University of Ghana http://ugspace.ug.edu.gh

UNIVERSITY OF GHANA

COLLEGE OF BASIC AND APPLIED SCIENCES

SCHOOL OF BIOLOGICAL SCIENCES

MOLECULAR EPIDEMIOLOGY OF *MYCOBACTERIUM TUBERCULOSIS* COMPLEX IN GHANA:
UNDERSTANDING TRANSMISSION DYNAMICS

BY

PRINCE ASARE

(10250691)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF DOCTOR OF PHILOSOPHY DEGREE IN MOLECULAR CELL BIOLOGY OF INFECTIOUS DISEASES

INTEGRI PROCEDAMUS

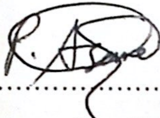
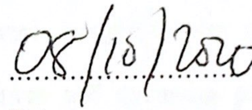
WEST AFRICAN CENTRE FOR CELL BIOLOGY OF INFECTIOUS PATHOGENS, DEPARTMENT OF BIOCHEMISTRY, CELL AND MOLECULAR BIOLOGY

JULY 2019

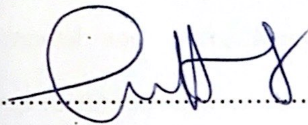
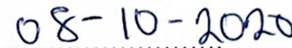
University of Ghana http://ugspace.ug.edu.gh

Declaration

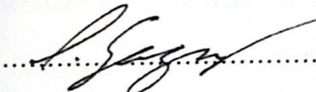
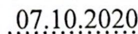
I, Prince Asare, do hereby declare that except for references to other people's work, for which I have duly acknowledged, this thesis is the product of my own research carried out under the supervision of Prof. Dorothy Yeboah-Manu, Prof. Sebastien Gagneux, Prof. Kwadwo A. Koram and Dr. Jonathan P. Adjimani. I do further declare that no part of this thesis has been previously submitted for a degree or any other qualification.

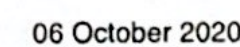
....
Prince Asare (Candidate) Date

....
Prof. Dorothy Yeboah-Manu (Supervisor) Date

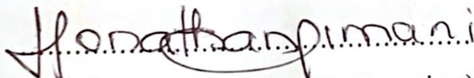
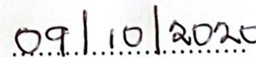



.... 07.10.2020..........
Prof. Sebastien Gagneux (Co-supervisor) Date



06 October 2020

....
Prof. Kwadwo A. Koram (Co-supervisor) Date

....
Dr. Jonathan P. Adjimani (Co-supervisor) Date

University of Ghana http://ugspace.ug.edu.gh

ABSTRACT

Tuberculosis (TB), a disease of antiquity, still infects millions of people annually. In 1993 the World Health Organization declared it a global health emergency, calling for more resources and studies to aid control this global menace. An objective under the research pillar of the End TB Strategy, is to investigate the transmission dynamics of the disease using molecular epidemiology (molepi) to understand factors leading to occurrence of the disease within distinct populations. Consequently, this PhD work sought to use a population-based molepi study to determine the transmission pattern and dynamics of *Mycobacterium tuberculosis* complex (MTBC) strains from two areas in Ghana (East Mamprusi and Accra Metropolitan Area) and identify associated risk factors to complement conventional control program efforts.

Mycobacterial species isolated from consecutively sampled pulmonary TB patients presenting at 13 selected health facilities between July 2012 and December 2015 were confirmed as MTBC by IS*6110* and rpo*ß* PCR and further assigned lineages and sub-lineages by spoligotyping and large sequence polymorphism PCR assays. Patient characteristics were obtained with a structured questionnaire. We used SaTScan and ArcMap analyses to identify and map significantly clustered MTBC lineages/sub-lineages. Isolates were strain typed using standard mycobacterial interspersed repetitive-unit variable-number tandem repeat (MIRU-VNTR) typing. Whole genome sequencing (WGS) was performed to resolve traditional genotype clusters and epidemiologically linked cases. Phylogenetic analysis was then used to assess strain relatedness.

Among 2,551 isolates genotyped for spatio-temporal analysis, 2,019 (79.1%), 516 (20.2%) and 16 (0.6%) were identified *as M. tuberculosis* sensu stricto (MTBss), *M. africanum*

University of Ghana http://ugspace.ug.edu.gh

(MAF) and animal strains respectively. MAF lineages were found to persist at approximately 20% over an 8-year period. Whereas the Cameroon sub-lineage was associated with Southern Ghana, the Beijing, Ghana and animal genotypes were significantly ($p<0.05$) associated with Northern Ghana. We estimated recent transmission rate of 24.7% using WGS and confirmed the existence of reduced recent transmission of MAF compared to MTBss L4. WGS confirmed a wide spread of a Cameroon sub-lineage clone mainly from the Ablekuma sub-district. More importantly, we identified a recent transmission cluster associated with isoniazid resistance belonging to the Ghana sub-lineage of L4. Risk factor analysis using logistic regression modeling, identified younger individuals (age <30 years) and male gender as significant risk factors for recent TB transmission. Majority (94.4%, 34/36) of individuals with recurring TB episodes were males and 58.6% (21/36) had TB recurrence within 12 months post treatment. Recurrent TB was attributed to inadequate treatment (relapse of same strain) in 75.0% (27/36) of participants with 25.0% (9/36) attributed to re-infection. Epidemiologically linked TB cases were likely the results of recent TB transmission within the house or from neighboring recent transmission events.

For the first time in Ghana using WGS, we confirm high recent TB transmission within the population driven largely by MTBss sub-lineages Cameroon and Ghana and show that unresolved previous infection due to inadequate treatment is largely the cause of recurring pulmonary TB. Observed reduced recent transmission of MAF suggests other factor(s) (host/environmental) may be responsible for its continuous presence in West Africa. We show that it is possible to monitor recurring TB cases and follow-up household related transmission in a resource-limited setting.

University of Ghana http://ugspace.ug.edu.gh

DEDICATION

To God Almighty, creator of the universe and all that is within it.

To my family and girlfriend who never understood my absence from their company.

To all participants of the study.

University of Ghana http://ugspace.ug.edu.gh

ACKNOWLEDGEMENT

I thank the Almighty God for grace, guidance and sustenance throughout the course of this PhD work.

My sincerest appreciation goes to my principal supervisor; Prof. Dorothy Yeboah-Manu of the Noguchi Memorial Institute for Medical Research (NMIMR) for admitting me into her research group and granting me the opportunity to undertake this PhD work on one of her projects. I am particularly grateful for her guidance through her collaborative research with Prof. Sebastien Gagneux (my co-supervisor and mentor) of Swiss Tropical and Public Health Institute (SwissTPH), University of Basel, Switzerland, for their partnership to help develop my scientific career towards the successful completion of this thesis.

I would like to show extreme gratitude to Prof. Kwadwo A. Koram of NMIMR and Dr. Jonathan P. Adjimani of department of Biochemistry, Cell and Molecular Biology (BCMB) whose inputs were invaluable. The precious time they spent helping me with this thesis will forever be appreciated.

I am also grateful to all members of Dorothy Yeboah-Manu's research group at bacteriology department, NMIMR (Dr. Adwoa Asante-Poku, Dr. Isaac Darko Otchere, Dr. Samuel Aboagye, Stephen Osei-Wusu, Emelia Danso, Doris Aboagyewaa Boampong, Esther Sarpong, Dina Prah, Edmund Bedeley, Diana Asandem, Portia Morgan, Pheonah Badu, Selessie Ameke and all National Service Personnel) for all the support and hard work. I am also extremely grateful to Sandra Adelaide King at parasitology department NMIMR for her support during my early years of intensive lab work.

University of Ghana http://ugspace.ug.edu.gh

To all members of the TB research unit at SwissTPH, I am also grateful for the time spent with you and the wealth of knowledge I acquired from you especially Sonia, Chloe and Daniela.

My deepest appreciation to the able staff and colleagues of WACCBIP and BCMB and all other members of the bacteriology department, NMIMR.

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. I was also supported by a West African Centre for Cell Biology of Infectious Pathogens (WACCBIP) World Bank African Centres of Excellence PhD fellowship.

I cannot forget my family, especially my dad, Mr. Seth Asare (Snr) who has invested a lot of resources in my education. This is how far I the Lord has brought me.

THANK YOU ALL.

University of Ghana http://ugspace.ug.edu.gh

TABLE OF CONTENTS

DECLARATION ... i
ABSTRACT ... ii
DEDICATION ... iv
ACKNOWLEDGEMENT ... v
TABLE OF CONTENTS ... vii
List of Figures ... xiii
List of Tables ... xvii
LIST OF ABBREVIATIONS ... xix
THESIS OUTLINE ... xxiii
Chapter One ... 1
INTRODUCTION ... 1
1.1 Background ... 1
1.2 Rationale ... 5
1.2.1 Problem statement ... 5
1.2.2 Justification ... 7
1.3 Hypothesis ... 8
1.3.1 Research questions ... 8
1.4 Aim ... 8
1.5 Specific objectives ... 9
Chapter two ... 10
LITERATURE REVIEW ... 10
2.1 Burden and epidemiology of TB ... 10
2.1.1 Global perspective ... 10
2.1.2 Tuberculosis Burden in Ghana ... 14
2.2 The causative agent ... 15
2.2.1 *Mycobacterium tuberculosis* complex ... 16

University of Ghana http://ugspace.ug.edu.gh

2.2.2 Genome of the *Mycobacterium tuberculosis* complex 16
2.3 Pathogenesis 20
2.4 Control of TB 25
2.4.1 Three pillars of the End TB strategy for Global TB control 26
2.4.1.1 *An integrated, patient-centered care and prevention* 26
2.4.1.2 *Bold policies and supportive systems* 26
2.4.1.3 *Intensified research and innovation* 26
2.4.2 Vaccination against TB 27
2.4.3 Diagnosis of tuberculosis 28
2.4.3.1 *Latent infection* 28
2.4.3.2 *Active TB* 28
2.4.4 Treatment 32
2.4.5 Relapse verses Re-infection 33
2.5 Molecular epidemiology of Infectious Diseases 36
2.5.1 Molecular Epidemiology of tuberculosis: Opportunities for TB control 36
2.5.2 Genotyping methods used for molecular epidemiological studies of MTBC 43
2.5.2.1 IS*6110* restriction fragment length polymorphism 46
2.5.2.2 Spoligotyping 48
2.5.2.3 MIRU-VNTR typing 51
2.5.2.4 Whole genome sequencing 54
2.5.3 Hunter-Gaston Discriminatory Index and recent TB transmission 56
CHAPTER THREE 59
3.0 Manuscript 1: Spatio-Temporal Distribution of *Mycobacterium tuberculosis* Complex Strains in Ghana 59
3.1 Abstract 59
3.2 Introduction 61

University of Ghana http://ugspace.ug.edu.gh

3.3 Materials and Methods 64
3.3.1 Ethics Statement 64
3.3.2 Study Locale and Participants Data 64
3.3.3 Mycobacterial Isolation, Species and Lineage Classification 67
3.3.4 Data Management and Analysis 67
3.3.5 Spatial and Space-time Analysis 68
3.3.6 Normalization of TB cases for within district comparison 69
3.4 Results..69
3.4.1 Characteristics of Patients Presenting with Tuberculosis 69
3.4.2 The Population Structure of MTBC causing pulmonary TB in Ghana............... 75
3.4.3 Spatial Distribution of MTBC Genotypes... 75
3.4.4 Spatial and space-time clustering analysis of MTBC cases at the district/sub-district level (2012–2014) 79
3.5 Discussion and conclusion...84
CHAPTER 4 87
4.0 Manuscript 2. Reduced Transmission of *Mycobacterium africanum* Compared to *Mycobacterium tuberculosis* in Urban West Africa. 87
4.1 Abstract..87
4.2 Introduction...88
4.3 Methods 90
4.3.1 Study design and population 90
4.3.2 Mycobacterial isolation, species identification and drug susceptibility testing.. 90
4.3.3 Lineage and strain classification 91
4.3.4 Clustering analysis and risk factor assessment 92
4.3.5 Data management and analysis 93
4.4.1 Characteristics of study participants 94

University of Ghana http://ugspace.ug.edu.gh

4.4.2 Population structure and recent transmission rate estimation 95

4.4.3 Exploring the diversity and clustering within the MTBC lineages 100

4.4.4 Molecular epidemiology and factors associated with clustering: Logistic regression modeling 101

4.5 Discussion and Conclusion 113

CHAPTER 5 117

5.0 Manuscript 3. Whole Genome Sequencing and Spatial Analysis Identifies Recent TB Transmission Hotspots in Ghana 117

5.1 Abstract 117

5.2 Introduction...118

5.3 Materials and methods 120

5.3.1 Study design and population 120

5.3.2 Isolate selection, DNA extraction and whole genome sequencing 120

5.3.3 Whole genome sequence analysis and variance calling.................................. 121

5.3.4 Phylogenetic analysis 122

5.3.5 Cluster definition and analysis 123

5.3.6 Recent TB transmission rate estimation.. 124

5.3.7 Mutation rates and within-host micro-evolution of longitudinal isolates 125

5.3.8 Data management and analysis 125

5.4 Results...126

5.4.1 Characteristics of the study population 126

5.4.2 Single nucleotide polymorphism threshold for clustering analysis 127

5.4.3 Recent transmission hotspots and characteristics of large clusters 130

5.4.3.1 Socio-demographic characteristics of individuals infected with a strain from the largest cluster (WGSC-5) 131

5.5 Discussion..135

University of Ghana http://ugspace.ug.edu.gh

CHAPTER 6 140
6.0 Manuscript 4. Whole Genome Sequencing and Epidemiological Analysis Identifies High Relapse Among Individuals with Recurring Tuberculosis in Ghana 140
6.1 Abstract 140
6.2 Introduction 141
6.3 Materials and methods 143
6.3.1 Ethical considerations 143
6.3.2 Study design and population 143
6.3.3 Mycobacterial isolation, characterization and DNA extraction 145
6.3.4 Conventional strain typing and drug susceptibility testing 145
6.3.5 Whole genome sequencing and analysis 146
6.3.6 Phylogenetic reconstruction 147
6.3.7 Case definitions 147
6.3.8 Data analysis and epidemiology 147
6.4 Results 148
6.4.1 Characteristics of individuals with recurring TB episodes 150
6.4.2 Whole genome sequence analysis identifies high relapse rate among recurring TB episodes 154
6.4.3 Characteristics of individuals belonging to household related TB transmission cases 159
6.4.4 Evidence of household recent transmission 163
6.4.5 15-locus MIRU-VNTR typing is enough for predicting the cause of recurring TB and identifying suspected household related TB transmission 163
6.5 Discussion 166
6.6 Conclusion 171
7.0 General Discussion, Conclusions and Recommendations 172

University of Ghana http://ugspace.ug.edu.gh

7.1 General Discussion 172

7.1.1 Phylogeography of *Mycobacterium tuberculosis* complex: Implications for TB control 173

7.1.2 Transmissibility and virulence of members of MTBC: Need for host-pathogen co-evolution studies 176

7.1.3 Molecular Epidemiology and TB control 178

7.1.4 Tools for studying TB transmission dynamics.................... 184

7.1.5 Limitations 186

7.2 Conclusions....................187

7.3 Recommendations....................189

APPENDICES....................221

APPENDIX 1 221

Supplementary data to chapter 3....................221

APPENDIX 2 231

Supplementary data to chapter 4....................231

APPENDIX 3 241

Supplementary data to chapter 5....................241

University of Ghana http://ugspace.ug.edu.gh

LIST OF FIGURES

Figure 2.1. Estimated tuberculosis incidence rates, 2017 10
Figure 2.2. Countries in the three high-burden country lists for TB, TB/HIV and MDR-TB being used by WHO during the period 2016–2020, and their areas of overlap 13
Figure 2.3 Projected incidence and mortality curves that are required to reach the End TB Strategy targets and milestones, 2015–2035. 13
Figure 2.4. A diagrammatic representation of the *M. tuberculosis* H37Rv genome...........19
Figure 2.5. *Mycobacterium tuberculosis* infection and pathogenesis. 24
Figure 2.6. The cellular composition of a representative *Mycobacterium tuberculosis* granuloma..25
Figure 2.7. Relapse and re-infection of *M. tuberculosis* strains...35
Figure 2.8. Phylogenetic relationship among members of the *M. tuberculosis* complex. ...38
Figure 2.9. Phylogenetic analysis of MTBC 39
Figure 2.10. Phylogeography of MTBC lineages 41
Figure 2.11. Prevalence of *M. africanum* in Western African countries 42
Figure 2.12. Schematic representation of the chromosome of a hypothetical MTBC isolate with marked repetitive elements as targets for different typing methods. 45
Figure 2.13. Comparison of MIRU-VNTR digitized result output verses IS*6110* RFLP profile 48
Figure 2.14. Presence (filled squares) or absence (empty squares) of spacer regions after spoligotyping...50
Figure 2.15. Location of 41 MIRU loci in *M. tuberculosis* H37Rv genome. 53
Figure 2.16. Hypothetical representation of the genotypic pattern of BCG, H37Rv and strain X based on MIRU-VNTR 53
Figure 3.1. Map of sampling sites and study area. 66
Figure 3.2. Spatial distribution and prevalence of identified *Mycobacterium tuberculosis* lineages...71
Figure 3.3. Temporal distribution and prevalence of human adapted *Mycobacterium tuberculosis* complex (MTBC)..76
Figure 3.4. Spatial distribution of human adapted MTBC lineages and major sub-lineages within the eight-year study period...77

University of Ghana http://ugspace.ug.edu.gh

Figure 3.5. Spatial distribution of isolated MTBC within selected districts (2012 – 2014) 80
Figure 3.6. Spatial distribution of lineage 4 and major lineage 4 sub-lineages within selected districts (2012 – 2014) 83
Figure 4.1. Pipeline for recruited participants and culture positive TB cases used for clustering analysis. 95
Figure 4.2. Temporal distribution of 2,309 MTBC isolates stratified by lineage 97
Figure 4.3. Cluster distribution and size stratified by lineage (panel A and C) and sub-lineage (panel B and D). 100
Figure 4.4. Minimum spanning tree (MST) representation of the clustering of 2,322 MTBC isolates from Accra Metropolitan Assembly and East Mamprusi district built with Bionumerics software. 101
Figure 4.5. Adjusted predictions of the probability of belonging to a clustered case with 95% confidence interval 112
Figure 5.1. Pipeline for sample selection and whole genome sequencing. 126
Figure 5.2. Phylogenetic relationship of 6 *M. tuberculosis* complex longitudinal isolates obtained from three participants 128
Figure 5.3. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 10-SNPs 129
Figure 5.4. Clustering analysis stratified by cluster definition. 130
Figure 5.5. Phylogenetic reconstruction of 146 *M. tuberculosis* complex isolates showing characteristics of the 8 identified large clusters as defined by a threshold of 10-SNPs.....133
Figure 5.6. Spatial distribution of pulmonary tuberculosis cases belonging to large clusters in and around Accra metropolis. 134
Figure 5.7. Frequency and relative geographical distribution of cases in all 8 large clusters identified by a threshold of 10-SNPs. 135
Figure 6.1. Relative geographical location of 13 health facilities and distribution of recurring and household related TB cases 144
Figure 6.2. Pipeline for sample selection and genotyping. 149
Figure 6.3. Phylogenetic reconstruction showing the genomic relationship between 60 *M. tuberculosis* complex isolates from 29 recurring TB cases 157

University of Ghana http://ugspace.ug.edu.gh

Figure 6.4. Phylogenetic reconstruction showing the genomic relationship between 19 *M. tuberculosis* complex isolates from 9 household transmission groups 162
Figure 6.5. Time till event of household related tuberculosis cases. The time of diagnosis for each household related TB case is shown 164
Figure 6.6. Number of MIRU-VNTR allelic variations between identified recurrent TB cases (black bars) and household related cases (grey bars)...165
Figure 6.7. Concordance between MIRU-VNTR allelic polymorphism and WGS SNP distances. 165
Figure S4.1. Map of study area showing the location of the 12 sampling sites (health facilities)...235
Figure S4.2. Minimum spanning tree (MST) representation of the clustering of sub-lineages of lineage 4 MTBC isolates built with Bionumerics software 236
Figure S4.3. MST representation of the clustering of lineage 1 - 3 MTBC isolates built with Bionumerics software...237
Figure S4.4. MST representation of the clustering of lineage 4 MTBC isolates built with Bionumerics software..237
Figure S4.5. MST representation of the clustering of lineage 5 MTBC isolates built with Bionumerics software..238
Figure S4.6. MST representation of the clustering of lineage 6 MTBC isolates built with Bionumerics software..238
Figure S4.7. MST representation of the clustering of lineage 4 MTBC isolates stratified by study setting...239
Figure S4.8. MST representation of the clustering of lineage 5 MTBC isolates stratified by study setting...239
Figure S4.9. MST representation of the clustering of lineage 6 MTBC isolates stratified by study setting...240
Figure S4.10. Time lapse between first and last diagnosed case within selected large and very large clusters...240
Figure S5.1. Clustering and phylogenomic relationship of 452 *M. tuberculosis* complex isolates 241

University of Ghana http://ugspace.ug.edu.gh

Figure S5.2. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 5-SNPs. 242
Figure S5.3. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 12-SNPs 243

University of Ghana http://ugspace.ug.edu.gh

LIST OF TABLES

Table 2.1. Next generation sequencing platforms commonly used for sequencing bacterial genomes........54
Table 3.1. Demographic and clinical data of 2551 TB cases........72
Table 4.1. Geographical distribution and population structure of MTBC in Ghana by spoligotyping........96
Table 4.2a. Clustering analysis stratified by lineages and major sub-lineage populations of MTBC........98
Table 4.2b. Clustering analysis stratified by study setting and lineages/major sub-lineage populations of MTBC........99
Table 4.3. Characteristics of large molecular clusters resulting from combined 15-MIRU and spoligotyping cluster analysis........102
Table 4.4. Logistic regression analysis of risk factors associated with TB clustering (recent TB transmission)........106
Table 4.5: Risk Factors associated with TB clustering: logistic regression analysis stratified by lineage........110
Table 6.1. Characteristics of cases with recurring TB episodes........151
Table 6.2. Demographic and clinical characteristics of participants with recurring TB episodes........158
Table 6.3. Characteristics of cases involved in household related transmission........161
Table S3.1. Districts and population statistics within designated time points within the study period........221
Table S3.2. Annual distribution of TB cases used for spatial or space-time analysis (2007 – 2014)........222
Table S3.3. Lineage distribution of isolated MTBC recruited and used for spatial or space-time analysis (2007 – 2014)........223
Table S3.4. Default parameters used in SaTScan for clustering analysis........224
Table S3.5. Primer Sequence for large sequence polymorphism LSP Assays........225
Table S3.6. Comparison of selected risk factors among the two regions........226
Table S3.7. Genotyping profile of 2551 MTBC isolates from Ghana........227

University of Ghana http://ugspace.ug.edu.gh

Table S3.8. Distribution of Species and Lineages of MTBC 230
Table S4.1. Summary of recruited cases 231
Table S4.2. Distribution of participants' characteristics stratified by the main MTBC lineages found in Ghana: A univariate and multivariate logistic regression analysis........232
Table S5.1. Clustering analysis stratified by location of residence for large clusters........244

University of Ghana http://ugspace.ug.edu.gh

LIST OF ABBREVIATIONS

AFB	Acid Fast Bacilli
AMA	Accra Metropolitan Area
BCG	Bacillus Calmette-Guerin
CLR	C-Type Lectin Receptors
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyl Trimethyl Ammonium Bromide
DC	Dendritic Cells
DC-SIGN	DC-Specific Intracellular Adhesion Molecule 3–Grabbing Non-Integrin
DNA	Deoxyribonucleic Acid
DOTs	Directly Observed Treatment short course
DR	Drug Resistant
EMB	Ethambutol
EPTB	Extra Pulmonary Tuberculosis
ESAT-6	Early Secretory Antigen Target 6
GIS	Geographic Information Systems
HBC	High Burden Country
HGDI	Hunter-Gaston Discriminatory Index
HIV	Human Immuno-Deficiency Virus
IGRA	Interferon-gamma Release Assay
IL	Interleukin
INH	Isoniazid
IQR	Interquartile Range

University of Ghana http://ugspace.ug.edu.gh

IR	Interspersed Repeats
IS	Insertion Sequence
KBTH	Korle-Bu Teaching Hospital
L1	Lineage 1
L2	Lineage 2
L3	Lineage 3
L4	Lineage 4
L5	Lineage 5
L6	Lineage 6
L7	Lineage 7
LAM	Lipoarabinomannan
LJ	Loweinstein-Jenson
LM	Lipomannan
LPA	Line Probe Assay
LSP	Large Sequence Polymorphism
LTBI	Latent Tuberculosis Infection
MAF	*Mycobacterium afrcanum*
MamE	East Mamprusi District
ManLAM	Mannose-Capped Lipoarabinomannan
MDR-TB	Multidrug Resistant tuberculosis
MIRU-VNTR	Mycobacteria Interspersed Repetitive Unit – Variable Number Tandem Repeats
MMR	Macrophage Mannose Receptor

University of Ghana http://ugspace.ug.edu.gh

molepi	Molecular Epidemiology
MOTT	Mycobacterium Other Than Tuberculosis
MST	Minimum Spanning Tree
MTBC	*Mycobacterium tuberculosis* Complex
MTBss	*Mycobacterium tuberculosis* sensu stricto
NAAT	Nucleic Acid Amplification Test
NF	Nuclear Transcription Factor
NGS	Next Generation Sequencing
NMIMR	Noguchi Memorial Institute for Medical Research
NTM	Non-Tuberculous Mycobacteria
NTP	National Tuberculosis control Program
oriC	Origin of Replication
P:N	Prevalence to Notification Ration
PCR	Polymerase Chain Reaction
PE	Proline-Glutamate
PPD	Purified Protein Derivative
PPE	Proline- Proline Glutamate
PTB	Pulmonary Tuberculosis
PZA	Pyrazinamide
*rc*TB	Recurring Tuberculosis
RD	Region of Difference
RFLP	Restriction Fragment Length Polymorphism
RIF	Rifampicin

University of Ghana http://ugspace.ug.edu.gh

rrn	Ribosomal RNA
RR-TB	Rifampicin Resistant Tuberculosis
SDG	Sustainable Development Goal
SIT	Spoligotype International Type
SNP	Single Nucleotide Polymorphism
Spoligotyping	Spacer Oligonucleotide Typing
STROME-ID	Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases
TB	Tuberculosis
TDR-TB	Totally Drug Resistant Tuberculosis
TLR	Toll-Like Receptor
TNF	Tumor Necrosis Factor
TR	Tandem Repeats
TST	Tuberculin Skin Test
UN	United Nations
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
VNTR	Variable Numbered Tandem Repeats
WA	West Africa
WGS	Whole Genome Sequencing
WHO	World Health Organization
XDR	Extensively Drug Resistant TB
ZN	Zeihl-Neelsen

University of Ghana http://ugspace.ug.edu.gh

THESIS OUTLINE

The thesis is organized into seven chapters as follows:

Chapter 1 provides a general introduction of the field of study.

