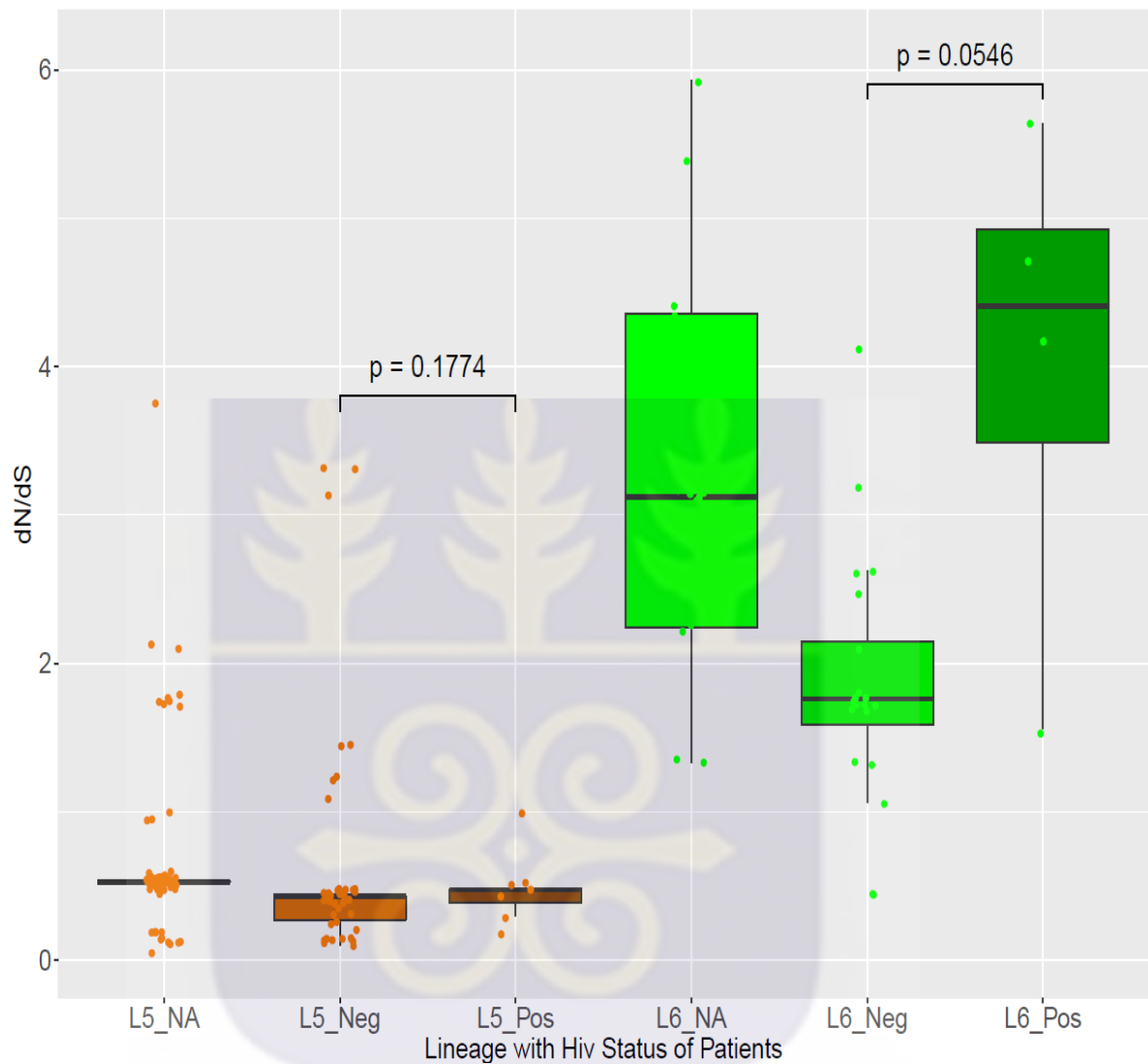


Figure 4.12: Pairwise dN/dS of T cell epitopes of L6 stratified by patient HIV status.

No significant difference ($p > 0.05$) between of dN/dS of T cell epitopes of L5 (Left panel in brown) and L6 (right panel in green) from HIV negative and positive TB patients.

4.7 COMPARATIVE GENOMIC ANALYSIS REVEALS MTBC LINEAGE-SPECIFIC GENOME DEGRADATION

There were 5,947 pan genes (total number of unique genes) among the MTBC genomes, out of which 3,215 (54.1%) were core genes (present in every genome among the genomes) whereas 719 (12.1%) genes were found in single genomes (genome specific genes). Duplication of genes such as of PE-PGRS/PPE genes (1281) and other genes (431) contributed to the large genome. Excluding these genes leaves a pan-genome of 4,235 genes comparable to the established genome size of the *Mtbss* H37Rv (Cole *et al.*, 1998). There were approximately 4015, 4035 and 3992 average number of genes found in Lineages 4, 5 and 6, respectively, with different number of lineage-specific core (genes shared among over 99% of genomes), soft-core (genes shared among 95% to 99% of genomes), shell (genes shared among 15% to 95% of genomes) and cloud (genes shared among less than 15% of genomes) genes (Fig 4.13). The genome degradation was more pronounced in L4 and L6 as compared to L5 (Fig 4.14) with a number of lineage-specific deletion of some genes.

For instance, 5 genes namely: *mldD* (encoding a mammalian cell entry protein), *mymT* (encoding a virulence, adaptation and detoxification protein Metallothionein MymT), *rho* (encoding transcription factor Rho), *bglB* (encoding a beta-glucosidase) and *mpt64* (an immunogenic protein MPT64) were specifically deleted from all Lineage 6 genomes. In a similar way, 5 genes: *apr* (encoding membrane-anchored mycosin mycP4), *desA3* (encoding linoleoyl-CoA desaturase), *cyp142b* (encoding cytochrome P450 monooxygenase), *aldA* (encoding L-alanine dehydrogenase) and *cobLa* (encoding

precorrin-6y methyltransferase) found in L5 and L6 were all missing from L 4 genomes. From the 93 L5 genomes analyzed, the only annotated gene that was uniquely absent was *ppe67* (encoding a PPE family protein with not yet known function) whereas all *Maf* genomes (L5 and L6) had *icL* (encoding a *citrate lyase*) deleted. However, there were deletions of several genes limited sub-lineage or individual strains among all the 3 lineages of the MTBC, a few of which have been summarized (Table 4.3; Appendix C).

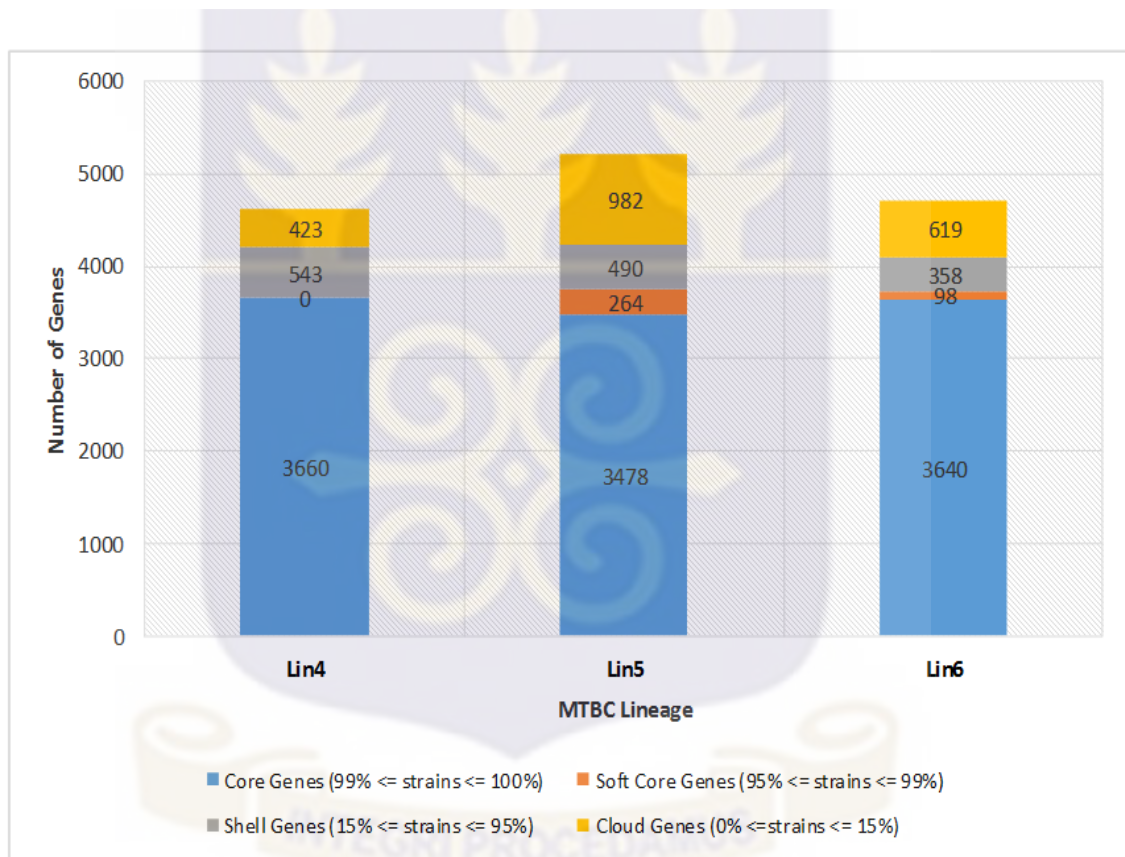


Figure 4.13: Comparative pan-genomes of the MTBC lineages.

Majority of the predicted genes were conserved or core genes among the three lineages (coloured blue).

Table 4.3: Lineage-specific deletion of genes

Gene Absent	Function	Essential	L4	L5	L6
<i>mlaD</i>	NA	No	Present	Present	Absent
<i>mymT</i>	virulence, detoxification, adaptation	No	Present	Present	Absent
<i>echA1</i>	lipid metabolism	No	Present	Present	Absent
<i>frdB_1</i>	intermediary metabolism and respiration	No	Present	Present	Absent
<i>PE32</i>	NA	No	Present	Present	Absent
<i>mce3D</i>	virulence, detoxification, adaptation	No	Present	Present	Absent
<i>rho</i>	information pathways	Yes	Present	Present	Absent
<i>mce3F</i>	virulence, detoxification, adaptation	No	Present	Present	Absent
<i>yrbE3A</i>	virulence, detoxification, adaptation	No	Present	Present	Absent
<i>bglB</i>	NA	No	Present	Present	Absent
<i>viuB</i>	intermediary metabolism and respiration	No	Present	Present	Absent
<i>arnB</i>	NA	No	Present	Present	Absent
<i>mpt64</i>	cell wall and cell processes	No	Present	Present	Absent
<i>icL</i>	intermediary metabolism and respiration	Yes	Present	Absent	Absent
<i>PPE67</i>	NA	No	Present	Absent	Present
<i>PPE56</i>	NA	Yes	Absent	Absent	Present

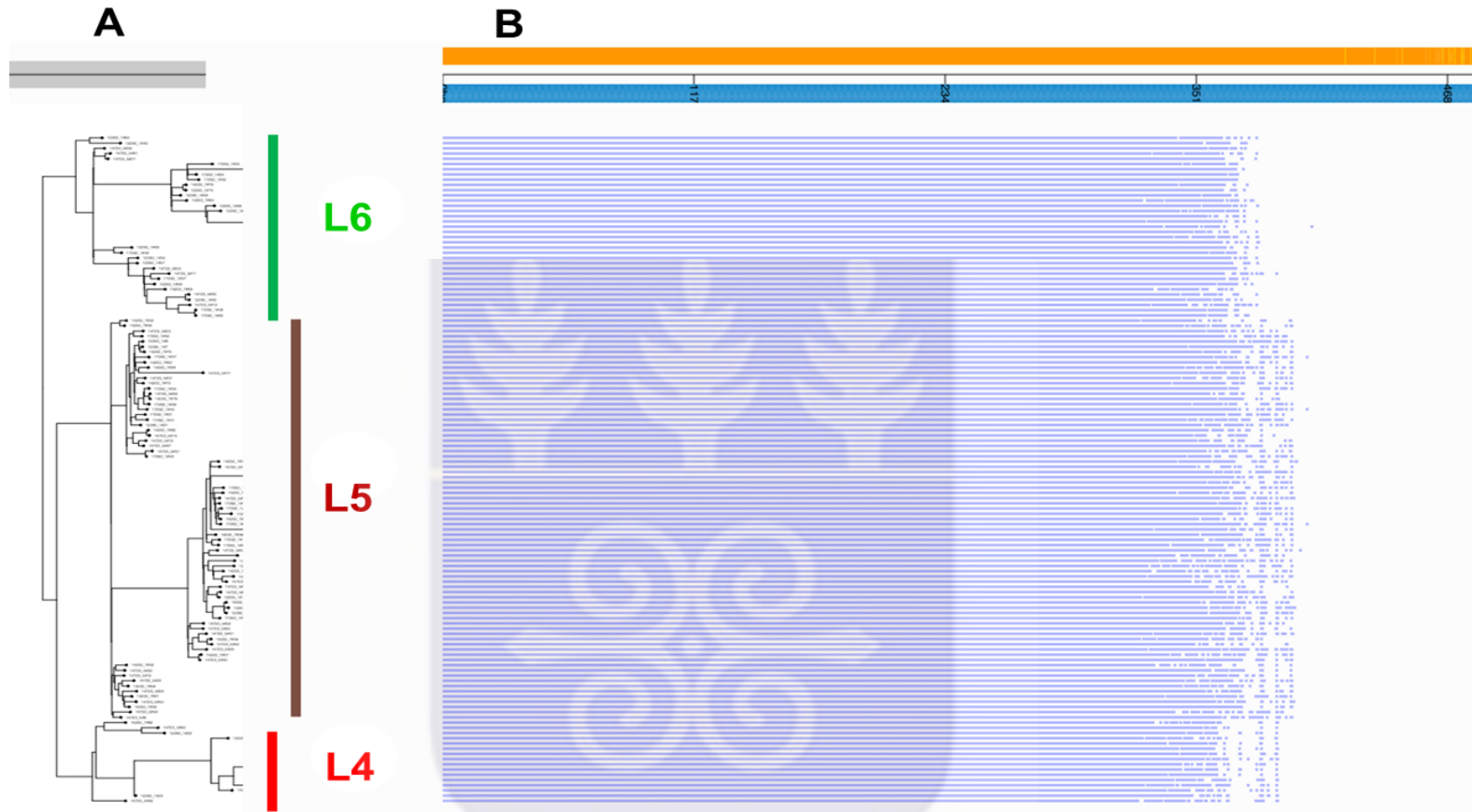


Figure 4.14: MTBC lineage-specific genome erosion.

A: Core genome maximum likelihood phylogeny of the three lineages. B: Plot of presence and absence of annotated genes along the genome. Blue dots are genes that are present in a genome whereas white spaces are absent genes. Genome erosion is more prominent in L6 (the green lineage) followed by L4 (the red lineage) whereas L5 (the brown lineage) has the biggest genome relative to the others.

4.8 DETECTION OF LINEAGE-SPECIFIC MUTATIONS WITHIN SOME FUNCTIONAL GENES

From the analyses of the selected genes (Table 3.2) previously reported to harbor lineage-specific mutations (Gehre *et al.*, 2013; Rengarajan *et al.*, 2005; Sasseti *et al.*, 2003), a number of mutations such as amino acid substitutions, insertions/deletions, frame-shifts and truncations of the peptide were identified (Appendix D). There were lineage-specific and cluster-specific mutations as well. When the mutations were grouped according to the function of the gene (Fig 4.15; Appendix E), there were relatively more mutations among L6 compared to L5 and L4. Notably, among the mutations were those among ESAT-6 secretory system II and virulence genes. Whereas there was no mutation within the ESAT-6 secretory system II that was common among L5 and L6, there was 1 mutation (G377-V378 ins AAAG) in *espE* which is part of the ESAT-6 secretory system I and 5 mutations in 3 virulence genes (I179T in *mce1B*, L370P in *mce1F* and 3 mutations I289fs, Q290fs and S291A in Rv0176) that were *Maf*-specific (Table 4.4). However, two mutations detected within an immunogenic protein mpt64 were restricted to L5. They were I43N found in all L5 strains and a frame shift mutation (P90fs) detected in 2 L5 strains.

Two lineage-specific truncations (R375trunc in *mmpL1* and S917trunc in *mmpL9*) detected were all found in L5 strains. Whereas there was no L6 specific indels within the genes analyzed, there was one L5-specific (V301-L302 del) and a *Maf* specific insertion of 4 amino acids (G377-V378 insAAAG) in *espE*. There was only one L6-specific frame-shift mutation (G488fs) which was detected within a polyketide synthase (*pks15*). However, there were two *Maf*-specific frame-shift mutations (I289fs and Q290fs) within virulence gene Rv0176. The remaining 3 frame-shift

mutations (Q458fs in Rv0171, G490fs in pks15 and P275fs in ctpJ) were all L5-specific (Table 4.4).

Lineage-specific mutations were further analyzed in detail (Fig 4.16). Most of the lineage specific mutations were among genes for cell wall and cell processes (66) followed by genes for lipid metabolism (26) whereas the group of genes with lesser number of lineage-specific mutations was genes involved with information pathways (3) followed by genes for intermediate metabolism and respiration (4). Amino acid substitutions were the most dominant mutations (107) followed by frame-shift mutations (10).

Table 4.4: Lineage-specific mutations

Gene	Name	Mutation	L4	L5	L6
<i>Rv0014c</i>	<i>pknB</i>	V251L	No	Yes	No
<i>Rv0048c</i>		G224S	No	No	No
<i>Rv0048c</i>		V250A	No	Yes	Yes
<i>Rv0048c</i>		Y269C	No	Yes	No
<i>Rv0092</i>	<i>ctpA</i>	T3A	No	No	Yes
<i>Rv0103c</i>	<i>ctpB</i>	P309L	No	No	Yes
<i>Rv0107c</i>	<i>ctpI</i>	A1327S	No	No	No
<i>Rv0107c</i>	<i>ctpI</i>	A136V	No	No	No
<i>Rv0107c</i>	<i>ctpI</i>	F157fs	No	No	No
<i>Rv0143c</i>		F34L	No	No	Yes
<i>Rv0169</i>	<i>mce1A</i>	P359S	No	No	No
<i>Rv0169</i>	<i>mce1A</i>	P401A	No	Yes	No
<i>Rv0170</i>	<i>mce1B</i>	I179T	No	Yes	Yes
<i>Rv0171</i>	<i>mce1C</i>	A433G	No	Yes	No
<i>Rv0171</i>	<i>mce1C</i>	E212D	No	No	Yes
<i>Rv0171</i>	<i>mce1C</i>	Q458fs	No	Yes	No
<i>Rv0173</i>	<i>lprK</i>	K84R	No	Yes	No
<i>Rv0174</i>	<i>mce1F</i>	L370P	No	Yes	Yes
<i>Rv0176</i>		I289fs	No	Yes	Yes
<i>Rv0176</i>		P318L	No	No	Yes
<i>Rv0176</i>		Q290fs	No	Yes	Yes

Gene	Name	Mutation	L4	L5	L6
<i>Rv0176</i>		S291A	No	Yes	Yes
<i>Rv0202c</i>	<i>mmpL11</i>	A421S	No	Yes	No
<i>Rv0206c</i>	<i>mmpL3</i>	D466E	No	No	Yes
<i>Rv0206c</i>	<i>mmpL3</i>	F384I	No	No	Yes
<i>Rv0362</i>	<i>mgtE</i>	A398S	No	Yes	No
<i>Rv0402c</i>	<i>mmpL1</i>	L451P	No	No	No
<i>Rv0402c</i>	<i>mmpL1</i>	P568fs	No	No	No
<i>Rv0402c</i>	<i>mmpL1</i>	R376trunc	No	Yes	No
<i>Rv0405</i>	<i>pks6</i>	A456fs	No	No	No
<i>Rv0405</i>	<i>pks6</i>	D578G	No	Yes	No
<i>Rv0405</i>	<i>pks6</i>	G1262R	No	No	Yes
<i>Rv0405</i>	<i>pks6</i>	R1402P	No	Yes	Yes
<i>Rv0410c</i>	<i>pknG</i>	A665T	No	Yes	No
<i>Rv0410c</i>	<i>pknG</i>	T736K	No	No	Yes
<i>Rv0425c</i>	<i>ctpH</i>	E222A	No	Yes	Yes
<i>Rv0425c</i>	<i>ctpH</i>	S888N	No	Yes	No
<i>Rv0450c</i>	<i>mmpL4</i>	H825Y	No	Yes	No
<i>Rv0507</i>	<i>mmpL2</i>	D509Y	No	No	Yes
<i>Rv0507</i>	<i>mmpL2</i>	S722C	No	No	Yes
<i>Rv0522</i>	<i>gabP</i>	R307L	No	Yes	No
<i>Rv0557</i>	<i>mgtA</i>	G304D	No	Yes	No
<i>Rv0557</i>	<i>mgtA</i>	T74M	No	Yes	No
<i>Rv0642c</i>	<i>mmaA4</i>	F95L	No	No	Yes
<i>Rv0676c</i>	<i>mmpL5</i>	T794I	No	Yes	Yes
<i>Rv0820</i>	<i>phoT</i>	F35L	No	No	No
<i>Rv0828c</i>		P62L	No	Yes	No
<i>Rv0862c</i>		D160E	No	No	Yes
<i>Rv0862c</i>		L749S	No	Yes	Yes
<i>Rv0931c</i>	<i>pknD</i>	H368Y	No	Yes	No
<i>Rv0931c</i>	<i>pknD</i>	L376fs	No	No	No
<i>Rv0912</i>		S115A	No	Yes	Yes
<i>Rv0933</i>	<i>pstB</i>	L64fs	No	No	No
<i>Rv0934</i>	<i>pstS1</i>	V352A	No	No	Yes
<i>Rv0969</i>	<i>ctpV</i>	D484N	No	Yes	Yes
<i>Rv1030</i>	<i>kdpB</i>	T66I	No	No	Yes
<i>Rv1030</i>	<i>kdpB</i>	V304A	No	No	Yes
<i>Rv1030</i>	<i>kdpB</i>	V699A	No	Yes	No
<i>Rv1096</i>		P272S	No	No	Yes
<i>Rv1125</i>		S101G	No	Yes	Yes

Gene	Name	Mutation	L4	L5	L6
<i>Rv1145</i>	<i>mmpL13a</i>	A276fs	No	Yes	Yes
<i>Rv1178</i>		A318V	No	Yes	Yes
<i>Rv1181</i>	<i>pks4</i>	D505A	No	No	No
<i>Rv1181</i>	<i>pks4</i>	L1253F	No	No	Yes
<i>Rv1181</i>	<i>pks4</i>	Q473R	No	Yes	No
<i>Rv1323</i>	<i>fadA4</i>	Q262H	No	No	Yes
<i>Rv1328</i>	<i>glgP</i>	D532G	No	No	No
<i>Rv1328</i>	<i>glgP</i>	V576F	No	No	No
<i>Rv1348</i>	<i>irtA</i>	A653T	No	No	Yes
<i>Rv1522c</i>	<i>mmpL12</i>	S947N	No	No	No
<i>Rv1522c</i>	<i>mmpL12</i>	T979A	No	No	Yes
<i>Rv1527c</i>	<i>pks5</i>	F1439L	No	No	No
<i>Rv1527c</i>	<i>pks5</i>	G29V	No	No	Yes
<i>Rv1527c</i>	<i>pks5</i>	T881I	No	No	Yes
<i>Rv1607</i>	<i>chaA</i>	P6T	No	No	No
<i>Rv1611</i>	<i>trpC</i>	T5I	No	No	Yes
<i>Rv1622c</i>	<i>cydB</i>	M98I	No	No	Yes
<i>Rv1660</i>	<i>pks10</i>	T276P	No	Yes	No
<i>Rv1661</i>	<i>pks7</i>	S1176P	No	No	No
<i>Rv1662</i>	<i>pks8</i>	A808V	No	No	No
<i>Rv1662</i>	<i>pks8</i>	D78Y	No	No	No
<i>Rv1662</i>	<i>pks8</i>	R1260G	No	No	Yes
<i>Rv1662</i>	<i>pks8</i>	Y1469C	No	No	No
<i>Rv1664</i>	<i>pks9</i>	A380P	No	Yes	Yes
<i>Rv1664</i>	<i>pks9</i>	A35T	No	Yes	Yes
<i>Rv1664</i>	<i>pks9</i>	L975V	No	Yes	No
<i>Rv1707</i>		R438Q	No	No	No
<i>Rv1707</i>		S272A	No	Yes	Yes
<i>Rv1811</i>	<i>mgtC</i>	A107V	No	No	Yes
<i>Rv1811</i>	<i>mgtC</i>	G187E	No	No	No
<i>Rv1859</i>	<i>modC</i>	T313M	No	No	Yes
<i>Rv1979c</i>		D286H	No	No	Yes
<i>Rv1979c</i>		T35A	No	No	No
<i>Rv1980c</i>	<i>mpt64</i>	I43N	No	Yes	No
<i>Rv2127</i>	<i>ansP1</i>	G44S	No	No	No
<i>Rv2127</i>	<i>ansP1</i>	G9D	No	Yes	Yes
<i>Rv2127</i>	<i>ansP1</i>	N361K	No	No	Yes
<i>Rv2241</i>	<i>aceE</i>	A777T	No	No	Yes
<i>Rv2287</i>	<i>yjcE</i>	V520A	No	Yes	No

Gene	Name	Mutation	L4	L5	L6
<i>Rv2329c</i>	<i>narK1</i>	S387A	No	No	No
<i>Rv2339</i>	<i>mmpL9</i>	A426E	No	Yes	No
<i>Rv2339</i>	<i>mmpL9</i>	A44V	No	No	No
<i>Rv2339</i>	<i>mmpL9</i>	R612L	No	Yes	No
<i>Rv2339</i>	<i>mmpL9</i>	S917trunc	No	Yes	No
<i>Rv2339</i>	<i>mmpL9</i>	W827trunc	No	No	No
<i>Rv2383c</i>	<i>mbtB</i>	L978F	No	No	Yes
<i>Rv2383c</i>	<i>mbtB</i>	R263Q	No	No	No
<i>Rv2400c</i>	<i>subI</i>	R166L	No	No	Yes
<i>Rv2524c</i>	<i>fas</i>	C2771R	No	Yes	Yes
<i>Rv2524c</i>	<i>fas</i>	G672A	No	No	Yes
<i>Rv2524c</i>	<i>fas</i>	L400F	No	No	No
<i>Rv2524c</i>	<i>fas</i>	P1951S	No	Yes	No
<i>Rv2691</i>	<i>ceoB</i>	Q132P	No	Yes	No
<i>Rv2691</i>	<i>ceoB</i>	W139S	No	Yes	No
<i>Rv2692</i>	<i>ceoC</i>	I133V	No	Yes	Yes
<i>Rv2737c</i>	<i>recA</i>	D355Y	No	Yes	No
<i>Rv2737c</i>	<i>recA</i>	Q566P	No	No	Yes
<i>Rv2883c</i>	<i>pyrH</i>	P174L	No	No	No
<i>Rv2883c</i>	<i>pyrH</i>	S111I	No	No	No
<i>Rv2869c</i>	<i>rip</i>	V259I	No	Yes	Yes
<i>Rv2877c</i>		Y35H	No	Yes	Yes
<i>Rv2942</i>	<i>mmpL7</i>	G574S	No	Yes	No
<i>Rv2946c</i>	<i>pks1</i>	A1360V	No	No	No
<i>Rv2946c</i>	<i>pks1</i>	V1283G	No	No	No
<i>Rv2947c</i>	<i>pks15</i>	G374R	No	Yes	Yes
<i>Rv2947c</i>	<i>pks15</i>	G488fs	No	No	Yes
<i>Rv2947c</i>	<i>pks15</i>	G490fs	No	Yes	No
<i>Rv2947c</i>	<i>pks15</i>	V301 - L302 del	No	Yes	No
<i>Rv2955c</i>		R27H	No	No	No
<i>Rv2956</i>		G135D	No	Yes	No
<i>Rv2956</i>		I237T	No	No	No
<i>Rv3082c</i>	<i>virS</i>	R322C	No	No	No
<i>Rv3200c</i>		H197R	No	Yes	No
<i>Rv3200c</i>		N12D	No	Yes	Yes
<i>Rv3270</i>	<i>ctpC</i>	S442A	No	Yes	No
<i>Rv3282</i>		A133S	No	No	No

fs: frame-shift mutation, *trunc*: truncation, *del*: deletion.

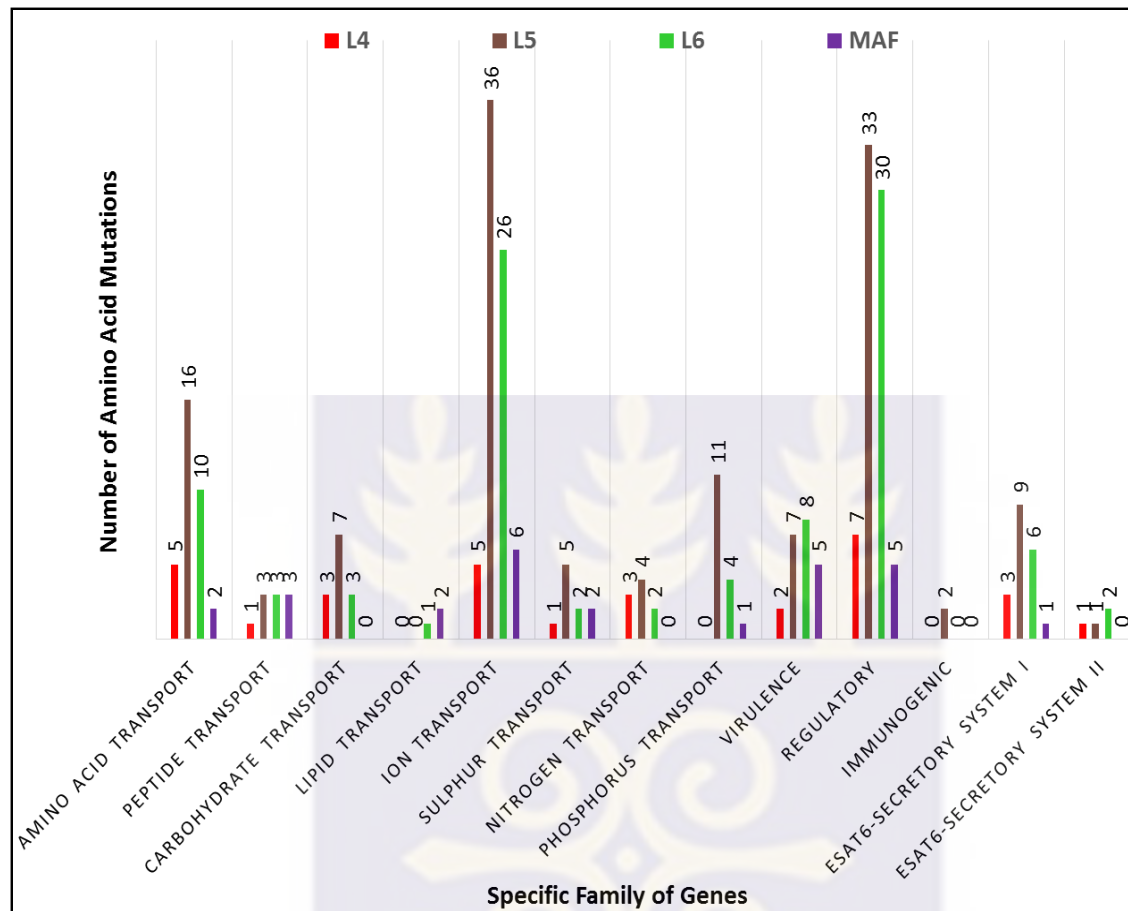


Figure 4.15: Distribution of lineage-specific mutations among genes of specific functional categories. *Maf* lineages 5 (brown) and 6 (green) had majority of the lineage-specific mutations especially among amino acid transporters, ion transporters and regulatory proteins compared to *Mtbss* L4 (red). There were limited number of *Maf*-specific mutations (purple) thus unique between L5 and L6

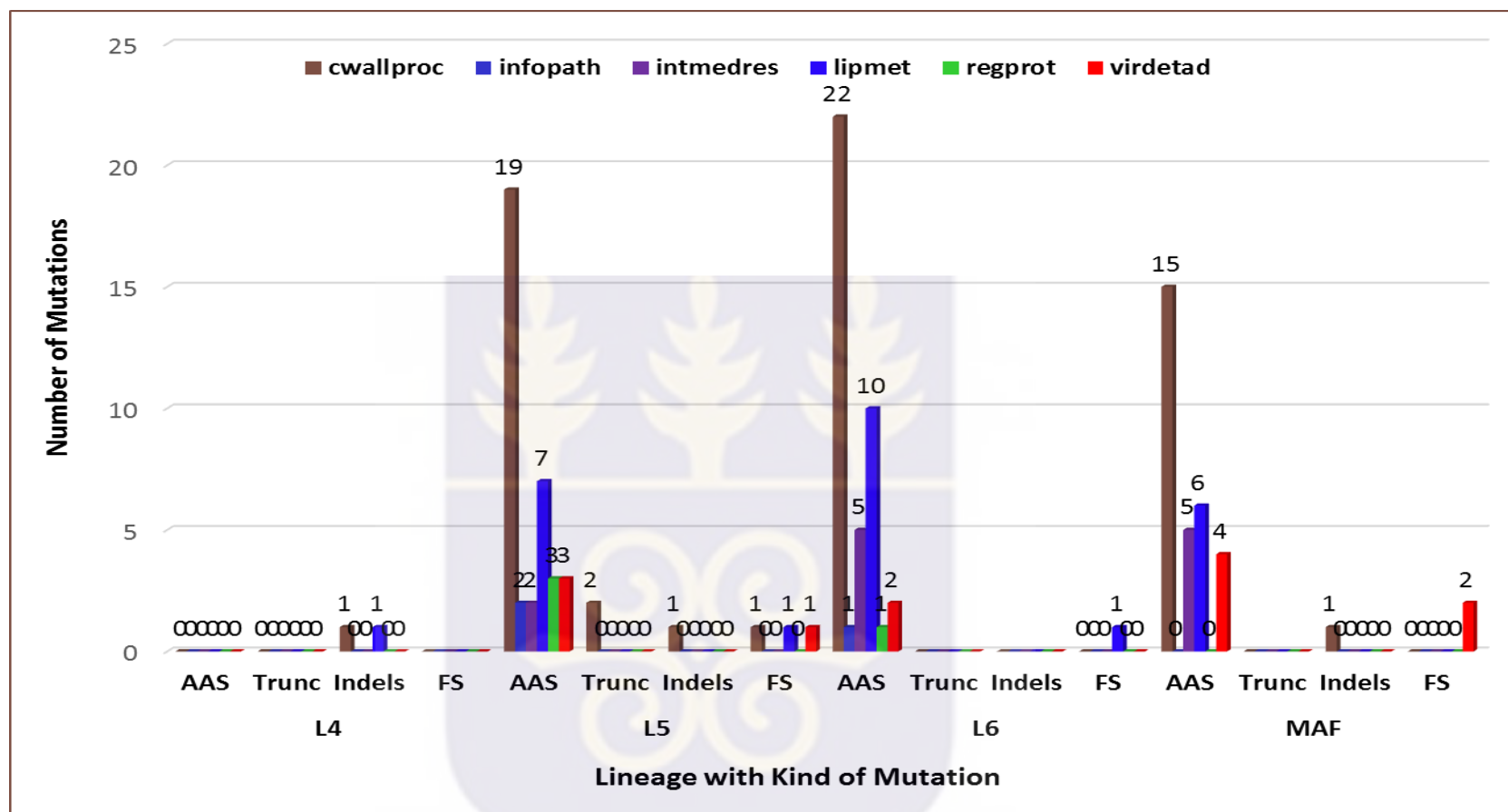


Figure 4.16: Abundance of specific mutations among the MTBC lineages.

AAS : amino acid substitutions, Trunc: truncations, Indels: Insertions and deletions, FS: Frame-shift mutations

cwallproc: genes for cell wall and cell processes

inforpath: information pathway genes

intmedres: genes for intermediate metabolism and respiration

lipmet: genes for lipid metabolism

regprot: regulatory protein genes

virdetad: genes for virulence, detoxification and adaptation

4.9 DRUG RESISTANT TB IN GHANA

One thousand four hundred and ninety isolates were obtained from 1,490 TB patients and genotyped as 1,208; 81.1% *Mtbss*, 272; 18.3% *Maf* and 10; 0.7% *M. bovis* followed by testing against INH and RIF, the two backbone drugs of the DOTS regimen. One hundred and eleven (7.4%), 10; 0.7% and 40; 2.6% of the isolates were INH mono-resistant (INH^r), RIF mono-resistant (RIF^r) and multi-drug resistant (MDR) respectively. In total, 161 (10.8%) of the isolates were resistant to at least one drug (ANY^r).

4.9.1 Characteristics of Patients with Drug Resistant MTBC Isolates

One hundred and sixty one (10.8%) out of the 1,490 patients recruited had an MTBC isolate resistant to at least one of the drugs. They were 114 (70.8%) males and 47 (29.2%) females. Drug-resistant TB patients under 50 years were 113 (70.2%) and 48 (29.8%) above 50 years. Bacterial burden of the DR-TB patients determined by direct smear microscopy were: 9 (5.6%) scanty (less than 10 AFB in 100 fields), 83 (51.6%) 1+ (between 10 and 99 bacilli in 100 fields), 29 (18.0%) 2+ (between 1 to 10 AFBs per field in at least 50 fields) and 40 (24.8%) 3+ (more than 10 AFB per field in at least 20 fields).

One hundred and forty (87.0%) of the DR-TB patients were from the South (Greater Accra region) representing 10.4% of the total TB patients recruited from the region compared to 21 (13.0%) from the North representing 15.0% of the patients recruited from the region. The likelihood of an isolate to be resistant to at least a single drug or mono-resistant to INH or RIF did not differ between the isolates from Greater-Accra and the

Northern Region of Ghana ($p = 0.0924$). However, the Northern region had a higher tendency for MDR-TB compared to the Greater Accra region (Table 4.5).

Table 4.5: Prevalence of DR-TB between the north and south of Ghana

Resistance	Total (1490) n;%	South (1350) n;%	North (140) n;%	<i>p-value</i>
INH ^r	111;7.5%	99;7.3%	12;8.6%	0.6108
RIF ^r	10;0.7%	8;0.6%	2;1.4%	0.2404
MDR	40;2.6%	33;2.4%	7;5.0%	0.0924
ANY ^r	161;10.8%	140;10.4%	21;15.0%	0.1139

Among the DR MTBC isolates, *Mtbss*, *Maf* and *M. bovis* that were INH^r were 90 (7.5%), 272 (7.0%) and 2 (10.0%), respectively. None of the *Maf* isolates were RIF^r, however, 9 (0.7%) and 1 (10.0%) of the *Mtbss* and *M. bovis* isolates, respectively, were RIF^r. On the other hand, there were 10 (3.7%) of *Maf* and 30 (2.5%) of *Mtbss* strains that were MDR. This study did not find any association between and MTBC species and resistance to any of the drugs tested (Table 4.6).

Table 4.6: The prevalence of drug resistance among MTBC species

Drug	Total (1480)	MTBSS (1208)	MAF (272)	<i>p-value</i>
INH ^r	109;7.4%	90;7.5%	19;6.9%	0.8979
RIF ^r	9;0.6%	9;0.7%	0;0.0%	0.3794
MDR	40;2.7%	30;2.5%	10;3.7%	0.2988
ANY ^r	158;10.7%	129;10.7%	29;10.7%	0.9999

INH^r: INH-mono resistant, RIF^r: RIF-mono resistant, ANY^r: Resistant to at least one drug.

Drug resistant TB was then stratified by lineages and sub-lineages of the MTBC and compared (Table 4.7). The Ghana sub-lineage of *Mtbss* L4, which is the second most dominant L4 sub-lineage in Ghana after the Cameroon sub-lineage, was found to be

highly associated with INH^r, MDR and RIF^r and ANY^r when compared to the Cameroon sub-lineage and *Maf* Lineage 6 with p-values less than 0.05 at 95% confidence level (Table 4.7). Comparing the Ghana sub-lineage to *Maf* L5, the previous was found to be statistically associated with resistance to at least a single drug ($p = 0.0472$). In a similar way, *Maf* L5 was found to be associated with INH^r, MDR and RIF^r and ANY^r when compared to *Mtbss* L4 sub-lineage Cameroon and *Maf* L6. Between the Cameroon sub-lineage of L4 and *Maf* Lineage 6, the Cameroon sub-lineage however, was more likely to be INH^r, MDR and RIF^r and ANY^r (Table 4.7).

4.9.3 Mutations in Genes Associated with INH and RIF Resistance

Mutations were found in all targets analysed but *rpoA* of RIF resistant isolates (Table 4.8). The most prominent INH-resistance conferring mutation found was *katG* S315T followed by the *inhApro* -15A/T. In addition, novel mutations G204D in *inhA* as well as K32R and V117I in *ndh* were also detected in INH resistant isolates. All the 16 isolates with the novel *inhA* mutation belonged to *Mtbss* Lineage 4 sub-lineages (11 Ghana, 3 Cameroon, 1 Uganda I and 1 LAM). As shown (Table 4.9), the proportion of *katG* mutations were significantly higher ($p = 0.0013$) among INH-resistant *Mtbss* as compared to INH-resistant *Maf*. In contrast, within the *inhApro* region, INH-resistant *Maf* strains had higher propensity to acquire mutations ($p = 0.0419$) compared to INH-resistant *Mtbss* strains. In addition, INH-resistant *ndh* mutants were more likely to be *Maf* compared to *Mtbss* ($p = 0.0011$). Among the RIF resistant strains analysed, mutations in the RRDR of the *rpoB* were identified in 95.9% (47/49) of the RIF resistant isolates. The most dominant mutation was S450L making up 51.0% (25/49) and D435V accounting for

20.4% of all the RIF resistant isolates. *Mtbss* strains with *rpoB* mutations were 38; 97.4% as compared to 8; 80.0% *Maf* ($p = 0.1018$). The only RIF^r bovine strain had two *rpoB* mutations (Q432P and I491S).

Table 4.7: Comparison of DR-TB among species lineages and sub-lineages of the MTBC

Resistance	Cameroon (806; 66.7%)	Ghana (195; 16.1%)	OR	95% CI	<i>p</i> -value
INH ^r	54; 6.7%	43; 22.05%*	0.2	0.2-0.4	0.0000
RIF ^r	19; 2.4%	12; 6.15%*	0.4	0.2-0.8	0.0103
MDR	13; 1.6%	9; 4.62%*	0.3	0.1-0.9	0.0240
ANY ^r	60; 7.4%	46; 23.59%*	0.3	0.2-0.4	0.0000
Resistance	<i>Maf</i> L5 (165; 11.1%)	Ghana (195; 16.1%)	OR	95% CI	<i>p</i> -value
INH ^r	25; 15.2%	43; 22.05%	0.6	0.3-1.1	0.1059
RIF ^r	10; 6.1%	12; 6.15%	0.9	0.4-2.6	0.9998
MDR	10; 6.1%	9; 4.62%	1.3	0.5-3.8	0.6384
ANY ^r	25; 15.2%	46; 23.59%*	0.6	0.3-1.0	0.0472
Resistance	<i>Maf</i> L6 (107; 7.2%)	Ghana (195; 16.1%)	OR	95% CI	<i>p</i> -value
INH ^r	6; 5.6%	43; 22.05%*	0.2	0.1-0.5	0.0001
RIF ^r	1; 0.9%	12; 6.15%*	0.1	0.0-1.0	0.0372
MDR	1; 0.9%	9; 4.62%	0.2	0.0-1.4	0.1040
ANY ^r	6; 5.6%	46; 23.59%*	0.2	0.1-0.5	0.0000
Resistance	<i>Maf</i> L5 (165; 11.1%)	Cameroon (806; 66.7%)	OR	95% CI	<i>p</i> -value
INH ^r	25; 9.1%	54; 6.70%	2.5	1.4-4.2	0.0008
RIF ^r	10; 6.1%	19; 2.36%	2.7	1.1-6.2	0.0202
MDR	10; 6.1%	13; 1.61%	3.9	1.5-9.9	0.0023
ANY ^r	25; 15.2%	60; 7.44%	2.2	1.3-3.7	0.0036
Resistance	<i>Maf</i> L5 (165; 11.1%)	<i>Maf</i> L6 (107; 7.2%)	OR	95% CI	<i>p</i> -value
INH ^r	25; 9.1%*	6; 5.6%	3.0	1.1-9.3	0.0183
RIF ^r	10; 6.1%	1; 0.9%	6.8	0.9-299.1	0.0546
MDR	10; 6.1%	1; 0.9%	6.8	0.9-299.1	0.0546
ANY ^r	25; 15.1%*	6; 5.6%	3.0	1.1-9.3	0.0183

INH^r: INH-mono-resistance

RIF^r: RIF-mono-resistance

MDR : Multidrug-resistance (resistance to INH and RIF)

ANY^r: Resistance to at least one drug

*Statistically significant

Table 4.8: Identified mutations among drug resistant MTBC

Drug (number of Resistant Strains)	Gene	SNP	Amino acid Change	Number of Isolates
INH (151)	<i>inhApro</i>	-8T/C	-	5;3.31%
		-15C/T	-	16;10.60%
		-15C/T & -102G/A	-	8;5.30%
		-17G/C	-	1;0.66%
		-102G/A	-	10;6.7%
		-47G/C	-	1;0.66%
	<i>katG</i>	G944C	S315T	107 (70.9%)
	<i>inhA</i>	G611A*	G204D	16;10.60%
		T233C	V78A	1;0.66%
	<i>ahpCpro</i>	-54C/T	-	1;0.66%
		-88G/A*	-	2;1.32%
		-142G/A*	-	1;0.66%
	<i>ndh</i>	G349A	V117I	4;2.64%
		C411T	G137G	1;0.66%
		A95G & G349A	K32R & V117I	1;0.66%
RIF (49)	<i>rpoB</i>	A1295C	Q432P	1;2.00%
		A1295C & T1472G	Q432P & I491S	1;2.00%
		C1294A	Q432K	1;2.00%
		A1304T	D435V	10;20.00%
		G1303T	D435Y	1;2.00%
		C1332T	S441L	1;2.00%
		A1334G	H445R	2;4.00%
		CA1333/4TG	H445C	2;4.00%
		C1333G	H445D	2;4.00%
		C1333T	H445Y	1;2.00%
		C1349T	S450L	25;49.02%
	<i>rpoC</i>	G994C*	G332R	2;4.00%
		T1448G*	V483G	3;6.00%

Novel Mutations are coloured red. *proposed compensatory mutations whereas the circled mutation may not be a stand-alone drug resistance conferring.

When the mutations were explored for lineage and sub-lineage associations, the Ghana sub-lineage had a higher propensity for acquiring *inhA* mutations (11; 25.6%) compared to 4; 7.4% of the Cameroon sub-lineage (OR = 0.24, 95%CI = 0.05-0.88, $p = 0.0220$). In addition, all the 3 strains with the novel *ndh* mutation (V117I) belonged to *Maf* L6. Furthermore, the 2 strains with the *rpoC* mutation (G332R) were of the Cameroon sub-lineage whereas the 3 with the *rpoC* (V483G) mutation were Ghana sub-lineage even though they all were RIF-resistant strains carrying the most dominant *rpoB* mutation S450L.

Table 4.9: Distinct drug target mutations and MTBC species

Drug	Target	MTBSS	MAF	OR	95% CI	p -value
INH	<i>katG</i>	91;75.8%	15;51.7%	3.73	1.56-9.09	0.0013
	<i>inhApro</i>	32;26.6%	14;48.3%	0.39	0.16-0.98	0.0419
	<i>inhA</i>	17;14.2%	1;3.45.0%	4.59	0.66-199.77	0.1999
	<i>ahpCpro</i>	2;1.7%	1;3.45.0%	0.48	0.02-28.99	0.4802
	<i>ndh</i>	1;0.8%	5;17.2%	0.04	0.00-0.39	0.0011
RIF	<i>rpoB</i>	38;97.4%	8;80.00%	8.88	0.42-572.05	0.1018
	<i>rpoC</i>	5;12.8%	0;0.0%	Inf	0.22-Inf	0.5834

NB: The RIF^r *M. bovis* had double *rpoB* mutations A1295C [Q432P] & T1472G [I491S]. The other two INH^r *M. bovis* had no mutation within the INH targets studied.

CHAPTER FIVE

DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

5.1 GENERAL DISCUSSION

The restriction of *Maf* to West Africa as opposed to the worldwide distribution of *Mtbss* remains an unsolved enigma. Findings from observational studies mainly in the Gambia where L6 is prevalent, suggested attenuation of *Maf* relative to *Mtbss* due to its prevalence among HIV patients, slow progression to disease, slow growth rate and attenuated ESAT-6 response (de Jong *et al.*, 2006, 2008; de Jong *et al.*, 2010; Gehre *et al.*, 2013; Reiling *et al.*, 2013) could mean *Mtbss* has competitive advantage over *Maf*. This relative attenuated virulence of *Maf* is thought to have contributed to its restriction to West Africa and that the more virulent *Mtbss* will eventually out-compete it. This thought coupled with the high nucleotide and amino acid similarity among the MTBC (Sreevatsan *et al.*, 1997) have contributed to the neglect of *Maf* in the search for new biomarkers for development of new drugs, vaccines and rapid diagnostic tools.

However, recent studies indicate *Maf* remains an important TB causing pathogen in West Africa causing appreciable proportions of all TB cases such as 20% in Ghana, 18% in Burkina Faso, 39% in Benin, 39% in the Gambia, 47% Guinea Bissau, 10% in Senegal and 24% in Sierra Leone (Affolabi *et al.*, 2009; de Jong *et al.*, 2010; Diallo *et al.*, 2016; Frothingham *et al.*, 1999; Groenheit *et al.*, 2011; Yeboah-Manu *et al.*, 2016; Zumla *et al.*, 2017). Furthermore, the limited efficacy of the BCG vaccine, the failure of the TB vaccine MVA85A (Tameris *et al.*, 2013) and variable sensitivity of current diagnostics, such as MPT64 based rapid kits across different geographical settings (Arora *et al.*, 2015;

Kumar *et al.*, 2011; Martin *et al.*, 2011; Ofori-Anyinam *et al.*, 2016) indicate the importance of diversity within the MTBC.

In this study, whole genome sequencing was used to analyze and compare the largest number of genomes (published to date) of the two *Maf* lineages for comparative analysis together with the L4 lineage of *Mtbss*. A number of innovative finds have come out of this study that may change the existing knowledge on MTBC, especially the 2 West African lineages. This study found for the first time that the genome of L5 relative to L6 is significantly less diverse. In addition, the study generated the first whole genome phylogeny of *Maf* in Ghana, which highlights the possibility of sub-lineages within these West African restricted pathogens. Furthermore, this study has shown that L5 has the biggest genome compared to L6 and L4 which does support L5 as an ancient lineage but challenges the classification of L6 as ancient lineage since evolution of the MTBC largely involves deletion of genetic material (Kato-Maeda *et al.*, 2001; Mostowy *et al.*, 2005; Tsolaki *et al.*, 2004). Finally, drug target analysis among the isolates established for the first time that the Ghana sub-lineage of *Mtbss* L4 is associated with drug resistant TB in the country.