Chapter 2 reviews literature on various tools and concepts of the study.

The results of the study are presented in 4 chapters (Chapters 3 to 6) organized into 4 manuscripts.

Chapter 3 (Yeboah-Manu *et al.,* 2016) and **chapter 4** (Asare *et al.,* 2018) constitute the first two manuscripts which have been published in peer reviewed journals (references below).

Chapter 5 is the third manuscript written with the objective of confirming recent TB transmission using whole genome sequencing.

Chapter 6 is the final manuscript written with the objective of investigating occurrence of relapse and re-infection among individuals with recurring TB episodes in Ghana.

Chapter 7, the final chapter, is a general discussion and conclusion of all the work presented.

1. Yeboah-Manu, D., Asare, P., Asante-Poku, A., Otchere, I. D., Osei-Wusu, S., Danso, E., . . . Gagneux, S. (2016). Spatio-Temporal Distribution of *Mycobacterium tuberculosis* Complex Strains in Ghana. PloS One, 11(8), e0161892. doi:10.1371/journal.pone.0161892.
2. Asare, P., Asante-Poku, A., Prah, D. A., Borrell, S., Osei-Wusu, S., Otchere, I. D., . . . Yeboah-Manu, D. (2018). Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. International Journal of Infectious Diseases. doi:10.1016/j.ijid.2018.05.014.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER ONE

INTRODUCTION

1.1 Background

Tuberculosis (TB), a disease of antiquity, affects millions of people yearly and in 1993 the World Health Organization (WHO) declared it a global health emergency, calling for more resources and studies for effective control. The WHO estimates that a fourth of the world's population (1.7 billion individuals) are latently infected with the causative bacteria (WHO, 2018b), creating a pool of future active cases. In 2017 alone, the WHO estimated that 10 million new TB cases occurred globally, out of which 1.6 million died of TB making TB the number one infectious disease killer and one of the ten killer diseases. There were TB cases in all countries and all age groups and out of the 1.6 million TB deaths; 1.3 million were among HIV-negative individuals with the remaining 300,000 among HIV-positive individuals. Although it is home to only 14% of the world's population, the WHO African region reports 25% (2,480,000) of the global TB incidence and has the highest HIV-associated TB cases (WHO, 2018b).

Challenges mitigating against the control of TB include socio-economic factors such as increased urbanization leading to overcrowding; lack of political will to commit funds and resources; pathogen related factors such as emergence of strains resistant to available antimycobacterials, threatening to make a treatable disease untreatable and the HIV epidemic. Other factors that hinder the control of TB include limited knowledge of the biology of the causative pathogen leading to the use of an ineffective vaccine and relatively lack of cheap and effective sensitive diagnostic tools.

University of Ghana http://ugspace.ug.edu.gh

The influx of individuals from rural areas to urban areas in search for 'greener pastures' has led to overcrowding of the urban sectors. The causative pathogen is transmitted mainly by inhalation of aerosolized droplet nuclei containing the bacilli; Thus, an overcrowded population provides an ideal means for increased transmission through frequent person-to-person contact with an infected person. Other factors which are prominent in resource poor settings like sub-Saharan Africa countries are: 1) the continual use and reliance on less sensitive diagnostic technique (microscopy) in peripheral health facilities; 2) the lack of political will to commit funds for control activities and 3) stigma leading to late reporting, that have led to increased spread of the pathogen. Sputum-smear microscopy has the ability to detect only 50% of TB cases implying that up to half the number of TB cases are missed who go back to their neighbourhoods and continue to spread the pathogen. Moreover, the continuous use of the only available but largely ineffective Bacillus Calmette-Guerin (BCG) vaccine (only effective against childhood TB), which was introduced in 1921, (Davenne & McShane, 2016; WHO, 2018a) is due to the lack of an in-depth understanding of the biology of the bacteria and its interaction with the host.

Low compliance to treatment regimens has led to the development of difficult-to-treat multidrug resistant TB/ rifampicin resistant TB cases (MDR/RR-TB) and extensively drug resistant TB cases (XDR-TB). Resistance to at least the two most potent first line anti-TB drugs (rifampicin and isoniazid) is termed MDR and patients harboring such MDR/RR-TB strains are put on second line drugs which are more toxic, expensive and have a longer treatment period. An MDR-TB strain resistant to at least two second-line anti-TB drugs including any of the fluoroquinolones and any of the injectable aminoglycosides is termed an XDR-TB strain. In 2017 alone, 558,000 TB cases were reported as being MDR/RR-TB

University of Ghana http://ugspace.ug.edu.gh

(WHO, 2018b), and those unreported cases continue to contribute to the global challenge of TB control. While the treatment outcome of sensitive strains globally is 85% success rate, the rates for MDR and XDR are 55% and 34% respectively (Alene *et al.*, 2017; WHO, 2018b).

Tuberculosis in mammals is mainly caused by 9 genetically related mycobacterial species comprising *Mycobacterium tuberculosis* sensu stricto, *M. africanum , M. bovis, M. mungi, M. microti, M. caprae, M. pinnipedii, , M. suricattae and M. orygis* together referred to as the *M. tuberculosis* complex (MTBC) (Brosch *et al.*, 2002; Huard *et al.*, 2003; Mostowy *et al.*, 2002; van Ingen *et al.*, 2012). Of the nine members of the MTBC, two are known to be the main cause of human TB, hence referred to as the human adapted MTBC species which include; *M. tuberculosis* sensu stricto (MTBss), and *M. africanum* (MAF).

Compared to MAF, MTBss has been found to have a competitive advantage and work done mainly in the Gambia, suggested that MAF is attenuated (de Jong, Adetifa*, et al.*, 2010; de Jong, Antonio*, et al.*, 2010). However, MAF seems endemic in West-African countries including Ghana and is responsible for about 50% of TB cases in some of the countries (Addo *et al.*, 2007; Asante-Poku *et al.*, 2016; Gagneux & Small, 2007; Yeboah-Manu *et al.*, 2016). Could it be that, within the human population in West-Africa, both MTBss and MAF actively transmit equally? Knowledge of the transmissibility of these two human-adapted lineages will not only guide TB control measures but also contribute to influence the advancement of active research involving both species, which is in line with the third pillar of the End TB Strategy.

University of Ghana http://ugspace.ug.edu.gh

The global TB control strategy aims at having a TB-free world by attaining zero deaths, disease and suffering due to TB. To achieve this aim, The End TB strategy for global TB control was instituted which outlines three pillars including; 1) an integrated, patient-centered care and prevention, 2) bold policies and supportive systems and 3) intensified research and innovation (WHO, 2015a). Generally, these pillars are aimed at finding and curing all persons with any form of active TB (be it a resistant or sensitive strain) as well as preventing new cases of TB infection and calls for combined efforts from the governments through to community care providers. In addition, the overall control strategy calls for improving diagnostic, intervention and research tools to facilitate achieving the set aim.

An objective under the research pillar, is to investigate the transmission dynamics of the disease (by applying molecular tools in combination with classical epidemiological data - so called molecular epidemiology (molepi)) to understand factors leading to occurrence of the disease within distinct population. Currently, the molecular tools that are widely being employed in molecular epidemiological studies include spacer oligonucleotide typing (spoligotyping) and mycobacteria interspersed repetitive unit – variable number of tandem repeats typing (MIRU-VNTR typing) (Anderson *et al.*, 2014; Barnes & Cave, 2003; Hamblion *et al.*, 2016). MIRU-VNTR and spoligotyping assays are both PCR-based assays, the later followed with hybridization and spacer detection steps.

Molecular epidemiological studies also involve the use of geographic information systems (GIS) (ESRI, 2010) and scan statistics (Kulldorff, 2015) to analyze and detect spatio-temporal variations and clustering of TB disease (Dangisso *et al.*, 2015). Spatial and spatio-temporal analysis through mapping of TB cases together with isolated strains has

University of Ghana http://ugspace.ug.edu.gh

enhanced the effectiveness of recent molepi studies in generating vital information about TB disease distribution and transmission pattern, detecting clusters, aid to identify both environmental and socio-economic risk factors enhancing the spread of the disease and also useful for the evaluation of intervention efforts (Bastida *et al.*, 2012; Dangisso *et al.*, 2015; Randremanana *et al.*, 2009; Touray *et al.*, 2010; Yakam *et al.*, 2014). Such analysis has however not been explored in Ghana and several other sub-Saharan countries due to limited expertise and resources. Some of the few studies, however, are only able to conduct spatial analysis with no temporal analysis due to the limited period of sampling.

1.2 Rationale

1.2.1 Problem statement

Tuberculosis remains a major public health concern in Ghana. Indicators mainly used to evaluate the performance of national programs in Africa are: the proportion of diagnosed cases that started treatment; how many sputa converted at 2/3 months; of the treated how many completed; how many cured and those that defaulted. These analyses are usually done quarterly at the sub-district and district levels, which are later pulled together as annual reports often showing TB distribution at regional and national administrative levels. These analyses leave understanding of very important indicators such as risk factors for improvement of control efforts including education and putting in place measures to improve infection prevention. Very important is the understanding of transmission dynamics, which has been studied extensively in developed countries of North America and Europe and has been useful for identification of outbreaks as well as most at risk groups (Hamblion *et al.*, 2016; Seraphin *et al.*, 2016; Vluggen *et al.*, 2017) for targeted

University of Ghana http://ugspace.ug.edu.gh

control activities. However due to expertise and infrastructural demand there is lack of data on population-based prospective molepi studies done in endemic countries of sub-Saharan Africa.

Existing data from several studies point to the fact that Ghana has six out of the seven human-adapted TB lineages circulating, which makes it a good hub for transmission dynamics studies (Addo *et al.*, 2007; Asante-Poku, Aning*, et al.*, 2014; Asante-Poku *et al.*, 2015; Yeboah-Manu *et al.*, 2012; Yeboah-Manu *et al.*, 2011). However, these studies were either not population-based and/or were not sampled for at least two years and did not use optimum discriminatory tools to allow for more detailed molecular epidemiological study which is needed to provide plausible reasons for disease distribution. This is crucial for planning control interventions.

The high prevalence (22%) of HIV among TB patients in Ghana coupled with the emergence of difficult to treat MDR-TB cases (Osei-Wusu *et al.*, 2018; WHO, 2018c) call for more research that will lead to understanding of risk factors leading to occurrence of drug resistance cases and also factors driving recent TB transmission in Ghana. In 2014 Ghana recorded a very low TB detection rate of 33% (14,668 notified cases out of estimated 44,000) which is way below the African and global targets of 50% and 70% respectively (WHO, 2015d). The reported trends in 2014 is still not significantly different from current reported trends; out of the estimated 44,000 cases only a third (14,550) was reported (WHO, 2018c). Importantly, patients report very late to the hospitals as indicated by most cases being diagnosed as sputum ≥2+ (Asante-Poku *et al.*, 2016), which means they are likely to transmit the bacilli to other susceptible individuals in the population

University of Ghana http://ugspace.ug.edu.gh

within their communities before reporting to the hospitals and there may be undetected outbreaks in some communities.

1.2.2 Justification

Tuberculosis is a global burden, however based on case: population ratio, sub-Saharan Africa is the most burdened region. Ghana has been ranked among the top 30 countries with a high TB-HIV burden. Thus, the NTP needs a more vigorous strategy to effectively control TB. Due to the lack of effective vaccine, the main control strategy, which is the Directly Observed Treatment short course (DOTs), though successful in improving treatment compliance, lacks the ability to detect at risk groups, provide explanation for relapse case/treatment failure or provide understanding on the transmission dynamics. To effectively control TB in Ghana, it is therefore paramount to conduct studies that will provide understanding of the transmission dynamics in a population-based prospective study, which will contribute to knowledge on factors that enhance spread of the disease. Such population-based prospective studies which has been employed in this current study are necessary to help in the detection of unsuspected recent TB outbreaks/ transmission chains. Moreover, molecular epidemiology is also helpful for tracking TB strains among recurring TB patients and provide indications of the cause of secondary case source, thus differentiating between exogenous (reinfection with new strain) versus endogenous re-infections (relapse) (Varghese *et al.*, 2013), for appropriate treatment, evaluation of performance and epidemiology (Barnes & Cave, 2003; van Soolingen *et al.*, 2003).

Molecular epidemiological studies are also useful to correctly identify and characterize bacterial lineages. The ability to readily and accurately distinguish between lineage and strains provides the opportunity to study the consequence of genetic diversity among the

University of Ghana http://ugspace.ug.edu.gh

different MTBC phylogenetic lineages on distinct pathogen traits including virulence and drug-resistance. Summing all together molepi study will complement conventional control efforts to allow the establishment of good preventive strategies, appropriate therapy and a better understanding of the biology of the pathogen which will contribute to development of future control tools.

1.3 Hypothesis

We hypothesize that;

1. *Mycobacterium africanum* is endemic in Ghana because it transmits within the Ghanaian population at the same rate as *M. tuberculosis sensu stricto*.
2. Individuals presenting with multiple TB episodes are due to exogenous re-infection and not endogenous re-infection.

1.3.1 Research questions

I will test the hypotheses by addressing the following questions.

1. Is *M. africanum* being truly outcompeted by *M. tuberculosis* sensu stricto?
2. Which will play a greater role towards developing new active TB; recent infection or reactivation of old infection?

1.4 Aim

The main goal of this PhD study is to determine the transmission pattern and dynamics of *M. tuberculosis* complex strains in Ghana, specifically from East Mamprusi district (rural)

University of Ghana http://ugspace.ug.edu.gh

and Accra Metro (urban), and to identify possible risk factors including drug resistance involved in recent TB transmission.

1.5 Specific objectives

1. To explore the spatio-temporal distribution of the different circulating *M. tuberculosis* genotypes and estimate their incidence in two study sites using large sequence polymorphism typing and spoligotyping analysis.

2. To compare recent transmission of *M. tuberculosis* sensu stricto and *M. africanum* and identify associated risk factors such as HIV and drug resistance in Ghana.

3. To perform a whole genome sequence analysis to confirm evidence of family related and molecular epidemiologically linked recent TB transmission in Ghana.

4. To investigate the occurrence of relapse and re-infection among individuals with recurring TB episodes in Ghana.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER TWO

LITERATURE REVIEW

2.1 Burden and epidemiology of TB

2.1.1 Global perspective

Tuberculosis (TB) is an important global public health problem. It has caused the most human deaths by an infectious disease (Paulson, 2013) and is among the top 10 causes of death globally. In the year 2017, 1.6 million people died of TB and 10 million people got sick with TB (WHO, 2018b), worldwide. Cases occurred in all age groups; about one million occurred among children and the ratio of male: female was approximated at 2:1 (5.8 million males: 3.2 million females). As indicated in Figure 2.1, TB cases were reported in all countries, though the brunt of cases occurred in the developing countries of Africa and Asia.

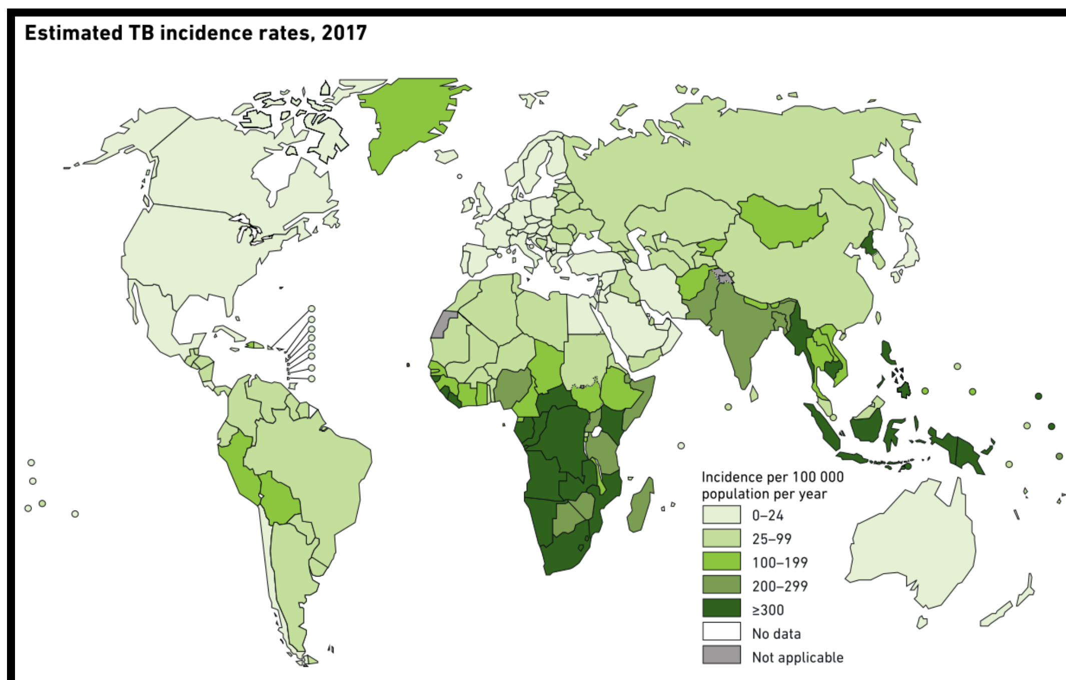


Figure 2.1. Estimated tuberculosis incidence rates, 2017 (WHO, 2018b).

University of Ghana http://ugspace.ug.edu.gh

The 30 most burdened TB countries that report the highest number of cases to the World Health Organisation (WHO), together contributed about 87% of the global TB cases in 2017 of which the top eight countries included India, China, Indonesia, Philippines, Pakistan, Nigeria, Bangladesh and South Africa (WHO, 2018b). Most of the 30 high burden countries recorded incidence of between 150 – 400 new cases per 100,000 population per year (WHO, 2018b). Although about 14% of the world population live in Africa, close to 25% (2,480,000) of TB cases are estimated to be in Africa, making Africa the most burdened region.

The continued public health importance of TB has been attributed in part to the co-epidemic with the Human Immunodeficiency Virus (HIV) and the alarming emergence of drug-resistant TB strains, which if not controlled may turn a treatable disease into an untreatable one. Drug resistant TB is defined as TB disease caused by bacteria that can tolerate clinically achievable concentrations of at least one of the main anti-TB drugs (WHO, 2015b). More than half a million of the TB cases that occurred in 2017 were caused by either multi-drug resistance TB bacilli (MDR-TB) defined as resistant to at least isoniazid (INH) and rifampicin (RIF) (82%); or resistant to rifampicin alone (RR-TB). Worldwide, 3.5% of new TB cases and 18% of previously treated cases had MDR/RR-TB (WHO, 2018b). Individuals infected with such strains are not treated with the standard first line anti-TB drugs and require longer treatment duration with more toxic second line anti-TB drugs with poor treatment outcomes. Some TB strains are also either extensively drug resistant (XDR-TB) or totally drug resistant (TDR-TB). The former is defined as an MDR strain which is additionally resistant to at least two of the second-line anti-TB drugs including any of the fluoroquinolones and any of the second–line anti-TB injectable drugs

University of Ghana http://ugspace.ug.edu.gh

(Velayati *et al.*, 2016). The latter, which makes TB virtually untreatable, refers to resistance to all known first and second line anti-TB drugs. About 5% and 2% of MDR-TB cases may be XDR or TDR respectively and have been reported in Italy, Iran, India and South Africa (Parida *et al.*, 2015; Velayati *et al.*, 2013).

HIV infection offers the highest risk for progression of the infected to TB disease. While 5-10% individuals infected with the pathogen will progress to active TB, the rate among people living with HIV is 20 to 30 times compared to HIV negative individuals. In 2017 it was estimated that 920,000 cases of TB were among people living with HIV representing 9% of the total TB burden of which 72% were in Africa. With an estimated 2,480,000 total TB cases from Africa, Africa's HIV/TB co-infection rate now stands at 26.7% (663,000/2,480,000). Of the 920,000 estimated HIV/TB comorbidity, only 51% (464,633) were notified.

To improve allocation of resource and for monitoring of performance, the WHO has defined two other high burden country (HBC) lists for the period 2016 - 2020, one for multidrug resistant TB (MDR-TB) and one for TB/HIV co-infection (Figure 2.2). Ghana is among the high burden TB/HIV country list.

The WHO has put in a strategy termed 'End TB Strategy' and the set targets are to reduce the absolute number of TB deaths and TB incidence by 90% and 80% respectively by 2030 and 95% and 90% respectively by 2035 (Figure 2.3) (WHO, 2018b). In May 2014, all members of the WHO and the United Nations (UN) unanimously endorsed the WHO's End TB Strategy and proceeded to adopting the UN Sustainable Development Goal (SDGs) in September 2015. The global TB incidence is averagely falling annually by

University of Ghana http://ugspace.ug.edu.gh

about 2%, and this is far lower than the end-TB strategy and SDG targets of 10%. The fastest regional declines have been in the WHO European Region estimated at 5%.