5.1.1 First Insight Into the Whole Genome Phylogeny and Nucleotide Identity and SNP Distance Of *Maf* Lineages in Ghana, Implications for Taxonomy and Geographical Restriction.

The phylogeny provided in this report is the first whole genome-wide phylogeny of *Maf* with the largest number of *Maf* genomes ever published in a single report. Clustering of this phylogeny indicates the obviously neglected diversity within *Maf* and makes it

imperative that representative genomes of *Maf* from across West Africa be pooled together for in-depth analysis of this unique pathogen to define sub-lineages within each of the *Maf* lineages as has been done for L4 (Stucki *et al.*, 2016). Such sub-lineages would have implications for current and future control tools and contribute to better understanding of the biology of the pathogen.

Within this study, the ANI between any *Maf* genome (L5 or L6) and any L4 genome was always >99% which is larger than the cut off 95% for speciation (Goris *et al.*, 2007; Konstantinidis & Tiedje, 2004). This finding agrees with the findings of a recent study that also found the ANI between members of the MTBC to be always above 99% (Winglee *et al.*, 2016). Comparing this estimated ANI among the MTBC with those of other bacterial species such as *E. coli*, *S. aureus*, *S. pyogenes*, *S. enterica*, *H. pylori* and *X. fastidiosa* which were all less than 98% (Goris *et al.*, 2007; Konstantinidis & Tiedje, 2004) as well as the closest mycobacteria *M. canettii* which was between 97-98% (Supply *et al.*, 2013), further highlights the homogenous nature of the MTBC relative to other bacteria. This finding supports the call for declassification the human adapted MTBC lineages as *Mtbss* or *Maf* species. It however, advocates for the classification of the MTBC species/lineages as different sub-species or biovars of the same species.

At the whole genome level, there was no difference between the number of het-SNPs between L5 and L6 relative to the MTBC ancestor indicating no difference in ongoing evolution or within host evolution between the two lineages. However, the significant number of homo-SNPs of L6 compared to L5 relative to the ancestor reference suggests that L6 has undergone significant diversification compared to L5 after the emergence of

their most recent common ancestor (MRCA) some 6,000 to 70,000 years ago (Bos *et al.*, 2014; Comas *et al.*, 2013; Kay *et al.*, 2015; Luo *et al.*, 2015).

The number of homo-SNPs was again significantly lower in L4 compared to L5 and L6. However, this observation could be confounded by the fact that the hypothetical ancestor (Comas *et al.*, 2010) used for this analysis was mainly based on H37Rv which is the L4 laboratory strain hence the close proximity. This assertion seems to be corroborated by the observed significantly lower genome-wide pairwise average nucleotide diversity of L5 compared to L6 but not L4 to L6 (Fig 4.1B). This observation contradicts the classification of L5 and L6 as one species and challenges their combined assumption as specialist pathogens due to their geographical restriction to West Africa (de Jong *et al.*, 2010; Stucki *et al.*, 2016).

5.1.2 Lineage Specific Antigenic Variation: Implications for Specialist Pathogens

Most of the T cell epitopes were highly conserved in L4, L5 and L6 which is in line with previous reports of conservation of T cell epitopes in the MTBC (Comas *et al.*, 2010; Coscolla & Gagneux, 2014; Stucki *et al.*, 2016), T cell epitopes in L4 and L5 were found to exhibit similarly less nucleotide diversity and under purifying selection compared to L6. The purifying selection of T cell epitopes mutations within L5 is comparable to that reported for the specialist sub-lineages of L4 (Stucki *et al.*, 2016). Interestingly, dN/dS within essential genes for growth within macrophages did not differ between L5 and L6 even though it was significantly higher in L4 (Fig 4.4). However, the estimates were all below 1.0 an indication of purifying selection. This suggests that the three lineages have

conserved genes for intracellular survival within macrophages; however, L4 and L5 may induce T cell response more effectively than L6. Since efficient T cell response is essential for both protective immunity and pathogenesis of TB (Cooper, 2009), the relative conservation of human T cell epitopes in L4 and L5 is an indication that the two lineages are more human-specific pathogens compared to L6. However, it does not account for the restriction of L5 to West Africa and the global presence of L4 (de Jong *et al.*, 2010; Gagneux *et al.*, 2006) as well as the association of L5 with the Ewe ethnicity in Ghana (Asante-Poku *et al.*, 2015, 2016).

Moreover, there was random distribution of L5 Ewe isolates among the L5 clade (Fig4.11). This thesis also found no variation between T cell epitopes of L5 isolated from Ewe TB patients L5 and those from non-Ewe ethnicity. These observations suggest that if L5 is indeed maintained in West Africa by its co-evolution with the native West African ethnic group, there is no specific sub-group of L5 that is responsible for this association but rather the whole clade.

This thesis also found T cell epitopes of L6 isolated from HIV TB patients to be likely under positive selection for mutations compared to T cell epitopes of L6 from HIV negative TB patients. However, the limited number of genomes with data on patient's HIV status limits the impact of this finding. Nevertheless, it makes it important that a larger data set with complete HIV data be used in the future to ascertain if T cell epitopes of L6 from HIV patients are indeed under positive selection for mutations compared to those from HIV negative TB patients or not since L6 has been independently associated

with HIV in Ghana and the Gambia (Asante-Poku *et al.*, 2016; de Jong *et al.*, 2007; Homolka *et al.*, 2010).

All mutated T cell epitopes found within the analysis were lineage-specific except two compared to 12 T cell antigens with mutations between L5 and L6. And among the two common T cell epitopes with IEDB numbers 178644 and 178609, the mutations were lineage specific (Table 4.2). The observed dominance of lineage-specificity of mutated T cell epitopes in this study compares with earlier comparative genomics studies (Comas *et al.*, 2013; Stucki *et al.*, 2016). However, the T cell epitopes reported to be mutated between lineages used in those studies differ from this study possibly because those studies involved limited number of 20 *Maf* genomes. This antigenic variation among the MTBC could be as a results of pathogen-host co-evolution and have implications for global phylogeography of the MTBC (Gagneux *et al.*, 2006). In addition, it could have relevance for the observed clinical phenotypes such as progression to active disease (de Jong *et al.*, 2007) and transmission (Asare *et al.*, unpublished data).

Regulatory proteins play an important role in the survival and propagation of the complex modulation of the host immune response in favor of the bacteria via the action of secretory proteins which are also regulatory proteins (Bigi *et al.*, 2016; Raman *et al.*, 2004). The observation that regulatory proteins in L5 are under positive selection (pairwise dN/dS > 1.0) compared to L6 and L4 is in contrast to what was observed among the T cell epitopes of L5. This suggests that L4 and L6 are able to regulate the host's immune response better than L5 due to the relative conservation of their regulatory proteins compared to L5. The observation is also reminiscent of lineage-specific

accumulation of regulatory proteins mutations which is comparable to the earlier report comparing mutations within regulatory proteins between *Mtbss* and *M. bovis* (Bigi *et al.*, 2016). This observation could also account for the reported attenuated virulence of *Maf* relative to *Mtbss* (de Jong *et al.*, 2009; Homolka *et al.*, 2010; Portevin *et al.*, 2011; Tientcheu *et al.*, 2015) as mutations within some regulatory proteins have been associated with attenuated virulence (Gonzalo-asensio *et al.*, 2014). This therefore calls for further comparative studies of regulatory proteins between L5 and other MTBC lineages especially the generalist lineages to ascertain the physiological effect of the positive selection of mutations within regulatory proteins.

5.1.3 Genomic Erosion and Accumulation of Mutations Among MTBC Lineages, Implications for Evolution and Development of Control Tools.

Evolution of the MTBC is mainly driven by genomic erosion (complete deletion or truncation of genes and other chromosomal elements) and accumulation of SNPs. Thus, “ancient” MTBC lineages are expected to have relatively bigger genomes compared to “modern” lineages albeit not necessarily. Genomic erosion could have some physiological benefit to the pathogen such as the deletion of *katG* in some MTBC strains which makes them innately resistant to INH (Rouse & Morris, 1995). On the other hand, such deletion could be to the disadvantage of the pathogen such as the deletion of RD1 from BCG vaccine strain of *M. bovis* (Pym *et al.*, 2002), which has led to its attenuation *in vivo* even though it does not affect its ability to grow *in vitro*. My PhD found among the three lineages studied, that L5 has the biggest genome and L6 having the smallest (Fig 4.14), suggesting that a number of genes in L6 have been either deleted, truncated,

become pseudogenes or are not *in silico* expressed due to mutations within the promoter region. This observation bolsters the classification of L5 as an ancient lineage, however, it contradicts the classification of L6 as an ancient lineage of the human adapted MTBC if genomic erosion is to be pursued as the mode of evolution of the MTBC (Comas *et al.*, 2013). This finding further supports the claim that L5 and L6 are c two very different pathogens and need to be considered as such.

Most important among the lineage-specific deletions is the *mpt64* encoding a secretory immunogenic protein MPT64. The *mpt64* is the biomarker used for the lateral flow diagnostic tools BD MGIT TBc Identification Test (BD TBc ID; Becton Dickinson Diagnostics, Becton, Dickinson and Company, Sparks, MD, USA); SD Bioline Ag MPT64 Rapid (SD Bioline; Standard Diagnostics, Inc., Yongin-si, Gyeonggi-do, Republic of Korea); and Capilia TB-Neo (TAUNS Laboratories, Inc., Numazu, Shizuoka, Japan) for rapid detection of members of the MTBC. This finding explains the report of Ofori-Anyinam and colleagues (Ofori-Anyinam *et al.*, 2016) who found the MPT64-based tests to detect some *Maf* strains as non-tuberculous mycobacteria (NTM). Further comparative study found individuals with untreated TB caused by *Mtbss* express *mpt64* gene in their sputa 2.5-fold more than those infected with *Maf* (Ofori-Anyinam *et al.*, 2016)

My finding therefore underscores the need to take into consideration the diversity within the MTBC, when novel control tools are developed in order to ensure universality of the use. In addition, I observed that *rho* was absent among all the L6 genomes but present in all L5 and L4 genomes. *rho* encodes transcription termination factor Rho whose function

is to activate rho'SRNA-dependent ATPase to release the mRNA from the DNA template activity by binding to the nascent RNA (Lew *et al.*, 2011). Even though this gene is essential (Griffin *et al.*, 2011; Sassetti *et al.*, 2003), there are other homologs that can do the same job but perhaps with less efficiency, which could contribute to some of the phenotypic features of L6 relative to *Mtbss* (de Jong *et al.*, 2008; Gehre *et al.*, 2013). Furthermore, the observed L6-specific absence of *mce3D* and *mce3F* encoding mammalian cell entry proteins thought to be associated with host cell invasion though nonessential due to redundancy (Griffin *et al.*, 2011; Panigada *et al.*, 2002; Tekaiia *et al.*, 1999) could possibly affect the efficiency of cell invasion by L6 and hence its ability to transmit among immune-competent humans compared to *Mtbss* (de Jong *et al.*, 2008; Gehre *et al.*, 2013).

Peptide mutations such as amino acid substitutions (non-synonymous SNPs), frame-shifts (insertion or deletion of nucleotides other than 3 or multiples of 3), truncations (generation of nonsense codons) and deletions/insertions (insertions or deletions of nucleotides in 3s or multiples of 3) within some proteins of the MTBC strains have previously been implicated for the observed relative attenuation of *Maf* compared to H37Rv (Gehre *et al.*, 2013; Winglee *et al.*, 2016). In this study, I observed that majority of the lineage-specific mutations within selected genes were among *Maf* (Table 4.4), which compares with earlier reports (Gehre *et al.*, 2013; Winglee *et al.*, 2016). However, our larger sample size of *Maf* (appreciable number of L5 and L6 genomes) made it possible to observe that L6 may be more attenuated compared to L5 evidenced by the number of L6-specific and L6 sub-group-specific mutations recorded (Table 4.4; Appendix F).

Among the lineage-specific mutations identified was the I43N within *mpt64* (the same gene missing from the annotated genome of all L6) in all L5 genomes. The substitution of a hydrophobic isoleucine by a hydrophilic asparagine could affect the conformation of the resulting mpt64 antigen which can have dire consequence on sensitivity of the mpt64 based-diagnostics for detection of L5 as well. This therefore calls for proper evaluation of the *mpt64*-based diagnostic kits in L5 endemic African settings as has been done in the Gambia (Ofori-Anyinam *et al.*, 2016) where L6 is the circulating *Maf* and bolsters the call for appreciating the genomic diversity of MTBC in any future efforts towards control of TB.

Apart from the consequence of these observed mutations on *Maf*, some of these lineage-specific mutations can be employed in PCR based SNP typing of *Maf* using the untranslated DNA information. For example, the observed L6-specific insertion of AAAG in *espE* equivalent of 12 nucleotides and L5-specific deletion of V301-L302 an equivalent of 6 nucleotides can serve as unique molecular markers for rapid SNP typing (Stucki & Gagneux, 2012). Most of the available genotyping tools for the MTBC are based on mobile and/or repetitive elements (Van Embden *et al.*, 2000; Van Soolingen *et al.*, 1993; Supply *et al.*, 2000) however, these elements are prone to homoplasy (Coll *et al.*, 2014; Comas *et al.*, 2009). Albeit whole genome sequencing is the best tool for genotyping, it is expensive and requires higher expertise that may not obtain in most routine labs in West Africa where *Maf* is endemic. Therefore, a simple SNP assay would be very useful and convenient for proper typing of *Maf* lineages in the sub-region.

5.1.4 Drug Resistance Mutations, Implication for Rapid Diagnostics and TB Control.

The prokaryotic DNA polymerases are error prone; most mutant progenies resulting from accumulated mutation(s) from DNA replication are relatively weaker (due to fitness cost associated with the acquired mutation) compared to wild-types under normal conditions (Heidel *et al.*, 2004). However, under selection pressure such as in the presence of drugs, some of the mutations may be beneficial. Such mutant strains can become even more successful than the wild-type progeny. Among the MTBC the Beijing genotype of L2 has been associated with outbreaks of DR-TB in different geographical settings such as Eastern Europe (Casali *et al.*, 2012), Southern America (Eldholm *et al.*, 2015) South Africa (Johnson *et al.*, 2006, 2010; Marais *et al.*, 2013), North America (Bifani *et al.*, 1999; Valway *et al.*, 1994), Western Europe (Hillemann *et al.*, 2005) and Asia (Anh *et al.*, 2000; Cox *et al.*, 2005; Almeida *et al.*, 2005; Iwamoto *et al.*, 2008). This global observation emphasizes the importance of strain genetic background on the propensity to acquire and transmit drug resistance. This study took advantage of the availability of well-characterized isolates (Yeboah-Manu *et al.*, 2016) to investigate the genetic forces driving drug resistance in the country.

The proportion of INH^r, RIF^r and MDR among the close to 1500 isolates analysed were comparable to those reported by two previous studies conducted in Ghana over 10 years ago, (Asante-Poku *et al.*, 2015; Homolka *et al.*, 2010). This suggests that the rates of MDR and RIF^r TB in the country have remained fairly constant and low over the years, which is reminiscent of the efforts of the National TB control programme (NTP) in the country. However, the 7.5% INH^r compared to 0.7% RIF^r indicates that acquisition of INH resistance predates that of RIF in the events that lead to emergence of MDR-TB. In

Ghana, there has been national implementation of GeneXpert in most public health labs across the country by the NTP for rapid diagnosis of TB and detection of RIF resistance as surrogate for diagnosis of MDR-TB. MTBDR_{plus} would have been the better choice because it diagnoses both RIF and INH resistance based on already known genetic alterations that have been experimentally shown to confer resistance to either INH or RIF. However, its operational requirement specially to prevent DNA contamination and hence false result does not make it feasible at point of care facilities. However, in geographical settings such as Ghana that have about 75% of all DR-TB been INH^r, it is advisable to include in the national diagnostic algorithm the detection of both INH and RIF resistance at specialized laboratory facilities to support management of treatment failures.

Generally, isolates from the Northern Ghana had a larger proportion of MDR (5%) compared to the South (2.4%) even though the difference was not statistically significant. However, the fact that the estimation is more than twice the national average of 2% calls for intensifying control activities in Northern Ghana and probably other rural communities as most of the Global fund activities are centered in the urban areas especially the Accra Metro Health Directorate. However, there was neither *Mtbss* nor *Maf* was associated with DR-TB relative to the other which was consistent with earlier report (Asante-Poku *et al.*, 2015). This seems to contradict a much earlier study in Ghana that found *Mtbss* to be more associated with drug resistance relative to *Maf* (Yeboah-manu *et al.*, 2011). However, resistance to streptomycin (which is no more included the current TB treatment regimen) drove that association.

The Ghana sub-lineage of *Mtbss* L4 designated as L4.3.1 (Niemann *et al.*, 2016) L4.1.3 (Stucki *et al.*, 2016) and is restricted to few West African states, which is the second most dominant L4 sub-lineage in Ghana (Yeboah-Manu *et al.*, 2016) was found to be significantly associated with drug resistance when compared to the most dominant Cameroon sub-lineage or *Maf*. This is the first report that associates the West Africa restricted pathogen with drug resistance. This association is similar to the association of the Beijing genotype of L2 in several geographical settings (Anh *et al.*, 2000; Casali *et al.*, 2012; Cox *et al.*, 2005; Eldholm *et al.*, 2015; Hillemann *et al.*, 2005; Johnson *et al.*, 2010). However, whereas the Beijing genotype has been studied extensively relative to other MTBC genotypes to understand the basis of its association with DR-TB, not much is known about the Ghana sub-lineage.

This thence calls for further studies into the biology of this sub-lineage. Since acquisition of drug resistance conferring mutations in most bacteria, including MTBC are often associated with a fitness cost, the ability of a strain to acquire and transmit drug resistance could be influenced by the genetic background that alleviates the fitness cost. For instance, the Beijing genotype is associated with increased virulence and frequency of compensatory mutations relative the virulent *Mtbss* L4 laboratory strain H37Rv (Couvin & Rastogi, 2015; Merker *et al.*, 2015; Niemann *et al.*, 2010; Ribeiro *et al.*, 2014; Saelens *et al.*, 2015). Therefore, learning from this observation, it would be in line to infer whether the Ghana sub-lineage has developed the capacity to develop drug resistance through some as-yet-unidentified evolutionary compensatory mechanism. Because drug resistance is somehow linked with virulence (Couvin & Rastogi, 2015; Merker *et al.*, 2015; Niemann *et al.*, 2010; Ribeiro *et al.*, 2014; Saelens *et al.*, 2015),

detailed comparative genomics coupled with phenotypic investigations involving this Ghana sub-lineage would highlight the likely evolutionary forces driving this association with drug resistance.

Relating the association of the Ghana sub-lineage to a recent report on epidemiology of TB in Ghana, gives the indication that the observed higher level of DR-TB observed in the Northern region of Ghana could probably be confounded by the relative abundance of this sub-lineage in that part of the country as compared to the South (Yeboah-Manu *et al.*, 2016). This observation however, could also be as a result of local transmission of DR strains of the Ghana sub-lineage instead of emergence of resistance among patients which would compare with report from an earlier study in Argentina involving a strain of the Beijing genotype (Eldholm *et al.*, 2015). Therefore, whole genome sequencing of the drug resistant strains would be required for clarity.

Corresponding to an earlier report from Ghana (Homolka *et al.*, 2010), INH-resistant *Mtbss* isolates were more likely to harbor *katG* mutations compared to *inhApro* mutations in *Maf*. This observation indicates possible genotype-specificity with regards to acquisition of DR-conferring mutations. However, whereas that study reported that *inhApro* mutation -102G/A which has not been associated with drug resistance was likely to be *Maf*-specific SNP, it was later detected in INH resistant *Mtbss* strains harboring either *katG* S315T or *inhApro* -15C/T which challenged the *Maf*-specificity of this SNP (Asante-Poku *et al.*, 2015). In this current study, the mutation was again found in both *Mtbss* and *Maf* (Appendix G), hence agreeing with the later study and refuting the averred *Maf*-specificity of the SNP. I therefore hypothesize that the -102G/A, as well as -

47G/C and -8T/C SNPs all within the *inhApro* are more likely to be compensatory mutations instead of specific phylogenetic markers. This hypothesis is supported by the concurrent detection of these three SNPs with either the experimentally proven INH resistance conferring mutations *katG* S315T or *inhApro* -15C/T which requires further work to confirm or otherwise. However, if this hypothesis is substantiated, it could negatively impact the use of the WHO approved MTBDR*plus* line probe assay which has the *inhApro* -8T/C SNP as one of the markers for INH resistance. That would be true if emergence of the compensatory mutation predates the acquisition of DR-conferring mutation. That would mean that some TB patients who are diagnosed as INH resistant because of the mere presence of -8T/C SNP would have to be reconsidered for proper management. Nevertheless, this marker would be invaluable if the acquisition of the DR-conferring mutation predates the emergence of the compensatory mutation. This is because; the presence of the compensatory mutation would be a surrogate maker for drug resistance.

In addition, this study proposes that a novel *inhA* mutation (G204D) identified among INH-resistant *Mtbss* L4 isolates harboring the *katG* S315T could be a compensatory mutation for the acquisition of *katG* S315T mutations in *Mtbss* whereas another novel mutation in *ndh* (K32R) could be a drug resistance mutation hence requiring experimental confirmation. The hypothesis is based on the observation that all the 16 INH resistant isolates with the *inhA* G204D mutation also had the *katG* S315T mutation whereas the only isolate with the *ndh* K32R mutation had no other mutation in all the targets analysed apart from another novel mutation V117I in the same gene. I further hypothesize that the second *ndh* mutation V117I may be compensatory because it was

also detected in three additional INH resistant strains with *katG* S315T. However, whereas all the 4 isolates with the *ndh* V117I mutation were *Maf* (L6), the *inhA* G204D was detected in 16 *Mtbss* L4 strains of four different sub-lineages with *katG* S315T hence ruling out the possibility of *inhA* G204D being a phylogenetic marker (Appendix G). Strikingly, this *inhA* G204D mutation albeit found in four different sub-lineages of L4 was significantly associated with the Ghana sub-lineage ($p = 0.0220$). Therefore, if this mutation is verified as compensatory, the association with the Ghana sub-lineage could explain the positive selection and fixation of INH resistance mutations in this genotype similar to the situation with the Beijing genotype (Ribeiro *et al.*, 2014).

In the case of RIF resistance, even though no association was found between either *Maf* and *Mtbss* or sub-groups within either of them and any specific RIF resistance conferring mutation, the different *rpoC* mutations identified among L4 sub-lineage Cameroon (V483G) and Ghana (G332R) isolates harboring *rpoB* S450L is in line with report by an earlier study (Eldholm *et al.*, 2015), which suggested the influence of genetic background of MTBC strain on selection and fixation of specific mutations. However, it is also possible that the physical environment of the strains may be the driving force for selection and fixation of these mutations regardless of the genotype (Björkman *et al.*, 2000). Since all these reports are based on laboratory generated mutant strains, both proposals could nevertheless be possible under clinical conditions.

Even though the *rpoC* mutations were found only in Ghana and Cameroon sub-lineages, they were not spoligotype-specific (Appendix G), hence ruling out the possibility of them being phylogenetic markers. In furtherance to that claim, the V483G mutation has been

shown to occur at the interface between the *rpoB*, *rpoA* and *rpoC* of the tertiary structure of the bacterial polymerase, hence predicted with high confidence to be a compensatory mutation (Comas *et al.*, 2013). This report also reaffirms the dominance of the *rpoB* S450L mutations among clinical RIF resistant strains and emphasizes the possible role played by compensatory mutations for the fixation of this high level RIF resistance associated mutation among clinical isolates (Comas *et al.*, 2013).

The findings reaffirm the importance of understanding the genomic diversity of the MTBC in the fight against drug resistant TB. This work indicates that even though two different genotypes may acquire the same drug resistant conferring mutation, they may harbor different compensatory mutations suggesting genotype-lineage specific epistatic interaction between drug resistance conferring and non-drug resistant conferring mutations. The study hence provides the first evidence of the genetic basis of drug resistance in Ghana which even though require further studies does provide germane information for the management of TB in the country.

5.2 LIMITATIONS OF THE STUDY

The first limitation of this work was the disparate number of genomes in the different groups for comparisons. However, bootstrapping was used to normalize the data where possible to reduce the effect of different sample sizes. For instance, a 1000 bootstrap sampling with replacement of L4, L5 and L6 of equal sample size (15) was used to limit any possible bias due to differences in sample size when comparing the 17 L4, 147 L5 and the 67 L6 genomes. Even though L4 was the least, the estimates within T cell

epitopes were comparable to reports from studies that used hundreds of L4 genomes (Coscolla & Gagneux, 2014; Stucki *et al.*, 2016). Another limitation was the large number of L5 genomes without data on patient ethnicity which affected the number of L5 isolated from patients of the Ewe ethnicity. In trying to account for the possible bias due to the limited number of L5 genomes from Ewe TB patients, averaged estimates of 10 random samples of same size as L5 genomes from Ewe TB patients drawn out of the L5 genomes from non-Ewe TB patients were used. Furthermore, all the targets explored for the drug resistance analysis were previously described hotspots (da Silva *et al.*, 2011) which could lead to missing of novel hotspots that could be associated with drug resistance. Nevertheless, the information gleaned albeit not exhaustive is germane for efficient management of TB in Ghana.

5.3 CONCLUSIONS

5.3.1 **Key Findings**

This PhD work was carried out using MTBC strains isolated from TB patients in Ghana. The work was mainly carried out at the Noguchi Memorial Institute for Medical Research (NMIMR), University of Ghana, Legon, Accra Ghana. However, other aspects were conducted at the Wellcome Trust Sanger Institute (WTSI), University of Cambridge, United Kingdom and the Swiss Tropical and Public Health Institute (Swiss TPH), University of Basel, Basel, Switzerland.

The most significant achievements of this PhD work were:

1. The genome of L6 is significantly diverse compared to L4 and L5 with L5 being the most conserved. This finding is supported by the significantly higher whole genome π of L6 relative to L5 and L4 and L6 having accumulated significantly more fixed SNPs compared to L4 and L5 since the divergence from the most common ancestor of the MTBC. In addition, L5 has the biggest genome compared to L6 and L4 challenging the classification of L5 as ancient lineage and L4 as modern considering the MTBC evolves largely via genomic decay.

2. Human T cell epitopes of L4 and L5 are under purifying selection ($dN/dS < 1.0$) compared to those of L6 which are under positive selective pressure for mutations ($dN/dS > 1.0$) in addition to higher significantly higher diversity of T cell epitopes of L6 compared to the L4 and L5. However, regulatory proteins of L5 are under positive selection for mutations compared to those of L4 and L6 which are under purifying selection.

3. T cell antigenic variation among the three MTBC lineages is mainly lineage-specific. However, the conservation of T cell epitopes of L5 is not affected by ethnic background of the human host. However, there is the tendency that among L6, it would be affected by the HIV status of the host.

4. Construction of the first whole genome phylogeny of *Maf* in Ghana, which highlights the possibility of sub-lineages within these West African restricted pathogens.

5. The discovery for the first time that the Ghana sub-lineage of *Mtbss* L4 is associated with drug resistant TB in Ghana and the confirmation that INH-resistant *Mtbss* isolates are more likely to be *katG* S315T mutants compared to INH-resistant *Maf* which are more likely to be *inhApro* -15C/T mutants.
6. The discovery that RIF-resistant *Mtbss* strains of the Cameroon and Ghana sub-lineages harbor sub-lineage-specific compensatory *rpoC* mutations even though they have the same RIF-resistance-conferring *rpoB* mutation.
7. The discovery of novel mutations within drug resistance associated genes (such *inhA* G204D, *ndh* K32R & V117I and *ahpCpro* -88G/A & -142G/A) which could be of public health importance

5.3.2 Summary

The human adapted *Mycobacterium tuberculosis* complex albeit closely related are genomically distinct. The West Africa restricted L5 and L6 together classified as *M. africanum* are so genomically different pathogens suggesting different biovars of the complex with potential sub-lineages. The genomic diversity of the TB pathogen has potential implications for virulence, emergence of drug resistance and applicability of existing and future control tools.

5.4 RECOMMENDATIONS

1. Findings from this study indicate MTBC lineages L5 and L6 defined together as *Maf* are two distinct pathogens irrespective of their restriction to West Africa (de Jong *et al.*, 2010; Gagneux *et al.*, 2006). Even their combined phylogenetic classification as ancient lineages of the MTBC is also challenged by the disparate gene content. I therefore call for the recognition of MTBC L5 and L6 as unique biovars or sub-species of the complex.

2. T cell epitopes of the MTBC are generally conserved compared to other genomic elements (Comas *et al.*, 2010; Coscolla & Gagneux, 2014; Stucki *et al.*, 2016). However, the significant conservation of T cell epitopes of L5 relative to L6 suggests diverse and restricted host-ranges for L6 and L5 respectively. The observed mutations among L5 and L6 within genes involved with virulence, detoxification and adaptation bolsters the reported attenuation of L5 and L6 relative to L4. Nevertheless, based on these findings coupled with the association of L5 with Ewe ethnicity of West Africa (Asante-Poku *et al.*, 2015, 2016) and phylogenetic placement of L6 among animal adapted members of the MTBC (Bos *et al.*, 2014; Comas *et al.*, 2013; Niemann *et al.*, 2016), I propose that L5 is maintained in West Africa by its co-evolution with the Ewe ethnicity whereas L6 potentially has an environmental reservoir possibly zoonotic. I therefore recommend further work involving larger number of L5 and L6 genomes from all endemic West African settings to properly define the population structure these propose biovars.

3. In addition, I recommend comparative host-pathogen genome-wise interactive analysis to look for possible genetic reasons for the association of L5 with patient

ethnicity. In the case of L6, my work highlights the need to reignite the search for the unknown proposed environmental reservoir (possible zoonotic) in West Africa to aid the fight against TB caused by L6.

4. Furthermore, the persistent importance of L5 and L6 in West Africa despite their expected eventual elimination by the more virulent lineages of the MTBC makes it imperative that they are taken into consideration in promulgation of new control tools to avoid inefficacy of such tools in West Africa. For instance, it has already become apparent that the novel *mpt64*-based diagnostics have reduced sensitivity in the Gambia which has over 30% of TB caused by L6. My PhD work has indicated a possible impaired expression of this gene in L6 possibly accounting for the Gambian occurrence. In addition, the L5-specific I43N in the same gene, potentially affects the conformation of the resulting protein and may negatively impact the sensitivity of the *mpt64*-based diagnostics in L5 endemic West African settings.

5. Furthermore, my work also reaffirms the relevance of genomic diversity of the MTBC in the fixation of drug resistant conferring mutations. The observed association of the Ghana sub-lineage with drug resistance calls for in-depth phenotypic, genomic, gene expression and proteomic analyses involving this sub-lineage and other standard MTBC strains to fully understand the basis of this association. This work finally recommends comprehensive evaluation of novel diagnostics in West Africa before they are rolled out to be used in public health labs. For instance, it would be appropriate to test all Xpert diagnosed RIF-resistant cases for susceptibility to INH before managing them as MDR-TB cases since INH could still be part of the regimen if those cases are INH sensitive.

BIBLIOGRAPHY

- Abdallah, A. M., Gey van Pittius, N. C., DiGiuseppe Champion, P. A., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E., *et al.*, (2007). Type VII secretion-mycobacteria show the way. *Nature Reviews Microbiology*, 5(11), 883–891. <http://doi.org/10.1038/nrmicro1773>
- Achtman, M. (2008). Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annual Review of Microbiology*, 62(1), 53–70. <http://doi.org/10.1146/annurev.micro.62.081307.162832>
- Affolabi, D., Faihun, F., Sanoussi, N., Anyo, G., Shamputa, I. C., Rigouts, L., *et al.*, (2009). Possible outbreak of streptomycin-resistant *Mycobacterium tuberculosis* Beijing in Benin. *Emerging Infectious Diseases*, 15(7), 1123–1125. <http://doi.org/10.3201/eid1507.080697>
- Amo-adjei, J., & Kumi-kyereme, A. (2013). Myths and misconceptions about tuberculosis transmission in Ghana. *BMC International Health and Human Rights*, 13(1), 38. <http://doi.org/10.1186/1472-698X-13-38>
- Ando, H., Kitao, T., Miyoshi-Akiyama, T., Kato, S., Mori, T., & Kirikae, T. (2011). Downregulation of katG expression is associated with isoniazid resistance in *Mycobacterium tuberculosis*. *Molecular Microbiology*, 79(6), 1615–1628. <http://doi.org/10.1111/j.1365-2958.2011.07547.x>
- Anh, D. D., Borgdorff, M. W., Van, L. N., Lan, N. T. N., Van Gorkom, T., Kremer, K., & Van Soolingen, D. (2000). *Mycobacterium tuberculosis* Beijing genotype emerging in vietnam. *Emerging Infectious Diseases*. <http://doi.org/10.3201/eid0603.000312>
- Antoine, D., French, C. E., Jones, J., & Watson, J. M. (2007). Tuberculosis treatment outcome monitoring in England, Wales and Northern Ireland for cases reported in 2001. *Journal of Epidemiology and Community Health*, 61(4), 302–307. <http://doi.org/10.1136/jech.2005.044404>
- Aranaz, A. (2003). Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz *et al.* 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 53(6), 1785–1789. <http://doi.org/10.1099/ijls.0.02532-0>
- Arora, J., Kumar, G., Verma, A. K., Bhalla, M., Sarin, R., & Myneedu, V. P. (2015). Utility of MPT64 Antigen Detection for Rapid Confirmation of *Mycobacterium tuberculosis* Complex. *Journal of Global Infectious Diseases*, 7(2), 66–69. <http://doi.org/10.4103/0974-777X.154443>
- Asante-Poku, A., Otchere, I. D., Danso, E., Mensah, D. D., Bonsu, F., Gagneux, S., & Yeboah-Manu, D. (2015). Evaluation of GenoTypeW MTBDRplus for the rapid detection of drug-resistant tuberculosis in Ghana. *International Journal of Tuberculosis and Lung Disease*, 19(8). <http://doi.org/10.5588/ijtld.14.0864>
- Asante-Poku, A., Otchere, I. D., Osei-Wusu, S., Sarpong, E., Baddoo, A., Forson, A., *et*

- al.*, (2016). Molecular epidemiology of *Mycobacterium africanum* in Ghana. *BMC Infectious Diseases*, 16(1), 385. <http://doi.org/10.1186/s12879-016-1725-6>
- Asante-Poku, A., Yeboah-Manu, D., Otchere, I. D., Aboagye, S. Y., Stucki, D., Hattendorf, *et al.*, (2015). *Mycobacterium africanum* Is Associated with Patient Ethnicity in Ghana. *PLoS Neglected Tropical Diseases*, 9(1), e3370. <http://doi.org/10.1371/journal.pntd.0003370>
- Baess, I. (1979). Deoxyribonucleic acid relatedness among species of slowly-growing mycobacteria. *Acta Pathologica et Microbiologica Scandinavica. Section B, Microbiology*, 87(4), 221–226. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/115225>
- Baker, L. V, Brown, T. J., Maxwell, O., Gibson, A. L., Fang, Z., Yates, M. D., & Drobniewski, F. A. (2005). Molecular analysis of isoniazid-resistant *Mycobacterium tuberculosis* isolates from England and Wales reveals phylogenetic significance of the *ahpC-46A* polymorphism. *Antimicrob Agents Chemother*, 49(4), 1455–1464. <http://doi.org/10.1128/AAC.49.4.1455>
- Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., *et al.*, (1994). *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science (New York, N.Y.)*, 263(5144), 227–230. <http://doi.org/10.1126/science.8284673>
- Barkan, D., Hedhli, D., Yan, H. G., Huygen, K., & Glickman, M. S. (2012). *Mycobacterium tuberculosis* lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice. *Infection and Immunity*, 80(6), 1958–1968. <http://doi.org/10.1128/IAI.00021-12>
- Barreto, M. L., Pereira, S. M., Pilger, D., Cruz, A. A., Cunha, S. S., Sant’Anna, C., *et al.*, (2011). Evidence of an effect of BCG revaccination on incidence of tuberculosis in school-aged children in Brazil: Second report of the BCG-REVAC cluster-randomised trial. *Vaccine*, 29(31), 4875–4877. <http://doi.org/10.1016/j.vaccine.2011.05.023>
- Barry, C. E., Boshoff, H. I., Dartois, V., Dick, T., Ehrt, S., Flynn, J., *et al.*, (2009). The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nature Reviews. Microbiology*, 7(12), 845–855. <http://doi.org/10.1038/nrmicro2236>
- Baysarowich, J., Koteva, K., Hughes, D. W., Ejim, L., Griffiths, E., Zhang, K., *et al.*, (2008). Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr. *Proceedings of the National Academy of Sciences*, 105(12), 4886–4891. <http://doi.org/10.1073/pnas.0711939105>
- Behr, M. A., Warren, S. A., Salamon, H., Hopewell, P. C., Ponce de Leon, A., Daley, C. L., & Small, P. M. (1999). Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet (London, England)*, 353(9151), 444–449. [http://doi.org/10.1016/S0140-6736\(98\)03406-0](http://doi.org/10.1016/S0140-6736(98)03406-0)
- Belisle, J. T., Vissa, V. D., Sievert, T., Takayama, K., Brennan, P. J., & Besra, G. S. (1997). Role of the major antigen of *Mycobacterium tuberculosis* in Cell Wall Biogenesis. *Science*, 276(5317), 1997. <http://doi.org/10.1126/science.276.5317.1420>

- Bentley, S. D., Comas, I. I., Bryant, J. M., Walker, D., Smith, N. H., Harris, S. R., *et al.*, (2012). The genome of *Mycobacterium africanum* West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. *PLoS Neglected Tropical Diseases*, 6(2), e1552. <http://doi.org/10.1371/journal.pntd.0001552>
- Beukes, M., Lemmer, Y., Deysel, M., Al Dulayymi, J. R., Juma'a, R. A. D., Baird, M. S., *et al.*, (2010). Structure-function relationships of the antigenicity of mycolic acids in tuberculosis patients. *Chemistry and Physics of Lipids*, 163(8), 800–808. <http://doi.org/10.1016/j.chemphyslip.2010.09.006>
- Bhamidi, S., Scherman, M. S., Jones, V., Crick, D. C., Belisle, J. T., Brennan, P. J., & Mcneil, M. R. (2011). Detailed Structural and Quantitative Analysis Reveals the Spatial Organization of the Cell Walls of in Vivo Grown *Mycobacterium leprae* and in Vitro Grown *Mycobacterium tuberculosis* *. *The Journal of Biological Chemistry*, 286(26), 23168–23177. <http://doi.org/10.1074/jbc.M110.210534>
- Bhatt, A., Brown, A. K., Singh, A., Minnikin, D. E., & Besra, G. S. (2008). Loss of a *Mycobacterial* Gene Encoding a Reductase Leads to an Altered Cell Wall Containing ??-oxo- Mycolic Acid Analogs and Accumulation of Ketones. *Chemistry and Biology*, 15(9), 930–939. <http://doi.org/10.1016/j.chembiol.2008.07.007>
- Bhatt, A., Molle, V., Besra, G. S., Jacobs, W. R., & Kremer, L. (2007). The *Mycobacterium tuberculosis* FAS-II condensing enzymes: Their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Molecular Microbiology*, 64(6), 1442–1454. <http://doi.org/10.1111/j.1365-2958.2007.05761.x>
- Bifani, P. J., Mathema, B., Liu, Z., Moghazeh, S. L., Shopsin, B., Tempalski, B., *et al.*, (1999). Identification of a W Variant Outbreak of *Mycobacterium tuberculosis* via Population-Based Molecular Epidemiology. *JAMA*, 282(24), 2321. <http://doi.org/10.1001/jama.282.24.2321>
- Bigi, M. M., Blanco, F. C., Araújo, F. R., Thacker, T. C., Zumárraga, M. J., Cataldi, A. A., *et al.*, (2016). Polymorphisms of 20 regulatory proteins between *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Microbiology and Immunology*. <http://doi.org/10.1111/1348-0421.12402>
- Björkman, J., Nagaev, I., Berg, O. G., Hughes, D., & Andersson, D. I. (2000). Effects of Environment on Compensatory Mutations to Ameliorate Costs of Antibiotic Resistance. *Science*, 287(5457). Retrieved from <http://science.sciencemag.org/content/287/5457/1479/tab-pdf>
- Blanchard, J. S. (1996). Molecular mechanisms of drug resistance in *Mycobacterium tuberculosis*. *Annual Review of Biochemistry*, 65, 215–239. <http://doi.org/10.1146/annurev.biochem.65.1.215>
- Blouin, Y., Cazajous, G., Dehan, C., Soler, C., Vong, R., Hassan, M. O., *et al.*, (2014). Progenitor “*Mycobacterium canettii*” clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerging Infectious Diseases*, 20(1), 21–28. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24520560>
- Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., Dehan, C., *et al.*, (2012).

- Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PloS One*, 7(12), e52841. <http://doi.org/10.1371/journal.pone.0052841>
- Bogaert, D., Valk, P. Van Der, Ramdin, R., Sluijter, M., Monninkhof, E., Hendrix, R., *et al.*, (2004). Host-Pathogen Interaction during Pneumococcal Infection in Patients with Chronic Obstructive Pulmonary Disease. *Society*, 72(2), 818–823. <http://doi.org/10.1128/IAI.72.2.818>
- Bold, T. D., Davis, D. C., Penberthy, K. K., Cox, L. M., Ernst, J. D., & de Jong, B. C. (2012). Impaired fitness of *Mycobacterium africanum* despite secretion of ESAT-6. *The Journal of Infectious Diseases*, 205(6), 984–990. <http://doi.org/10.1093/infdis/jir883>
- Bonah, C. (2005). The “experimental stable” of the BCG vaccine: Safety, efficacy, proof, and standards, 1921-1933. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(4), 696–721. <http://doi.org/10.1016/j.shpsc.2005.09.003>
- Boritsch, E. C., Khanna, V., Pawlik, A., Honoré, N., Navas, V. H., Ma, L., *et al.*, (2016). Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 113(35), 9876–9881. <http://doi.org/10.1073/pnas.1604921113>
- Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., *et al.*, (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, 514(7523), 494–497. <http://doi.org/10.1038/nature13591>
- Bradley, S. G. (1972). Reassociation of Deoxyribonucleic Acid from Selected *Mycobacteria* with that from *Mycobacterium Bovis* and *Mycobacterium Farcinica*¹. *American Review of Respiratory Disease*, 106(1), 122–124. <http://doi.org/10.1164/arrd.1972.106.1.122>
- Brennan, P. J. (2003). Structure , function , and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis* (2003), 83, 91–97. [http://doi.org/10.1016/S1472-9792\(02\)00089-6](http://doi.org/10.1016/S1472-9792(02)00089-6)
- Brisse, S., Supply, P., Brosch, R., Vincent, V., & Gutierrez, M. C. (2006). Continuing the Debate, 2(9). <http://doi.org/10.1371/journal.ppat.0020095>
- Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., *et al.*, (2002). A new evolutionary scenario for the *Mycobacterium*. *Proceedings of the National Academy of Sciences*, 99, 3684–3689. <http://doi.org/10.1073/pnas.052548299>
- Brudey, K., Driscoll, J. J. R., Rigouts, L., Prodinger, W. M., Gori, A., Al-Hajoj, S. A., *et al.*, (2006). *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*, 6(1), 23. <http://doi.org/10.1186/1471-2180-6-23>

- Brzostek, A., Dziadek, B., Rumijowska-Galewicz, A., Pawelczyk, J., & Dziadek, J. (2007). Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*. *FEMS Microbiology Letters*, 275(1), 106–112. <http://doi.org/10.1111/j.1574-6968.2007.00865.x>
- Buboltz, A. M., Nicholson, T. L., Karanikas, A. T., Preston, A., & Harvill, E. T. (2009). Evidence for horizontal gene transfer of two antigenically distinct O antigens in *Bordetella bronchiseptica*. *Infection and Immunity*, 77(8), 3249–3257. <http://doi.org/10.1128/IAI.01448-08>
- Bynum, H. (2012). Spitting Blood: the history of tuberculosis. *Oxford: Oxford University Press*.
- Cambau, E., & Drancourt, M. (2014). Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clinical Microbiology and Infection*, 20(3), 196–201. <http://doi.org/10.1111/1469-0691.12555>
- Casali, N., Nikolayevskyy, V., Balabanova, Y., Ignatyeva, O., Kontsevaya, I., Harris, S. R., *et al.*, (2012). Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome Research*, 22(4), 735–745. <http://doi.org/10.1101/gr.128678.111>
- Cavanagh, R., Begon, M., Bennett, M., Ergon, T., Graham, I. M., De Haas, P. E. W., *et al.*, (2002). *Mycobacterium microti* infection (vole tuberculosis) in wild rodent populations. *Journal of Clinical Microbiology*, 40(9), 3281–3285. <http://doi.org/10.1128/JCM.40.9.3281-3285.2002>
- Chackerian, A. A., Alt, J. M., Perera, T. V., Dascher, C. C., & Behar, S. M. (2002). Dissemination of *Mycobacterium tuberculosis* is influenced by host factors and precedes the initiation of T cell immunity. *Infection and Immunity*, 70(8), 4501–4509. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=128141&tool=pmcentrez&rendertype=abstract>
- Charif D, L. J. (2007). *SeqinR 1.0-2: a contributed package to the R-project for statistical computing devoted to biological sequences retrieval and analysis*. In: Bastolla U, Porto MERH, Vendruscolo M, editors. *Structural approaches to sequence evolution: Molecules, networks, populations*. Springer Verlag. New York. Retrieved from <http://dx.plos.org/10.1371/journal.pone.0006456>
- Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., & Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution*, 30(5), 1224–1228. <http://doi.org/10.1093/molbev/mst028>
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.*, (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393(6685), 537–344. <http://doi.org/10.1038/31159>
- Colgrove, J. (2002). The McKeown thesis: A historical controversy and its enduring influence. *American Journal of Public Health*, 92(5), 725–729. <http://doi.org/10.2105/AJPH.92.5.725>
- Coll, F., Preston, M., Guerra-Assunção, J. A., Hill-Cawthorn, G., Harris, D., Perdigião, J.,

- et al.*, (2014). PolyTB: a genomic variation map for *Mycobacterium tuberculosis*. *Tuberculosis (Edinburgh, Scotland)*, 94(3), 346–354. <http://doi.org/10.1016/j.tube.2014.02.005>
- Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-maeda, M., *et al.*, (2011). Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nature Genetics*, 44(1), 106–110. <http://doi.org/10.1038/ng.1038>
- Comas, I., Chakravartti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., *et al.*, (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature Genetics*, 42(6), 498–503. <http://doi.org/10.1038/ng.590>
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-maeda, M., *et al.*, (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature Genetics*, 45(10), 1176–1182. <http://doi.org/10.1038/ng.2744>
- Comas, I., Homolka, S., Niemann, S., & Gagneux, S. (2009). Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PLoS ONE*, 4(11), e7815. <http://doi.org/10.1371/journal.pone.0007815>
- Comstock, G. W. (1994). The International Tuberculosis Campaign: a pioneering venture in mass vaccination and research. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 19(3), 528–540. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7811874>
- Cooper, A. M. (2009). Cell mediated immune responses in Tuberculosis. *Annu Rev Immunol*, 27, 393–422. <http://doi.org/10.1146/annurev.immunol.021908.132703>
- Coscolla, M., & Gagneux, S. (2010). Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discovery Today. Disease Mechanisms*, 7(1), e43–e59. <http://doi.org/10.1016/j.ddmec.2010.09.004>
- Coscolla, M., & Gagneux, S. (2014). Consequences of genomic diversity in mycobacterium tuberculosis. *Seminars in Immunology*, 26(6), 431–444. <http://doi.org/10.1016/j.smim.2014.09.012>
- Costa, E. R. D., Ribeiro, M. O., Silva, M. S. N., Arnold, L. S., Rostirolla, D. C., Cafrune, P. I., *et al.*, (2009). Correlations of mutations in *katG*, *oxyR-ahpC* and *inhA* genes and in vitro susceptibility in *Mycobacterium tuberculosis* clinical strains segregated by spoligotype families from tuberculosis prevalent countries in South America. *BMC Microbiology*, 9, 39. <http://doi.org/10.1186/1471-2180-9-39>
- Cousins, D. V., Bastida, R., Cataldi, A., Quse, V., Redrobe, S., Dow, S., *et al.*, (2003). Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 53(5), 1305–1314. <http://doi.org/10.1099/ijs.0.02401-0>

- Couvin, D., & Rastogi, N. (2015). Tuberculosis - A global emergency: Tools and methods to monitor, understand, and control the epidemic with specific example of the Beijing lineage. *Tuberculosis*, 95(S1), S177–S189. <http://doi.org/10.1016/j.tube.2015.02.023>
- Cox, H. S., Kubica, T., Doshetov, D., Kebede, Y., Rüsç-Gerdess, S., & Niemann, S. (2005). The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. *Respiratory Research*, 6(6), 134. <http://doi.org/10.1186/1465-9921-6-134>
- Crofton, J., & Mitchison, D. . A. (1948). Streptomycin Resistance in Pulmonary Tuberculosis. *British Medical Journal*, 2(Dec 11), 1009–1015. [http://doi.org/10.1016/S0140-6736\(48\)91127-1](http://doi.org/10.1016/S0140-6736(48)91127-1)
- Cudahy, P., Haven, N., Shenoi, S., & Haven, N. (2016). Diagnostics for pulmonary tuberculosis. *Postgraduate Medical Journal*, 92(1086), 187–193. <http://doi.org/10.1136/postgradmedj-2015-133278.Diagnostics>
- da Silva, P. E. A., Von Groll, A., Martin, A., & Palomino, J. C. (2011). Efflux as a mechanism for drug resistance in *Mycobacterium tuberculosis*: Table 1. *FEMS Immunology & Medical Microbiology*, 63(1), 1–9. <http://doi.org/10.1111/j.1574-695X.2011.00831.x>
- Daniel, T. M. (2000). *The origins and precolonial epidemiology of tuberculosis in the Americas: Can we figure them out? International Journal of Tuberculosis and Lung Disease* (Vol. 4).
- Daniel, T. M. (2006). The history of tuberculosis. *Respiratory Medicine*, 100(11), 1862–1870. <http://doi.org/10.1016/j.rmed.2006.08.006>
- Daugelat, S., Kowall, J., Mattow, J., Bumann, D., Winter, R., Hurwitz, R., & Kaufmann, S. H. E. (2003). The RD1 proteins of *Mycobacterium tuberculosis*: Expression in *Mycobacterium smegmatis* and biochemical characterization. *Microbes and Infection*, 5(12), 1082–1095. [http://doi.org/10.1016/S1286-4579\(03\)00205-3](http://doi.org/10.1016/S1286-4579(03)00205-3)
- David, H. L., Selin, M., Gontijo Filho, P. P., Beam, E., & Sud, M. (1978). Induction kinetics of delayed hypersensitivity towards *Mycobacterium tuberculosis* and *M. avium*: analysis of cross-reactions elicited by PPD-tuberculin and PPD-B. *Revista Brasileira de Pesquisas Médicas E Biológicas*, 11(6), 313–9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/108772>
- Davies, P. D. (2001). Drug-resistant tuberculosis. *Journal of the Royal Society of Medicine*, 94(6), 261–263.
- Davis, J. L. (2014). Diagnostic accuracy of same-day microscopy versus standard microscopy for pulmonary tuberculosis: a systematic review and meta-analysis. *Lancet Infectious Disease*, 13(2), 1–16. [http://doi.org/10.1016/S1473-3099\(12\)70232-3.Diagnostic](http://doi.org/10.1016/S1473-3099(12)70232-3.Diagnostic)
- de Jong, B. C., Adetifa, I., Walther, B., Hill, P. C., Antonio, M., Ota, M., & Adegbola, R. A. (2010). Differences between tuberculosis cases infected with *Mycobacterium africanum*, West African type 2, relative to Euro-American *Mycobacterium*

- tuberculosis: an update. *FEMS Immunology and Medical Microbiology*, 58(1), 102–5. <http://doi.org/10.1111/j.1574-695X.2009.00628.x>
- de Jong, B. C., Antonio, M., Awine, T., Ogungbemi, K., de Jong, Y. P., Gagneux, S., *et al.*, (2009). Use of spoligotyping and large sequence polymorphisms to study the population structure of the Mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. *Journal of Clinical Microbiology*, 47(4), 994–1001. <http://doi.org/10.1128/JCM.01216-08>
- de Jong, B. C., Antonio, M., & Gagneux, S. (2010). Mycobacterium africanum-review of an important cause of human tuberculosis in West Africa. *PLoS Neglected Tropical Diseases*, 4(9), e744. <http://doi.org/10.1371/journal.pntd.0000744>
- de Jong, B. C., Hill, P. C., Aiken, A., Awine, T., Antonio, M., Adetifa, I. M., *et al.*, (2008). Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in The Gambia. *The Journal of Infectious Diseases*, 198(7), 1037–43. <http://doi.org/10.1086/591504>
- de Jong, B. C., Hill, P. C., Aiken, A., Jeffries, D. J., Onipede, A., Small, P. M., *et al.*, (2007). Clinical presentation and outcome of tuberculosis patients infected by M. africanum versus M. tuberculosis. *The International Journal of Tuberculosis and Lung Disease : The Official Journal of the International Union against Tuberculosis and Lung Disease*, 11(4), 450–6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17394693>
- de Jong, B. C., Hill, P. C., Brookes, R. H., Gagneux, S., Jeffries, D. J., Otu, J. K., *et al.*, (2006). Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. *The Journal of Infectious Diseases*, 193(9), 1279–86. <http://doi.org/10.1086/502977>
- de Mello, K. G. C., Mello, F. C. Q., Borga, L., Rolla, V., Duarte, R. S., Sampaio, E. P., *et al.*, (2013). Clinical and therapeutic features of pulmonary nontuberculous mycobacterial disease, Brazil, 1993-2011. *Emerging Infectious Diseases*, 19(3), 393–9. <http://doi.org/10.3201/eid1903.120735>
- Deaton, A., & Stone, A. A. (2013). Two happiness puzzles. *The American Economic Review*, 103(3), 591–597. <http://doi.org/10.1257/aer.103.3.591>
- de Vos, M., Muller, B., Borrell, S., Black, P. A., van Helden, P. D., Warren, R. M., *et al.*, (2013). Putative Compensatory Mutations in the rpoC Gene of Rifampin-Resistant Mycobacterium tuberculosis Are Associated with Ongoing Transmission. *Antimicrobial Agents and Chemotherapy*, 57(2), 827–832. <http://doi.org/10.1128/AAC.01541-12>
- Deepak Almeida, Camilla Rodrigues, Tester F. Ashavaid, Ajit Lalvani, Zarir F. Udwadia, and A. M. (2005). High Incidence of the Beijing Genotype among Multidrug-Resistant Isolates of Mycobacterium tuberculosis in a Tertiary Care Center in Mumbai, India. *Clinical Infectious Diseases*, 40, 881–886.
- Derbyshire, K. M., & Gray, T. A. (2014). Distributive Conjugal Transfer: New Insights into Horizontal Gene Transfer and Genetic Exchange in Mycobacteria.