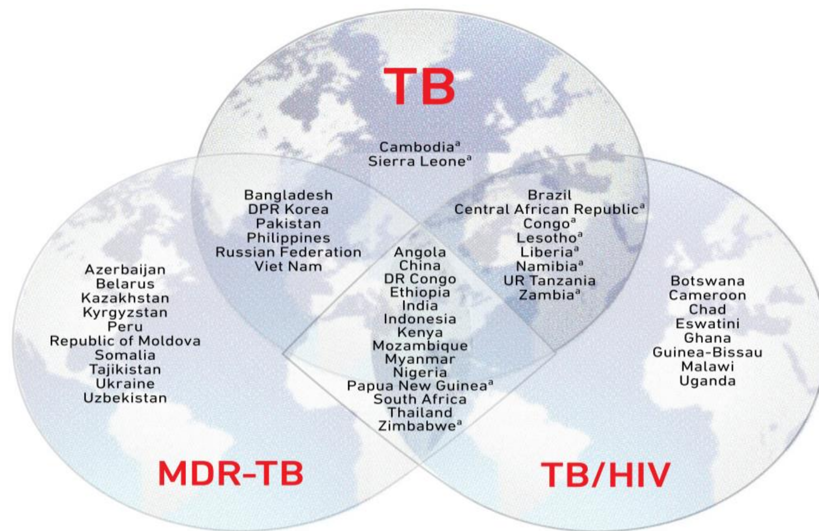


Figure 2.2. Countries in the three high-burden country lists for TB, TB/HIV and MDR-TB being used by WHO during the period 2016–2020, and their areas of overlap (WHO, 2018b).

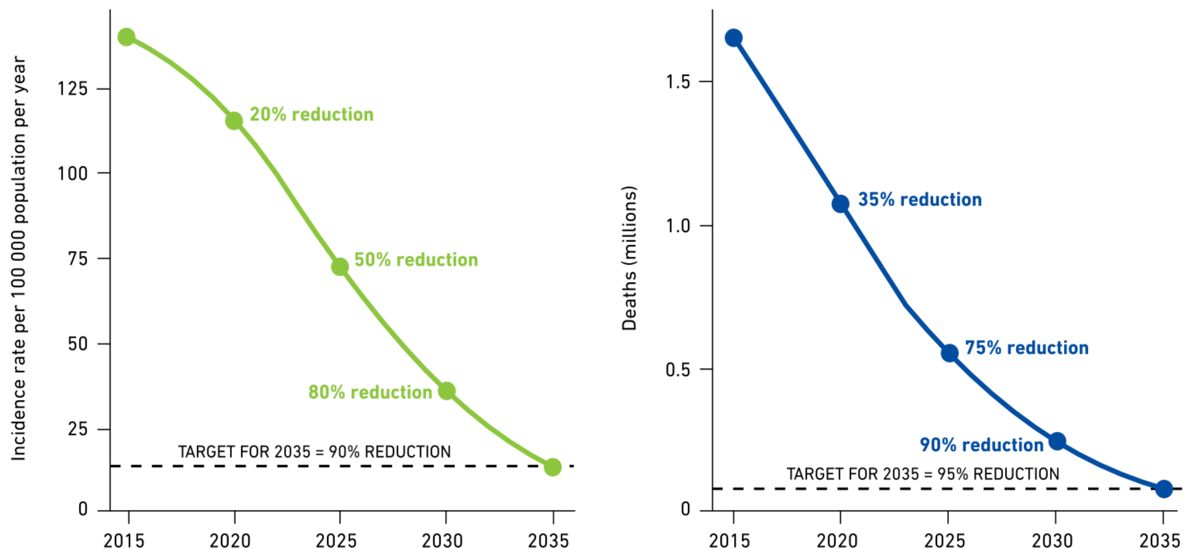


Figure 2.3 Projected incidence and mortality curves that are required to reach the End TB Strategy targets and milestones, 2015–2035.

University of Ghana http://ugspace.ug.edu.gh

2.1.2 Tuberculosis Burden in Ghana

Tuberculosis is of public health importance in Ghana; recent incidence is estimated at 152/100,000 populations (WHO, 2018c). In 2017, 44,000 new TB cases were estimated out of which only a third was notified (14,550) which is way below the African and global targets of 50% and 70% respectively giving a prevalence to notification (P:N) rate of approximately 3:1 (WHO, 2018c). The P:N rate is higher in males (>3:1) compared to females (<2:1). This low TB detection rate implies that a greater proportion of infectious individuals are left undetected and continue to serve as reservoirs for on-going transmission.

A nationwide prevalence survey conducted in 2013 however, had a prevalence of 252 cases/100,000 population, against the previously estimated prevalence of 72/100,000 computed from reported cases (WHO, 2014, 2015b, 2015d). TB is found in all age groups, however, similar to other countries like Zimbabwe and Malawi, the prevalence of TB disease increases with age (WHO, 2015b, 2015d, 2018b).

The TB epidemiology in Ghana seems to be highly influenced by HIV infection. While HIV prevalence in Ghana is estimated at approximately 1.5% and there are about only 4.7% of newly diagnosed HIV patients who are TB positive, currently, the total number of TB patients with known HIV status is about 21% (Osei *et al.*, 2017; WHO, 2015d, 2018c). Ghana contributes at least 1,000 TB/HIV cases per year to the global pool and in 2017 reported 2,759 cases (WHO, 2018c). The high HIV/TB co-infection burden also adversely affects TB treatment outcome; in 2017, 34% (5,200/15,200) estimated deaths (mortality rate of 18/100,000 population) were attributed to HIV/TB co-infection (WHO, 2018c).

University of Ghana http://ugspace.ug.edu.gh

In 2017, the estimated number of TB patients in Ghana who were harboring an MDR/RR-TB was 1000 of which only 330 were notified among pulmonary TB cases. The first case of extensive drug resistant TB (XDR-TB) was reported in 2017 (Osei-Wusu *et al.*, 2018). There has been reported increase in treatment success among notified cases over the past decade from 44% to 87% (Amo-Adjei & Awusabo-Asare, 2013; GHS, 2015; WHO, 2018c), however, it was estimated that 15,200 individuals (both HIV negative and positive) died of TB in 2017 due to late reporting.

2.2 The causative agent

Tuberculosis is caused mainly by a group of genetically closely related pathogens together known as the *Mycobacterium tuberculosis* complex (MTBC). The MTBC belongs to the Mycobacteriaceae family and the genus *Mycobacterium*. About 170 distinct species of the genus have been characterized (Falkinham, 2015; Nunes-Costa *et al.*, 2016). This high species numbers are partly due to improved microbiological isolation techniques and advancement in molecular biology techniques for characterizing. The species of this genus are grouped based on growth rate, colonial morphology and ability to cause disease. All species except the main TB causing pathogens (the MTBC) and *M. leprae*, are collectively referred to as non-tuberculous mycobacteria (NTM), also known as "atypical" mycobacteria or mycobacterium other than tuberculosis (MOTT). The NTMs are widely found in the environment including domestic water sources, soil, vegetation and animals (Aboagye *et al.*, 2016; Wolinsky, 1979).

University of Ghana http://ugspace.ug.edu.gh

2.2.1 *Mycobacterium tuberculosis* complex

The MTBC include: *M. tuberculosis* sensu stricto (MTBss) *M. africanum* (MAF)*, M. bovis, M. mungi, M. microti, M. caprae, M. pinnipedii, M. suricattae and M. orygis* (Brosch *et al.*, 2002; Huard *et al.*, 2003; Mostowy *et al.*, 2002; van Ingen *et al.*, 2012). The members of the MTBC are 2 – 4 µm in length and generally slow growers with some species taking as long as 4 to 6 weeks to observe macroscopic growth on solid microbiological medium (Loweinstein-Jenson medium) even under optimal conditions. The slow growing nature of this pathogen is largely due to its long generation time, about 20 - 24 hours (Smith, 2003). This is partially responsible for the chronic nature of the disease, the long period taken to respond to treatment and a huge challenge for researchers.

The members of the MTBC are genetically similar (characterized by 99.9% similarity at the nucleotide level), yet they show several phenotypic differences including host preference; they are adapted to infect specific hosts but cross-species infections do occur occasionally (Gutierrez *et al.*, 1997; Horstkotte *et al.*, 2001; van Soolingen *et al.*, 1997; Yeboah-Manu *et al.*, 2016). MTBss and MAF frequently infect humans thus consequently will be referred as human adapted MTBC; *M. bovis* infects cattle, *M. microti* infects voles (small rodents), *M. pinnipedii* infects seals/sea lions, *M. caprae* infects goats/sheep, *M. mungi* infects mongoose, *M. suricattae* infects meerkats and *M. orygis* infects antelope/deer (Esteban & Munoz-Egea, 2016; van Soolingen *et al.*, 1997).

2.2.2 Genome of the *Mycobacterium tuberculosis* complex

The laboratory reference strain *M. tuberculosis* H37Rv has been used as a proxy for all investigations of the human adapted MTBC, so all the descriptions of MTBC genome provided here are based on this reference strain. The H37Rv chromosome was completely

University of Ghana http://ugspace.ug.edu.gh

sequenced and published in 1998 showing a circular genome consisting of 4,411,529 base pairs (~4.4Mbp) with 3,924 open reading frames representing approximately 91% of its potential coding capacity (Figure 2.4) (Cole *et al.*, 1998). Currently about 4,018 genes in the complete genome of *M. tuberculosis* H37Rv have been predicted (Ramakrishnan *et al.*, 2015). The genome of MTBC is rich in guanine and cytosine nucleotides (GC rich, 65.6%) which is reflected in the biased amino-acid content of its proteins and possess many functional genes that are involved in cellular processes such as lipogenesis, lipolysis, glycolysis as well as enzymes that are used in aerobic and microaerophilic environments and this enables its survival in different environments including the oxygen rich lung, the macrophage and at the center of caseous granuloma (Cole *et al.*, 1998). The GC content is fairly constant throughout the entire genome and this is an indication of lack of horizontally transferred genetic elements (like pathogenicity islands) which mostly do have atypical base composition. A very large portion of the coding capacity of MTBC is devoted to the production of enzymes involved in lipogenesis and lipolysis which make it differ drastically from other bacteria, and this is evidenced in the unique lipid-rich cell wall (Cole *et al.*, 1998).

The genome of the MTBC contains two large acidic, glycine-rich protein families known as the PE (proline-glutamate) and PPE (proline- proline glutamate) some of which are associated with antigenic variability and virulence (Cole *et al.*, 1998). The genome is rich in repetitive DNA, particularly insertion sequences, and in new multigene families and duplicated housekeeping genes. For instance, the insertion sequences IS*6110* and IS*1081* are present with sixteen and eight copies respectively although one of the IS*1081* element is truncated (Cole *et al.*, 1998).

University of Ghana http://ugspace.ug.edu.gh

The only ribosomal RNA (*rrn*) operon is unusually situated at about 1.5Mbp away from the putative origin of replication (*oriC*) and may account for its slow growing nature (Verma *et al.*, 1999). Almost 60% of its genes are transcribed in the same direction as the replication fork. Naturally, MTBC is resistant to a number of antibiotics mainly due to its highly hydrophobic cell wall, but also, many potential resistance determinants are present in the genome including hydrolytic or drug-modifying enzymes such as β-lactamases and aminoglycoside acetyl transferases, and many potential drug–efflux systems, such as 14 members of the major facilitator family and numerous ABC transporters (Cole *et al.*, 1998).

University of Ghana http://ugspace.ug.edu.gh

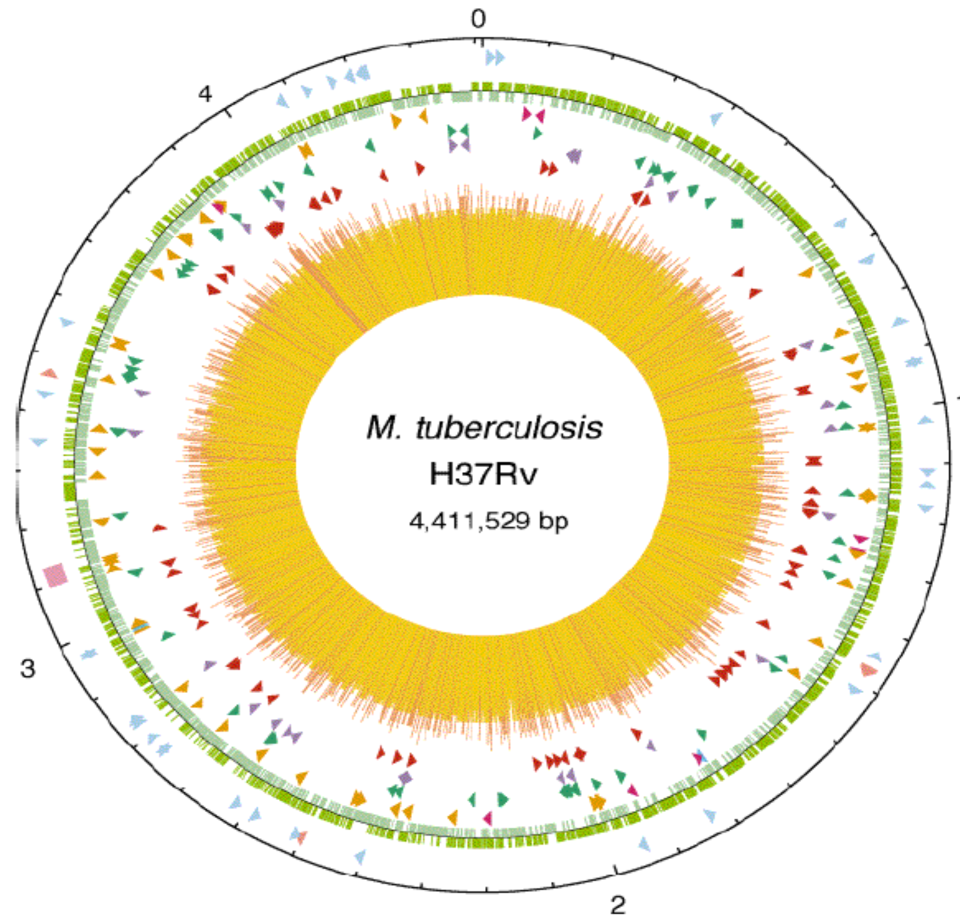


Figure 2.4. A diagrammatic representation of the *M. tuberculosis* H37Rv genome. "The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes (tRNAs are blue, others are pink) and the direct repeat region (pink cube); the second ring inwards 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); the fifth ring shows the PE family members (purple, excluding PGRS); and the sixth ring shows the positions of the PGRS sequences (dark red). The histogram (center) represents G + C content, with <65% G + C in yellow, and >65% G + C in red" adapted from (Cole *et al.*, 1998).

University of Ghana http://ugspace.ug.edu.gh

2.3 Pathogenesis

Tuberculosis (TB) is transmitted primarily via inhalation of aerosolized droplet nuclei containing the TB bacilli. From a single bacillus up to 200 bacilli may be needed to establish an infection (Balasubramanian *et al.*, 1994; Sakamoto, 2012). Inhalation of aerosol droplet containing MTBC bacilli can lead to three different outcomes primarily depending on the interplay between the immunity of the host and genetic background of the infecting bacteria; 1) In some few individuals, the bacilli is eliminated by the innate immune system with no immunological memory, 2) In about 90% of infected individuals, an effective cell-mediated immune response develops 2–8 weeks after infection that stops further multiplication of the tubercle bacilli (Ahmad, 2011) in a phenomenon known as latent TB infection (LTBI), 3) Yet, in some 5-10% of individuals, there is the establishment of infection which rapidly progress to an active TB disease which may culminate in death if not managed well.

Upon inhalation of droplet nuclei containing the bacilli, infection is established usually at the distal alveoli part of the lungs where the bacilli encounter the innate immune cells such as neutrophils, resident macrophages, and dendritic cells (DC) that phagocytose the bacteria and play important roles in the pathogen's initial clearance or establishment of infection (Kang *et al.*, 2011; Pai *et al.*, 2016; Philips & Ernst, 2012).

The phagocytic cells express receptors including C-type lectin receptors (CLR), scavenger receptors, and complement receptors that bind the mycobacteria and enable the initiation of its entry. The bacilli's ability to infect macrophages through several entry routes appears to be an important determinant of its dissemination with CLR largely implicated in the binding and uptake of the mycobacteria. The CLRs include the macrophage mannose

University of Ghana http://ugspace.ug.edu.gh

receptor (MMR) and DC-specific intracellular adhesion molecule 3–grabbing non-integrin (DC-SIGN). The MMR and DC-SIGN do not only aid in bacterial uptake but are known to influence phagosome maturation as well as cytokine signaling by recognizing a potential virulent factor called mannose-capped lipoarabinomannan (ManLAM), which is abundantly available in the mycobacteria cell wall (Philips & Ernst, 2012). Toll-like receptor (TLR) signaling (particularly TLR2 and TLR4), which is one of the main arms of the innate immune response during MTBC infection, also interacts with ligands including the 19 and 27 kDa lipoproteins, 38 kDa glycolipoprotein, the lipomannan (LM) and ManLAM (Ahmad, 2011). This interaction eventually culminates in activation of nuclear transcription factor (NF)-κB and production of pro-inflammatory cytokines such as tumor necrosis factor (TNF)-α, interleukin (IL)-1, IL-12, chemokines, and nitric oxide which together aid in killing the pathogen. Initially, the phagocytic cells are not able to contain the bacilli and dendritic cells which later engulf extracellular bacilli released from lysed macrophages migrate to and prime both CD4+ and CD8+ T cells at the regional lymph node (Bodnar *et al.*, 2001). Meanwhile the infected macrophage and other neutrophils continue to produce chemokines at the site of infection which guides the primed T cells and B cells to migrate to the site of infection to help keep the bacilli under check leading to the formation of the hallmark of TB infection, granuloma (Figure 2.5 and 2.6) (Kang *et al.*, 2011; Pai *et al.*, 2016; van Crevel *et al.*, 2002).

Granulomas are the classical histopathological lesions in TB and are formed by the fusion of the plasma membrane of several macrophages including mature macrophages, differentiated or epithelioid macrophages, foamy macrophages, multinucleated giant cells as well as T cells, and other host cells (dendritic cells, fibroblasts, endothelial cells, and

University of Ghana http://ugspace.ug.edu.gh

stromal cells) surrounded by fibroblasts (Philips & Ernst, 2012) (Figure 2.6). A well-formed granuloma forms the bases for latency in tuberculosis infection. This stage of infection where the bacilli is dormant and the infected individual does not show any clinical evidence of active TB is what characterizes the LTBI (Smith, 2003).

Recent studies have shown that differences exist in the immunological response mounted by different individuals that lead to the formation of physiologically distinct granulomatous lesions in individuals exposed to the TB bacilli. Some of the lesions are known to result in sterilizing immunity while others promote the persistence of viable bacilli in the microenvironment (Ahmad, 2011). Up to a quarter of the world's population are latently infected with the TB bacilli and the only indication of infection is the presence of immune responsiveness to mycobacterial antigens (Philips & Ernst, 2012; WHO, 2018a).

The granuloma forms equilibrium between host containment of infection and protection of MTBC from the host killing mechanism. The elimination of the pathogen mainly involves the success of the interaction between the infected macrophages, B and T lymphocytes (van Crevel *et al.*, 2002). The exact processes involved with progression from LTBI to active TB is not well understood, however the active disease mainly develops when the immune system is compromised, and the bacilli becomes reactivated and replicates within the growing granuloma (Pai *et al.*, 2016). In addition to HIV infection, host factors including diabetes and malnutrition have been found to be critical risk factors for reactivation of LTBI (Taylor *et al.*, 2005; Zheng *et al.*, 2017).

If the bacterial load becomes too great, the granuloma protection fails to contain the infection and its center undergoes necrosis with time resulting in a caseous necrotic

University of Ghana http://ugspace.ug.edu.gh

appearance consisting of dead macrophages which serves as nourishment for the metabolically active bacilli. Such individuals proceed to active TB disease. At this stage, the bacteria disseminate into the alveoli and the airways leading to the induction of a productive cough and associated signs and symptoms of TB.

The general symptoms of active TB or TB disease are chronic wet/productive cough, night sweats, weight loss and chest pains. The phlegm produced by such individuals are laden with high bacteria burden, thus can transmit the bacilli which normally reside in the lungs or larynx to others through coughing, singing, talking and spitting. TB affect mainly the lungs in the condition called pulmonary TB (PTB) which is the most common form of TB. Extra-pulmonary TB (EPTB) (TB affecting organs or tissues other than the lungs), which is a non-contagious form of TB, also exist and may either be localized or disseminated (miliary TB).

University of Ghana http://ugspace.ug.edu.gh

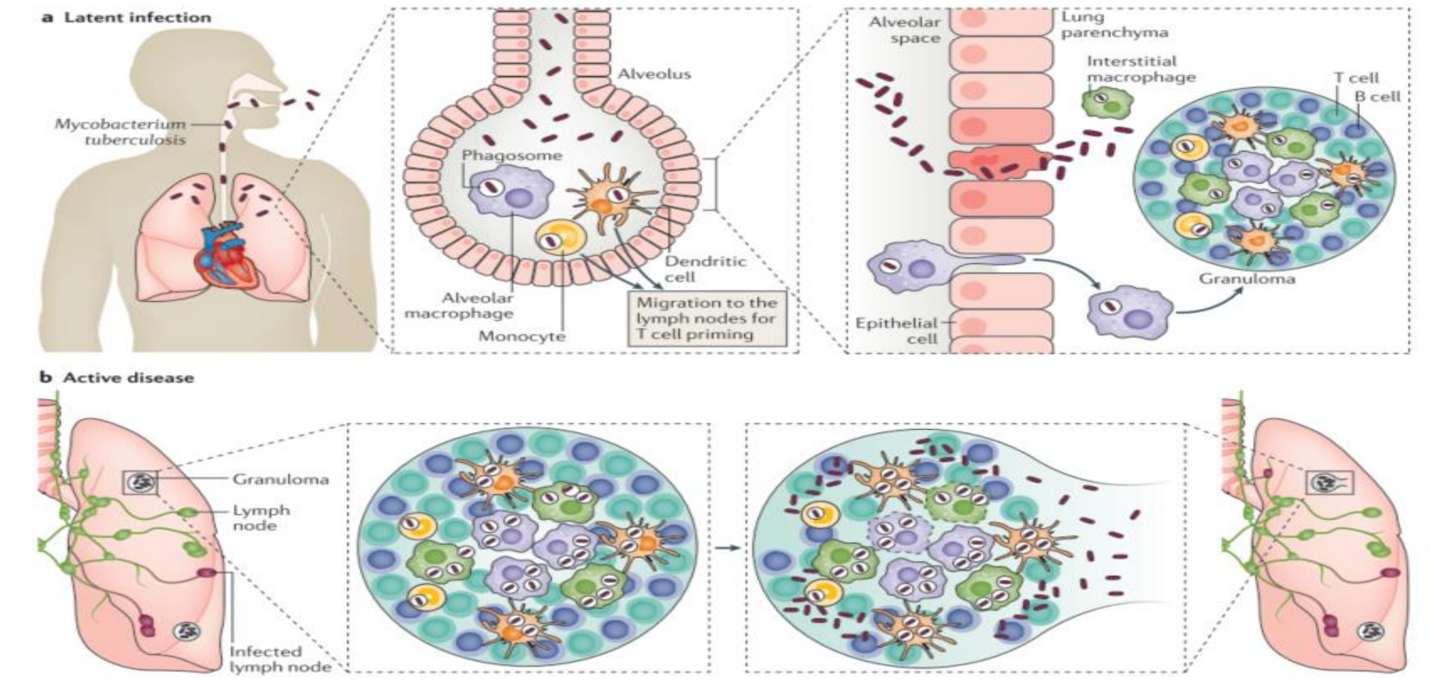


Figure 2.5. *Mycobacterium tuberculosis* infection and pathogenesis. “a) Infection begins when *Mycobacterium tuberculosis* enters the lungs via inhalation, reaches the alveolar space and encounters the resident alveolar macrophages. If this first line of defense fails to eliminate the bacteria, *M. tuberculosis* invades 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 *M. tuberculosis* 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 will disseminate eventually to other organs, including the brain. At this phase, the bacteria can enter the bloodstream or re-enter the respiratory tract to be released — the infected host is now infectious, symptomatic and is said to have active TB disease.” Adapted from (Pai *et al.*, 2016).