- Microbiology Spectrum*, 2(1), :04. <http://doi.org/10.1128/microbiolspec.MGM2-0022-2013>
- Diallo, A. B., Walbang, O. G., Camara, M., Lo, S., Ayorinde, A., Niang, A., *et al.*, (2016). Molecular genotypes of *Mycobacterium tuberculosis* strains circulating in Dakar, Senegal. *African Journal of Microbiology Research*, 10(35), 1460–1466. <http://doi.org/10.5897/AJMR2016.8221>
- Dubos, R. ; Dubos, J. (1990). The white plague: Tuberculosis, man, and society, 2nd ed. New Brunswick, N.J.: Rutgers University Press, 1987; origina. *Journal of the History of the Behavioral Sciences*, 26(3), 307–307. [http://doi.org/10.1002/1520-6696\(199007\)26:3<307::AID-JHBS2300260345>3.0.CO;2-1](http://doi.org/10.1002/1520-6696(199007)26:3<307::AID-JHBS2300260345>3.0.CO;2-1)
- Edinburgh-Anti-Tuberculosis-Scheme. (1914). The Edinburgh Anti-Tuberculosis Scheme. *The British Journal of Nursing*, 117.
- Eldholm, V., Monteserin, J., Rieux, A., Lopez, B., Sobkowiak, B., Ritacco, V., & Balloux, F. (2015). Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. *Nature Communications*, 6(May), 7119. <http://doi.org/10.1038/ncomms8119>
- Embden, J. D. A. Van, Gorkom, T. Van, Kremer, K., Jansen, R., Zeijst, B. A. M. Van Der, & Schouls, L. M. (2000). Genetic Variation and Evolutionary Origin of the Direct Repeat Locus of *Mycobacterium tuberculosis* Complex Bacteria Genetic Variation and Evolutionary Origin of the Direct Repeat Locus of *Mycobacterium tuberculosis* Complex Bacteria. <http://doi.org/10.1128/JB.182.9.2393-2401.2000>.Updated
- Fenner, L., Egger, M., Bodmer, T., Furrer, H., Ballif, M., Battegay, M., *et al.*, (2013). HIV infection disrupts the sympatric host-pathogen relationship in human tuberculosis. *PLoS Genetics*, 9(3), e1003318. <http://doi.org/10.1371/journal.pgen.1003318>
- Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., *et al.*, (2013). *Mycobacterial* lineages causing pulmonary and extrapulmonary Tuberculosis, Ethiopia. *Emerging Infectious Diseases*, 19(3), 460–463. <http://doi.org/10.3201/eid1903.120256>
- Flandrois, J.-P., Lina, G., & Dumitrescu, O. (2014). MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. *BMC Bioinformatics*, 15(1), 107. <http://doi.org/10.1186/1471-2105-15-107>
- Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., *et al.*, (2002). Whole-Genome Comparison of *Mycobacterium tuberculosis* Clinical and Laboratory Strains, 184(19), 5479–5490. <http://doi.org/10.1128/JB.184.19.5479>
- Forrellad, M. A., Klepp, L. I., Gioffré, A., Sabio y García, J., Morbidoni, H. R., Santangelo, M. de la P., *et al.*, (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*, 4(1), 3–66. <http://doi.org/10.4161/viru.22329>
- Franzblau, S. G., Witzig, R. S., Mclaughlin, J. C., Torres, P., Madico, G., Hernandez, A., *et al.*, (1998). Rapid, low-technology MIC determination with clinical

- Mycobacterium tuberculosis isolates by using the microplate Alamar Blue assay. *Journal of Clinical Microbiology*, 36(2), 362–366.
- Frota, C. C., Hunt, D. M., Buxton, R. S., Rickman, L., Hinds, J., Kremer, K., *et al.*, (2004). Genome structure in the vole bacillus, *Mycobacterium microti*, a member of the *Mycobacterium tuberculosis* complex with a low virulence for humans. *Microbiology*, 150(5), 1519–1527. <http://doi.org/10.1099/mic.0.26660-0>
- Frothingham, R., Strickland, P. L., Bretzel, G., Ramaswamy, S., Musser, J. M., & Williams, D. L. (1999). Phenotypic and genotypic characterization of *Mycobacterium africanum* isolates from West Africa. *Journal of Clinical Microbiology*, 37(6), 1921–1926. Retrieved from <https://www.scopus.com/record/display.uri?eid=2-s2.0-0032963021&origin=inward&txGid=F10679F8BA8FDB23705131DFDF4D7932.wsnAw8kcdt7IPYLO0V48gA%3A1>
- Gagneux, S. (2012). Host-pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society of London*, 367(1590), 850–859. <http://doi.org/10.1098/rstb.2011.0316>
- Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B. C., Narayanan, S., *et al.*, (2006). Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), 2869–2873. <http://doi.org/10.1073/pnas.0511240103>
- Gagneux, S., Long, C. D., Small, P. M., Van, T., Schoolnik, G. K., & Bohannon, B. J. M. (2006). The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science (New York, N.Y.)*, 312(5782), 1944–1946. <http://doi.org/10.1126/science.1124410>
- Gagneux, S. S. S., Burgos, M. V., Deriemer, K., Enciso, A., Muñoz, S., Hopewell, P. C., *et al.*, (2006). Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS Pathogens*, 2(6), e61. <http://doi.org/10.1371/journal.ppat.0020061>
- Gagneux, S., & Small, P. M. (2007). Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *The Lancet Infectious Diseases*, 7(5), 328–337. [http://doi.org/10.1016/S1473-3099\(07\)70108-1](http://doi.org/10.1016/S1473-3099(07)70108-1)
- Galagan, J. E. (2014). Genomic insights into tuberculosis. *Nature Reviews. Genetics*, 15(5), 307–20. <http://doi.org/10.1038/nrg3664>
- Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., *et al.*, (2003). The complete genome sequence of *Mycobacterium bovis*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), 7877–7882. <http://doi.org/10.1073/pnas.1130426100>
- Garrity, G. M. (2004). *Bergey's manual of systematic bacteriology. Vol. 2, The proteobacteria.* Springer. Retrieved from <http://www.springer.com/gp/book/9780387950402>
- Gehre, F., Antonio, M., Faihun, F., Odoun, M., Uwizeye, C., de Rijk, P., *et al.*, (2013).

- The first phylogeographic population structure and analysis of transmission dynamics of *M. africanum* West African 1--combining molecular data from Benin, Nigeria and Sierra Leone. *PloS One*, 8(10), e77000. <http://doi.org/10.1371/journal.pone.0077000>
- Gehre, F., Otu, J., Deriemer, K., Sessions, P. F. De, Hibberd, M. L., Mulders, W., *et al.*, (2013). Deciphering the growth behaviour of *Mycobacterium africanum*. *PLoS Neglected Tropical Diseases*, 7(5), e2220. <http://doi.org/10.1371/journal.pntd.0002220>
- Ghose, T. K. . P. G. S. K. B. (2003). *Biotechnology in India*. Springer.
- Glickman, M. S., Cox, J. S., & Jacobs, W. R. (2000). A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of *Mycobacterium tuberculosis*. *Molecular Cell*, 5(4), 717–727. [http://doi.org/10.1016/S1097-2765\(00\)80250-6](http://doi.org/10.1016/S1097-2765(00)80250-6)
- Gonzalo-asensio, J., Malaga, W., Pawlik, A., & Astarie-dequeker, C. (2014). Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator, 111(31). <http://doi.org/10.1073/pnas.1406693111>
- Gordon, S. V, Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., & Stewart, T. (1999). Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays, 32, 643–655.
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57(1), 81–91. <http://doi.org/10.1099/ijs.0.64483-0>
- Griffin, J. E., Gawronski, J. D., DeJesus, M. a., Ioerger, T. R., Akerley, B. J., & Sassetti, C. M. (2011). High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens*, 7(9), 1–9. <http://doi.org/10.1371/journal.ppat.1002251>
- Griffin, J. E., Pandey, A. K., Gilmore, S. A., Mizrahi, V., McKinney, J. D., Bertozzi, C. R., & Sassetti, C. M. (2012). Cholesterol catabolism by *Mycobacterium tuberculosis* requires transcriptional and metabolic adaptations. *Chemistry and Biology*, 19(2), 218–227. <http://doi.org/10.1016/j.chembiol.2011.12.016>
- Groenheit, R., Ghebremichael, S., Svensson, J., Rabna, P., Colombatti, R., Riccardi, F., *et al.*, (2011). The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. *PloS One*, 6(4), e18601. <http://doi.org/10.1371/journal.pone.0018601>
- Gs Kumar, V., Urs, T. A., & Ranganath, R. R. (2011). MPT 64 Antigen detection for Rapid confirmation of *M.tuberculosis* isolates. *BMC Research Notes*, 4, 79. <http://doi.org/10.1186/1756-0500-4-79>
- Guinn, K. M., Hickey, M. J., Mathur, S. K., Zakel, K. L., Grotzke, J. E., Lewinsohn, D. M., *et al.*, (2004). Individual RD1-region genes are required for export of ESAT-6 / CFP-10 and for virulence of *Mycobacterium tuberculosis*, 51, 359–370. <http://doi.org/10.1046/j.1365-2958.2003.03844.x>

- Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., *et al.*, (2005). Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathogens*, 1(1), e5. <http://doi.org/10.1371/journal.ppat.0010005>
- H Collins, B. C., Yates, M. D., Grange, J. M., Collins, C. H., Yates, M. D., & Grange, J. M. (1982). Subdivision of Mycobacterium tuberculosis into five variants for epidemiological purposes: methods and nomenclature. *The Journal of Hygiene*, 89(2), 235–242. <http://doi.org/10.1017/S0022172400070765>
- Hadfield, J., Croucher, N. J., Goater, R. J., Abudahab, K., Aanensen, D. M., & Harris, S. R. (2017). Phandango: an interactive viewer for bacterial population genomics. *bioRxiv*. Retrieved from <http://www.biorxiv.org/content/early/2017/03/22/119545>
- Hadley Wickham. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Retrieved from <http://ggplot2.org>
- Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics1 relevance of population genetics 4 1.2 phenotypic variation in natural populations 5* (Fourth). Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers. Retrieved from <http://www.sinauer.com/media/wysiwyg/tocs/PrinciplesPopulationGenetics.pdf>
- Hazbon, M. H., Brimacombe, M., Bobadilla del Valle, M., Cavatore, M., Guerrero, M. I., Varma-Basil, M., *et al.*, (2006). Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*, 50(8), 2640–2649. <http://doi.org/10.1128/AAC.00112-06>
- Heidel, A. J., Clarke, J. D., Antonovics, J., & Dong, X. (2004). Fitness costs of mutations affecting the systemic acquired resistance pathway in Arabidopsis thaliana. *Genetics*, 168(4), 2197–2206. <http://doi.org/10.1534/genetics.104.032193>
- Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., *et al.*, (2008). High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biology*, 6(12), e311. <http://doi.org/10.1371/journal.pbio.0060311>
- Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., Besra, G. S., Lee, O. Y., Gernaey, A. M., *et al.*, (2008). Detection and molecular characterization of 9,000-year-old Mycobacterium tuberculosis from a Neolithic settlement in the Eastern Mediterranean. *PLoS One*, 3(10), e3426. <http://doi.org/10.1371/journal.pone.0003426>
- Herzog, H. (1998). History of Tuberculosis. *Respiration*, 65, 5–15.
- Hetherington, S. V., Watson, A. S., & Patrick, C. C. (1995). Sequence and analysis of the rpoB gene of Mycobacterium smegmatis. *Antimicrobial Agents and Chemotherapy*, 39(9), 2164–2166. <http://doi.org/10.1128/AAC.39.9.2164>
- Heym, B., Saint-Joanis, B., & Cole, S. T. (1999). The molecular basis of isoniazid resistance in Mycobacterium tuberculosis. *Tubercle and Lung Disease*, 79(4), 267–271. <http://doi.org/10.1054/tuld.1998.0208>
- Hillemann, D., Kubica, T., Ru, S., & Niemann, S. (2005). Disequilibrium in Distribution

- of Resistance Mutations among Mycobacterium tuberculosis Beijing and Non-Beijing Strains Isolated from Patients in Germany, 49(3), 1229–1231. <http://doi.org/10.1128/AAC.49.3.1229>
- Ho, Y. M., Sun, Y. J., Wong, S. Y., Lee, A. S. G., Yu, M. H., Sun, Y. J., *et al.*, (2009). Contribution of dfrA and inhA mutations to the detection of isoniazid-resistant Mycobacterium tuberculosis isolates. *Antimicrobial Agents and Chemotherapy*, 53(9), 4010–4012. <http://doi.org/10.1128/AAC.00433-09>
- Homolka, S., Meyer, C. G., Hillemann, D., Owusu-dabo, E., Adjei, O., Horstmann, R. D., *et al.*, (2010). Unequal distribution of resistance-conferring mutations among Mycobacterium tuberculosis and Mycobacterium africanum strains from Ghana. *International Journal of Medical Microbiology: IJMM*, 300(7), 489–495. <http://doi.org/10.1016/j.ijmm.2010.04.019>
- Homolka, S., Niemann, S., Russell, D. G., & Rohde, K. H. (2010). Functional genetic diversity among Mycobacterium tuberculosis complex clinical isolates: delineation of conserved core and lineage-specific transcriptomes during intracellular survival. *PLoS Pathogens*, 6(7), e1000988. <http://doi.org/10.1371/journal.ppat.1000988>
- Homolka, S., Projahn, M., Feuerriegel, S., Ubben, T., Diel, R., Nübel, U., *et al.*, (2012). High resolution discrimination of clinical Mycobacterium tuberculosis complex strains based on single nucleotide polymorphisms. *PloS One*, 7(7), e39855. <http://doi.org/10.1371/journal.pone.0039855>
- Homolka, S., Ubben, T., & Niemann, S. (2016). High sequence variability of the PPE18 gene of clinical Mycobacterium tuberculosis complex strains potentially impacts effectivity of vaccine candidate M72/AS01E. *PLoS ONE*, 11(3), 1–10. <http://doi.org/10.1371/journal.pone.0152200>
- Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeón, L., Abdallah, A. M., *et al.*, (2012). ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cellular Microbiology*, 14(8), 1287–1298. <http://doi.org/10.1111/j.1462-5822.2012.01799.x>
- Hunt, M., Mather, A. E., Sánchez-Busó, L., Page, A. J., Parkhill, J., Keane, J. A., & Harris, S. R. (2017). ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *bioRxiv*. Retrieved from <http://www.biorxiv.org/content/early/2017/03/18/118000>
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119. <http://doi.org/10.1186/1471-2105-11-119>
- Ioerger, T. R., Feng, Y., Ganesula, K., Chen, X., Dobos, K. M., Fortune, S., *et al.*, (2010). Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. *Journal of Bacteriology*, 192(14), 3645–3653. <http://doi.org/10.1128/JB.00166-10>
- Iwamoto, T., Yoshida, S., Suzuki, K., & Wada, T. (2008). Population structure analysis of the Mycobacterium tuberculosis Beijing family indicates an association between

- certain sublineages and multidrug resistance. *Antimicrobial Agents and Chemotherapy*. <http://doi.org/10.1128/AAC.00579-08>
- Jay, V. (2000). A Portrait in History: The Legacy of Laënnec. *Arch Pathol Lab Med—Vol*, 124. Retrieved from <http://www.archivesofpathology.org/doi/pdf/10.1043/0003-9985%282000%29124%3C1420%3ATLLOLN%3E2.0.CO%3B2>
- Johnson, R., Warren, R. M., Van Der Spuy, G. D., Van Pittius, N. C. G., Theron, D., Streicher, E. M., *et al.*, (2010). Drug-resistant tuberculosis epidemic in the Western Cape driven by a virulent Beijing genotype strain. *International Journal of Tuberculosis and Lung Disease*. <http://doi.org/10.1128/JCM.01938-07>
- Johnson, R., Warren, R., Strauss, O. J., Jordaan, A. M., Falmer, A. A., Beyers, N., *et al.*, (2006). An outbreak of drug-resistant tuberculosis caused by a Beijing strain in the Western Cape, South Africa. *International Journal of Tuberculosis and Lung Disease*.
- Käser, M., Ruf, M.-T., Hauser, J., Marsollier, L., & Pluschke, G. (2009). Optimized method for preparation of DNA from pathogenic and environmental mycobacteria. *Applied and Environmental Microbiology*, 75(2), 414–418. <http://doi.org/10.1128/AEM.01358-08>
- Kasperbauer, S. H., & De Groote, M. A. (2015). The Treatment of Rapidly Growing Mycobacterial Infections. *Clinics in Chest Medicine*, 36(1), 67–78. <http://doi.org/10.1016/j.ccm.2014.10.004>
- Kato-Maeda, M., Kim, E. Y., Flores, L., Jarlsberg, L. G., Osmond, D., & Hopewell, P. C. (2010). Differences among sublineages of the East-Asian lineage of Mycobacterium tuberculosis in genotypic clustering. *International Journal of Tuberculosis and Lung Disease*, 14(5), 538–544. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625672&tool=pmcentrez&rendertype=abstract>
- Kato-Maeda, M., Rhee, J. T., Gingeras, T. R., Salamon, H., Drenkow, J., Smittipat, N., & Small, P. M. (2001). Comparing genomes within the species Mycobacterium tuberculosis. *Genome Research*, 11(4), 547–554. <http://doi.org/10.1101/gr166401>
- Kato-Maeda, M., Shanley, C. A., Ackart, D., Jarlsberg, L. G., Shang, S., Obregon-Henao, A., *et al.*, (2012). Beijing sublineages of Mycobacterium tuberculosis differ in pathogenicity in the guinea pig. *Clinical and Vaccine Immunology*. <http://doi.org/10.1128/CVI.00250-12>
- Kay, G. L., Sergeant, M. J., Zhou, Z., Chan, J. Z.-M., Millard, A., Quick, J., *et al.*, (2015). Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature Communications*, 6, 6717. <http://doi.org/10.1038/ncomms7717>
- Keating, L. A., Wheeler, P. R., Mansoor, H., Inwald, J. K., Dale, J., Hewinson, R. G., & Gordon, S. V. (2005). The pyruvate requirement of some members of the Mycobacterium tuberculosis complex is due to an inactive pyruvate kinase: Implications for in vivo growth. *Molecular Microbiology*, 56(1), 163–174.

<http://doi.org/10.1111/j.1365-2958.2005.04524.x>

- Kirimuhuzya, C. (2007). Multi-Drug / Extensively Drug Resistant Tuberculosis (Mdr / Xdr-Tb): Renewed Global Battle Against Tuberculosis ?
- Koch, A. B. P. W., & Marolda, L. (1961). Tuberculosis prevention in Ghana: A school children BCG campaign in Accra. *Tubercle*, 42(3), 366–370. [http://doi.org/10.1016/S0041-3879\(61\)80125-6](http://doi.org/10.1016/S0041-3879(61)80125-6)
- Koivula, T., Svenson, S. B., & Källenius, G. (2002). The mtp40 gene is not present in *Mycobacterium bovis*. *Tuberculosis (Edinburgh, Scotland)*, 82(4–5), 183–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12464490>
- Konstantinidis, K. T., & Tiedje, J. M. (2004). Genomic insights that advance the species definition for prokaryotes. Retrieved from <http://www.pnas.org/content/102/7/2567.full.pdf>
- Koro, F. K., Simo, Y. K., Piam, F. F., Noeske, J., Gutierrez, C., Kuaban, C., & Eyangoh, S. I. (2013). Population dynamics of tuberculous bacilli in cameroon as assessed by spoligotyping. *Journal of Clinical Microbiology*, 51(1), 299–302. <http://doi.org/10.1128/JCM.01196-12>
- Krishnan, N., Malaga, W., Constant, P., Caws, M., Tran, T. H. C., Salmons, J., *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. <http://doi.org/10.1371/journal.pone.0023870>
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H.-H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100–8. <http://doi.org/10.1093/nar/gkm160>
- Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1), 11–16. <http://doi.org/10.1093/nar/gkh152>
- Lawson, L., Zhang, J., Gomgnimbou, M. K., Abdurrahman, S. T., Le Moullec, S., Mohamed, F., *et al.*, (2012). A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. *PloS One*, 7(6), e38409. <http://doi.org/10.1371/journal.pone.0038409>
- Lázár-Molnár, E., Chen, B., Sweeney, K. A., Wang, E. J., Liu, W., Lin, J., *et al.*, (2010). Programmed death-1 (PD-1)-deficient mice are extraordinarily sensitive to tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30), 13402–13407. <http://doi.org/10.1073/pnas.1007394107>
- Lee, A. S. G., Teo, A. S. M., & Wang, S. Y. (2001). Novel mutations in *ndh* in isoniazid-resistant *Mycobacterium tuberculosis* isolates. *Antimicrobial Agents and Chemotherapy*, 45(7), 2157–2159. <http://doi.org/10.1128/AAC.45.7.2157-2159.2001>
- Leung, E. T. Y., Ho, P. L., Yuen, K. Y., Woo, W. L., Lam, T. H., Kao, R. Y., *et al.*, Molecular characterization of isoniazid resistance in *Mycobacterium tuberculosis*: Identification of a novel mutation in *inhA*. *Antimicrobial Agents and Chemotherapy*,

- 50(3), 1075–1078. <http://doi.org/10.1128/AAC.50.3.1075-1078.2006>
- Lew, J. M., Kapopoulou, A., Jones, L. M., & Cole, S. T. (2011). TubercuList - 10 years after. *Tuberculosis*, 91, 1–7. <http://doi.org/10.1016/j.tube.2010.09.008>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <http://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.*, (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <http://doi.org/10.1093/bioinformatics/btp352>
- Lipsitch, M. (1999). Bacterial vaccines and serotype replacement: Lessons from Haemophilus influenzae and prospects for Streptococcus pneumoniae. *Emerging Infectious Diseases*, 5(3), 336–345. <http://doi.org/10.3201/eid0503.990304>
- Liu, T. (1971). “The Taoists” Knowledge of Tuberculosis in the Twelfth Century’. *T’oung Pao*, 57(5), 285–301.
- López, B., Aguilar, D., Orozco, H., Burger, M., Espitia, C., Ritacco, V., *et al.*, (2003). A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. *Clinical and Experimental Immunology*, 133(1), 30–37. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1808750&tool=pmcentrez&rendertype=abstract>
- Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L., *et al.*, (2015). Southern East Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. *Proceedings of the National Academy of Sciences of the United States of America*, 112(26), 8136–8141. <http://doi.org/10.1073/pnas.1424063112>
- Macpherson, G. (1997). The Health Services since the War: Volume II. Government and Health Care: British National Health Service 1958-1979. *BMJ*, 314(7087). Retrieved from <http://www.bmj.com/content/314/7087/1136>
- Madkour M. Monir, Al-Otaibi E. Kitab, Al Swailem, R. (2004). *Tuberculosis*. Springer-Verlag (Vol. Chapter 2). Berlin Heidelberg.
- Madoff, L. C., Michel, J. L., Gong, E. W., Kling, D. E., & Kasper, D. L. (1996). Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proceedings of the National Academy of Sciences of the United States of America*, 93(9), 4131–4413. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8633028> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC39499>
- Magner, L. N. (2002). *A History of Infectious Diseases and the Microbial World*. Praeger.
- Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C., & Stover, C. K. (1996). Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M . bovis . Molecular Analysis of Genetic Differences between Mycobacterium bovis BCG and Virulent M . bovis, 178(5).

- Makinoshima, H., & Glickman, M. S. (2005). Regulation of *Mycobacterium tuberculosis* cell envelope composition and virulence by intramembrane proteolysis. *Nature*, 436(7049), 406–409. <http://doi.org/10.1038/nature03713>
- Malik, Z. A., Denning, G. M., & Kusner, D. J. (2000). Inhibition of Ca(2+) signaling by *Mycobacterium tuberculosis* is associated with reduced phagosome-lysosome fusion and increased survival within human macrophages. *The Journal of Experimental Medicine*, 191(2), 287–302. <http://doi.org/10.1084/jem.191.2.287>
- Malla, B. (2013). *Molecular Epidemiology of Mycobacterium tuberculosis in Nepal*. PhD Dissertation, University of Basel.
- Manganelli, R., Voskuil, M. I., Schoolnik, G. K., Dubnau, E., Gomez, M., & Smith, I. (2002). Role of the extracytoplasmic-function sigma factor sigma(H) in *Mycobacterium tuberculosis* global gene expression. *Molecular Microbiology*, 45(2), 365–374. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12123450>
- Mangtani, P., Abubakar, I., Ariti, C., Beynon, R., Pimpin, L., Fine, P. E. M., *et al.*, (2014). Protection by BCG vaccine against tuberculosis: A systematic review of randomized controlled trials. *Clinical Infectious Diseases*, 58(4), 470–480. <http://doi.org/10.1093/cid/cit790>
- Manzanillo, P. S., Shiloh, M. U., Portnoy, D. A., & Cox, J. S. (2012). *Mycobacterium tuberculosis* activates the DNA-dependent cytosolic surveillance pathway within macrophages. *Cell Host and Microbe*, 11(5), 469–480. <http://doi.org/10.1016/j.chom.2012.03.007>
- Marais, B. J., Mlambo, C. K., Rastogi, N., Zozio, T., Duse, A. G., Victor, T. C., *et al.*, (2013). Epidemic spread of multidrug-resistant tuberculosis in Johannesburg, South Africa. *Journal of Clinical Microbiology*. <http://doi.org/10.1128/JCM.00200-13>
- Martin, A., Bombeeck, D., Mulders, W., Fissette, K., De Rijk, P., & Palomino, J. C. (2011). Evaluation of the TB Ag MPT64 Rapid test for the identification of *Mycobacterium tuberculosis* complex [Technical note]. *The International Journal of Tuberculosis and Lung Disease*, 15(5), 703–705. <http://doi.org/10.5588/ijtld.10.0474>
- Mashta, A., Mishra, P., Philipose, S., Tamilzhalagan, S., Mahmud, H., Bhaskar, S., & Upadhyay, P. (2011). Diagnosis of tuberculosis: the experience at a specialized diagnostic laboratory. *Journal of Negative Results in BioMedicine*, 10(1), 16. <http://doi.org/10.1186/1477-5751-10-16>
- Mayer-Barber, K. D., Andrade, B. B., Barber, D. L., Hieny, S., Feng, C. G., Caspar, P., *et al.*, (2011). Innate and Adaptive Interferons Suppress IL-1 α and IL-1 β Production by Distinct Pulmonary Myeloid Subsets during *Mycobacterium tuberculosis* Infection. *Immunity*, 35(6), 1023–1034. <http://doi.org/10.1016/j.immuni.2011.12.002>
- Medjahed, H., Gaillard, J. L., & Reyrat, J. M. (2010). *Mycobacterium abscessus*: a new player in the mycobacterial field. *Trends in Microbiology*, 18(3), 117–123. <http://doi.org/10.1016/j.tim.2009.12.007>
- Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., *et al.*,

- (2015). Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nature Genetics*, 47(3), 242–249. <http://doi.org/10.1038/ng.3195>
- Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews. Genetics*, 11(1), 31–46. <http://doi.org/10.1038/nrg2626>
- Meyer, C. G., Scarisbrick, G., Niemann, S., Browne, E. N. L., Chinbuah, M. A., Gyapong, J., *et al.*, (2008). Pulmonary tuberculosis: Virulence of *Mycobacterium africanum* and relevance in HIV co-infection. *Tuberculosis*, 88(5), 482–489. <http://doi.org/10.1016/j.tube.2008.05.004>
- Mihret, A., Bekele, Y., Loxton, A. G., Aseffa, A., Howe, R., & Walzl, G. (2012). Plasma level of IL-4 differs in patients infected with different modern lineages of *M. tuberculosis*. *Journal of Tropical Medicine*, 2012. <http://doi.org/10.1155/2012/518564>
- Miner, M. D., Chang, J. C., Pandey, A. K., Sasseti, C. M., & Sherman, D. R. (2009). Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian Journal of Experimental Biology*, 47(6 SPEC. ISS.), 407–411. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19634704>
- Moghazeh, S. L., Pan, X., Arain, T., Stover, C. K., Musser, J. M., & Kreiswirth, B. N. (1996). Comparative antimycobacterial activities of rifampin, rifapentine, and KRM-1648 against a collection of rifampin-resistant *Mycobacterium tuberculosis* isolates with known *rpoB* mutations. *Antimicrobial Agents and Chemotherapy*, 40(11), 2655–2657. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=163595&tool=pmcentrez&rendertype=abstract>
- Mokrousov, I., Narvskaya, O., Limeschenko, E., Vyazovaya, A., Otten, T., & Vyshnevskiy, B. (2004). Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in *Mycobacterium tuberculosis* strains of the Beijing family: Practical implications and evolutionary considerations. *Journal of Clinical Microbiology*, 42(6), 2438–2444. <http://doi.org/10.1128/JCM.42.6.2438-2444.2004>
- Mortimer, T. D., & Pepperell, C. S. (2014). Genomic Signatures of Distributive Conjugal Transfer among *Mycobacteria*. *Genome Biology and Evolution*, 6(9), 2489–2500. <http://doi.org/10.1093/gbe/evu175>
- Mostowy, S., Inwald, J., Gordon, S., Warren, R., Kremer, K., Cousins, D., *et al.*, (2005). Revisiting the Evolution of *Mycobacterium bovis* Revisiting the Evolution of *Mycobacterium bovis*. *Journal of Bacteriology*, 187(18), 6386–6395. <http://doi.org/10.1128/JB.187.18.6386>
- Mostowy, S., Onipede, A., Gagneux, S., Niemann, S., Kremer, K., Desmond, E. P., *et al.*, (2004). Genomic analysis distinguishes *Mycobacterium africanum*. *Journal of Clinical Microbiology*, 42(8), 3594–3599. <http://doi.org/10.1128/JCM.42.8.3594-3599.2004>
- Musser, J. M., Amin, A., & Ramaswamy, S. (2000). Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure. *Genetics*, 155(1), 7–16. Retrieved from

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461055&tool=pmcentrez&rendertype=abstract>

- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics (Oxford, England)*, 29(22), 2933–2935. <http://doi.org/10.1093/bioinformatics/btt509>
- Neyrolles, O., & Guilhot, C. (2011). Recent advances in deciphering the contribution of Mycobacterium tuberculosis lipids to pathogenesis. *Tuberculosis*, 91(3), 187–195. <http://doi.org/10.1016/j.tube.2011.01.002>
- Nhu, N. T. Q., Heemskerk, D., Thu, D. D. A., Chau, T. T. H., Mai, N. T. H., Nghia, H. D. T., *et al.*, (2014). Evaluation of genexpert MTB/RIF for diagnosis of tuberculous meningitis. *Journal of Clinical Microbiology*, 52(1), 226–233. <http://doi.org/10.1128/JCM.01834-13>
- Niemann, S., Diel, R., Khechinashvili, G., Gegia, M., Mdivani, N., & Tang, Y. W. (2010). Mycobacterium tuberculosis Beijing lineage favors the spread of multidrug-resistant tuberculosis in the Republic of Georgia. *Journal of Clinical Microbiology*, 48(10), 3544–3550. <http://doi.org/10.1128/JCM.00715-10>
- Niemann, S., Kubica, T., Bange, F. C., Adjei, O., Browne, E. N., Chinbuah, M. A., *et al.*, (2004). The species Mycobacterium africanum in the light of new molecular markers. *Journal of Clinical Microbiology*, 42(9), 3958–3962. <http://doi.org/10.1128/JCM.42.9.3958-3962.2004>
- Niemann, S., Merker, M., Kohl, T., & Supply, P. (2016). Impact of Genetic Diversity on the Biology of Mycobacterium tuberculosis Complex Strains. *Microbiology Spectrum*, 4(6). <http://doi.org/10.1128/microbiolspec.TBTB2-0022-2016>
- Niemann, S., Rüsche-Gerdes, S., Joloba, M. L., Whalen, C. C., Guwatudde, D., Ellner, J. J., *et al.*, (2002). Mycobacterium africanum subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. *Journal of Clinical Microbiology*, 40(9), 3398–3405. <http://doi.org/10.1128/JCM.40.9.3398-3405.2002>
- Niobe-Eyangoh, S. N., Kuaban, C., Sorlin, P., Cunin, P., Thonnon, J., Sola, C., *et al.*, (2003). Genetic biodiversity of Mycobacterium tuberculosis complex strains from patients with pulmonary tuberculosis in Cameroon. *Journal of Clinical Microbiology*, 41(6), 2547–53. <http://doi.org/10.1128/JCM.41.6.2547-2553.2003>
- Nunes-Alves, C., Booty, M. G., Carpenter, S. M., Jayaraman, P., Rothchild, A. C., & Behar, S. M. (2014). In search of a new paradigm for protective immunity to TB. *Nature Publishing Group*, 12. <http://doi.org/10.1038/nrmicro3230>
- Nunes-Costa, D., Alarico, S., Dalcolmo, M. P., Correia-Neves, M., & Empadinhas, N. (2016). The looming tide of nontuberculous mycobacterial infections in Portugal and Brazil. *Tuberculosis*, 96. <http://doi.org/10.1016/j.tube.2015.09.006>
- Ofori-Anyinam, B., Kanuteh, F., Agbla, S. C., Adetifa, I., Okoi, C., Dolganov, G., *et al.*, (2016). Impact of the Mycobacterium africanum West Africa 2 Lineage on TB Diagnostics in West Africa: Decreased Sensitivity of Rapid Identification Tests in

- The Gambia. *PLoS Neglected Tropical Diseases*, 10(7), e0004801. <http://doi.org/10.1371/journal.pntd.0004801>
- Ohno, H., Koga, H., Kohno, S., Tashiro, T., & Hara, K. (1996). Relationship between rifampin MICs for and *rpoB* mutations of *Mycobacterium tuberculosis* strains isolated in Japan. *Antimicrobial Agents and Chemotherapy*, 40(4), 1053–1056. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=163262&tool=pmcentrez&rendertype=abstract>
- Orme, I. M., Robinson, R. T., & Cooper, A. M. (2015). The balance between protective and pathogenic immune responses in the TB-infected lung. *Nature Immunology*, 16(1), 57–63. <http://doi.org/10.1038/ni.3048>
- Otchere, I. D., Asante-poku, A., Osei-wusu, S., Baddoo, A., Sarpong, E., Ganiyu, A. H. H., *et al.*, (2016). Detection and characterization of drug-resistant conferring genes in *Mycobacterium tuberculosis* complex strains: A prospective study in two distant regions of. *Tuberculosis*, 99, 147–154. <http://doi.org/10.1016/j.tube.2016.05.014>
- Ozcaglar, C., Shabbeer, A., Vandenberg, S., Yener, B., & Bennett, K. P. (2011). Sublineage structure analysis of *Mycobacterium tuberculosis* complex strains using multiple- biomarker tensors. *BMC Genomics*, 12(Suppl 2), S1. <http://doi.org/10.1186/1471-2164-12-S2-S1>
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., *et al.*, (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22), 3691–3693. <http://doi.org/10.1093/bioinformatics/btv421>
- Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., *et al.*, (2016). Tuberculosis. *Nature Reviews Disease Primers*, 2, 16076. <http://doi.org/10.1038/nrdp.2016.76>
- Pandey, A. K., & Sassetti, C. M. (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), 4376–4380. <http://doi.org/10.1073/pnas.0711159105>
- Pandey, A. K., Yang, Y., Jiang, Z., Fortune, S. M., Coulombe, F., Behr, M. A., *et al.*, (2009). NOD2, RIP2 and IRF5 Play a Critical Role in the Type I Interferon Response to *Mycobacterium tuberculosis*. *PLoS Pathogens*, 5(7), :e1000500. <http://doi.org/10.1371/journal.ppat.1000500>
- Pang, Y., Lu, J., Wang, Y., Song, Y., Wang, S., & Zhao, Y. (2013). Study of the rifampin monoresistance mechanism in *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy*, 57(2), 893–900. <http://doi.org/10.1128/AAC.01024-12>
- Panigada, M., Sturniolo, T., Besozzi, G., Bocchieri, M. G., Sinigaglia, F., Grassi, G., *et al.*, (2002). Identification of a Promiscuous T cell Epitope in *Mycobacterium tuberculosis* Mce Proteins Identification of a Promiscuous T cell Epitope in *Mycobacterium tuberculosis* Mce Proteins. *Society*, 70(1), 79–85. <http://doi.org/10.1128/IAI.70.1.79>

- Paolo, W. F. J., & Nosanchuk, J. D. (2004). Tuberculosis in New York city: recent lessons and a look ahead. *The Lancet. Infectious Diseases*, 4(5), 287–293. [http://doi.org/10.1016/S1473-3099\(04\)01004-7](http://doi.org/10.1016/S1473-3099(04)01004-7)
- Perkins, M. D., & Cunningham, J. (2007). Facing the crisis: improving the diagnosis of tuberculosis in the HIV era. *The Journal of Infectious Diseases*, 196 Suppl, S15–S27. <http://doi.org/10.1086/518656>
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785–786. <http://doi.org/10.1038/nmeth.1701>
- Phillely, J. V., & Griffith, D. E. (2015). Treatment of Slowly Growing Mycobacteria. *Clinics in Chest Medicine*, 36(1), 79–90. <http://doi.org/10.1016/j.ccm.2014.10.005>
- Piccini, P., Chiappini, E., Tortoli, E., de Martino, M., & Galli, L. (2014). Clinical peculiarities of tuberculosis. *BMC Infectious Diseases*, 14 Suppl 1(Suppl 1), S4. <http://doi.org/10.1186/1471-2334-14-S1-S4>
- Popescu, A.-A., Huber, K. T., & Paradis, E. (2012). ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, 28(11), 1536–1537. <http://doi.org/10.1093/bioinformatics/bts184>
- Portevin, D., Gagneux, S., Comas, I., Young, D., Belardelli, F., Discriminate, C., *et al.*, (2011). Human macrophage responses to clinical isolates from the Mycobacterium tuberculosis complex discriminate between ancient and modern lineages. *PLoS Pathogens*, 7(3), e1001307. <http://doi.org/10.1371/journal.ppat.1001307>
- Pym, A. S., Brodin, P., Brosch, R., Huerre, M., & Cole, S. T. (2002). Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. *Molecular Microbiology*, 46(3), 709–717. <http://doi.org/10.1046/j.1365-2958.2002.03237.x>
- Quan, S., Venter, H., & Dabbs, E. R. (1997). Ribosylative inactivation of rifampin by Mycobacterium smegmatis is a principal contributor to its low susceptibility to this antibiotic. *Antimicrobial Agents and Chemotherapy*, 41(11), 2456–2460. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=164144&tool=pmcentrez&rendertype=abstract>
- R Core Team. (2015). R: A language and environment for statistical computing version 3.2.3. Retrieved July 26, 2017, from <https://www.coursehero.com/file/pe12cv/Code-R-version-323-2015-12-10-Wooden-Christmas-Tree-Copyright-C-2015-The-R/>
- Raman, S., Hazra, R., Dascher, C. C., & Husson, R. N. (2004). Transcription regulation by the Mycobacterium tuberculosis alternative sigma factor SigD and its role in virulence. *Journal of Bacteriology*. <http://doi.org/10.1128/JB.186.19.6605-6616.2004>
- Ramaswamy, S., & Musser, J. M. (1998). Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. *Tuber Lung Dis*, 79(1), 3–29. <http://doi.org/10.1054/tuld.1998.0002> rS0962-8479(98)90002-8 [pii]

- Ramaswamy, S. V, Reich, R., Dou, S., Jasperse, L., Pan, X., Wanger, A., *et al.*, (2003). Single Nucleotide Polymorphisms in Genes Associated with Isoniazid Resistance in Mycobacterium tuberculosis, 47(4), 1241–1250. <http://doi.org/10.1128/AAC.47.4.1241>
- Rao, V., Fujiwara, N., Porcelli, S. A., & Glickman, M. S. (2005). Mycobacterium tuberculosis controls host innate immune activation through cyclopropane modification of a glycolipid effector molecule. *J. Exp. Med.*, 201(4), 535–543. <http://doi.org/10.1084/jem.20041668>
- Rattan, A., Kalia, A., & Ahmad, N. (1998). Multidrug-resistant Mycobacterium tuberculosis: Molecular perspectives. *Emerging Infectious Diseases*, 4(2), 195–209. <http://doi.org/10.3201/eid0402.980207>
- Reiling, N., Homolka, S., & Walter, K. (2013). Clade-Specific Virulence Patterns of Mycobacterium tuberculosis. <http://doi.org/10.1128/mBio.00250-13.Editor>
- Rengarajan, J., Bloom, B. R., & Rubin, E. J. (2005). Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), 8327–8332. <http://doi.org/10.1073/pnas.0503272102>
- Ribeiro, S. C. M., Gomes, L. L., Amaral, E. P., Andrade, M. R. M., Almeida, F. M., Rezende, A. L., *et al.*, (2014). Mycobacterium tuberculosis strains of the modern sublineage of the beijing family are more likely to display increased virulence than strains of the ancient sublineage. *Journal of Clinical Microbiology*, 52(7), 2615–2624. <http://doi.org/10.1128/JCM.00498-14>
- Rinder, H., Thomschke, A., Rüsche-Gerdes, S., Bretzel, G., Feldmann, K., Rifai, M., & Löscher, T. (1998). Significance of ahpC promoter mutations for the prediction of isoniazid resistance in Mycobacterium tuberculosis. *European Journal of Clinical Microbiology and Infectious Diseases*, 17(7), 508–511. <http://doi.org/10.1007/s100960050116>
- Roguin, A. (2006). Rene Theophile Hyacinthe Laënnec (1781-1826): the man behind the stethoscope. *Clinical Medicine & Research*, 4(3), 230–235. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17048358>
- Rouse, D. A., & Morris, S. L. (1995). Molecular mechanisms of isoniazid resistance in Mycobacterium tuberculosis and Mycobacterium bovis. *Infection and Immunity*, 63(4), 1427–1433.
- Roy, A., Eisenhut, M., Harris, R. J., Rodrigues, L. C., Sridhar, S., Habermann, S., *et al.*, (2014). Effect of BCG vaccination against Mycobacterium tuberculosis infection in children: systematic review and meta-analysis. *BMJ (Clinical Research Ed.)*, 349(aug04_5), g4643. <http://doi.org/10.1136/bmj.g4643>
- Russell, D. G. (2011). Mycobacterium tuberculosis and the intimate discourse of a chronic infection. *Immunological Reviews*, 240(1), 252–268. <http://doi.org/10.1111/j.1600-065X.2010.00984.x>
- Saelens, J. W., Lau-Bonilla, D., Moller, A., Medina, N., Guzmán, B., Calderón, M., *et*

- al.*, (2015). Whole genome sequencing identifies circulating Beijing-lineage *Mycobacterium tuberculosis* strains in Guatemala and an associated urban outbreak. *Tuberculosis*, 95(6), 810–816. <http://doi.org/10.1016/j.tube.2015.09.001>
- Samstein, M., Schreiber, H. A., Leiner, I. M., Susac, B., Glickman, M. S., & Pamer, E. G. (2013). Essential yet limited role for CCR2+ inflammatory monocytes during *Mycobacterium tuberculosis*-specific T cell priming. *eLife*, 2, e01086. <http://doi.org/10.7554/eLife.01086>
- Santos, L. C. (2012). Review: The Molecular Basis of Resistance in *Mycobacterium tuberculosis*. *Open Journal of Medical Microbiology*, 2(1), 24–36. <http://doi.org/10.4236/ojmm.2012.21004>
- Sarkar, R., Lenders, L., Wilkinson, K. A., Wilkinson, R. J., & Nicol, M. P. (2012). Modern Lineages of *Mycobacterium tuberculosis* Exhibit Lineage-Specific Patterns of Growth and Cytokine Induction in Human Monocyte-Derived Macrophages. <http://doi.org/10.1371/journal.pone.0043170>
- Sassetti, C. M., Boyd, D. H., & Rubin, E. J. (2003). Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology*, 48(1), 77–84. <http://doi.org/10.1046/j.1365-2958.2003.03425.x>
- Schlesinger, L. S. (1993). Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. *The Journal of Immunology*, 150(7). Retrieved from <http://www.jimmunol.org/content/150/7/2920.long>
- Schouls, L., van der Heide, H., Witteveen, S., Zomer, B., van der Ende, A., Burger, M., & Schot, C. (2008). Two variants among *Haemophilus influenzae* serotype b strains with distinct *bcs4*, *hcsA* and *hcsB* genes display differences in expression of the polysaccharide capsule. *BMC Microbiology*, 8, 35. <http://doi.org/10.1186/1471-2180-8-35>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*. <http://doi.org/10.1093/bioinformatics/btu153>
- Segen, J. (1992). The dictionary of modern medicine. *Park Ridge, NJ: Parthenon*.
- Sherman, D. R., Mdluli, K., Hickey, M. J., Arain, T. M., Morris, S. L., Barry, C. E., & Stover, C. K. (1996). Compensatory *ahpC* Gene Expression in Isoniazid-Resistant *Mycobacterium tuberculosis*. *Science*, 272(5268), 1641–1643. <http://doi.org/10.1126/science.272.5268.1641>
- Sinha, P., Prakash, P., Patne, S. C. U., Anupurba, S., Gupta, S., & Srivastava, G. N. (2017). Performance of nested multiplex PCR assay targeting MTP40 and IS6110 gene sequences for the diagnosis of tubercular lymphadenitis. *Journal of Microbiology*, 55(1), 63–67. <http://doi.org/10.1007/s12275-017-6127-y>
- Sinsimer, D., Huet, G., Manca, C., Tsenova, L., Koo, M. S., Kurepina, N., *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), 3027–3036. <http://doi.org/10.1128/IAI.01663-07>

- Smith, N. H., Gordon, S. V., de la Rua-Domenech, R., Clifton-Hadley, R. S., & Hewinson, R. G. (2006). Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nature Reviews Microbiology*, 4(9), 670–681. <http://doi.org/10.1038/nrmicro1472>
- Somoskovi, a, Parsons, L. M., & Salfinger, M. (2001). The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*. *Respiratory Research*, 2(3), 164–168. <http://doi.org/10.1186/rr54>
- Soolingen, D. V. A. N., Hermans, P. W. M., Haas, P. E. W. D. E., Soll, D. R., Embden, J. A. N. D. A. V. A. N., van Soolingen, D., *et al.*, (1991). Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *Journal of Clinical Microbiology*, 29(11), 2578–2586. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=270376&tool=pmcentrez&rendertype=abstract>
- Soolingen, D. Van, Haas, P. E. De, Hermans, P. W., Groenen, P. M., & Embden, D. Van. (1993). as genetic markers for strain differentiation and Comparison of Various Repetitive DNA Elements as Genetic Markers for Strain Differentiation and Epidemiology of *Mycobacterium tuberculosis*.
- Spratt, B. G., Greenwood, B. M., Dagan, R., Dixon, J., Eykyn, S., & Fenoll, A. (2000). Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet (London, England)*, 356(9237), 1210–1. [http://doi.org/10.1016/S0140-6736\(00\)02779-3](http://doi.org/10.1016/S0140-6736(00)02779-3)
- Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S., & Musser, J. M. (1997). Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences of the United States of America*, 94(18), 9869–9874. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=23284&tool=pmcentrez&rendertype=abstract>
- Staden, R., Beal, K. F., & Bonfield, J. K. (2000). The Staden package, 1998. *Methods in Molecular Biology (Clifton, N.J.)*, 132, 115–130. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10547834>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <http://doi.org/10.1093/bioinformatics/btu033>
- Stanford, J. L., Grange, J. M., Gernez-Rieux, C., Jenkins, P. A., Kappler, W., Kubica, G. P., *et al.*, (1974). The meaning and structure of species as applied to mycobacteria. *Tubercle*, 55(2), 143–152. [http://doi.org/10.1016/0041-3879\(74\)90008-7](http://doi.org/10.1016/0041-3879(74)90008-7)
- Stanley, S. A., Raghavan, S., Hwang, W. W., & Cox, J. S. (2003). Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. *Proceedings of the National Academy of Sciences of the United*