University of Ghana http://ugspace.ug.edu.gh

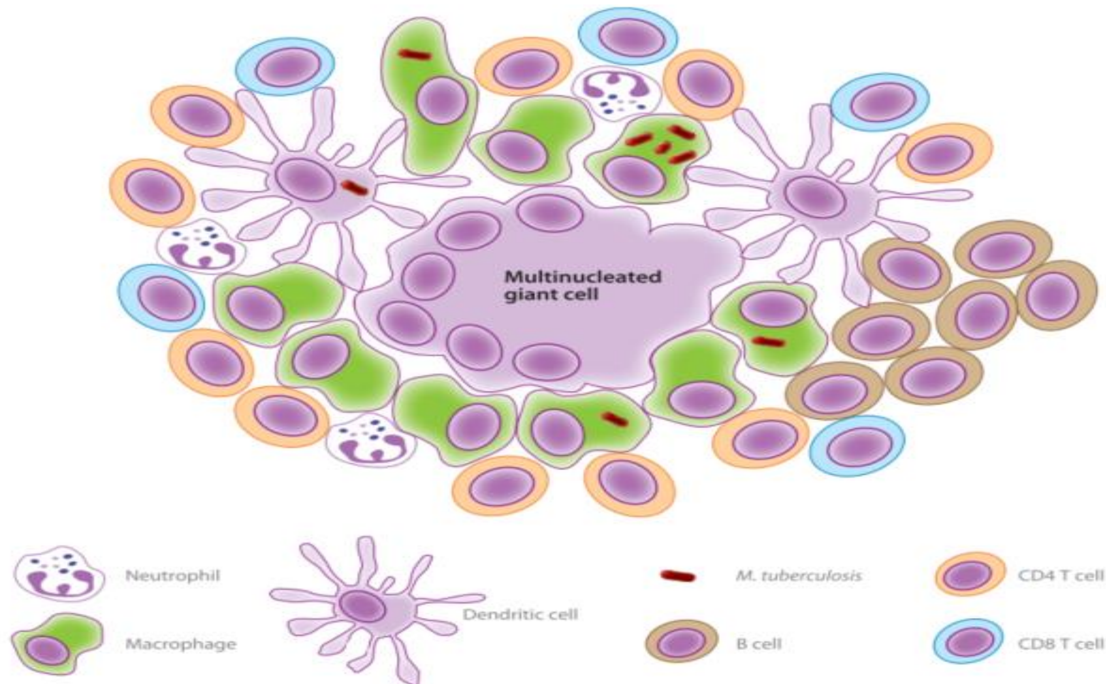


Figure 2.6. The cellular composition of a representative *Mycobacterium tuberculosis* granuloma. "Individual granulomas can exhibit diverse morphologies, including striking areas of central caseous necrosis. The cells depicted have been identified in human granulomas, as well as in those of experimentally infected animals. A characteristic finding is that a small minority of the macrophages and dendritic cells in granulomas are infected by *M. tuberculosis*" Adapted from (Philips & Ernst, 2012).

2.4 Control of TB

The primary mode of protection against TB is by vaccination of uninfected individuals; in the case of infected individuals, early case detection using both clinical and laboratory-based tests followed by antibiotic treatment of confirmed cases are employed. These control measures have nevertheless not reduced the TB incidence to the expected levels and as such new and better measures are needed to be able to meet the expectation of the End TB strategy by 2035. Currently the global control strategy aims at having a world free

University of Ghana http://ugspace.ug.edu.gh

of TB (zero deaths, disease and suffering due to TB) and in view of this has established the End TB strategy for Global TB control which outlines three pillars to achieve the set goal; 1) an integrated, patient-centered care and prevention, 2) bold policies and supportive systems and 3) intensified research and innovation (WHO, 2015a).

2.4.1 Three pillars of the End TB strategy for Global TB control

2.4.1.1 *An integrated, patient-centered care and prevention*

This pillar is made of four main components namely; A) Early diagnosis of tuberculosis including universal drug-susceptibility testing, and systematic screening of contacts and high-risk groups, B) Treatment of all people with tuberculosis including drug-resistant tuberculosis, and patient support, C) Collaborative tuberculosis/HIV activities, and management of co-morbidities, D) Preventive treatment of persons at high risk, and vaccination against tuberculosis (WHO, 2015a).

2.4.1.2 *Bold policies and supportive systems*

The components of this pillar include; A) Political commitment with adequate resources for tuberculosis care and prevention, B) Engagement of communities, civil society organizations, and public and private care providers, C) Universal health coverage policy, and regulatory frameworks for case notification, vital registration, quality and rational use of medicines, and infection control, D) Social protection, poverty alleviation and actions on other determinants of tuberculosis (WHO, 2015a).

2.4.1.3 *Intensified research and innovation*

This final pillar also bears the following components; A) Discovery, development and rapid uptake of new tools, interventions and strategies, B) Research to optimize implementation and impact, and promote innovations (WHO, 2015a).

University of Ghana http://ugspace.ug.edu.gh

2.4.2 Vaccination against TB

Vaccination is the best option for fighting against communicable diseases. However, this has largely failed with the fight against TB because the only WHO approved *M. bovis* – bacille Calmette-Guérin (BCG) vaccine used since 1921 is only effective in preventing disseminated TB in neonates and its efficacy wanes with time hence does not offer protection later in life (Davenne & McShane, 2016). BCG vaccination is administered once at birth and although over 90% of newborns are vaccinated annually it has not contributed much in reducing TB incidence (Zwerling *et al.*, 2011). Nevertheless, the usefulness of the BCG vaccine cannot be undervalued because without vaccination of neonates, the morbidity and mortality associated with childhood (<5 years) TB can be so devastating. The current TB epidemic could be drastically reduced with a potent vaccine that is able to either protect against new infections or induce clearance of latent infections or both.

Current efforts towards TB vaccines are focused on developing new vaccines that would be effective against TB irrespective of the type of TB and the age group of the affected person using recombinant BCGs, protein/adjuvant combinations or attenuated MTBC strains (Ndiaye *et al.*, 2015; WHO, 2015b). Despite the numerous challenges faced in the development of novel TB vaccines, at least 13 vaccine candidates (whole-cell or lysates of mycobacteria, viral vector vaccines and adjuvanted recombinant protein vaccines) are currently in different phases of clinical trials (Pai *et al.*, 2016). A recent modeling study showed that by the end of the year 2050 up to 50 million new TB cases can be avoided by administering adolescent and adult vaccines (with about 60% efficacy) to 20% of the population at risk (Aeras, 2014).

University of Ghana http://ugspace.ug.edu.gh

2.4.3 Diagnosis of tuberculosis

2.4.3.1 *Latent infection*

Two tests are available for the diagnosis of latent TB infection (even though not very useful in high TB burdened areas); the tuberculin skin test (TST) and the interferon-γ (IFNγ) release assay (IGRA). The TST is performed using the Mantoux technique which consists of an intradermal injection of 5 tuberculin units (5 TU) of purified protein derivative (PPD) S or 2 TU of PPD RT23 (Pai *et al.*, 2016). If an individual has had cell-mediated immunity to the antigens, resulting in memory T cell, a delayed-type hypersensitivity reaction will occur within 48–72 hours and the results of the test is interpreted considering the size of induration. The IGRA, on the other hand, is an *in-vitro* blood test of cell-mediated immune response which measures T cell release of IFNγ following stimulation by region of difference (RD) 1-encoded antigens (namely, the 6 kDa early secretory antigenic target and culture filtrate protein 10) (Pai *et al.*, 2016). One advantage of the IGRA over the TST is its ability to distinguish between BCG-induced and other MTBC infection-induced positive TST responses. Because of the development of memory T cell responses, a positive IGRA or TST test may not outrightly imply latent TB infection because people who have successfully eliminated the infection might still test positive. These assays can also be used to diagnose active TB in some individuals, such as in children who may not readily produce sputum or extrapulmonary cases.

2.4.3.2 *Active TB*

Active TB diagnosis usually include; i) an initial clinical diagnosis by observation of symptoms suggestive of PTB such as chronic cough lasting more than two weeks with associated weight loss and night sweats ii) detection of lung abnormalities through a chest

University of Ghana http://ugspace.ug.edu.gh

X-ray or PET-CT scan suggestive of MTBC infection and iii) once clinical presentations and X-ray are suggestive of MTBC infection, the presumptive TB patient will now proceed to perform laboratory-based tests for confirmation. Compared to PTB, EPTB diagnosis is difficult due to the varying clinical presentation, paucity of bacilli and use of invasive procedures for sampling and mainly involves direct (laboratory-based methods) and indirect (measuring host's humoral and cellular response against mycobacterium) means; the recommended diagnosis being made on the basis of a culture-positive specimen or caseating granuloma on biopsy (Purohit & Mustafa, 2015; WHO, 2014).

Currently, three categories of laboratory-based tests exist for TB diagnosis.

1) Sputum smear microscopy: This diagnostic technique has been in existence for over a century and requires the physical examination of stained sputum samples using a specialized staining technique known as Zeihl-Neelsen (ZN) staining and thereafter using a microscope to observe the presence of bacteria. The TB bacilli, like other members of the mycobacteria family, have a thick waxy cell wall made up of peptidoglycans and mycolic acids which resists decolorization by acid-alcohols following ZN staining, and hence are termed acid-fast bacilli (AFB). An observation of AFB from a presumptive TB case is a confirmation of active TB. The technique is cheap and rapid (results obtained in a day) but lacks sensitivity, specificity and is unable to detect drug resistance. It detects about 50% of TB cases requiring about 10^5 bacilli per ml to give a positive result and not all AFBs are members of the MTBC (Singhal & Myneedu, 2015). Although this technique has many limitations, it remains the most widely used test for diagnosing active TB disease in most low and middle-income countries.

University of Ghana http://ugspace.ug.edu.gh

2) Rapid molecular tests: The detection of unique DNA segments (including the insertions sequence 6110 (IS*6110*) and bacteria RNA polymerase gene (*rpoB*) in the genome of the MTBC is the bases for molecular diagnosis and mainly involves amplifying the genetic segment using nucleic acid amplification test (NAAT) such as polymerase chain reaction (PCR). Several rapid molecular tests have been developed, but the Gene Xpert® MTB/RIF assay (Cepheid, USA), an automated, cartridge- based real-time PCR tool, which has existed close to a decade now, is currently the only WHO recommended rapid test for diagnosis of active TB in the general population. This test uses direct sputum samples and can provide results on both MTBC detection (identifying unique regions of the *rpoB*) and rifampicin resistance (clinically relevant mutations localized within the 81 bp core region in *rpoB*) within 2 hours. The Gene Xpert test has much better accuracy than sputum smear microscopy and it is recommended for diagnosing active TB in both adults and children (Togun *et al.*, 2015; WHO, 2018b). Other rapid TB diagnostic tests include; TB-LAMP assay and lateral flow urine LAM assay (LF-LAM) (Na Songkhla *et al.*, 2017; Pham *et al.*, 2018). The LF-LAM has been recommended for the diagnosis and screening of active tuberculosis in people living with HIV whereas the TB-LAMP is recommended for TB screening in adults and a good replacement to microscopy. The LF-LAM has an added advantage of being able to detect extra-pulmonary TB. An added advantage of these DNA-based assays is the ability to differentiate MTBC from NTMs. The limitations of these rapid diagnostic tests mainly stem from the huge cost involved to set one up. The use of molecular-based tests is gradually increasing globally with several countries phasing out on the use of smear microscopy for diagnosis.

University of Ghana http://ugspace.ug.edu.gh

In addition to the Gene Xpert MTB/RIF mentioned earlier, the most widely used DNA-based assays for drug resistance detection is the line probe assay (LPA) developed by Hain life sciences (GmbH, Nehren, Germany). The LPA offers a wider spectrum of test including the ability to test for resistance to rifampicin and isoniazid (referred to as first–line LPAs developed in 2008, GenoType MTBDR*plus*) as well as to test for resistance to fluoroquinolones and injectable second-line anti-TB drugs (referred to as a second-line LPA also developed in 2016, GenoType MTBDR*sl*) (WHO, 2018b; Zumla *et al.*, 2012) by detecting resistant conferring mutations in anti-TB drug resistance associated genes. The GenoType MTBDR*plus* contains probes specific for MTBC, as well as probes for common RIF resistance-conferring mutations (in *rpoB*) and a subset of the mutations conferring INH resistance (*katG* gene coding for the catalase peroxidase; and the promoter region of the *inhA* gene coding for the NADH enoyl ACP reductase).

3) Culture-based methods: Bacteriological cultivation and isolation of the TB bacilli is the gold standard for TB diagnosis because it is highly sensitive and compared to microscopy requires few bacilli to form colonies. It is also the foolproof method for confirming the cause of the pathology. The isolation process usually involves cultivation on selective egg-based solid medium known as Löwenstein-Jensen medium or liquid based medium such as the Middlebrook 7H9 broth. Members of the MTBC usually appears rough and buff on solid media and macroscopic observation followed with AFB staining is usually used to confirm a TB case in addition to the molecular tests. This method is limited by the fact that it requires more sophisticated equipment and laboratory capacity and usually takes between 1 to 12 weeks to provide results depending on the method used and hence not optimum for use for case diagnosis. The liquid cultures are quicker and takes 10

University of Ghana http://ugspace.ug.edu.gh

to 21 days to produce results. The cultivation process also requires the use of an expensive Biosafety level 3 laboratory and trained personnel hence the capacity is limited to few high laboratories and reserved for research and treatment monitoring purposes (WHO, 2018b; Zumla *et al.*, 2012).

2.4.4 Treatment

Anti-TB drugs have been in use since the early 1940s. The WHO recommended treatment regimens for LTBI include 6–9 months of isoniazid, 3 months of rifapentine plus isoniazid, 3–4 months of isoniazid plus rifampicin or 3–4 months of rifampicin alone (WHO, 2015c). Treatment of LTBI is however often done in only low TB burden countries.

The first line regimen for drug sensitive cases is a 6-month multi-drug therapy made up of two phases (WHO, 2015c). The first phase (intensive phase) involves four first-line anti-TB drugs, which includes isoniazid (INH), rifampicin (RIF), pyrazinamide (PZA) and ethambutol (EMB) for two months. The second phase (continuation phase) which spans the remaining four months, the patient is put on only INH and RIF treatment. Individual patient response to treatment is monitored at months 2, 5 and 6 months using both clinical and microbiological indicators. The main indicator is smear microscopy results. Smear conversion at 2 months is a good prognostic; if the patients remains negative at 5 months and 6 months, the patient is then declared cured. The treatment success of drug sensitive cases is approximately 85% (WHO, 2018b). By 2025 at the latest, WHO recommends that, the treatment coverage and treatment success rate should be $\geq$ 90% (WHO, 2018b). However, patients' compliance to treatment is sometimes undermined by the long treatment duration. Therefore, to ensure optimal adherence, the DOTs strategy, where patients are supervised to take their medication has been instituted by the WHO and

University of Ghana http://ugspace.ug.edu.gh

national programs. Other supportive measures or enablers such as provision of transport fares and food are necessary as relapse or drug resistance can emerge due to lack of treatment adherence. Using adherence interventions such as patient education and counseling, psychological interventions, incentives and enablers and digital health technologies, it is possible to improve TB treatment outcomes (Alipanah *et al.*, 2018; Amo-Adjei & Awusabo-Asare, 2013).

Treatment of drug resistant cases such as MDR/RIF-resistance requires longer treatment periods (9 – 20 months) with more expensive and toxic second-line drugs such as fluoroquinolones (Moxifloxacin, Levofloxacin, gatifloxacin) and injectable aminoglycosides (Amikacin, Kanamycin and Capreomycin). Other drugs such as ethionamide, prothionamide, cycloserine, terizidone, clofazimine or linezolid can be added to some second line regimens (Pai *et al.*, 2016). There are currently up to 20 TB drugs in various phase of clinical trials aiming at shortening the long treatment regimens (WHO, 2018a). Regimens for treating XDR-TB include linezolid, bedaquiline and delamanid combinations even though some of the drugs are still under clinical trials. Treatment success for XDR-TB globally is not well documented though the indicators show relatively poor success rate. Therefore, the best way to control drug resistant TB is to focus on early case detection and ensuring compliance to avoid emergence of drug resistant.

2.4.5 Relapse verses Re-infection

Like many other infectious diseases, having a first episode of active TB does not guarantee you a lifetime protection against TB. About 4% of previously treated TB cases in Ghana is estimated to develop another episode of active TB (WHO, 2015d, 2018c). This recurring TB falls into one of two categories (Figure 2.7); either being a relapse case (endogenous

University of Ghana http://ugspace.ug.edu.gh

reactivation) or a re-infection case. The former occurs in patients who may not have been adequately treated during their primary disease and as such the original strain may have rebounded. Situations where the patient was adequately treated but still presents with a second episode implies that the current episode is due to re-infection especially among HIV co-infected patients and those living in high TB burdened countries (Narayanan *et al.*, 2010; Verver *et al.*, 2005). For instance, in South India, about 88% of HIV individuals with secondary case of TB were due to re-infection (Narayanan *et al.*, 2010). Relapse cases are more likely to be associated with drug resistance and a poor prognosis (Alipanah *et al.*, 2018).

It is difficult for clinicians to decipher which category a TB patient presenting a second episode will fall into, and this may have far reaching consequences with regards to accurate treatment. The introduction of modern genotyping tools has made it possible in most instances to distinguish between relapse and re-infection. The assumption is that, if strains from both episodes are genotypically indistinguishable (clustered), it suggests relapse whereas a distinguishable strain suggests re-infection. Many studies have employed genotyping tools like MIRU-VNTR, IS*6110* RFLP and the ultimate whole genome sequencing technique to explore relapse and re-infection among TB patients (Bryant *et al.*, 2013; Guerra-Assuncao *et al.*, 2015; McIvor *et al.*, 2017; Unis *et al.*, 2014; Varghese *et al.*, 2013; Zong *et al.*, 2018). Varghese and colleagues have employed such tools for the successive detection of endogenous reactivation followed by exogenous reinfection with drug resistant strains found among a Saudi Arabian cohort (Varghese *et al.*, 2013). However, these large DNA-sequence based typing assays can potentially be confounded by convergence evolution. This means that established clusters may not actually reflect

University of Ghana http://ugspace.ug.edu.gh

transmission events. This therefore advocates for the use of more robust but stable assays relatively free from convergence evolution such as SNP typing or whole genome sequencing.

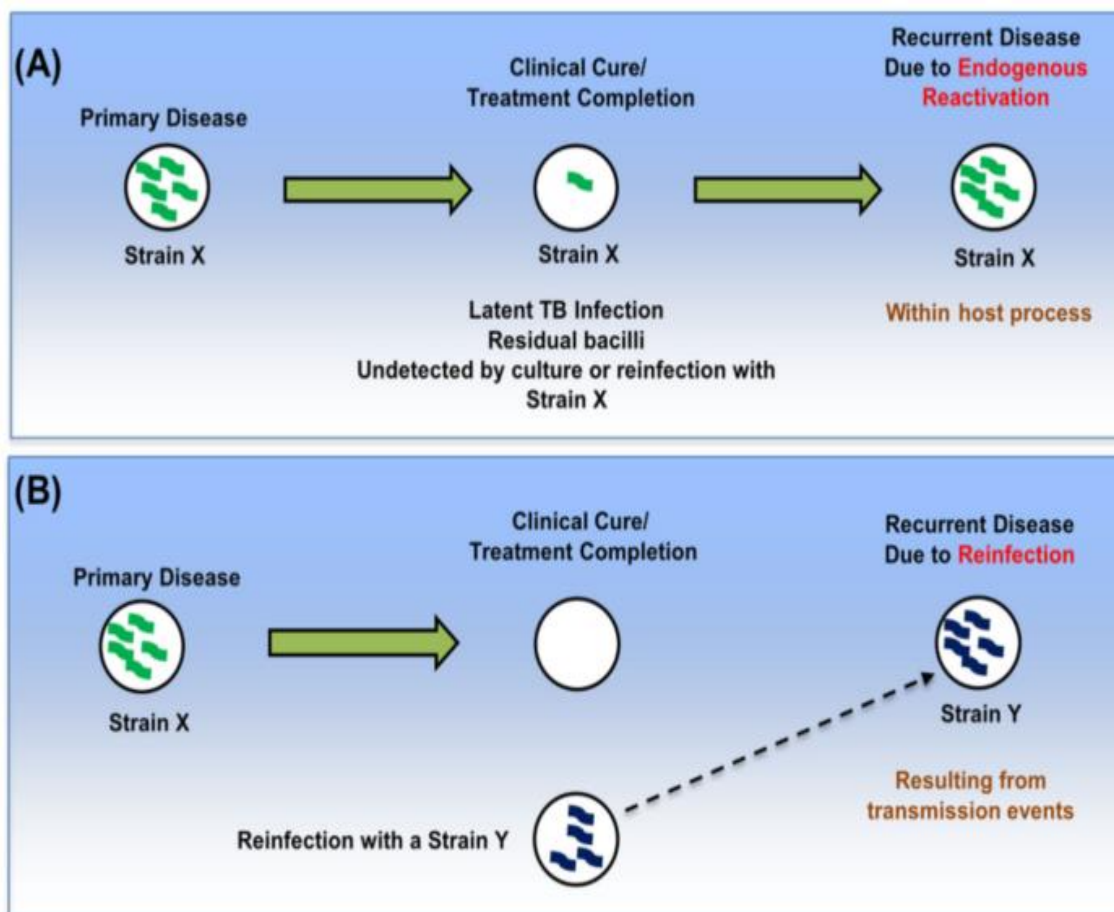


Figure 2.7. Relapse and re-infection of *M. tuberculosis* strains. "Shown in the illustration are the two scenarios that give rise to recurrent TB disease. (A) Relapse disease occurs due to the re-emergence of an *M. tuberculosis* strain that caused the original infection, suggesting that complete eradication of tubercle bacteria was not achieved during the primary disease episode. In this case, the second disease episode results from the presence of persisting organisms that emerge when treatment is stopped, and environmental factors are favorable for bacillary survival. (B) Re-

University of Ghana http://ugspace.ug.edu.gh

infection occurs when a patient is infected with an *M. tuberculosis* strain that is distinct from the strain that caused the original infection. High rates of re-infection can be attributed to both environmental and host factors. Alternatively, an individual can be functionally cured, with no surviving bacteria but can then be re-infected with the same strain that caused the primary disease episode. In the absence of molecular tools to distinguish these events, this would result in the incorrect classification of re-activation." Adapted from (McIvor *et al.*, 2017).

2.5 Molecular epidemiology of Infectious Diseases

2.5.1 Molecular Epidemiology of tuberculosis: Opportunities for TB control

Epidemiology is a branch of medicine which aims to study the incidence, distribution, determinants, and causes of health-related problems in populations in order to effectively prevent and control them together with some associated health factors (Last, 2001).

Molecular epidemiology is a discipline that stems from epidemiology and medical science and focuses on the contribution of potential genetic and environmental risk factors, identified at the molecular level (using molecular or genetic markers), to the etiology, distribution, and prevention of disease within families and across populations. It offers a potent understanding of transmission as well as the population structure and evolution of pathogens and has emerged from the integration of molecular biology into classical epidemiological research.

It makes use of molecular and epidemiological data of either the pathogen or the host or both. A comprehensive epidemiological data for any infectious disease related investigation may include a wide spectrum of data ranging from clinical (from as early as

University of Ghana http://ugspace.ug.edu.gh

the date of onset of disease symptoms), demographical (place of residence, work, ethnicity, etc.), socio-economic (social habits, travel history, salary) among others. Molecular epidemiology is not just genotypic characterization of isolates; the collection of comprehensive epidemiological data is critical for any investigation as it forms the backbone in outbreak studies and very useful to establish epidemiological associations. Nevertheless, epidemiological data are often challenging to obtain (Walker, Monk*, et al.*, 2013).

Current understanding and revision of previous dogmas have been possible due to some molecular epidemiology and pathogen genomics studies in advanced countries of the north, though there is lack of studies in sub-Saharan Africa. Such studies involving the MTBC have allowed the study of evolutionary relationships and helped identify the transmission routes of the pathogen between hosts and for investigating the phylogeny of the MTBC (Figure 2.8 and 2.9). The broad classification and evolutionary scenario of members of the MTBC was made possible from earlier studies that utilized genetic classification tools based on the presence or absence of some genomic regions known as region of difference (RD) (Figure 2.8) (Brosch *et al.*, 2002; van Ingen *et al.*, 2012).

Based on specific genomic deletion pattern, the MAF are further characterized into MAF West African 1 (WA-1) and West African 2 (WA-2) (Figure 2.8). Both MAF WA-1 and WA-2 have RD9 deleted which distinguishes them from the MTBss. *M. africanum* WA-1 and WA-2 are also further distinguished from each other by the deletion of RD7, RD8 and RD10 in MAF WA-2. Additionally, RD 711 and RD702 have been found to differentiate MAF WA-1 (RD711 deleted) from WA-2 (RD702 deleted) (de Jong, Antonio*, et al.*, 2010; Gagneux *et al.*, 2006).

University of Ghana http://ugspace.ug.edu.gh

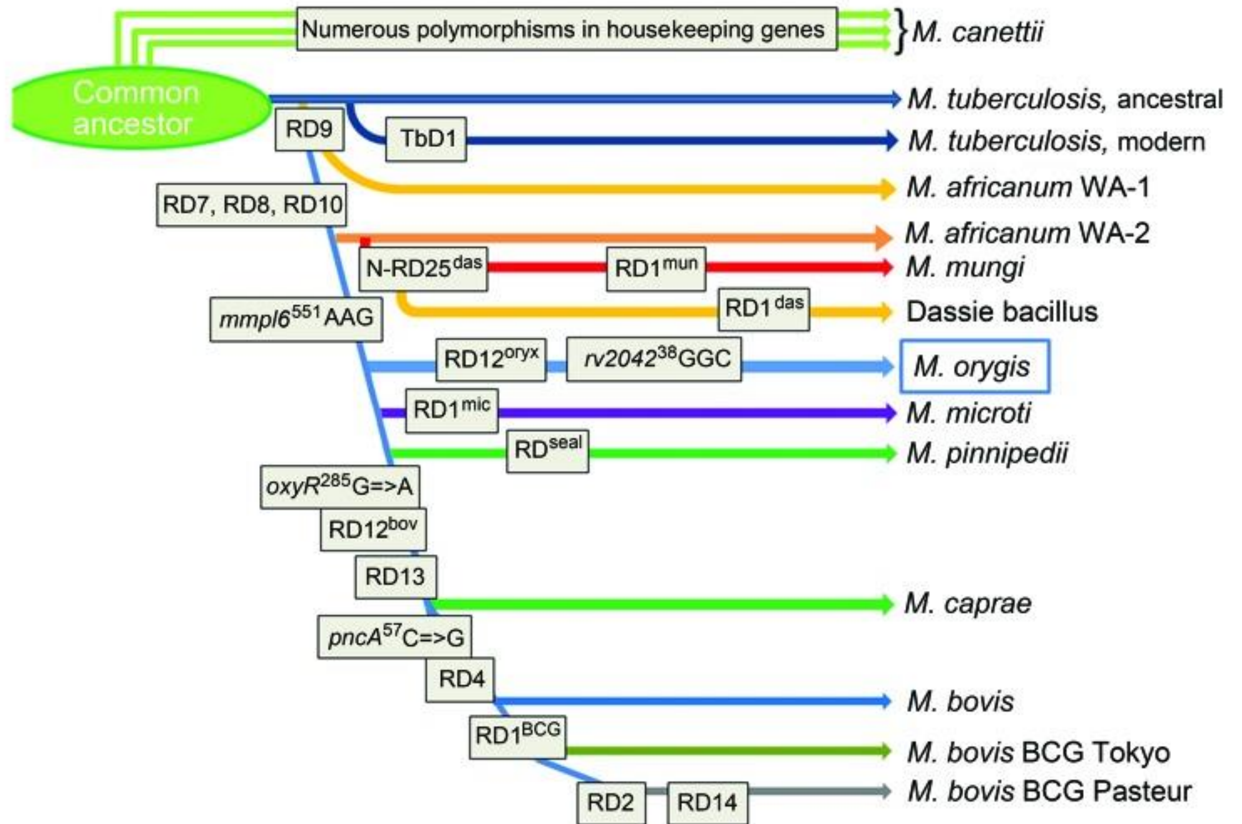


Figure 2.8. Phylogenetic relationship among members of the *M. tuberculosis* complex.

Adapted from (van Ingen *et al.*, 2012)

University of Ghana http://ugspace.ug.edu.gh

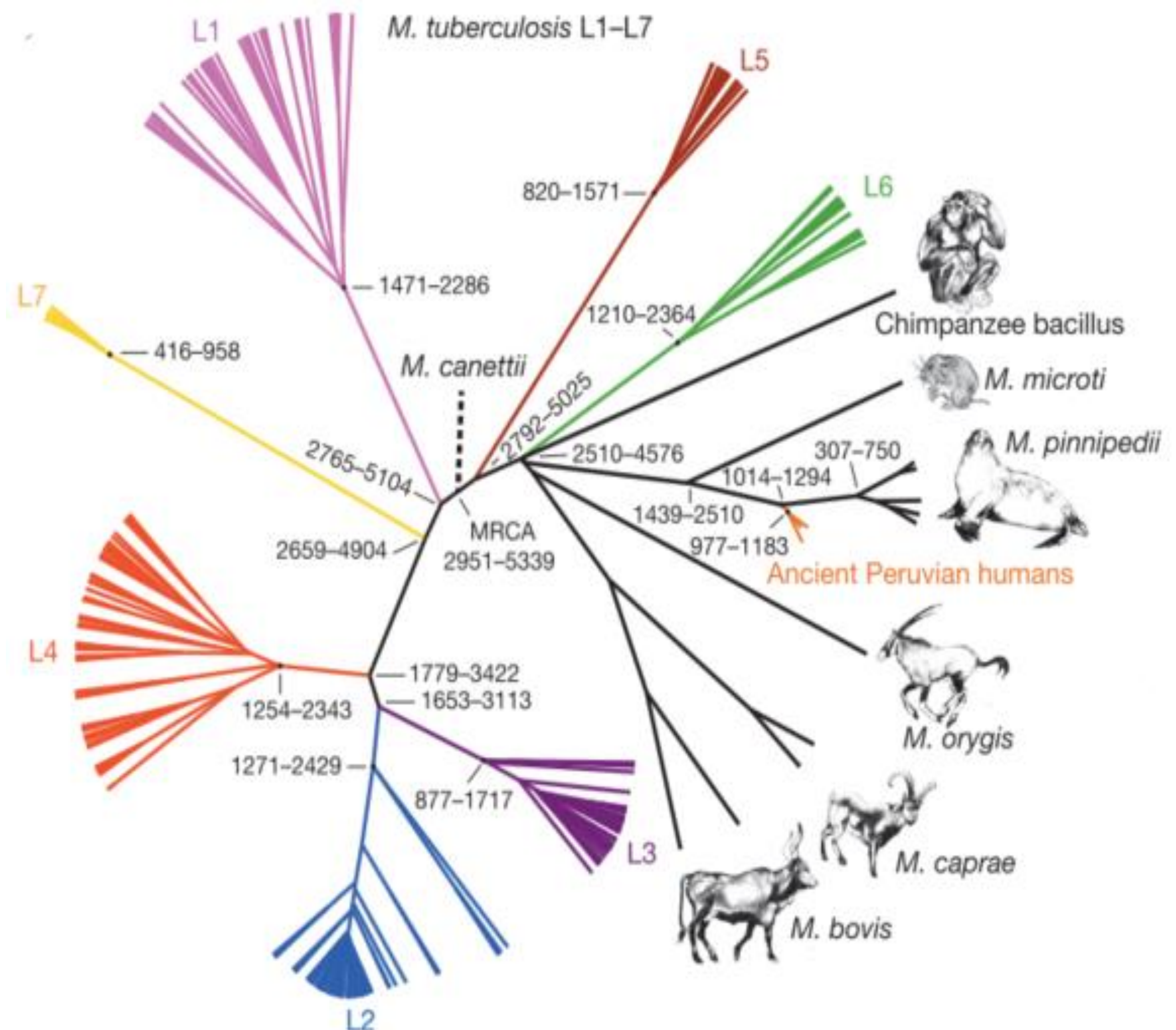


Figure 2.9. Phylogenetic analysis of MTBC. Bayesian maximum clade credibility tree of 261 MTBC genomes, with estimated divergence dates shown in years before present using a model of population. Each clade represents an MTBC lineage and the clade size corresponds to the number of individual strain genomes. The clades are coloured according to the Gagneux 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. Adapted from (Bos *et al.*, 2014).

University of Ghana http://ugspace.ug.edu.gh

Using genetic tools like LSP and SNP, seven lineages have been delineated between MTBss and MAF; 5 among the MTBss and 2 MAF and the lineages show a phylogeographic distribution. Thus each of the 7 human adapted MTBC lineages is associated with distinct geographical areas; the lineage 1 (Indo-Oceanic lineage) are predominantly found in the Philippines & Rim of Indian Ocean; lineage 2 (East-Asian lineage) in East Asia, Russia and South Africa; lineage 3 (East-Africa-Indian lineage) in North India and East Africa; lineage 4 (Euro-America lineage) in Europe, America, Africa and Middle East; lineages 5 and 6 (West-African lineage 1 and 2 respectively) in West Africa and lineage 7 (recently identified lineage) in the Horn of Africa (Blouin *et al.*, 2012; Gagneux *et al.*, 2006; Gagneux & Small, 2007). Although these phylogenetic lineages are associated with specific geographical areas, all the seven lineages can be found in Africa (Figure 2.10). While all the lineages of the MTBss with the exception of L7 (restricted to the horn of Africa) is globally distributed, MAF is limited to West-Africa causing over 50% of human TB in some of the countries (de Jong, Antonio*, et al.*, 2010) (Figure 2.11). In Ghana for instance, MAF is responsible for approximately 20% of all TB cases (Yeboah-Manu *et al.*, 2011; Yeboah-Manu *et al.*, 2016). Ghana is one of the few countries that have 6 out of the 7 lineages and hence serves a good hub for phylogenetic investigations of the human adapted MTBC.

University of Ghana http://ugspace.ug.edu.gh

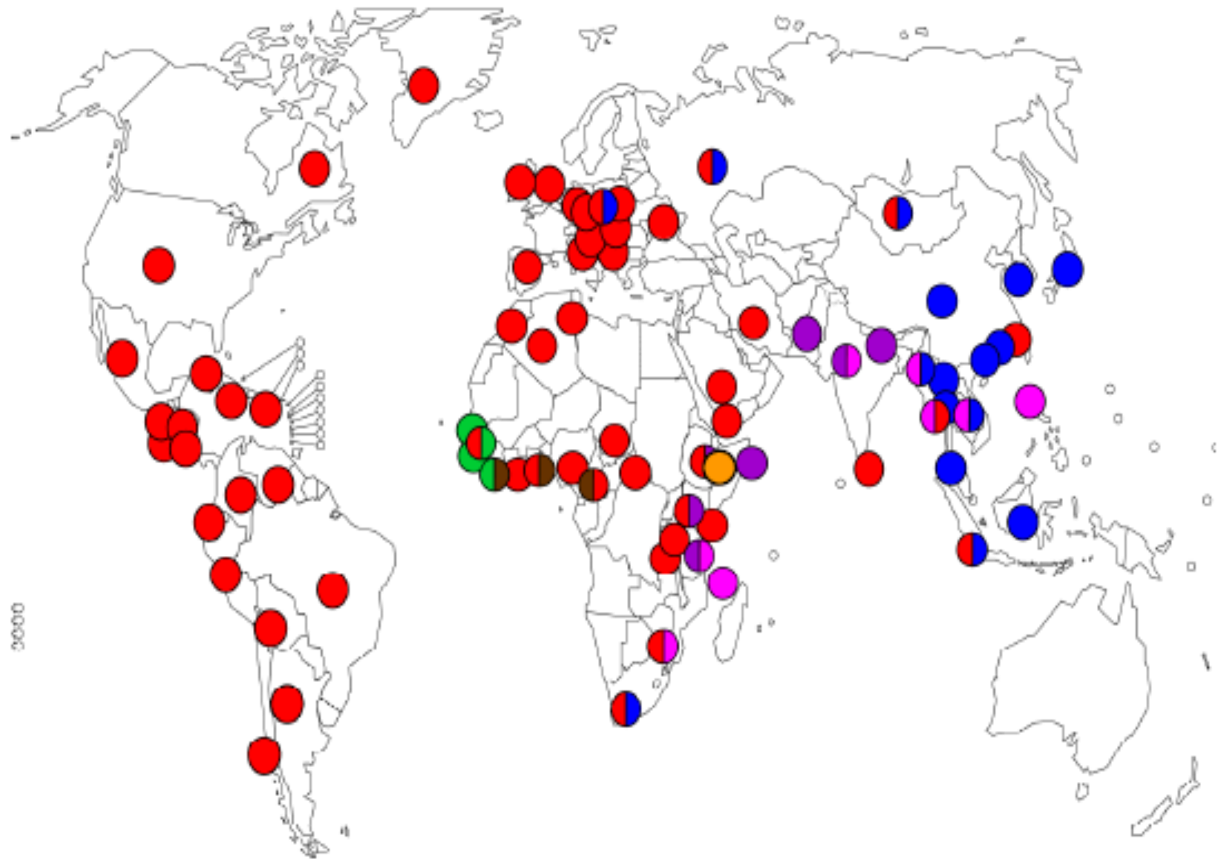


Figure 2.10. Phylogeography of MTBC lineages. The MTBC exhibits a phylogeographic nature. The colours of the dots relate to the Gagneux lineages and indicate the dominant lineage(s) in the respective geographical setting. Lineages 1 - 7 are represented by the colours pink, blue, purple, red, brown, green and yellow respectively. Adapted and modified from Gagneux *et al.,* 2006 (Gagneux *et al.*, 2006).

University of Ghana http://ugspace.ug.edu.gh

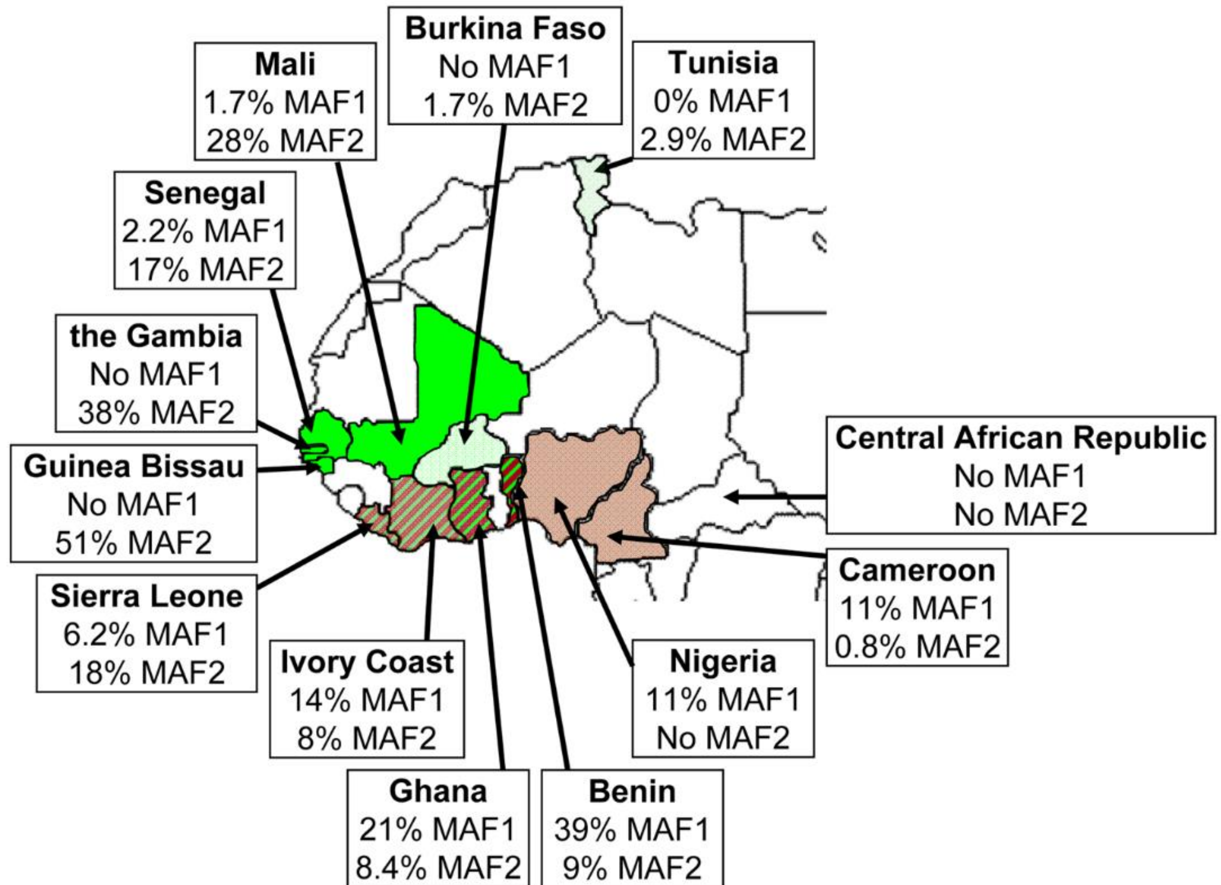


Figure 2.11. Prevalence of *M. africanum* in Western African countries. Adapted from (de Jong, Antonio*, et al.*, 2010)

University of Ghana http://ugspace.ug.edu.gh

2.5.2 Genotyping methods used for molecular epidemiological studies of MTBC

Molecular typing techniques that have been used to study the biology and epidemiology of the MTBC and TB disease respectively include large sequence polymorphism (LSP-typing), single nucleotide polymorphism typing (SNP-typing), mycobacterium interspersed repetitive unit – variable number of tandem repeat typing (MIRU-VNTR-typing), spacer oligonucleotide typing (spoligotyping), insertion sequence *6110* typing (IS*6110*-typing) and more recently, whole genome sequence typing (WGS-typing) (Afaghi-Gharamaleki *et al.*, 2017; Affolabi *et al.*, 2009; Alexander *et al.*, 2009; Jagielski *et al.*, 2014; Moström *et al.*, 2002; Small *et al.*, 1994; Soolingen *et al.*, 1999; Walker, Monk*, et al.*, 2013). These molecular tools have varying discriminatory power (Barnes & Cave, 2003; Jagielski *et al.*, 2014; Jamieson *et al.*, 2014; Kanduma *et al.*, 2003; Mendez *et al.*, 2016; Murray & Alland, 2002). When considering a genotyping method, several factors come into play; reliability, reproducibility, data portability, discriminatory/resolution power and molecular clock of the genetic marker being used and the intended use of the results.

Reliability refers to the quality of the genotyping tool being worthy of being depended on in such a way that it does not fail to produce the desired results. Reproducibility refers to the capacity of the tool to produce consistent results when performed in different laboratories over different geographic settings. However, simplicity/difficulty of tool can potentially affect its reproducibility. Data portability also refers to the nature of the final output from the genotyping tool and the quality of being handy, shared and interpreted the same way by different users. The level of discriminatory power is very critical for interpreting data and it refers to the ability of the genotyping tool to accurately distinguish

University of Ghana http://ugspace.ug.edu.gh

between two unrelated strains. A better discriminatory tool will subdivide an unrelated cluster of strains into distinct groups than one with a lower discriminatory power.

The molecular clock of a genetic marker employed for the assay simply refers to the rate of change of that element or how frequently it evolves (evolutionary rate). The molecular clock of a good genotyping marker should be fast enough to be able to accurately distinguish unrelated strains and yet sufficiently slow enough to capture epidemiologically linked cases. Genetic markers that evolve at a faster or slower rate may either underestimate or overestimate the rate of recent transmission events.

The genome sequence of the reference strain *M. tuberculosis* H37Rv (Cole *et al.*, 1998) has large amount of repetitive DNA sequences which vary in length, localization and structure (Figure 2.12), that are commonly used as markers for MTBC genotyping. These repetitive sequences can be broadly categorized into two groups: tandem repeats (TR) such as the variable number of tandem repeat (VNTR), and interspersed repeats (IR) such as the direct repeat (DR) locus and the *IS6110* used for restriction fragment length polymorphism typing (IS*6110*-RFLP) (Stavrum *et al.*, 2009; Supply *et al.*, 2006) all widely used as genotyping markers for MTBC (Jagielski *et al.*, 2014). The TRs are short monomeric sequences (up to 100 bp) organized as one directly after the other in arrays, whereas the IRs are distributed as individual units.

University of Ghana http://ugspace.ug.edu.gh

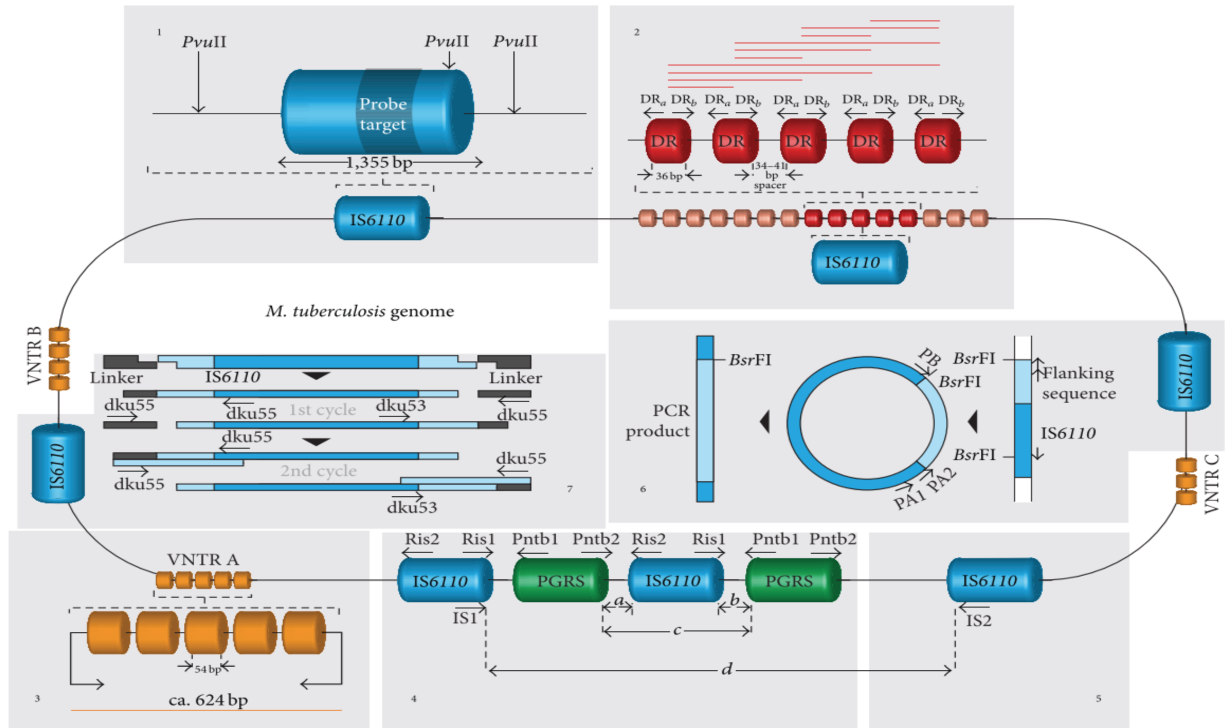


Figure 2.12. Schematic representation of the chromosome of a hypothetical MTBC isolate with marked repetitive elements as targets for different typing methods. "The principle of those methods is pictorially outlined. (1) In IS*6110*-RFLP typing, mycobacterial DNA is cleaved with the restriction endonuclease PvuII, and the resulting fragments are separated electrophoretically on an agarose gel, transferred onto a nylon membrane by Southern blotting, and hybridized to a probe complementary to the 3'end of the IS*6110* (probe target) yielding a characteristic banding pattern, in which every band represents a single IS*6110* element. (2) Spoligotyping relies upon PCR amplification of a single direct repeat (DR) locus which harbors 36 bp direct repeats interspersed with unique 34–41 bp spacer sequences. The PCR products (red horizontal lines) are hybridized to a membrane containing 43 oligonucleotides corresponding to the spacers from *M. tuberculosis* H37Rv and *Mycobacterium bovis* BCG. The presence or absence of each of those 43 spacers in the DR region of the analyzed isolate will be represented as the pattern of positive or negative hybridization signals. (3) The variable numbers of tandem repeat loci (VNTR) or mycobacterial interspersed repetitive units (MIRU) are PCR-amplified and the obtained products (yellow horizontal line) are sized on agarose gels to deduce the number of repeats in each individual locus". Adapted from (Jagielski *et al.*, 2014).