- States of America*, 100(22), 13001–13006. <http://doi.org/10.1073/pnas.2235593100>
- Steingart, K. R., Ng, V., Henry, M., Hopewell, P. C., Ramsay, A., Cunningham, J., *et al.*, (2006). Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *The Lancet. Infectious Diseases*, 6(10), 664–674. [http://doi.org/10.1016/S1473-3099\(06\)70602-8](http://doi.org/10.1016/S1473-3099(06)70602-8)
- Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., *et al.*, (2016). Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature Genetics*, 48(12), 1535–1543. <http://doi.org/10.1038/ng.3704>
- Stucki, D., & Gagneux, S. (2012). Single nucleotide polymorphisms in Mycobacterium tuberculosis and the need for a curated database. *Tuberculosis*, 93(1), 30–39. <http://doi.org/10.1016/j.tube.2012.11.002>
- Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., *et al.*, (2013). Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. *Nature Genetics*, 45(2), 172–179. <http://doi.org/10.1038/ng.2517>
- Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B., & Locht, C. (2000). Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Molecular Microbiology*, 36(3), 762–771. <http://doi.org/10.1046/j.1365-2958.2000.01905.x>
- Tameris, M. D., Hatherill, M., Landry, B. S., Scriba, T. J., Snowden, M. A., Lockhart, S., *et al.*, (2013). Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Www.thelancet.com*, 381(23). [http://doi.org/10.1016/S0140-6736\(13\)60177-4](http://doi.org/10.1016/S0140-6736(13)60177-4)
- TB CARE I. (2014). *International Standards for Tuberculosis Care, Edition 3*. The Hague.
- Tekaia, F., Gordon, S. V., Garnier, T., Brosch, R., Barrell, B. G., & Cole, S. T. (1999). Analysis of the proteome of Mycobacterium tuberculosis in silico. *Tubercle and Lung Disease : The Official Journal of the International Union against Tuberculosis and Lung Disease*, 79(6), 329–342. <http://doi.org/10.1054/tuld.1999.0220>
- Telenti, A., Imboden, P., Marchesi, F., Schmidheini, T., & Bodmer, T. (1993). Direct, automated detection of rifampin-resistant Mycobacterium tuberculosis by polymerase chain reaction and single-strand conformation polymorphism analysis. *Antimicrobial Agents and Chemotherapy*, 37(10), 2054–2058. <http://doi.org/10.1128/AAC.37.10.2054>
- Tientcheu, L. D., Maertzdorf, J., Weiner, J., Adetifa, I. M., Mollenkopf, H.-J., Sutherland, J. S., *et al.*, (2015). Differential transcriptomic and metabolic profiles of M. africanum- and M. tuberculosis-infected patients after, but not before, drug treatment. *Genes and Immunity*, 16(5), 347–355. <http://doi.org/10.1038/gene.2015.21>
- Trunz, B. B., Fine, P., & Dye, C. (2006). Effect of BCG vaccination on childhood

- tuberculous meningitis and miliary tuberculosis worldwide: a meta-analysis and assessment of cost-effectiveness. *Lancet*, 367(9517), 1173–1180. [http://doi.org/10.1016/S0140-6736\(06\)68507-3](http://doi.org/10.1016/S0140-6736(06)68507-3)
- Tsolaki, A. G., Hirsh, A. E., DeRiemer, K., Enciso, J. A., Wong, M. Z., Hannan, M., *et al.*, (2004). Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14), 4865–4870. <http://doi.org/10.1073/pnas.0305634101>
- Tsukamura, M. (1976). An Approach to Numerical Identification of Bacterial Species. *Journal of General Microbiology*, 95(2), 207–212. <http://doi.org/10.1099/00221287-95-2-207>
- Tuberculosis research centre. (1999). Fifteen year follow up of trial of BCG vaccines in south India for tuberculosis prevention. Tuberculosis Research Centre (ICMR), Chennai. *The Indian Journal of Medical Research*, 110, 56–69. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10573656>
- Valway, S. E., Richards, S. B., Kovacovich, J., Greifinger, R. B., Crawford, J. T., & Dooley, S. W. (1994). Outbreak of multi-drug-resistant tuberculosis in a New York State prison, 1991. *American Journal of Epidemiology*, 140(2), 113–122. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8023800>
- van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., *et al.*, (2007). M. tuberculosis and M. leprae Translocate from the Phagolysosome to the Cytosol in Myeloid Cells. *Cell*, 129(7), 1287–1298. <http://doi.org/10.1016/j.cell.2007.05.059>
- van Embden, J. D., Cave, M. D., Crawford, J. T., Dale, J. W., Eisenach, K. D., Gicquel, B., *et al.*, (1993). Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *Journal of Clinical Microbiology*, 31(2), 406–409. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=262774&tool=pmcentrez&rendertype=abstract>
- Van Ingen, J., Boeree, M. J., Van Soolingen, D., & Mouton, J. W. (2012). Resistance mechanisms and drug susceptibility testing of nontuberculous mycobacteria. *Drug Resistance Updates*, 15(3), 149–161. <http://doi.org/10.1016/j.drug.2012.04.001>
- van Soolingen, D., Hoogenboezem, T., de Haas, P. E., Hermans, P. W., Koedam, M., Teppema, K. S., *et al.*, (1997). A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *International Journal of Systematic Bacteriology*, 47(4), 1236–1245. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9336935>
- Viana-niero, C., Gutierrez, C., Sola, C., Filliol, I., Rastogi, N., & Boulahbal, F. (2001). Genetic Diversity of. *Society*, 39(1), 57–65. <http://doi.org/10.1128/JCM.39.1.57>
- Vilchèze, C., & Jacobs, W. R. (2007). The mechanism of isoniazid killing: clarity through the scope of genetics. *Annual Review of Microbiology*, 61, 35–50. <http://doi.org/10.1146/annurev.micro.61.111606.122346>

- Vilchèze, C., Weisbrod, T. R., Chen, B., Kremer, L., Hazbón, M. H., Wang, F., *et al.*, (2005). Altered NADH/NAD⁺ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrobial Agents and Chemotherapy*, 49(2), 708–720. <http://doi.org/10.1128/AAC.49.2.708-720.2005>
- Walters, E., Cotton, M. F., Rabie, H., Schaaf, H. S., Walters, L. O., & Marais, B. J. (2008). Clinical presentation and outcome of Tuberculosis in Human Immunodeficiency Virus infected children on anti-retroviral therapy. *BMC Pediatrics*, 8(1), 1. <http://doi.org/10.1186/1471-2431-8-1>
- Wei, C. J., Lei, B., Musser, J. M., & Tu, S. C. (2003). Isoniazid activation defects in recombinant Mycobacterium tuberculosis catalase-peroxidase (KatG) mutants evident in InhA inhibitor production. *Antimicrobial Agents and Chemotherapy*, 47(2), 670–675. <http://doi.org/10.1128/AAC.47.2.670-675.2003>
- Wells, A. Q. (1937). TUBERCULOSIS IN WILD VOLES. *The Lancet*, 229(5934), 1221. [http://doi.org/10.1016/S0140-6736\(00\)83505-9](http://doi.org/10.1016/S0140-6736(00)83505-9)
- Who. (2010). Global Tuberculosis Control 2010. *Control*, WHO/HTM/TB, 218. <http://doi.org/WHO/HTM/TB 2010.7>
- WHO. (2011). Global tuberculosis control: WHO report 2011. *Who*, 258. <http://doi.org/WHO/HTM/TB/2011.16>
- WHO. (2015). WHO Global tuberculosis report 2015, 1, 1689–1699. <http://doi.org/10.1017/CBO9781107415324.004>
- WHO. (2016). *Global Tuberculosis Report 2016. Tuberculosis Report*. Geniva. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0>
- Wickham, H. (2007). Reshaping Data with the \pkg{reshape} Package. *Journal of Statistical Software*, 21(12), 1–20. [http://doi.org/10.1016/S0142-1123\(99\)00007-9](http://doi.org/10.1016/S0142-1123(99)00007-9)
- Winglee, K., Manson McGuire, A., Maiga, M., Abeel, T., Shea, T., Desjardins, C. A., *et al.*, (2016). Whole Genome Sequencing of Mycobacterium africanum Strains from Mali Provides Insights into the Mechanisms of Geographic Restriction. *PLoS Neglected Tropical Diseases*, 10(1), e0004332. <http://doi.org/10.1371/journal.pntd.0004332>
- Wolf, A. J., Desvignes, L., Linas, B., Banaiee, N., Tamura, T., Takatsu, K., & Ernst, J. D. (2008). Initiation of the adaptive immune response to Mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. *The Journal of Experimental Medicine*, 205(1), 105–115. <http://doi.org/10.1084/jem.20071367>
- World Health Organization. (2016). *THE SHORTER MDR-TB REGIMEN FEATURES OF THE SHORTER MDR-TB REGIMEN*. Geniva. Retrieved from http://www.who.int/tb/Short_MDR_regimen_factsheet.pdf
- Yang, X., Gao, J., Smith, I., Dubnau, E., & Sampson, N. S. (2011). Cholesterol is not an essential source of nutrition for Mycobacterium tuberculosis during infection. *Journal of Bacteriology*, 193(6), 1473–1476. <http://doi.org/10.1128/JB.01210-10>
- Yeboah-manu, D., Asante-poku, A., Bodmer, T., Stucki, D., Koram, K., Bonsu, F., *et al.*,

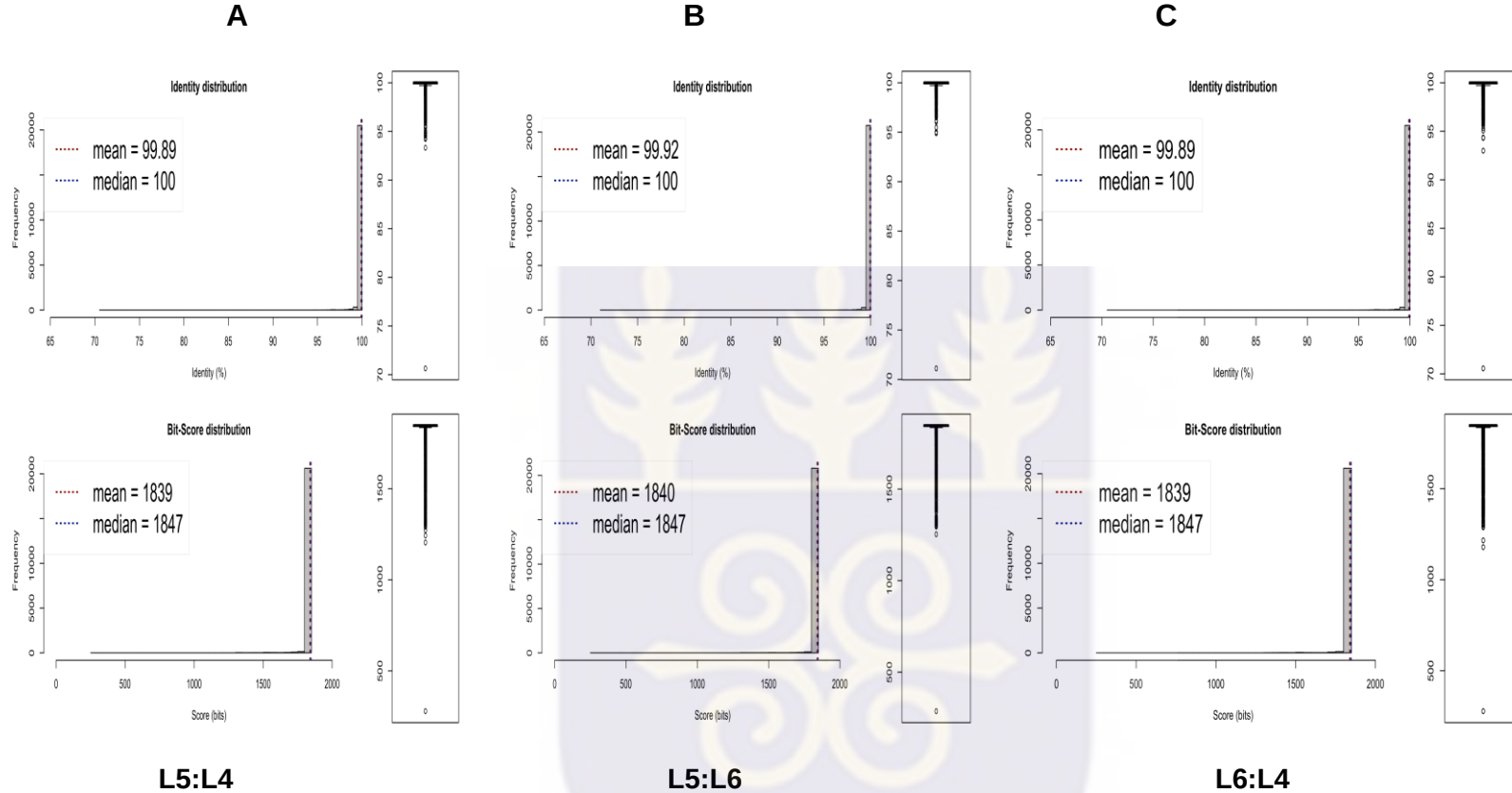
- (2011). Genotypic diversity and drug susceptibility patterns among *M. tuberculosis* complex isolates from South-Western Ghana. *PloS One*, 6(7), e21906. <http://doi.org/10.1371/journal.pone.0021906>
- Yeboah-Manu, D., Asare, P., Asante-Poku, A., Otchere, I. D., Osei-Wusu, S., Danso, E., *et al.*, (2016). Spatio-Temporal Distribution of *Mycobacterium tuberculosis* Complex Strains in Ghana. *Plos One*, 11(8), e0161892. <http://doi.org/10.1371/journal.pone.0161892>
- Yeboah-manu, D., Bodmer, T., Owusu, S., Ofori-adjei, D., Pluschke, G., & Mensah-quainoo, E. (2004). Evaluation of Decontamination Methods and Growth Media for Primary Isolation of *Mycobacterium ulcerans* from Surgical Specimens, 5875–5877. <http://doi.org/10.1128/JCM.42.12.5875>
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. <http://doi.org/10.1101/gr.074492.107>
- Zhou, P., & Xie, J. (2014). Comparative genomics of the *Mycobacterium* signaling architecture and implications for a novel live attenuated Tuberculosis vaccine. *Human Vaccines & Immunotherapeutics*, 10(1), 159–163. <http://doi.org/10.4161/hv.26268>
- Zink, A. R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., & Nerlich, A. G. (2003). Characterization of *Mycobacterium tuberculosis* Complex DNAs from Egyptian Mummies by Spoligotyping. *Society*, 41(1), 359–367. <http://doi.org/10.1128/JCM.41.1.359>
- Zumla, A., Otchere, I. D., Mensah, G. I., Asante-Poku, A., Gehre, F., Maeurer, *et al.*, (2017). Learning from epidemiological, clinical, and immunological studies on *Mycobacterium africanum* for improving current understanding of host–pathogen interactions, and for the development and evaluation of diagnostics, host-directed therapies, and vaccines f. *International Journal of Infectious Diseases*, 56, 126–129. <http://doi.org/10.1016/j.ijid.2016.12.003>
- Zwerling, A., Behr, M. A., Verma, A., Brewer, T. F., Menzies, D., & Pai, M. (2011). The BCG world atlas: A database of global BCG vaccination policies and practices. *PLoS Medicine*, 8(3). <http://doi.org/10.1371/journal.pmed.1001012>
- Zysk, G. K. (1998). Medicine in the Veda: Religious Healing in the Veda: with Translations and Annotations of Medical Hymns from the Rgveda and the Atharvaveda and Renderings from the Corresponding Ritual Texts. *Nanarsidass Publications*.

APPENDICES



Appendix A: Phylogeny of Ghanaian MTBC with Global Collection rooted on *M. canettii*.

Each lineage is represented by the universal colour codes: L1 (7 genomes):pink, L2 (5 genomes):blue, L3 (2 genomes):purple, L4 (21 genomes): red, L5 (177 genomes):brown and L6 (80 genomes):green.



Appendix B: Average nucleotide identity (ANI) among the MTBC lineages

The average nucleotide identity calculated between L4 and L5 (Pannel A), L5 and L6 (Pannel B) and L4 and L6 (Pannel C). The ANI between any two lineages is always greater than 99%

Appendix C: Specific deletion of genes among MTBC lineages and strains

Gene Absent	Annotation	Function	L4	L5	L6
<i>lppt</i>	lipoprotein LppT	cell wall and cell processes	Absent (11)		
<i>apr</i>	membrane-anchored mycosin mycP4	Acid and phagosome regulated protein	Absent		
<i>desA3</i>	linoleoyl-CoA desaturase	lipid metabolism	Absent		
<i>cyp142b</i>	cytochrome P450 monooxygenase	intermediary metabolism and respiration	Absent		
<i>cobLa</i>	precorrin-6y methyltransferase	intermediary metabolism and respiration	Absent		
<i>alda</i>	L-alanine dehydrogenase	intermediary metabolism and respiration	Absent		
<i>eccC</i>	ESX conserved component EccC2	cell wall and cell processes	Absent (15)	Absent (11)	
<i>pds</i>		not found in H37Rv	Absent		Absent (15, L6.3)
<i>pks7</i>	polyketide synthase	lipid metabolism	Absent	Absent (91)	Absent (28)
<i>icL</i>	isocitrate lyase	intermediary metabolism and respiration		Absent	Absent
<i>tal</i>	transaldolase	intermediary metabolism and respiration	Absent	Absent	Absent (26; but 10:L6.2)
<i>esxO_2</i>	Esat-6 like protein EsxO	cell wall and cell processes	Absent (6)	Absent (70)	Absent (26)
<i>arsb2</i>	arsenical pump integral membrane protein ARSB2	cell wall and cell processes		Absent (26)	Absent (21)
<i>PPE67</i>	PPE family protein			Absent	
<i>PPE56</i>	PPE family protein		Absent	Absent	
<i>PE_PGSR4</i>	PE-PGRS family protein		Absent	Absent (62)	Absent (23)

Gene Absent	Annotation	Function	L4	L5	L6
<i>pknH2</i>	Ser/Thr protein kinase	regulatory proteins	Absent	Absent (72)	Absent (24)
<i>pknH_2</i>	putative transmembrane serine/threonine-protein kinase H pknH	regulatory proteins	Absent	Absent (70)	Absent (26)
<i>pknH_1</i>	putative transmembrane serine/threonine-protein kinase H pknH	regulatory proteins	Absent	Absent (81)	Absent (23)
<i>galTb</i>	Probable galactose-1-phosphate uridylyltransferase GalTb	intermediary metabolism and respiration		Absent	Absent (24)
<i>stp_3</i>	drug transporter	cell wall and cell processes (efflux of drugs)	Absent	Absent (91)	Absent (32)
<i>dxs_3</i>	1-deoxy-D-xylulose-5-phosphate synthase	intermediary metabolism and respiration	Absent (9)	Absent	Absent (26)
<i>pyrH</i>	uridylate kinase	intermediary metabolism and respiration			Absent (12)
<i>PPE50</i>	PPE family protein		Absent (15)		
<i>frdB_2</i>	fumarate reductase	intermediary metabolism and respiration	Absent (15)		
<i>mmpL2</i>	transporter	cell wall and cell processes		Absent (8)	Absent (24)
<i>mlaD</i>	mammalian cell entry protein				Absent
<i>mymT</i>	Metallothionein, MymT	virulence, detoxification, adaptation			Absent
<i>echA1</i>	putative enoyl-CoA hydratase ECHA1	lipid metabolism			Absent
<i>frdB_1</i>	succinate dehydrogenase	intermediary metabolism and resp.			Absent

Gene Absent	Annotation	Function	L4	L5	L6
<i>PE32</i>	PE family protein				Absent
<i>mce3D</i>	MCE-family protein MCE3D	virulence, detoxification , adaptation			Absent
<i>rho</i>	transcription termination factor Rho	information pathways			Absent
<i>mce3F</i>	MCE-family protein MCE3F	virulence, detoxification , adaptation			Absent
<i>yrbE3A</i>	putative integral membrane protein YrbE3A	virulence, detoxification , adaptation			Absent
<i>PE_PGRS61</i>	PE-PGRS family protein			Absent (29)	Absent (8)
<i>bglB</i>	beta-glucosidase bglS				Absent
<i>viuB</i>	siderophore utilization protein	intermediary metabolism and respiration			Absent
<i>arnB</i>	TDP-4-oxo-6- deoxy-D-glucose transaminase				Absent
<i>mpt64</i>	immunogenic protein MPT64	cell wall and cell processes			Absent
<i>yjce</i>	transporter YjcE	cell wall and cell processes		Absent (37)	

Appendix D: Specific mutations identified among the individual lineages

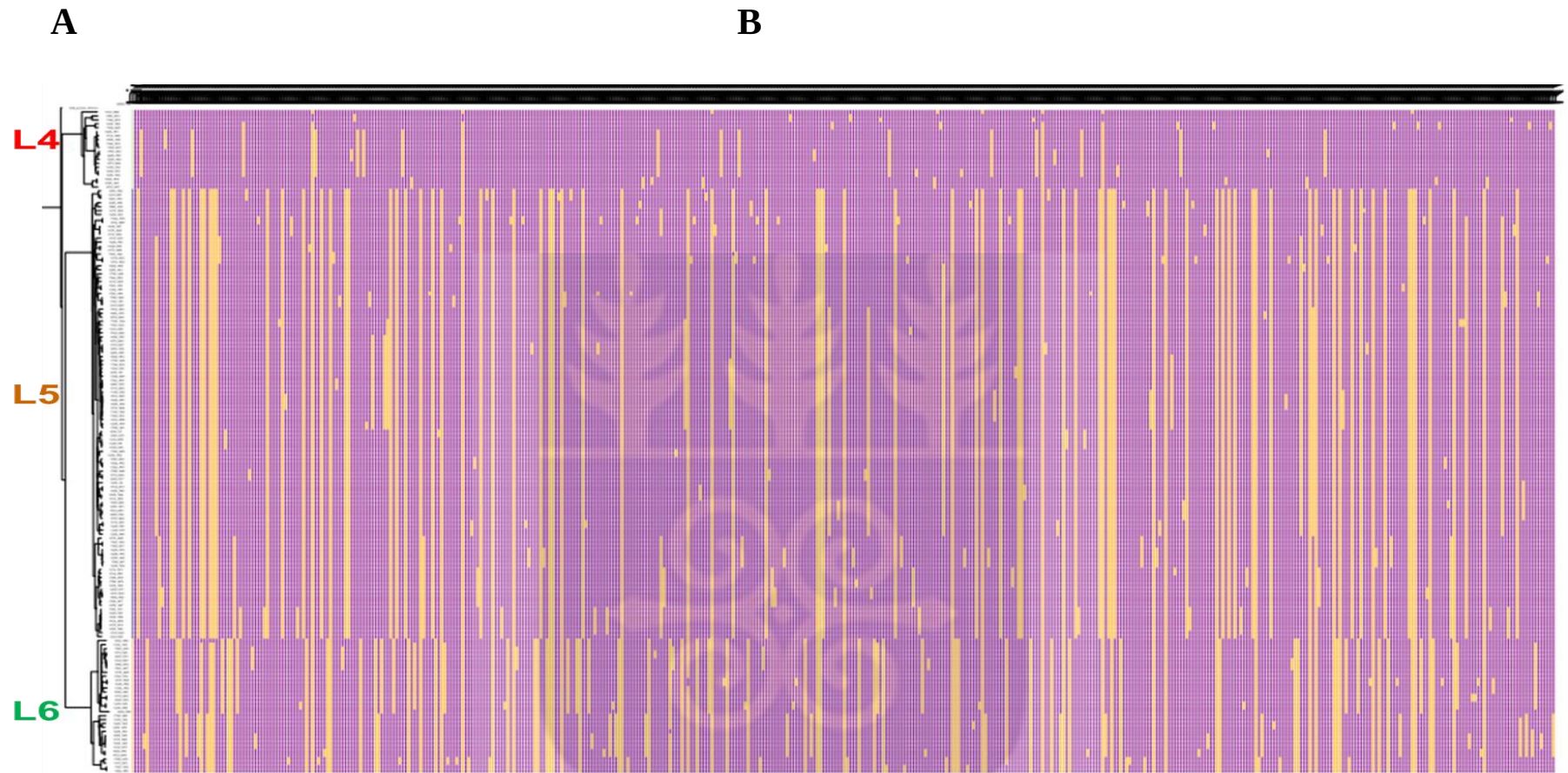
Gene	Common Name	Substitution	insertion	deletion	Lineage	cluster
<i>Rv0048c</i>	<i>Rv0048c</i>	G131A			L4	17 strains
<i>Rv0092</i>	<i>ctpA</i>	G331T			L4	L4.1
<i>Rv0107c</i>	<i>ctpI</i>	G2426A			L5	77 strains
<i>Rv0171</i>	<i>mce1C</i>			del AG 1408/9	L5	all
<i>Rv0171</i>	<i>mce1C</i>	T1236A			L6	L6.3(21)
<i>Rv0172</i>	<i>mce1D</i>	C839T			L5	23 not L5.5 & L5.4
<i>Rv0172</i>	<i>mce1D</i>	A275G, G276A			L6	L6.2(11)
<i>Rv0174</i>	<i>mce1F</i>	G1258C			L4	11(L4.1)
<i>Rv0174</i>	<i>mce1F</i>	T393C			L6	L6.2(11)
<i>Rv0177</i>	<i>mce-like</i>	C23T			L6	L6.1/2(15)
<i>Rv0202c</i>	<i>mmpL11</i>	A1566G			L6	L6.3(21)
<i>Rv0206c</i>	<i>mmpL3</i>	C2771T			L6	L6.3(21)
<i>Rv0261</i>	<i>narK3</i>	C280T			L4	L4.1/deut
<i>Rv0265c</i>	<i>Rv0265c</i>	C470T			L4	L4.1
<i>Rv0425c</i>	<i>ctpH</i>	C4395T			L6	L6.21(10)
<i>Rv0450c</i>	<i>mmpL4</i>	C1878T			L4	L4.1
<i>Rv0450c</i>	<i>mmpL4</i>	C291T			L6	L6.2/1(15)
<i>Rv0470c</i>	<i>pcaA</i>	G290A			L4	11(L4.1)
<i>Rv0507</i>	<i>mmpL2</i>			del T2119	L4	L4.1
<i>Rv0507</i>	<i>mmpL2</i>	C1162T			L6	L6.2/1(15)
<i>Rv0507</i>	<i>mmpL2</i>	G1606T			L5	L5.5
<i>Rv0522</i>	<i>gabP</i>	C579A and G1205T			L6	L6.3(21)
<i>Rv0642c</i>	<i>mmaA4</i>	C779T			L6	L6.3(21)
<i>Rv0862c</i>	<i>Rv0862c</i>	C530T			L6	L6.3(21)
<i>Rv0931</i>	<i>pknD</i>	C1570G			L6	L6.1(4)
<i>Rv1146</i>	<i>mmpL13b</i>		Ins A232		L5	all
<i>Rv1178</i>	<i>Rv1179</i>	C1079T			L6	L6.1(4)
<i>Rv1183</i>	<i>mmpL10</i>	C379T			L6	L6.3(21)
<i>Rv1242</i>	<i>vapC33</i>	G74T			L6	L6.1(2)(15)
<i>Rv1246c</i>	<i>relE</i>	G265A			L6	L6.2.1(7)
<i>Rv1280c</i>	<i>oppA</i>	G318A			L6	L6.3(21)
<i>Rv1281c</i>	<i>oppD</i>	C1570T			L4	L4.1
<i>Rv1281c</i>	<i>oppD</i>	G478A			L6	L6.1(4)
<i>Rv1281c</i>	<i>oppD</i>	C1115A			L6	L6.2(11)
<i>Rv1328</i>	<i>glgP</i>	G2553A			L6	L6.1(4)
<i>Rv1345</i>	<i>mbtM</i>	G1075A			L6	L6.3(21)