University of Ghana http://ugspace.ug.edu.gh

2.5.2.1 IS*6110* restriction fragment length polymorphism

An important class of interspersed repeat sequences is insertion sequences (IS), which are mobile genetic elements and the best described being IS*6110* (Jagielski *et al.*, 2014), identified by Eisenach *et al*. (Eisenach *et al.*, 1988). Two years onwards, Thierry and colleagues described the insertion sequence IS*6110* as being a member of the IS*3* family and is specific to the MTBC and first indicated its applicability in epidemiological studies (Otal *et al.*, 1991; Thierry *et al.*, 1990). They were the first group to determine the molecular size of the IS*6110* (1,355 bp long) with a unique 28bp terminal inverted repeat (Figure 2.12 (1)). The IS*6110* as a mobile genetic element can transpose into various regions of the genome giving it the ability to alter the bacteria phenotype and has an estimated molecular clock of between 3 – 4 years (de Boer *et al.*, 1999). A single MTBC strain can harbor as many as between 0 and 25 IS*6110* copies (van Soolingen *et al.*, 1991).

Restriction fragment length polymorphism (RFLP) is a technique that exploits polymorphisms in homologous DNA sequences in order to distinguish strains, populations or even species. The principle behind RFLP analysis is that, a restriction enzyme which has affinity to cut certain restricted sites of a DNA molecule based on its sequence is used to cut the DNA molecule to generate fragments with varying lengths. Thus, polymorphisms generated depends on the copy number of the IS and the location(s) within the genome of distinct MTBC genome.

The IS*6110* RFLP methodology involves using one restriction enzyme (*Pvu*II) which cut at only one site within the IS*6110* element to digest the DNA to be investigated and the resulting restriction fragments are then separated by gel electrophoresis (Figure 2.13) according to their size and thereafter are transferred onto a membrane followed by

University of Ghana http://ugspace.ug.edu.gh

hybridization. The resulting hybridized products are visualized as fragments which represent a single copy of IS*6110* surrounded by flanking DNA of different lengths. The ability of the enzyme to cut only one site within the IS*6110* element implies that a strain with only one IS*6110* element will have a single band on the agarose gel because of the circular nature of the bacterial genome.

The use of restriction enzymes implies that the technique requires good quality genomic DNA. This assay involves growing the bacteria to get a good bacteria suspension to start with and due to the slow growing nature of the MTBC, together with other required steps such as agarose gel electrophoresis, Southern blotting and detection of the IS elements with a peroxidase-labelled probe, the IS*6110* RFLP is laborious and requires elaborate infrastructure and expertise resulting in a questionable reproducibility. Strains with low copy numbers (<5 copies of IS*6110*), are not well differentiated with this technique and because it requires large amounts of DNA, its applicability in real-time outbreak investigations is limited. Despite these drawbacks, the technique offers a very good discriminatory power for strains with high IS*6110* copy number and has been widely used for MTBC strain typing to investigate several outbreak cases and for a long time until recent was considered the gold standard (Alexander *et al.*, 2009; Ano *et al.*, 2006; Maguire *et al.*, 2002; Masala *et al.*, 2010).

University of Ghana http://ugspace.ug.edu.gh

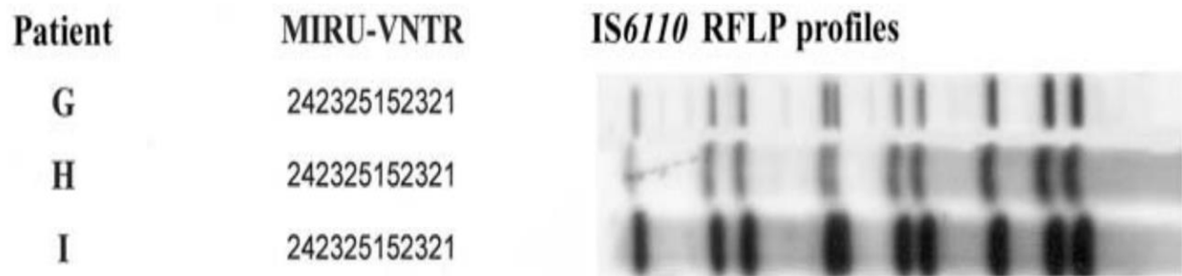


Figure 2.13. Comparison of MIRU-VNTR digitized result output verses IS*6110* RFLP profile. The banding pattern is similar for all three strains and can be confirmed by the allelic pattern of the MIRU-VNTR. Ten distinct bands imply that the *PvuII* cut at 9 positions and therefore each strain possess at least 9 uninterrupted IS6110 elements. Adapted from (Allix *et al.*, 2004).

2.5.2.2 Spoligotyping

Spacer-oligonucleotide typing (spoligotyping) is a PCR-based technique and it relies on detecting polymorphisms in the spacer units at the direct repeat locus of MTBC genome (Figure 2.12 (2)). Spoligotyping is currently one of the most widely used PCR-based techniques for MTBC phylogeographic studies. The DR locus is made up of conserved 36 bp DR's interspersed with unique non-repetitive 34 – 41 bp spacer sequences and it belongs to the clustered regularly interspaced short palindromic repeats (CRISPRs) family of repetitive DNA (Jagielski *et al.*, 2014). The spacer regions are over 60, but 43 have been selected for this typing technique. Polymorphism among the MTBC is due to the presence or absence of distinct spacer regions and this is detected by PCR- amplification of the whole DR locus using primers complimentary to the flanking sequences followed by reverse hybridization (Kamerbeek *et al.*, 1997). For instance, whereas MTBss spoligotypes lacks spacers 33–36, *M. bovis* usually lacks spacers 39–43 and *M. bovis* BCG lack spacers

University of Ghana http://ugspace.ug.edu.gh

3, 9, and 16. Most of the "Beijing" spoligo-family of lineage 2 lacks spacers 1 – 34 and only reacts with the last 9 spacers (35–43). The lack of spacers is most probably the result of deletions/alterations mediated by various genetic mechanisms, such as homologous recombination between adjacent or distinct DRs, transformation due to the insertion of an IS*6110* element, strand slippage during replication, point mutation or transposition events (Groenen *et al.*, 1993; Warren *et al.*, 2002).

The hybridization procedure takes a whole day and involves prior PCR amplification of the DR locus. After PCR amplification, the amplicons of varying sizes are hybridized to a nitrocellulose membrane with 43 covalently bound synthetic oligonucleotides representing the polymorphic spacer regions identified in the reference strain H37Rv (spacers 1–19, 22–32, and 37–43) and *M. bovis* BCG (spacers 20-21 and 33–36). The hybridization signals are detected by chemiluminescence through biotin labeling of the amplicons and a streptavidin-peroxidase conjugate system and then visualized by autoradiography (Figure 2.14). The output is then converted to binary data which is easily interpreted, computerized, and can be compared between different laboratories. A large database of spoligotyping patterns has been created and is available freely online for such comparison (Demay *et al.*, 2012; Weniger *et al.*, 2012)

Spoligotpying has a relatively low discriminatory power because it targets only a single genetic locus, covering less than 0.1% of the MTBC genome. Compared to IS*6110* RFLP, spoligotyping is relatively simple, highly sensitive, cost-effective, high throughput, gives accurate and reproducible results and has short turnaround time (within 2 days at most). The reliability of the spoligotyping technique is linked to the high stability of the DR locus. Spoligotyping is very sensitive requiring about 3 bacilli DNA molecules. As it is

PCR-based, it can be performed directly on sputum samples, Ziehl-Neelsen smear slides and paraffin-embedded tissue sections without requiring prior isolation of the bacteria (Niemann *et al.*, 1999).

This method is usually used together with other more discriminatory tool for transmission studies and to characterize members of the MTBC at the lineage and sub-lineage (spoligo-family) level.

SIT	Spoligotype pattern (spacers 1–43)	Isolate	Family	Lineage
SIT817	■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□■□□□□■■■□■■■	5	Ha3	EA
SIT272	□□□□■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	5	T1	EA
SITOrphan	■■■■■■■■■□□□□□□□□□□□□■■■■■■■■■■■■□□□□■■■■■■■	5	T3	EA
NEW	□□□□□□□□□□■■■■■■■■■■■■■■■■■■■■■□□□□□□□□■■□□	1	S	EA
SIT25	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□■■□□■■■■■	22	CAS	CAS
SIT1227	■■■■■■■■■■■■□■■■■■■□■■□■■■■■■■■■□□□□■■■■■■■	1	T1	EA
SIT1729	■■■□□□□□□□□□□□□□□□□□□□□□■□□□□■■■■■■■■■■■■■■	7	F36	IO
SITOrphan	■□■□□□□■■■■■■■■■■■■■■□□□□□□□□□□□□□■■■■■■■■■	1	CAS	CAS
SIT53	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	14	T1	EA
SIT1200	■■■□□□□■■■■■■■■■■■■■■■□□■■■■■■■■■■■■■■■■■■■	1	F33	IO
SIT1312	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□■■□□■□■■■	1	CAS	CAS
SIT613	■■■■■■■■■□□■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	1	T1	EA
SIT41	■■■■■■■■■■■■■■■■■■■□□□□□■□□■■■■■□□□□■■■■■■■	2	LAM9	EA
SIT523	■■■	4	F33	IO
SIT37	■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	4	T1	EA
SIT1198	■■■□□□□■■■■■□■■■■■■■■■□□□□□□□□□□□□■■□□■■■■■	1	CAS	CAS
SIT137	■■■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■□□□□■■□□□□■	1	X2	EA
SIT910	■■■□□□□□□□□□□□□□□□□□□□□□■■■□□■■■■■■■■■■■■■■	3	F36	IO
SIT1166	■■■■■■■■■□■■■■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	2	T1	EA
SIT149	■■■■■■■■■□□□□□□□□□□■■■■■■■■■■■■■□□□□■■■■■■■	7	T3	EA
SIT289	■■■□□□□■■■■■■■■■■■■■■■□□□□□□□□□□□□■■■□■■■■■	7	CAS	CAS
SIT21	■■■□□□□■■□■■■■■■■■■□□□□□□□□□□□□□□□□■■■■■■■■	5	CAS	CAS
NEW	■■■■■■■■■□□□□■■■■■■■■■■■■■■■■■■■□□□□■■■■■■■	1	T1	EA
NEW	■■■■■■□□□□□■□■■■□□□□□□□□■■■■■■■■□□□□■■■■■■■	1	LAM9	EA
SIT49	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■■□■■■	2	Ha3	EA
NEW	■■■■■■■■□■■■□□□□□□■□□■□□■■■■■■■■□□□□■■■■■■■	1	H37Rv	EA
SIT134	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■□□■■■	3	Ha3	EA
SIT373	■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■□□□□■■■■■■■	1	T1	EA
SIT1360	■□■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■□□■■■	1	Ha3	EA
SIT153	■■■■□■■■■■■■■■■■■■■■■■■■■■■■■■■■□□□□■■■□■■■	2	T1	EA
SIT54	■■■□□■□□■□□□□■■■■■□■■■■■■■□□■■■■■■■■■■■■■■■	1	F33	IO
SIT912	■■■□□□□□□□□□□□□□□□□□□□□□■■■□□■■■■■■■■■■■■■■	1	F36	IO
SIT764	■■■■■■■■■■■■■□■■■■■■■■■■■■■■■■□■□□□□■■■■■■■	1	Ha3	EA
SIT159	■■■■■■■■■■■■■□□□□□□□□□□■■■■■■■■■□□□□■■■■■■■	1	T4	EA
SIT1134	■■■■■■■■■■■■□■■■■■■■■■■■■■■■□□□■□□□□■■■□■■■	1	Ha3	EA
SIT1430	■□■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■□□□□■■□□■■■	1	Ha3	EA

Figure 2.14. Presence (filled squares) or absence (empty squares) of spacer regions after spoligotyping. Adapted from (Debebe *et al.*, 2014).

University of Ghana http://ugspace.ug.edu.gh

2.5.2.3 MIRU-VNTR typing

Mycobacterium tuberculosis was among the first bacterial species in which tandem repeat loci were identified (Supply *et al.*, 2000). The tandem repeat loci described in MTBC is analogous to the minisatellite loci described in eukaryotic genomes known as variable number of tandem repeat (VNTR).

Several bacteria including *Escherichia coli, Clostridium difficile, Yersinia pestis* and *Shigella spp.* also harbor repetitive units and have been genotyped using similar VNTR techniques (Gorge *et al.*, 2008; Klevytska *et al.*, 2001; Manges *et al.*, 2009; Marsh *et al.*, 2006). MIRU-VNTR typing is a PCR-based technique first proposed and named by Supply and colleagues who found up to 41 different repetitive units distributed throughout the genome of the MTBC (Figure 2.12 (3) and Figure 2.15) (Supply *et al.*, 2001; Supply *et al.*, 1997; Supply *et al.*, 2000). The repeat units are mainly intergenic sequences, consists of 46–101 bp sequences in length, are tandemly repeated and scattered throughout the genome.

The principle behind the typing system is based on the cumulative polymorphisms in the number of tandem repeats at distinct locus. Here, each VNTR locus is amplified by PCR with specific primers complementary to the flanking regions, and the resulting PCR products are visualized by standard gel electrophoresis (Figure 2.16), or more recently, analyzed on an automated, fluorescence-based sequencer after running multiplex PCRs (Gauthier *et al.*, 2015). Since the length of each repeat units is known, the calculated sizes reflect the numbers of the amplified MIRU copies. The result is a multi-digit numerical code corresponding to the repeat number at each analyzed locus.

University of Ghana http://ugspace.ug.edu.gh

The individual 41 identified MIRU-VNTR loci exhibit disparate discriminatory powers depending on the geographical setting which has led to the establishment of several different combinations of reduced sets of loci for use under different geographical settings (Allix-Beguec, Supply*, et al.*, 2008; Allix-Béguec *et al.*, 2014; Asante-Poku, Nyaho*, et al.*, 2014; Comas *et al.*, 2009; Roring *et al.*, 2004; Supply *et al.*, 2006; Surikova *et al.*, 2005). However, the currently standardized MIRU-VNTR technique examines up to 24 independent and well-calibrated loci and is a relatively more reliable and efficient typing system, which when combined with spoligotyping has a discriminatory power that exceeds that of IS*6110* RFLP typing and has replaced it as the gold standard for TB transmission studies.

This method is currently adopted and used in a variety of TB molecular epidemiological studies due to its reproducibility, portability, high discriminatory power and standardization (Hamblion *et al.*, 2016; Norheim *et al.*, 2017; Streit *et al.*, 2015; Vluggen *et al.*, 2017; Zamani *et al.*, 2016). In general, the discriminatory power of MIRU-VNTR analysis increases with the number of loci evaluated. The advantage of MIRU-VNTR over other genotyping tools is that it is fast, easy-to-perform, sensitive and highly reproducible and discriminatory. Like spoligotyping, a particular advantage compared to the IS*6110* RFLP typing, is its portability due to digitalization of the generated allelic patterns and therefore allows inter and intra laboratory comparisons around the globe and this facilitates the data to be deposited in global readily available online databases (http://www.miru-vntrplus.org/) (Weniger *et al.*, 2012). Despite the numerous advantages, the method is laborious, and its discriminatory power cannot compare with whole genome sequencing.

University of Ghana http://ugspace.ug.edu.gh

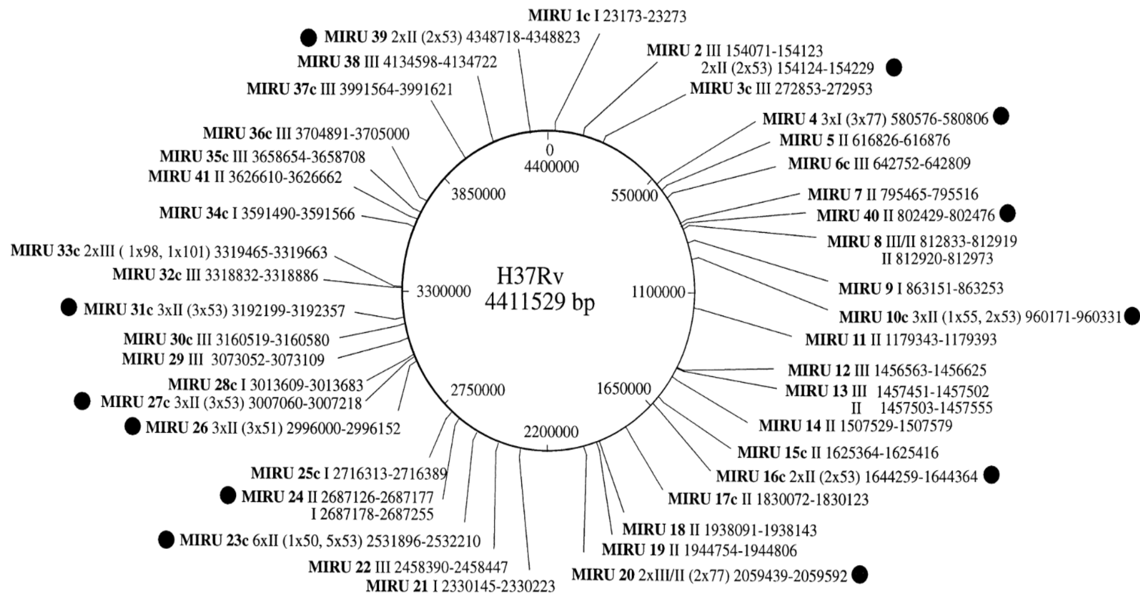


Figure 2.15. Location of 41 MIRU loci in *M. tuberculosis* H37Rv genome. Adapted from (Supply *et al.*, 2000)

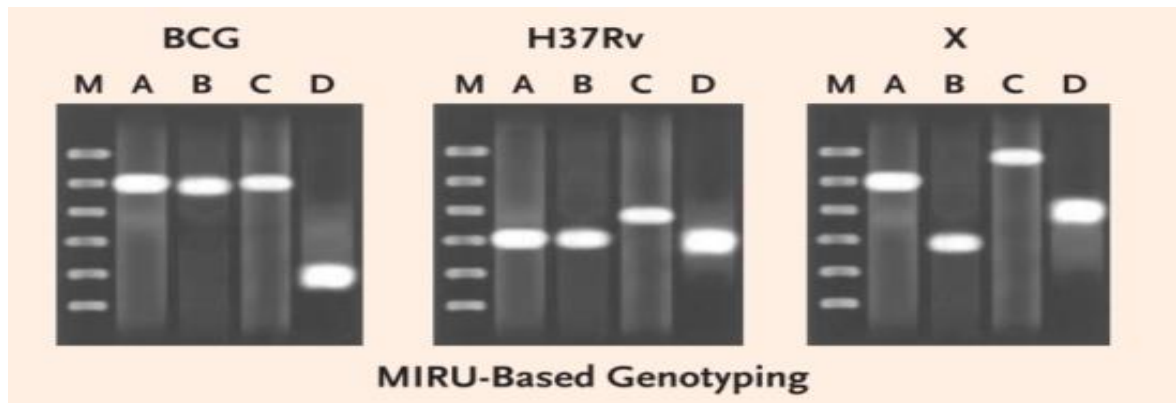


Figure 2.16. Hypothetical representation of the genotypic pattern of BCG, H37Rv and strain X based on MIRU-VNTR. MIRUs contain repeat units, and MIRU analysis involves the use of polymerase-chain-reaction (PCR) amplification and gel electrophoresis to categorize the number and size of repeats in 12 – 24 independent loci, each of which has a unique repeated sequence. The sizes of molecular-weight markers (M) and PCR products for the loci A, B, C, and D in BCG, H37Rv, and X are shown. The specific sizes of the various MIRUs in each strain result in a distinctive fingerprint for the strain" Adapted from (Barnes & Cave, 2003).

University of Ghana http://ugspace.ug.edu.gh

2.5.2.4 Whole genome sequencing

Whole genome sequencing (WGS) offer an attempt to read the entire length of base sequences in the order in which they occur in a given genome and is by far the ultimate tool for molecular epidemiological studies. Several sequencing technologies are currently available; however, bacterial genomes are mainly sequenced by the illumina platform (HiSeq and MiSeq), PacBio and Ion Torrent PGM, which are all next generation sequencing (NGS) platforms (Table 2.1). The main difference in these platforms is their ability to either produce short DNA sequence reads (illumina HiSeq, illumina MiSeq and Ion Torent) or long reads (PacBio); the latter of which is mostly used for de novo genome assembly and generation of reference sequences.

Table 2.1. Next generation sequencing platforms commonly used for sequencing bacterial genomes

Sequencing platform	Typical read length	Yield per run	Paired ends	Error rate (%)
Illumina HiSeq 2000	100 bp	40 Gb	Yes	~0.1
Illimina MiSeq	250 bp	10 Gb	Yes	~0.1
PacBio	6 – 8 kbp	350 – 500 Mb	No	~13
Ion Torrent PGM	200 – 400 bp	10 – 1,000 Mb	No	~1

Table adapted from (Benjak *et al.*, 2015)

The recent increase in NGS technologies have made WGS an attractive molecular tool used in many investigations. The use of WGS technology offers the ability to perform advanced analysis such as a comprehensive comparative genomics analysis for genome-

University of Ghana http://ugspace.ug.edu.gh

wide identification of mutations to detect small and rare differences between different strains as well as *de novo* assembly of entire genomes. Comparative genomics analysis is a very powerful tool and has not only been used to unravel the mechanism of bacterial resistance to various antimicrobials but also used widely in comparing different microbial genomes, for phylogeographic and phylogenetic studies, estimation of mutation rates as well as in transmission studies (Benjak *et al.*, 2015; Otchere *et al.*, 2018; Walker *et al.*, 2014; Walker *et al.*, 2018).

The use of WGS has however not been readily utilized globally due to the huge cost and expertise needed to analyze the generated data. However, due to the increased competition in available sequencing platforms and availability of simpler data analysis tools, WGS is gradually becoming less expensive and analysis of data generated becoming simpler. The discriminatory power of previously used molecular tool were limited by the highly monomorphic nature of the MTBC but WGS technology circumvent this challenge and enables isolates to be differentiated with much greater resolution.

Depending on the circulating strains for a given geographic setting, the strain resolution power of the above mentioned techniques may not be enough to distinguish between unrelated strains (Jamieson *et al.*, 2014). Due to this issue, it is advised that WGS be performed to make decisive conclusions regarding TB transmission studies. Whole genome sequence data does not only give us the power to identify recent TB transmission but also the ability to trace the route/direction of transmission between such epidemiologically linked TB cases (Walker, Ip*, et al.*, 2013). Since backwards mutations are rare in MTBC, the ability of WGS to trace the direction of transmission in an outbreak is made possible through the pattern of accumulated mutations.

University of Ghana http://ugspace.ug.edu.gh

In 2011, WGS was first effectively used to delineate two unrelated transmission events among a cohort of drug users with identical MIRU-VNTR profiles from Vancouver and ever since has been used in a large array of studies (Gardy *et al.*, 2011; Lalor *et al.*, 2018; Walker *et al.*, 2017; Walker, Monk, *et al.*, 2013).

2.5.3 Hunter-Gaston Discriminatory Index and recent TB transmission

The Hunter-Gaston Discriminatory Index (HGDI) is a powerful statistical tool that is widely used to calculate the discriminatory power of each locus of a genetic marker (or a typing tool) (Alonso-Rodriguez *et al.*, 2008; Hunter & Gaston, 1988; Ravansalar *et al.*, 2016). The average probability that a typing system will differentiate two randomly unrelated strains in a microbial population of a given taxon is referred as the discriminatory power and is denoted by **D**. In simple terms, the discriminatory power is the ability of a typing technique to distinguish between unrelated strains and it is defined by the number of genotypes detected by technique being used.

The discriminatory power is calculated using the formula (Hunter & Gaston, 1988);

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} x_j(x_j - 1)$$

"Where D is the index of discriminatory power, N the number of unrelated strains tested, S the number of different types, and xj the number of strains belonging to the jth type, assuming that strains will be classified into mutually exclusive categories" (Hunter & Gaston, 1988).

University of Ghana http://ugspace.ug.edu.gh

Interpretation of Discriminatory Power

- A **D** value of 1.0 indicates that a typing method was able to distinguish each member of a strain population from all other members of that population.
- Conversely, an index of 0.0 indicates that all members of a strain population were of an identical type.
- An index of 0.5 means that if one strain was chosen at random from a strain population, then there would be a 50% probability that the next strain chosen at random would be indistinguishable from the first.

The assumption in molecular epidemiology is that strains with similar pattern are linked and cluster together using the available typing tools. The number of clustered strains in a population may be used to calculate a percentage of cases due to recent transmission using an index known as the recent transmission rate using the n-1 formula described by Glynn (Crampin *et al.*, 2006; Glynn *et al.*, 1999).

Recent transmission rate = (nc-c)/n

Where; *nc* is the total number of clustered cases, *c* is the number of clusters, and *n* is the total number of cases in the sample.

Interpretation of Recent transmission rate

A high recent transmission rate suggests that most of the TB case is due to recent infection and not reactivation of latent TB infection while low recent transmission rate is an indication of the reverse.

University of Ghana http://ugspace.ug.edu.gh

Nevertheless, it is worth noting that there are situations that could lead to false conclusions, such that identification of genetic relatedness of two strains from separate individuals may not imply a recent transmission event. Examples are: the simultaneous reactivation of a previously acquired infection with the same organism (coincidence of time), the regional predominance of a particular strain, circulating over a long time, or a laboratory cross contamination (Jagielski *et al.*, 2014).

University of Ghana http://ugspace.ug.edu.gh

CHAPTER THREE

3.0 Manuscript 1: Spatio-Temporal Distribution of *Mycobacterium tuberculosis* Complex Strains in Ghana

(Published in PlosOne, 2016 doi:10.1371/journal.pone.0161892)

3.1 Abstract

Background: There is a perception that genomic differences in the species/lineages of the nine species making the *Mycobacterium tuberculosis* complex (MTBC) may affect the efficacy of distinct control tools in certain geographical areas. We therefore analyzed the prevalence and spatial distribution of MTBC species and lineages among isolates from pulmonary TB cases over an 8-year period, 2007-2014.

Methodology: Mycobacterial species isolated by culture from consecutively recruited pulmonary tuberculosis patients presenting at selected district/sub-district health facilities were confirmed as MTBC by IS*6110* and rpo*β* PCR and further assigned lineages and sub lineages by spoligotyping and large sequence polymorphism PCR (RDs 4, 9, 12, 702, 711) assays. Patient characteristics, residency, and risks were obtained with a structured questionnaire. We used SaTScan and ArcMap analyses to identify significantly clustered MTBC lineages within selected districts and spatial display, respectively.

Results: Among 2,551 isolates, 2,019 (79.1%), 516 (20.2%) and 16 (0.6%) were identified as *M. tuberculosis* sensu stricto (MTBss), *M. africanum* (MAF), 15 *M. bovis* and 1 *M. caprae,* respectively. The proportions of MTBss and MAF were fairly constant within the study period. MAF spoligotypes were dominated by Spoligotype International Type (SIT)

University of Ghana http://ugspace.ug.edu.gh

331 (25.42%), SIT 326 (15.25%) and SIT 181 (14.12%). We found *M. bovis* to be significantly higher in Northern Ghana (1.9% of 212) than Southern Ghana (0.5% of 2339) ($p=0.020$). Using the purely spatial and space-time analysis, seven significant MTBC lineage clusters ($p< 0.05$) were identified. Notable among the clusters were Ghana and Cameroon sub-lineages found to be associated with north and south, respectively.

Conclusion: This study demonstrated that overall, 79.1% of TB in Ghana is caused by MTBss and 20% by *M. africanum*. Unlike some West African Countries, we did not observe a decline of MAF prevalence in Ghana.

University of Ghana http://ugspace.ug.edu.gh

3.2 Introduction

One of the major threats to tuberculosis (TB) control is the emergence of strains that are resistant to most of the anti-TB drugs, which could make a treatable disease untreatable (WHO, 2014). Other factors that limit current TB control efforts are lack of effective vaccines, lack of cheap but effective rapid diagnostics, emergence of HIV/AIDS pandemic and limited understanding of the diversity of circulating strains (WHO, 2014). The increase in TB cases globally requires a concerted effort to control this global public health problem. This calls for improved understanding of the disease pathogenesis, epidemiology, and genetic variability within the causative agent.

TB is caused by a group of closely related acid-fast gram-positive bacteria, together referred to as the *Mycobacterium tuberculosis* complex (MTBC) (Brosch *et al.*, 2002; Gagneux & Small, 2007). The MTBC comprises *M. tuberculosis* sensu stricto (MTBss), *M. africanum* (MAF), *M. microti*, *M. bovis*, *M. caprae*, *M. mungi*, *M. suricattae, M. orygis* and *M. pinnipedii.* They have varying host ranges: *Mycobacterium microti* affects voles, (Frota *et al.*, 2004; Wells & Oxon, 1937) *M. caprae* a pathogen of goats and sheep (Aranaz *et al.*, 1999). *M. mungi:* Mangoose pathogen, *M. orygis* a pathogen of antelope (Jakko van *et al.*, 2012), *M. pinnipedii* a pathogen of seals and sea lions (Cousins *et al.*, 2003). *Mycobacterium bovis* displays the broadest spectrum of host affecting humans and animals (Garnier *et al.*, 2003). *Mycobacterium tuberculosis* sensu stricto and *M. africanum* are the main causative agents of TB in humans; referred to as human adapted MTBC and the remaining seven species as animal adapted (Gagneux & Small, 2007). The human adapted MTBC comprises seven main phylogenetic lineages, which have been confirmed by single nucleotide polymorphisms (SNPs) and whole genome sequencing (Comas *et al.*, 2009;

University of Ghana http://ugspace.ug.edu.gh

Gagneux *et al.*, 2006; Gagneux & Small, 2007; Rebuma *et al.*, 2013). These lineages were further found to exhibit a phylogeographical structure, which means that specific lineages are closely associated with specific geographic regions, and preferentially infect persons originating from these regions. Importantly, findings from recent genomic analysis indicate that some of these human MTBC lineages are as genetically distinct from each other as from the animal-adapted forms of MTBC (Gagneux *et al.*, 2006) and have genomic differences that may influence host-pathogen interaction as well as applicability of control tools such as diagnostics and vaccine. Thus, the lineages distribution needs to be taken into account in the development and testing of new control tools such as vaccines to account for any possible differential phenotypes. West Africa shows a unique mycobacterial population structure, as it is the only region worldwide where lineages of MAF are endemic (de Jong, Adetifa, *et al.*, 2010).

Work done mainly in the Gambia, suggested that MAF is attenuated compared to MTB (de Jong, Adetifa, *et al.*, 2010; de Jong, Antonio, *et al.*, 2010). While both transmit equally, the rate of progression to disease was slower in MAF infected contacts. Furthermore, MTBC lineage 6 (also known as MAF West Africa 2) was found to be associated with HIV co-infection and reduced ESAT6 secretion (de Jong *et al.*, 2008; de Jong *et al.*, 2006). Thus, MTBss seems to have a competitive advantage that could lead to a replacement of MAF with the more virulent MTBss. This might be particularly likely due to the large population increases in West African cities (UNICEF, 2014). One could also argue that with the HIV pandemic and other immune suppression diseases, MAF will still be an important pathogen in West Africa. Recent publications from various countries, however, observed an interesting trend: the slow replacement of MAF with MTBss, especially the Cameroun

University of Ghana http://ugspace.ug.edu.gh

sub-lineage of lineage 4. This phenomenon was first described in Guinea-Bissau, where lineage 6 decreased from 51% to 39% in about 2 decades (Kallenius *et al.*, 1999). Declines in prevalence of the other *M. africanum* lineage have also been observed in Côte d'Ivoire, and Cameroon (Dosso *et al.*, 1999; Huet *et al.*, 1971; Koro Koro *et al.*, 2013; Niobe-Eyangoh *et al.*, 2003).

At the same time, understanding of the genetic population structure of circulating MTBC strains is increasingly becoming important for TB control. Current genomic studies have revealed that substantial strain genetic diversity exists among the different members and genotypes of MTBC, which may have implications for the development and deployment of new TB vaccines and diagnostics (Hershberg *et al.*, 2008). In this study, we analyzed the distribution of MTBC lineages and sub-lineages in Ghana, a country harboring six of the seven identified MTBC (Asante-Poku *et al.*, 2015; Yeboah-Manu *et al.*, 2011) over an 8-year period. Our findings indicated a fairly constant distribution of the two main MTBC species and lineages over time. In addition, we observed clustering of some MTBC lineages at specific geographical locations.

University of Ghana http://ugspace.ug.edu.gh

3.3 Materials and Methods

3.3.1 Ethics Statement

The Scientific and Technical Committee and then the Institutional Review Board (IRB) of the Noguchi Memorial Institute for Medical Research with a federal wide assurance number FWA00001824 reviewed the protocols and procedures for this study and approved them. Written informed consent was obtained from participants using a designed form which was approved by the IRB. Methods for sputum sampling conformed to WHO guidelines (two sputa per patient) and patients' identity was protected.

3.3.2 Study Locale and Participants Data

The study was conducted from July 2007 to December 2014 in Ghana, involving sputum smear positive TB cases. From July 2007 to December 2011, patients were recruited from five health facilities; Korle-Bu Teaching Hospital (KBTH) in the Greater Accra region, Agona Swedru Government Municipal Hospital (ASH), Winneba Government Hospital (WGH), St. Gregory Catholic Clinic from the Central Region, all in the southern section of Ghana (Figure 3.1). Between 2012 and 2014, based on an on-going prospective study; sputum was collected from suspected TB cases reporting to the selected health facilities in the Accra Metropolitan Area (AMA) and 2 districts (Mamprusi East (MamE) and Tamale Metropolis (TamM)) in the northern region of Ghana (Figure 3.1) after informed consent. The study sampling sites span 13 administrative districts with a combined population of 4,024,810 (GSS, 2012, 2013a, 2013b, 2013c) in three regions according to the current administrative district division status created in 2013 (Table S3.1). The AMA administrative district is made up of 10 sub-districts; Ablekuma South, Ablekuma North, Ablekuma Central, Ashiedu Keteke, Okai Koi South, Okai Koi North, Osu Klotey,

University of Ghana http://ugspace.ug.edu.gh

Ayawaso East, Ayawaso Central and Ayawaso West Wogon with a combined population of 1,665,086 according to the 2010 population and housing census conducted in Ghana (GSS, 2012, 2014). The ten sub-districts within the AMA for the purposes of this study were merged into 5 sub-districts (Figure 3.1), based on the geographical demarcation existing as at 2007. These 5 sub-districts were: Ablekuma (Able), Ashiedu Keteke (AshK), Ayawaso (Ayaw), Okaikoi (Okai) and Osu Klottey (OsuK). The AMA covers a total land area of 136.674 square kilometres. Kpeshi, a former administrative sub-district of AMA is located on the eastern boundary of AMA and has been currently broken down into two districts; La Dade Kotopon Municipal and Ledzokuku/Krowor Municipal. We included the former Kpeshie demarcation in all targeted analysis involving AMA as TB patients still access facilities within AMA. Sites in the northern region, TamM and MamE, covers a land area of 790.5 and 1,823.6 square kilometres respectively with a combined population of 492,360 (TamM: 371,351 and MamE: 121,009). Together, TamM and MamE constitute 19.9% of the total population in the northern region of Ghana (GSS, 2013c).

Information on age, sex, nationality, ethnicity, employment status, previous history of TB, crowding, substance abuse and duration of symptoms were obtained from the patients with a structured questionnaire.

University of Ghana http://ugspace.ug.edu.gh

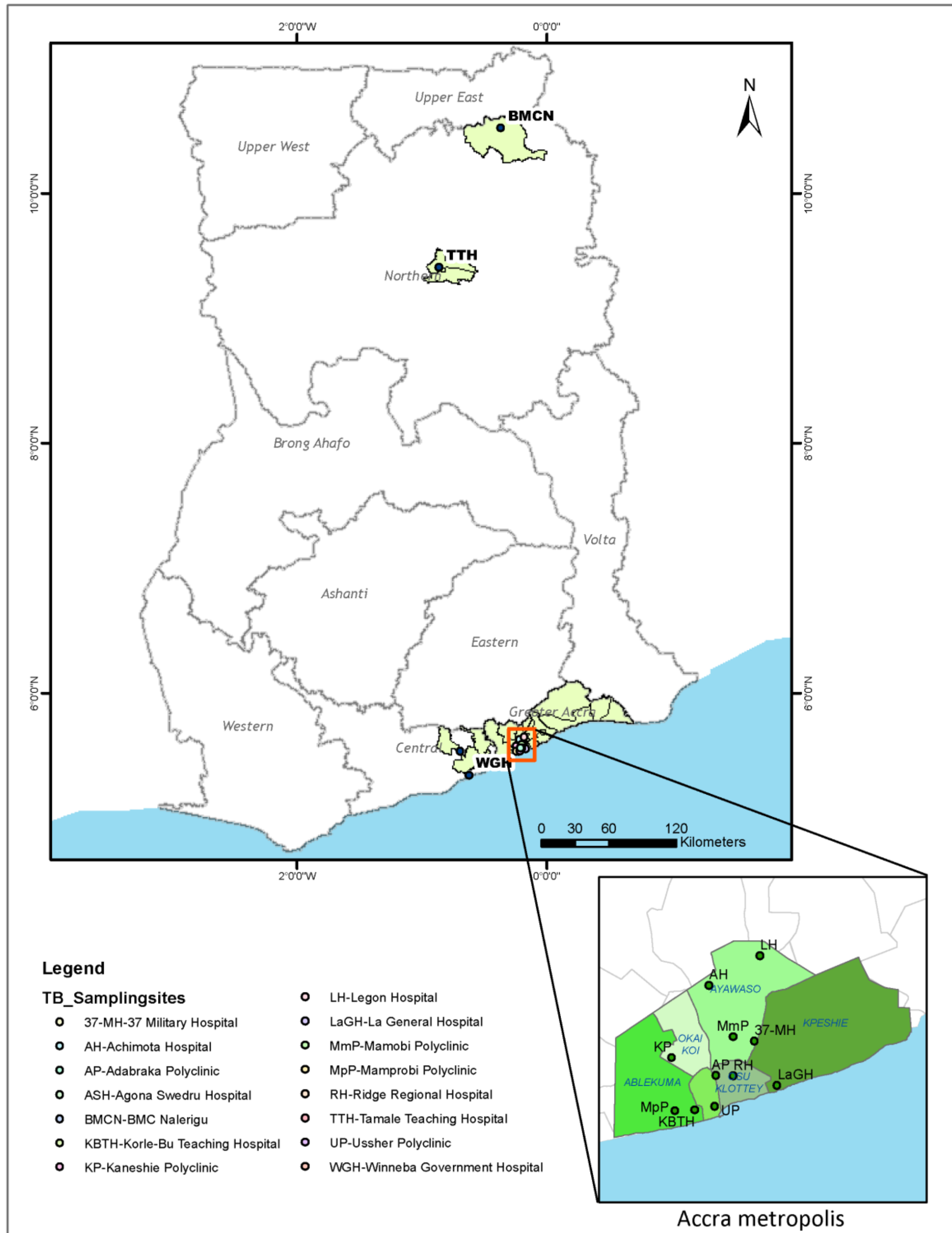


Figure 3.1. Map of sampling sites and study area.

University of Ghana http://ugspace.ug.edu.gh

Sputum samples were obtained from fifteen sampling sites (health facilities) all located within three regions in Ghana; Greater Accra, Central and Northern regions. During the period September 2012 – December 2014, samples were obtained mainly from all the 13 diagnostic centres within the Accra metropolis, (serving more than 46% of the Greater Accra region populace) and the two health facilities located in the northern region (Tamale Teaching Hospital and BMC Nalerigu). The ArcMap program in ArcGIS v. 10.0 was used to create the map.

3.3.3 Mycobacterial Isolation, Species and Lineage Classification

Mycobacterial species were isolated by decontaminating sputum samples with equal amount of 5% oxalic acid solution, then inoculation on Lowenstein Jensen media and incubated as previously described (Yeboah-Manu *et al.*, 2004). Members of the MTBC were confirmed by PCR detection of the insertion sequence IS*6110* and rpoβ as previously described (Yeboah-Manu *et al.*, 2001). Classification into the main phylogenetic lineages was achieved by large sequence polymorphism typing assay identifying regions of difference (RD) 4, 9, 12, 702, 711 (Brosch *et al.*, 2002; de Jong, Antonio*, et al.*, 2010; Gagneux & Small, 2007) and also by spoligotyping following manufacturer's directions (Isogen Bioscience BV Maarssen, The Netherlands).

3.3.4 Data Management and Analysis

Data obtained using the structured questionnaire was double entered using Microsoft Access and validated to correct entry errors. The questionnaire data primarily provided us with the year of diagnosis and residential address (location) of each TB case for the spatio-temporal analysis. In addition to these data, other demographic and clinical characteristics of each participant as indicated above were generated. The association of specific lineages

University of Ghana http://ugspace.ug.edu.gh

and/or sub-lineages of the MTBC with time and/or geographical locations were explored with Fishers exact test using the Stata statistical package version 14.1 (Stata Corp., College Station, TX, USA). All analyses were run with significance level pegged at $p < 0.05$.

To determine the TB case notification rates for the period 2012 – 2014, we obtained the projected population of the individual districts using the exponential growth rate formula; $\mathbf{P_t = P_o e^{rt}}$ (based on the assumption of constant population growth similar to compounded interest) (Shryock *et al.*, 1980). Where; P_t = projected population, P_o = initial population, e = base of the natural logarithm, r = intercensal growth rate and t = time elapsed after last census. The intercensal growth rates (Table S3.1) used for the various regions were obtained from the Ghana statistical service 2010 population and housing census data (GSS, 2012).

The GIS co-ordinates of the participants' self-reported district of residency was used to construct a pictorial plot of the distribution of the MTBC lineages analyzed using the ArcMap (Economic and Social Research Institute, version 10.0) (ESRI, 2010). The district allocation data generated was linked to a molecular data of all TB isolates and was used for TB lineage clustering analysis.

3.3.5 Spatial and Space-time Analysis

Kulldorff's scan statistics (SaTScan™ 9.4.2) tool (Kulldorff, 2015), a commonly used tool for spatial and space-time cluster analysis for diseases in a wide variety of settings (Dangisso *et al.*, 2015; Kammerer *et al.*, 2013; Kulldorff & Nagarwalla, 1995; Yakam *et al.*, 2014; Zhao *et al.*, 2013) was used for analysis of spatio-temporal clustering of TB cases using data obtained only within the time period; September 2012 to December 2014. TamM was excluded from analysis where 2012 data was used since we recorded no TB

University of Ghana http://ugspace.ug.edu.gh

case in 2012. The Kulldorff's scan statistics tool was used to detect significant MTBC clusters using the Monte Carlo simulations (Kulldorff, 1997). Three input files (cases, population and coordinates) were built using excel and saved in the required format for upload into the SaTScan software. The discrete Poisson model was used for the analysis with the assumption that the number of cases at each district had Poisson distribution with a known population at risk (Kulldorff, 1997). All other parameters were set at default for both spatial and space-time analysis (Table S3.4). The results of the analyses were tabulated to add statistical significance to the inferences made using ArcMap.

3.3.6 Normalization of TB cases for within district comparison

To analyse the spatial and space-time distribution of MTBC cases at the district/sub-district level, we normalized the relative case frequencies against their respective reference population obtained from the Ghana Statistical Service (GSS, 2012, 2013b, 2013c, 2014) Also, records of specific genotypes (or sub-lineages) were normalized using all recorded cases of the specified genotype (or lineage) within the specified time. For example, all Ghana and Cameroon sub-lineages per district were normalized using all Lineage 4 cases as the denominator.

3.4 Results

3.4.1 Characteristics of Patients Presenting with Tuberculosis

Sputum smear positive patients from whom MTBC strains were isolated were 2551/3110 (82.0%) cases, comprising 70% (1789/2551) males and 30% (762/2551) females. Participants' age ranged between 2 to 91 and a median age of 39 years. Ninety-one point

University of Ghana http://ugspace.ug.edu.gh

seven percent (2339/2551) of the patients were from Southern Ghana and the remaining 8.3% (212/2551) from Northern Ghana (Figure 3.2). The HIV status of 1613 patients was indicated, of which 15.5% (250/1613) were HIV positive. The additional demographic and clinical characteristics of the cases are indicated in Table 3.1.