Gene	Common Name	Substitution	insertion	deletion	Lineage	cluster
<i>Rv1347</i>	<i>mbtK</i>	G375A			L6	L6.2(11)
<i>Rv1350</i>	<i>fabG2</i>	C250T			L5	31(L5.5)
<i>Rv1592c</i>	<i>Rv1592c</i>	C822T and C823T			L4	17 strains
<i>Rv1592c</i>	<i>Rv1592c</i>	C141T and G1121A			L6	L6.3(21)
<i>Rv1607</i>	<i>chaA</i>	G994T			L4	L4.1
<i>Rv1704c</i>	<i>cycA</i>	C1578T			L6	L6.2.1.1(5)
<i>Rv1739c</i>	<i>Rv1739c</i>	T813C			L6	L6.1(4)
<i>Rv1747</i>	<i>Rv1748</i>	T447C			L4	11(L4.1)
<i>Rv1811</i>	<i>mgtC</i>	C543T			L6	L6.2(15)
<i>Rv1979c</i>	<i>Rv1979c</i>		Ins CGG 638(40)		L4	11(L4.1)
<i>Rv1999c</i>	<i>Rv1999c</i>	C1259T			L5	Sub of L5.5(II)
<i>Rv1999c</i>	<i>Rv1999c</i>	C802A			L6	L6.3(21)
<i>Rv2041c</i>	<i>Rv2041c</i>	G368A			L6	L6.1(4)
<i>Rv2112c</i>	<i>dop</i>	G841A			L6	L6.1(4)
<i>Rv2316</i>	<i>uspA</i>	G276A			L6	L6.1(4)
<i>Rv2320c</i>	<i>rocE</i>	A82G			L6	L6.2.1(7)
<i>Rv2329c</i>	<i>narK1</i>	G950A			L6	L6.3(21)
<i>Rv2383</i>	<i>mbtB</i>	G1945A			L5	77
<i>Rv2383c</i>	<i>mbtB</i>			del GGT(2399(401))	L4	11(L4.1)
<i>Rv2383c</i>	<i>mbtB</i>	C814T			L6	L6.3(21)
<i>Rv2397c</i>	<i>cysA1</i>	G191T			L6	L6.3(21)
<i>Rv2524c</i>	<i>fas</i>	T8619C			L4	11(L4.1)
<i>Rv2524c</i>	<i>fas</i>	G3249A			L6	L6.2(4)
<i>Rv2563</i>	<i>Rv2564</i>			del G81	L4, L5 and L6	L4 (19/22), all L5 and all L6
<i>Rv2585c</i>	<i>Rv2585c</i>	G1365A			L4	L4.1
<i>Rv2692</i>	<i>ceoC</i>	G294A			L6	L6.2(11)
<i>Rv2737c</i>	<i>recA</i>	C819A			L6	L6.3(21)
<i>Rv2833c</i>	<i>ugpB</i>	G1199T			L4	L4.1
<i>Rv2833c</i>	<i>ugpB</i>	G1056A			L6	L6.1(4)
<i>Rv2856</i>	<i>nicT</i>	T157C			L4	L4.1
<i>Rv2856</i>	<i>nicT</i>	C232T			L6	L6.3(21)
<i>Rv2869c</i>	<i>rip</i>	A823C			L6	L6.1(4)
<i>Rv2902c</i>	<i>rnhB</i>	A1316G			L4	L4.1/2
<i>Rv3041c</i>	<i>Rv3041c</i>	C471A			L6	L6.2.1
<i>Rv3041c</i>	<i>Rv3041c</i>	C284A			L6	L6.2(15)

Gene	Common Name	Substitution	insertion	deletion	Lineage	cluster
<i>Rv3044</i>	<i>fecB</i>	G103C			L4	L4.1
<i>Rv3200c</i>	<i>Rv3200c</i>	G89A			L6	L6.3
<i>Rv3236</i>	<i>kefB</i>	G888A			L6	L6.3(3)(4)
<i>Rv3236c</i>	<i>Rv3236c</i>	C1095T			L6	L6.1(4)
<i>Rv3499c</i>	<i>mce4A</i>	T1064C			L6	L6.1(4)
<i>Rv3540</i>	<i>ltp2</i>	G924A			L6	L6.2.1.1.1(4)
<i>Rv3543c</i>	<i>fadE29</i>	G520A			L5	77 strains
<i>Rv3545c</i>	<i>cyp125</i>	C113T			L6	L6.2(11)
<i>Rv3567c</i>	<i>hsaB</i>	C163A			L5	72 strains
<i>Rv3567c</i>	<i>hsaB</i>	A552G			L6	L6.3(21)
<i>Rv3570c</i>	<i>hsaA</i>	A964C			L5	57 strains
<i>Rv3663c</i>	<i>dppD</i>	G315A			L6	L6.3(21)
<i>Rv3665c</i>	<i>dppB</i>	C571A			L4	11(L4.1)
<i>Rv3679</i>	<i>Rv3679</i>	G300T			L4	L4.1
<i>Rv3680</i>	<i>Rv3680</i>	T736A			L6	L6.3
<i>Rv3682</i>	<i>ponA2</i>	T2197C			L6	L6.3
<i>Rv3743c</i>	<i>ctpJ</i>	G1619T			L6	L6.21(7)
<i>Rv3743c</i>	<i>ctpJ</i>	G430A			L6	L6.1(4)
<i>Rv3757c</i>	<i>proW</i>	C622T			L6	L6.3.1(7)
<i>Rv3759c</i>	<i>proX</i>	C87T			L6	L6.2(11)
<i>Rv3823c</i>	<i>mmpL8</i>	G2721A			L6	L6.3(21)
<i>Rv3864</i>	<i>espE</i>	A1124C			L4	11(L4.1)
<i>Rv3864</i>	<i>espE</i>	G1112			L6	L6.1(4)
<i>Rv3864</i>	<i>espE</i>	C771T			L6	L6.1(2)(15)
<i>Rv3864</i>	<i>espE</i>	T1098A			L6	L6.3(21)
<i>Rv3866</i>	<i>espG1</i>		Ins CGG CCG CAG GCG 6(84)95		L5	all
<i>Rv3868</i>	<i>eccA1</i>	G15/3A			L5	84
<i>Rv3870</i>	<i>eccCa1</i>	C1359T			L6	L6.1(4)
<i>Rv3870</i>	<i>eccCa1</i>	T1412C			L6	L6.2(11)
<i>Rv3870</i>	<i>eccCa1</i>	G1705T			L6	L6.2.1
<i>Rv3871</i>	<i>eccCb1</i>	A45C			L6	L6.2(11)
<i>Rv3871</i>	<i>eccCb1</i>	G1116A			L6	L6.1(4)
<i>Rv3871</i>	<i>eccCb1</i>	C1440T			L6	L6.3(21)
<i>Rv3880c</i>	<i>espL</i>	C146G			L4	18 strains

Appendix E: Number of accumulated mutation per lineages

Cluster of Genes	Lin 4 (18)	Lin 5 (93)	Lin 6 (38)
Sulfur transporters	3	14	9
Phosphorus transporters	6	31	7
Nitrogen transporters	4	19	5
Ion transporters	42	145	85
Carbohydrate transporters	14	24	32
lipid / fat transporters	29	59	42
Peptide transporters	13	22	15
Amino acid transporters	31	63	43
Transport / cell invasion operon	7	24	18
Tipid catabolism operon	2	11	7
Cholesterol catabolism operon	2	8	1
Phosphorus transport operon	3	1	1
Carbohydrate transport operon	1	5	8
ESAT6 ESX1 secretion system1	5	41	23
ESAT6 ESX1 secretion system2	2	6	33
Growth attenuation genes	30	93	40
Total Mutations	194	556	369
Total non-synonymous	59	242	186
Total synonymous	135	314	183
dN/dS	0.44	0.77	1.02



Appendix F: Lineage and sub-group specific mutations

Panel A: Midpoint rooted core genome phylogeny of L4 (red), L5 (brown) and L6 (green).

Panel B: Distribution of the complete specific amino acid mutations present (yellow dots) and absent (purple dots) among specific genes (black condensed upper block).

Appendix G: Mutations and Genotypic Data of Drug Resistant Strains

ID	Species	Region	Sensitivity		MIC (ug/mL)		Gene			
			INH	RIF	INH	RIF	katG	inhA	ndh	rpoB
2089	MTBss	Greater Accra	R	S	>1.00	<0.06				
3267	MTBss	Greater Accra	R	S	1	<0.06				
2659	MTBss	Northern	R	R	1	1				
2692	MTBss	Greater Accra	R	R	>1.00	2				
3265	MTBss	Greater Accra	R	S	>1.00	<0.06				
1018	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1035	MTBss	Greater Accra	R	S	>1.00	<0.06				
1040	MTBss	Greater Accra	R	R	>1.00	2				
1062	MTBss	Greater Accra	R	R	>1.00	2				
1104	MTBss	Greater Accra	R	R	>1.00	>2.00	S315T			S450L
1113	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	(-8T/C)	V117I	
1128	MTBss	Greater Accra	S	R	<0.03	1				
1187	MTBss	Greater Accra	R	S	0.5	<0.06				
1216	MTBss	Greater Accra	R	R	>1.00	>2.00	S315T			S450L
1238	MTBss	Northern	R	S	0.5	<0.06				
1241	MTBss	Greater Accra	S	R	<0.03	>2.00				
1245	MTBss	Greater Accra	R	R	>1.00	>2.00				
1252	MTBss	Northern	R	S	>1.00	<0.06				
1256	MTBss	Greater Accra	R	S	>1.00	<0.06				
1263	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	(-8T/C)	V117I	
1287	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1324	MTBss	Greater Accra	R	S	>1.00	<0.06				
1326	MTBss	Greater Accra	R	S	1	0.25				
1340	MTBss	Greater Accra	R	R	>1.00	>2.00				
1371	MTBss	Northern	R	S	>1.00	<0.06	S315T	G204D	V117I	
3255	MTBss	Greater Accra	R	R	>1.00	>2.00				
1403	MTBss	Greater Accra	R	S	1	<0.06				
1426	MTBss	Greater Accra	R	S	>1.00	<0.06				
1427	MTBss	Greater Accra	R	S	>1.00	0.25				
1460	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1462	MTBss	Greater Accra	R	R	>1.00	>2.00	S315T			S450L
1468	MTBss	Greater Accra	R	S	>1.00	<0.06				
1495	MTBss	Greater Accra	R	S	1	<0.06				
1503	MTBss	Greater Accra	R	S	0.5	<0.06				
1535	MTBss	Greater Accra	R	S	0.5	<0.06				
1542	MTBss	Greater Accra	S	R	<0.03	2				
1545	MTBss	Greater Accra	S	R	0.06	2				
1553	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	(-102G/A)		
1581	MTBss	Greater Accra	R	S	0.5	<0.06				
1584	MTBss	Greater Accra	R	S	>1.00	<0.06				
1636	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1652	MTBss	Greater Accra	R	S	0.5	<0.06				
1665	MTBss	Greater Accra	R	S	0.5	<0.06	S315T	(-102G/A)		
1680	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1683	MTBss	Greater Accra	R	S	>1.00	<0.06				
1687	MTBss	Greater Accra	R	S	0.5	<0.06				
1690	MTBss	Greater Accra	R	S	>1.00	<0.06				
1694	MTBss	Greater Accra	R	S	0.5	<0.06				
1695	MTBss	Greater Accra	R	S	0.5	0.13				
1709	MTBss	Northern	R	R	>1.00	>2.00	S315T			S450L
1759	MTBss	Greater Accra	R	S	>1.00	<0.06				
1777	MTBss	Northern	S	R	<0.03	1				
1791	MTBss	Greater Accra	R	S	>1.00	<0.06				
1794	MTBss	Greater Accra	R	S	>1.00	<0.06				
1796	MTBss	Greater Accra	R	R	>1.00	>2.00				
1824	MTBss	Greater Accra	R	S	0.5	<0.06				
1840	MTBss	Greater Accra	R	S	>1.00	<0.06				
1852	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1853	MTBss	Greater Accra	R	S	0.5	<0.06				
1867	MTBss	Greater Accra	R	S	>1.00	<0.06				
1958	MTBss	Greater Accra	R	S	1	<0.06				
1965	MTBss	Greater Accra	R	R	>1.00	2	S315T	G204D		

1998	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1999	MTBss	Greater Accra	R	S	>1.00	<0.06				
2032	MTBss	Greater Accra	R	R	>1.00	>2.00				
2061	MTBss	Greater Accra	R	S	>1.00	<0.06				
2087	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2121	MTBss	Greater Accra	R	S	>1.00	<0.06				
2132	MTBss	Greater Accra	S	R	<0.03	>2.00				
2152	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2165	MTBss	Greater Accra	R	S	0.5	<0.06				
2167	MTBss	Greater Accra	R	S	>1.00	<0.06				
2184	MTBss	Northern	R	S	0.5	<0.06				
2189	MTBss	Northern	R	R	0.5	1				
2204	MTBss	Greater Accra	R	S	>1.00	<0.06				
2243	MTBss	Greater Accra	R	S	0.5	<0.06				
2245	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2246	MTBss	Greater Accra	R	S	>1.00	<0.06				
2277	MTBss	Greater Accra	R	S	1	<0.06				
2295	MTBss	Greater Accra	R	S	>1.00	<0.06				
2308	MTBss	Northern	R	S	1	<0.06	S315T	(-47G/C)		
2320	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2338	MTBss	Greater Accra	R	S	>1.00	0.25				
2343	MTBss	Greater Accra	R	S	0.5	<0.06				
2378	MTBss	Greater Accra	R	S	0.5	<0.06				
2400	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2405	MTBss	Greater Accra	S	R	0.06	>2.00				
2411	MTBss	Greater Accra	R	S	0.5	<0.06				
2419	MTBss	Greater Accra	R	S	>1.00	<0.06				
2463	MTBss	Greater Accra	R	S	>1.00	<0.06				
2468	MTBss	Greater Accra	R	S	>1.00	<0.06				
2489	MTBss	Greater Accra	R	R	>1.00	1				
2492	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
2515	MTBss	Northern	R	R	>1.00	1				
2520	MTBss	Northern	R	S	>1.00	0.25				
2594	MTBss	Greater Accra	R	S	>1.00	<0.06				
2620	MTBss	Greater Accra	R	R	>1.00	>2.00				
2628	MTBss	Northern	R	S	1	<0.06				
2629	MTBss	Northern	R	S	>1.00	<0.06				
2630	MTBss	Northern	S	R	<0.03	2				
2649	MTBss	Greater Accra	R	S	1	<0.06				
2651	MTBss	Greater Accra	R	S	>1.00	<0.06				
2652	MTBss	Greater Accra	R	S	0.5	<0.06				
2653	MTBss	Greater Accra	R	S	>1.00	<0.06				
2677	MTBss	Northern	R	S	>1.00	<0.06				
2686	MTBss	Greater Accra	R	S	>1.00	<0.06				
2762	MTBss	Greater Accra	R	S	>1.00	<0.06				
2763	MTBss	Greater Accra	R	S	>1.00	<0.06				
2776	MTBss	Greater Accra	R	R	>1.00	>2.00				
2802	MTBss	Greater Accra	R	R	>1.00	0.5				
2820	MTBss	Greater Accra	R	S	>1.00	<0.06				
2870	MTBss	Greater Accra	R	R	>1.00	>2.00				
2889	MTBss	Greater Accra	R	S	>1.00	<0.06				
2902	MTBss	Greater Accra	R	R	>1.00	>2.00				
2926	MTBss	Greater Accra	R	R	>1.00	>2.00				
2942	MTBss	Greater Accra	R	S	>1.00	<0.06				
2946	MTBss	Greater Accra	R	S	>1.00	<0.06				
2964	MTBss	Greater Accra	R	S	1	0.13				
3002	MTBss	Greater Accra	R	R	0.5	0.5				
3003	MTBss	Northern	R	R	>1.00	>2.00				
3019	MTBss	Greater Accra	R	R	1	2				
3131	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	(-8T/C)		
3272	MTBss	Greater Accra	S	R	0.06	0.5				
3280	MTBss	Greater Accra	R	S	1	<0.06				
3347	MTBss	Greater Accra	R	S	0.5	<0.06				
2769B	MTBss	Greater Accra	R	R	>1.00	>2.00				
D35	MTBss	Greater Accra	R	R	>1.00	>2.00	S315T			S450L
1256B	MTBss	Greater Accra	R	S	>1.00	<0.06				
1273 P	MTBss	Greater Accra	R	S	>1.00	<0.06	S315T	G204D		
1019	Mafric	Greater Accra	R	S	>1.00	<0.06				

1083	Mafric	Greater Accra	R	S	>1.00	<0.06	S315T	(-102G/A)		
1615	Mafric	Greater Accra	R	S	0.5	<0.06				
1642	Mafric	Greater Accra	R	S	0.5	<0.06				
1803	Mafric	Greater Accra	R	R	1	0.5				
2096	Mafric	Greater Accra	R	R	>1.00	2	S315T	(-102G/A)		
2144	Mafric	Greater Accra	R	S	0.5	<0.06				
2263	Mafric	Greater Accra	R	R	>1.00	0.5	S315T	(-102G/A)		
2351	Mafric	Greater Accra	R	S	>1.00	<0.06	S315T	(-102G/A)		
2359	Mafric	Greater Accra	R	S	1	0.25				
2384	Mafric	Greater Accra	R	S	1	<0.06	S315T	(-102G/A)		
2394	Mafric	Greater Accra	R	S	1	<0.06				
2428	Mafric	Northern	R	R	>1.00	>2.00				
2452	Mafric	Greater Accra	R	R	0.5	1				
2519	Mafric	Greater Accra	R	S	>1.00	<0.06	S315T	(-102G/A)		
2603	Mafric	Greater Accra	R	R	>1.00	2				
2660	Mafric	Northern	R	S	1	<0.06				
2661	Mafric	Greater Accra	R	S	>1.00	<0.06				
2667	Mafric	Northern	R	S	0.5	0.25				
3064	Mafric	Greater Accra	R	S	0.5	<0.06				
3082	Mafric	Greater Accra	R	R	>1.00	>2.00				
3110	Mafric	Greater Accra	R	R	>1.00	1				
3132	Mafric	Greater Accra	R	S	0.5	0.25				
2428B	Mafric	Greater Accra	R	R	>1.00	0.5				
1861	Mafric	Greater Accra	R	S	0.5	<0.06				
1916	Mafric	Greater Accra	R	R	1	1				
1358	Mafric	Northern	R	R	>1.00	>2.00				
1401	Mafric	Greater Accra	R	S	>1.00	0.25	S315T	(-8T/C)	V117I	
3145	Mafric	Greater Accra	R	S	>1.00	<0.06				
2137	Mafric	Greater Accra	R	S	>1.00	<0.06	S315T	(-8T/C)	V117I	
2656	M.bovis	Northern	R	S	>1.00	<0.06				
2785	M.bovis	Greater Accra	R	R	1	>2.00				
1898	Mtbss	Greater Accra	S	S	<0.03	<0.06		(-102G/A)		
3286	Mtbss	Greater Accra	S	S	<0.03	<0.06		(-47G/C)		

Spoligotyping and SNP-typing data of all MTBC strains resistant to at least one drug is shown in the table above. However, Mutations are shown for only those isolates with hypothesized compensatory mutations. Isolates with specific mutations are coloured to show that hypothesized lineage/sub-lineage specific compensatory mutations are not phylogenetic markers (Olive green: Ghana sub-lineage with *katG* S315T and G204D, Bright green : Cameroon sub-lineage with *katG* S315T, *inhA* -8T/C and *ndh* V117I , Pink: Cameroon sub-lineage with *rpoB* S450L and *rpoC* G322R, Light Yellow: Cameroon sub-lineage with *katG* S315T and *inhA* G204D, Orange: Ghana sub-lineage with *katG* S315T, *rpoB* S315T and *rpoC* V483G, Dark Blue: LAM sub-lineage with *katG* S315T and *inhA* G204D, Bright Yellow: Uganda 1 sub-lineage with *katG* S315T and *inhA* G204D, Purple: MAF Lineage 6 with *katG* S315T, *inhA* -8T/C and *ndh* V117I).

Appendix H: Ethical Clearance for the Study

NOGUCHI MEMORIAL INSTITUTE FOR MEDICAL RESEARCH
Established 1979 *A Constituent of the College of Health Sciences*
University of Ghana

Phone: +233-302-916438 (Direct)
+233-289-522574
Fax: +233-302-502182/513202
E-mail: nirb@noguchi.mimcom.org
Telex No: 2556 UGL GH

INSTITUTIONAL REVIEW BOARD



Post Office Box LG 581
Legon, Accra
Ghana

My Ref. No: DF.22
Your Ref. No:

7th March, 2012

ETHICAL CLEARANCE

FEDERALWIDE ASSURANCE FWA 00001824

IRB 00001276

NMIMR-IRB CPN 067/11-12

IORG 0000908

On 7th March, 2012, the Noguchi Memorial Institute for Medical Research (NMIMR) Institutional Review Board (IRB) at a full board meeting reviewed and approved your protocol titled:

TITLE OF PROTOCOL : **Understanding the genetic diversity between *Mycobacterium africanum* and *Mycobacterium tuberculosis***

PRINCIPAL INVESTIGATOR : **Dr. Dorothy Yeboah-Manu, PhD**

CO-INVESTIGATORS : **Dr. Audrey Forson & Dr. Frank Bonsu**

Please note that a final review report must be submitted to the Board at the completion of the study. Your research records may be audited at any time during or after the implementation.

Any modification of this research project must be submitted to the IRB for review and approval prior to implementation.

Please report all serious adverse events related to this study to NMIMR-IRB within seven days verbally and fourteen days in writing.

This certificate is valid till 6th March, 2013. You are to submit annual reports for continuing review.

Signature of Chairman: 
Rev. Dr. Samuel Ayete-Nyampong
(NMIMR – IRB, Chairman)

cc: Professor Alexander K. Nyarko
Director, Noguchi Memorial Institute
for Medical Research, University of Ghana, Legon

NOGUCHI MEMORIAL INSTITUTE FOR MEDICAL RESEARCH
Established 1979 *A Constituent of the College of Health Sciences*
University of Ghana

INSTITUTIONAL REVIEW BOARD

Phone: +233-302-916438 (Direct)
+233-289-522574
Fax: +233-302-502182/513202
E-mail: nirb@noguchi.mimcom.org
Telex No: 2556 UGL GH



Post Office Box LG 581
Legon, Accra
Ghana

My Ref. No: DF.22
Your Ref. No:

- 6th May, 2015

ETHICAL CLEARANCE

FEDERALWIDE ASSURANCE FWA 00001824

IRB 00001276

NMIMR-IRB CPN 067/11-12 revd. 2015

IORG 0000908

On 6th May 2015, the Noguchi Memorial Institute for Medical Research (NMIMR) Institutional Review Board (IRB) at a full board meeting conducted continuing review and renewed your protocol titled:

TITLE OF PROTOCOL : Understanding the genetic diversity between Mycobacterium africanum and Mycobacterium tuberculosis

PRINCIPAL INVESTIGATOR : Prof. Dorothy Yeboah-Manu

CO – INVESTIGATORS : Dr. Audery Forson, & Dr. Frank Bonsu

Please note that a final review report must be submitted to the Board at the completion of the study. Your research records may be audited at any time during or after the implementation.

Any modification of this research project must be submitted to the IRB for review and approval prior to implementation.

Please report all serious adverse events related to this study to NMIMR-IRB within seven days verbally and fourteen days in writing.

This certificate is valid till 5th May, 2016. You are to submit annual reports for continuing review.

Signature of Chair:

Mrs. Chris Dadzie
(NMIMR – IRB, Chair)

cc: Professor Kwdwo Koram
Director, Noguchi Memorial Institute
for Medical Research, University of Ghana, Legon