University of Ghana http://ugspace.ug.edu.gh

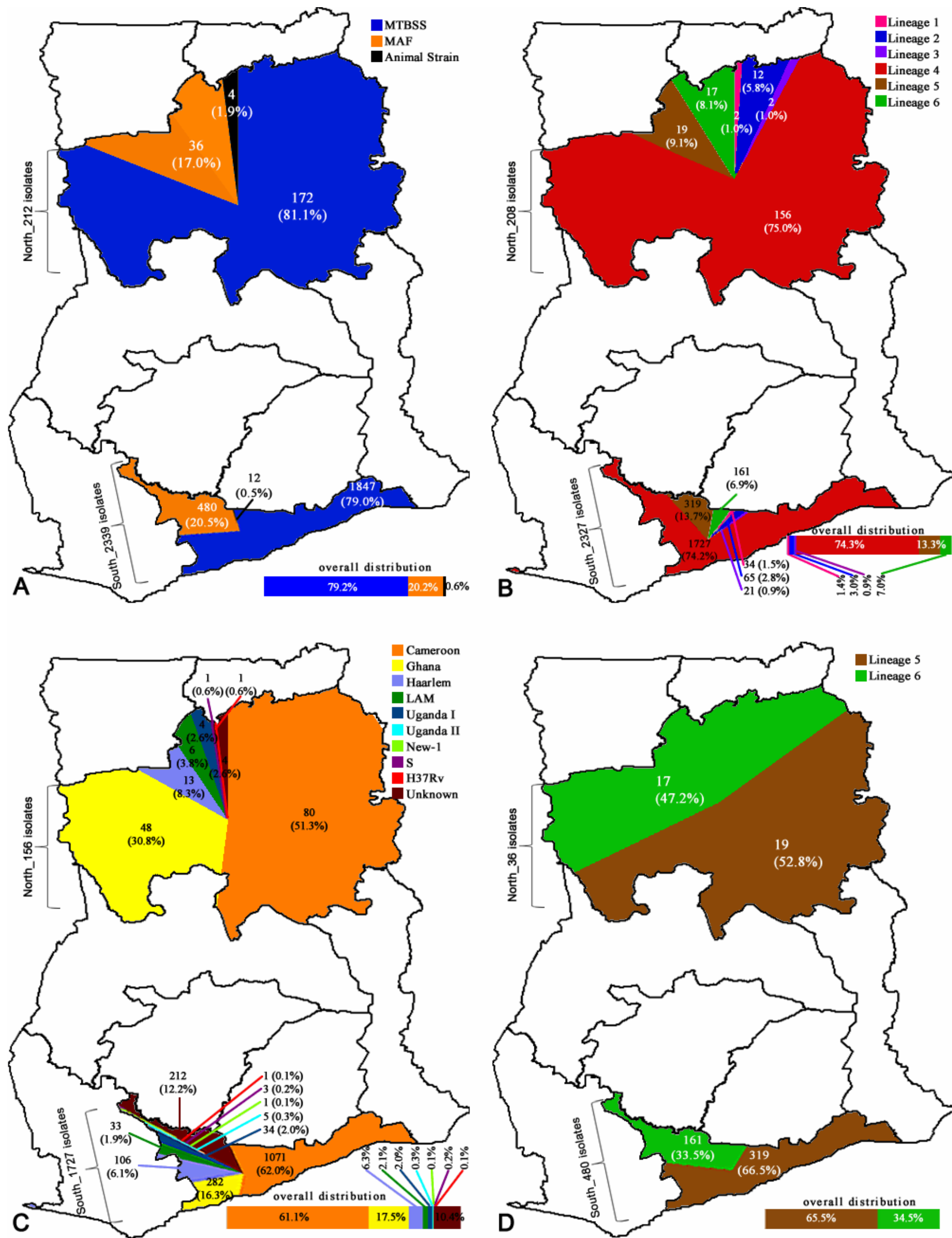


Figure 3.2. Spatial distribution and prevalence of identified *Mycobacterium tuberculosis* lineages.

University of Ghana http://ugspace.ug.edu.gh

Diagram shows the spatial distribution of (A) 2551 *Mycobacterium tuberculosis* complex (MTBC) strains; (B) 2535 human adapted MTBC; (C) regional prevalence of 1883 Lineage 4 sub lineages; (D) regional prevalence of 516 MAF isolates from the geographical regions served by the health facility where sampling was carried out in Ghana. Animal strains were found to be associated with the North (p=0.0389). Similarly, Lineage 2 was found to be associated with the North (p=0.0006). The most dominant Lineage 4 sub-lineage in the North is Ghana (p=0.0000) whereas in the South is Cameroon, even though the association is not statistically significant. The unknown sub-type of Lineage 4 is associated with South (p=0.0001).

Table 3.1. Demographic and clinical data of 2551 TB cases

Variable (Total number analyzed)	**Number (Percentage)**
***Sex* (2551)**	
Male	1789 (70.1)
Female	762 (29.9)
***Age category* (2551)**	
08-25	459 (18.0)
26-40	1082 (42.4)
41-77	980 (38.4)
>77	30 (1.2)
***Residency* (2551)**	
North	212 (8.3)
South	2339 (91.7)
***Occupation* (2530)**	
Skilled	490 (19.2)
Unskilled	1890 (74.1)
Unemployed	150 (5.9)
***Settlement* (2550)**	
Urban	2283 (89.5)
Rural	267 (10.5)
HIV status[1] (1613)	
Yes	250 (15.5)
No	1363 (84.5)
Presence of BCG scar [2](1817)	
Yes	904 (49.8)
No	913 (50.2)

University of Ghana http://ugspace.ug.edu.gh

Table 3.1continued

Variable (Total number analyzed)	Number (Percentage)
Income* (2551)	
None	771 (30.2)
Low	1500 (58.8)
High	280 (11.0)
Drinking status[3] (2071)	
Yes	556 (26.8)
No	1515 (73.2)
TB in the past[4] (2029)	
Yes	218 (10.7)
No	1811 (89.3)
Education Level (2551)	
None	300 (11.7)
Primary	605 (23.7)
Secondary	1551 (60.7)
Tertiary	95 (3.7)
Smear Grade (2551)	
Scanty	235 (9.2)
1	985 (38.6)
2	531 (20.8)
3	800 (31.4)
In Household number (2551)	
<5	589 (23.1)
>5	1962 (76.9)
Ethnicity (2551)	
Akan	800 (31.4)
Ewe	339 (13.3)
Ga	595 (23.3)
Mole -Dagbon	36 (1.4)
Gruma	5 (0.2)
Guan	9 (0.4)
Others	767 (30.0)
Marital status (2551)	
Single	859 (33.7)
Married	1037 (40.7)
Divorced	221 (8.7)
Widowed	104 (4.1)
Co habiting	330 (12.9)
Cough (2551)	
< 2 weeks	2117 (82.9)
> 2 weeks	122 (4.8)
Symptoms other than cough	312 (12.2)
Night Sweat [5] (2219)	
Yes	1372 (61.8)
No	847 (33.2)

University of Ghana http://ugspace.ug.edu.gh

Table 3.1continued

Variable (Total number analyzed)	**Number (Percentage)**
Hemoptysis [6] (2227)	
Yes	480 (21.6)
No	1747 (78.4)
House type (2551)	
Self-contained	486 (19.1)
Compound House	1707 (66.9)
Others	358 (14.0)
Swollen glands [7] (2211)	
Yes	210 (9.5)
No	2001 (91.5)
Chest Pain [8] (2234)	
Yes	1717 (76.9)
No	517 (23.1)
Nationality [9] (2251)	
Ghana	2187 (97.2)
Nigeria	21 (0.9)
Togo	12 (0.5)
Niger	11 (0.4)
Ivory Coast	7 (0.3)
Others West African Nationals	13 (0.7)
Contact with TB patient (2551)	
Yes	316 (12.4)
No	2234 (87.6)
Smoking [10](2419)	
Yes	500 (20.7)
No	1919 (75.3)

1 = 938 missed data for HIV status,

2 = 734 missed data for Presence of BCG scar,

3 = 480 missed data for Drinking status,

4 = 522 missed data for TB in the past,

5 = 332 missed data for Night Sweat,

6 = 324 missed data for Night Sweat,

7 = 340 missed data for Swollen glands,

8 = 317 missed data for Chest Pain,

9 = 300 missed data for Nationality,

10 =132 missed data for Smoking.

* Income below 1,000GH₵ was defined as low whilst those above 1,000GH₵ as high

University of Ghana http://ugspace.ug.edu.gh

3.4.2 The Population Structure of MTBC causing pulmonary TB in Ghana

Two thousand six hundred and three mycobacterial isolates were obtained from 3110 samples giving a cumulative isolation rate of 83.7%. We identified 2551 of the isolates as members of the MTBC and 52 as non-tuberculous mycobacteria (NTM) (as well as those with negative mycobacteria) which were excluded from further analysis. Among those confirmed as MTBC, 2019 (79.1%) were MTBss, 516 (20.2%) were MAF, and 16 (0.6%) animal strains (15 *M. bovis* (SIT 1037, 482*)* and 1 *M. caprae*) (Figure 3.2A). Six of the seven lineages of the human adapted MTBC (MAF and MTBss) were identified in the following proportions: L1 (36; 1.4%), L2 (77; 3.0%), L3 (23; 0.9%), L4 (1883; 74.3%), L5 (338; 13.3%) and L6 (178; 7.0%), respectively (Figure 3.2B). The sub-lineages identified within the L4 were the Cameroon (1151; 61.1%) followed by the Ghana (330; 17.5%), then Haarlem (119; 6.3%), LAM (39; 2.1%), Uganda I (38; 2.0%), Uganda II (5; 0.3%), New-1 (1; 0.1%), S (3; 0.2%) and H37Rv-like (2; 0.1%) (Figure. 2C).

3.4.3 Spatial Distribution of MTBC Genotypes

The combined number of isolates analyzed from the different geographical areas, identified species, lineages and sub-lineages are indicated in Figure 3.2 A-D, respectively. As shown in Figure 3.2D, there was no statistical difference in the MAF proportion between the north (17.0%; 36/212) and the south (21.9%; 378/1726) ($p=0.1099$). However, we found the proportion of animal-adapted species (MTBC other than MTBss and MAF) in the north (1.9%; 4/212) to be more than twice the proportion in the south (0.7%; 12/1726) ($p<0.0884$; OR=2.74). There was unequal spatial distribution of L4 sub-lineages and spoligotypes. The proportion of the Ghana sub-lineage was statistically higher in Northern Ghana (32.3%) compared to 20.1% in the south ($p=0.0016$, OR=1.9, 95%CI=1.3-2.9).

University of Ghana http://ugspace.ug.edu.gh

Whereas the Spoligotype international type (SIT) 61 was more likely to be found in the south (p=0.0330; OR=0.7; 95% CI=0.4-0.9), the SIT 53 was more likely to be found in the north (p=0.0015; OR=2.0; 95% CI=1.3-3.1). In addition, L2 was proportionally higher in the north (5.7%; 12/212) compared to the south (3.1%; 53/1726).

Even though the sample size changed over time due to increase in case study sites, the proportion of distinct species did not change over time (Figure 3.3). The species/lineages/sub-lineages distributions of MTBC within the 13 administrative districts where participants resided are displayed in Figure 3.4.

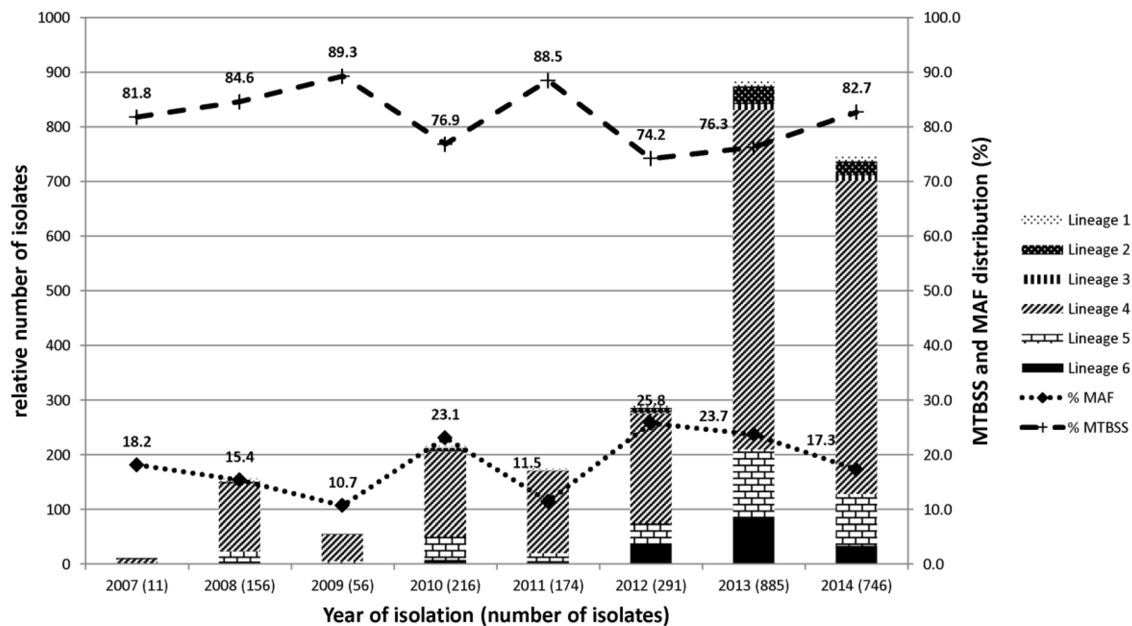


Figure 3.3. Temporal distribution and prevalence of human adapted *Mycobacterium tuberculosis* complex (MTBC). Figure displays a stacked graph showing the temporal distribution of human adapted MTBC (left y-axis) and a linear graph showing the prevalence of *Mycobacterium tuberculosis* sensu stricto (MTBss) and *Mycobacterium africanum* (MAF) (right y-axis) over the entire 8-year study period.

University of Ghana http://ugspace.ug.edu.gh

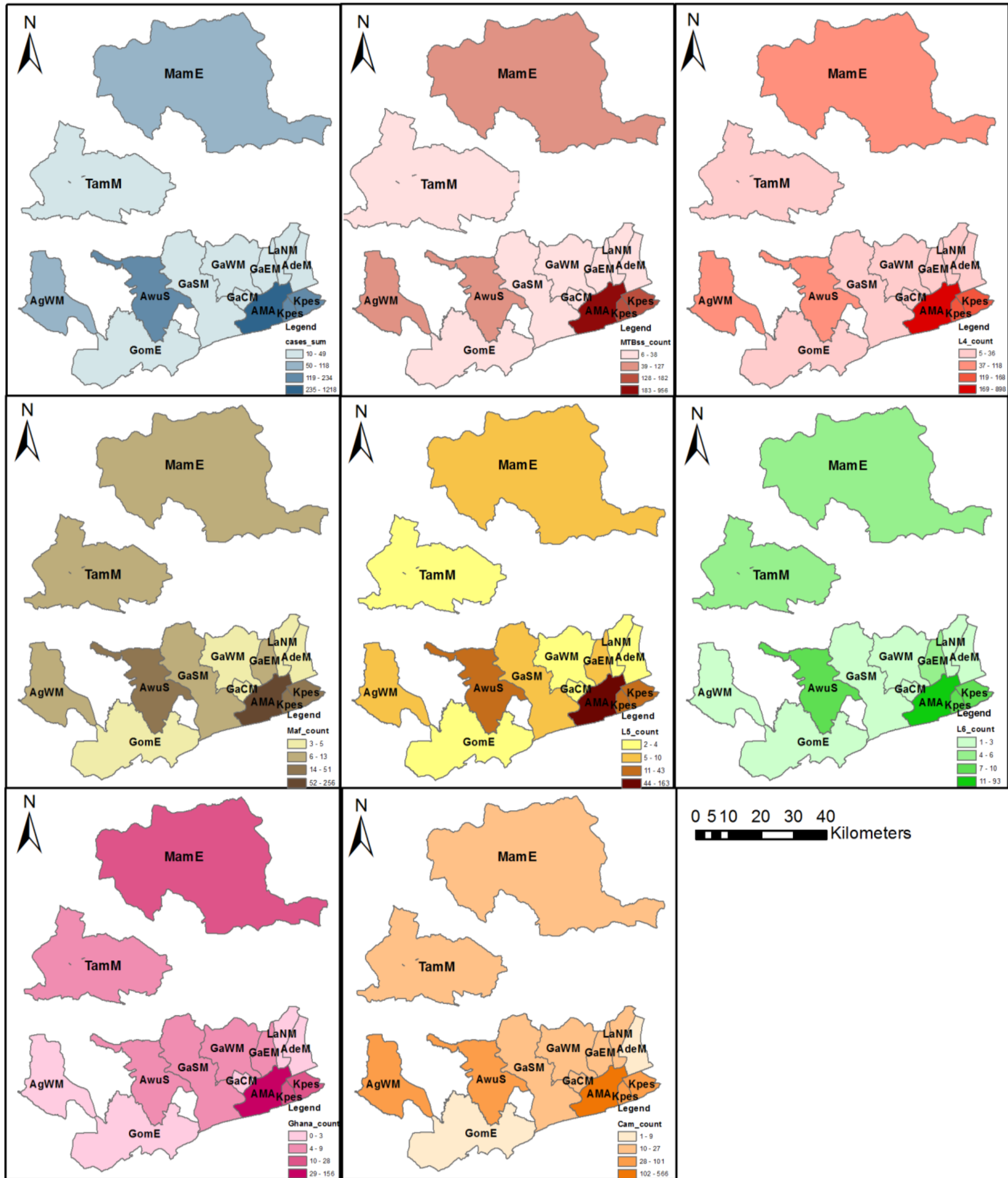


Figure 3.4. Spatial distribution of human adapted MTBC lineages and major sub-lineages within the eight-year study period.

University of Ghana http://ugspace.ug.edu.gh

The Figure shows the distribution of MTBC species/lineages/sub-lineages within the 13 districts where participants resided. The blue colored panel shows the distribution of all the tuberculosis cases recruited with well-defined residential status. The red, brown and green colored panels show the distribution of lineage 4, lineage 5 and lineage 6 respectively. All other sub-lineages/species have been indicated in the respective legends. This Figure was created using the ArcMap program in ArcGIS v. 10.0. Abbreviations: MTBC, *Mycobacterium tuberculosis* complex; MTBss, *Mycobacterium tuberculosis* sensu stricto; MAF, *Mycobacterium africanum*; L4, Lineage 4; L5, Lineage 5; L6, Lineage 6; Ghana, Ghana genotypes (Ghana sub-lineage); Cam, Cameroon sub-lineage; MamE, Mamprusi East district; TamM, Tamale Metropolis; AgWM, Agona West Municipal; GomE, Gomoa East; AwuS, Awutu Senya; GaSM, Ga South Municipal; GaWM, Ga West Municipal; GaCM, Ga Central Municipal; GaEM, Ga East Municipal; AMA, Accra Metropolis; LaNM, La-Nkwantanang Madina Municipal; AdeM, Adenta Municipal; Kpes, Kpeshie Municipal.

University of Ghana http://ugspace.ug.edu.gh

3.4.4 Spatial and space-time clustering analysis of MTBC cases at the district/sub-district level (2012–2014)

Spatial and space-time analyses were carried out for districts where sampling was performed within the time period; September 2012 to December 2014. These districts were AMA (sub-divided into sub-districts due to the population density) in the south and in the north, MamE. The TB case notification rate ranged from 3 to 52 cases/100,000 individuals at risk within the districts/sub-district analyzed (Figure. 3.5A) with the highest case notification rate occurring in 2013 (52 cases/100,000). In a purely spatial analysis, we found two significant clusters within the study period based on cases notified. The most likely cluster consisted of two sub-districts, AshK and OsuK ($p = 0.0000$, $RR = 3.99$) with a secondary cluster occurring at MamE, ($p = 0.0000$, $RR = 2.16$) (Table 3.2). Similar observations were made using a space-time analysis with the likely clusters occurring in 2013. To analyze the spatial and space-time distribution of the two human adapted MTBC (MTBss and MAF), we normalized the relative district case frequencies to that for all cases obtained per district/year (Figure 3.5B and 3.5C). We found that the normalized distribution of both MTBss and MAF fluctuated over the three-year period, and no particular district/sub-district showed constant high values (Figure 3.5B/3.5C). In a purely spatial analysis, significant clusters ($p=0.0000$, Table 3.2) were observed in 6 of the 8 districts (MamE, TamM, Ayaw, Okai, Kpes and OsuK).

University of Ghana http://ugspace.ug.edu.gh

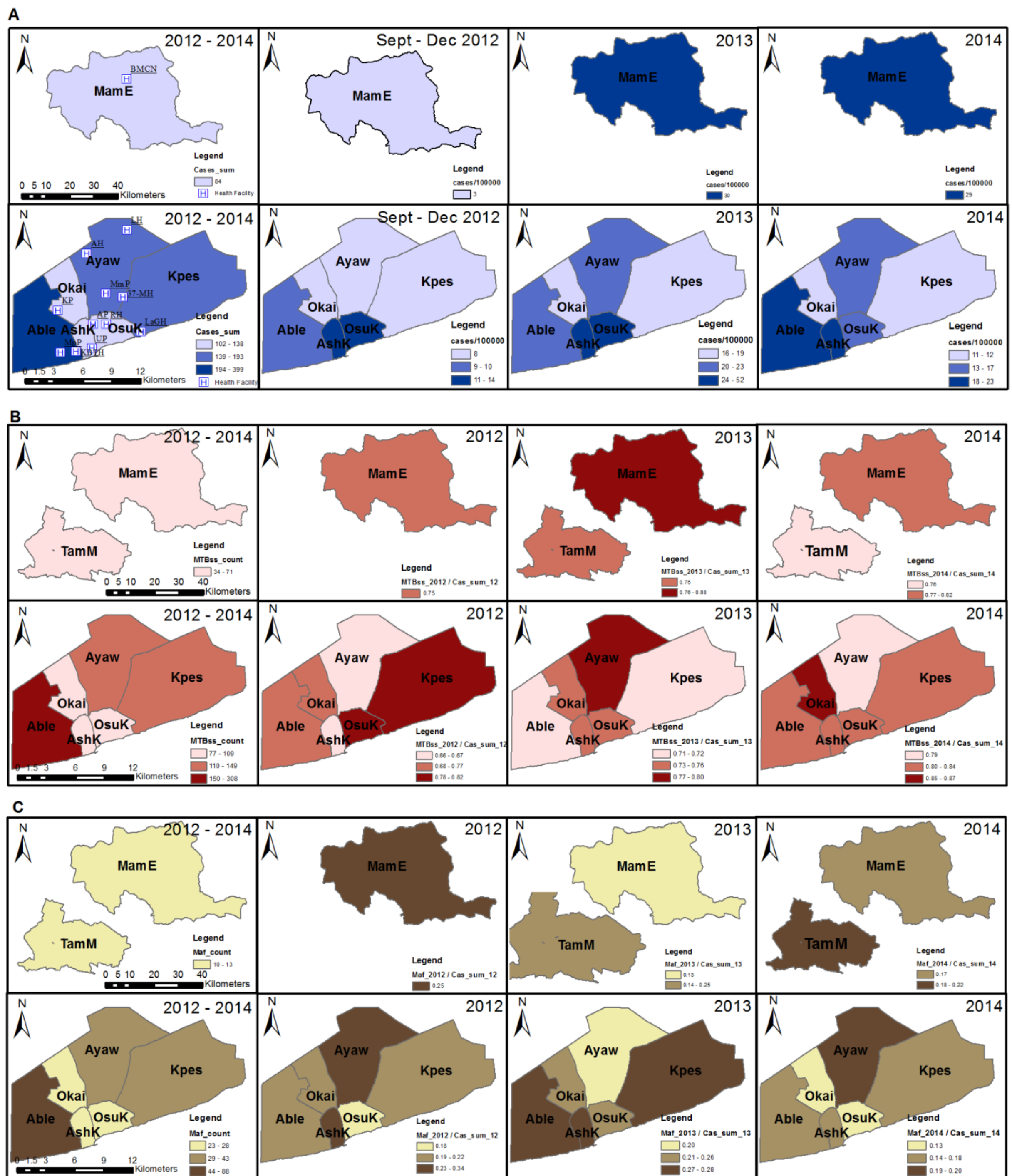


Figure 3.5. Spatial distribution of isolated MTBC within selected districts (2012 – 2014). This figure shows the; (A) Sum and case notification rate of all TB cases from September 2012 to December 2014, (B) Sum and normalized distribution of MTBss cases,

University of Ghana http://ugspace.ug.edu.gh

(C) Sum and normalized distribution of MAF. The total number of cases per year was used as the denominator for normalization. Sampling from TamM did not meet our criteria for being included in analyses for case notification rate and so was excluded in all columns of panel A. Likewise we also recorded no cases in 2012 as such TamM was excluded from 2012 analysis (panel B and C). This figure was created using the ArcMap program in ArcGIS v. 10.0. Abbreviations: MTBss, *Mycobacterium tuberculosis* sensu stricto; MAF, *Mycobacterium africanum*; MamE, Mamprusi East district; TamM, Tamale Metropolis; AshK, Ashiedu Keteke; Ayaw, Ayawaso; Able, Ablekuma; OsuK, Osu Klottey; Okai, Okaikoi; Kpes, Kpeshie.

The Ghana sub-lineage was found to significantly cluster in the north ($p = 0.013$, $RR = 2.27$). A space-time analysis revealed two significant clusters for the Cameroon sub-lineage with the most likely cluster occurring in 2014 ($p = 0.000$, $RR = 1.68$) consisting of five districts (Kpes, Okai, Ayaw, TamM and MamE). The second Cameroon sub-lineage cluster involved AshK and OsuK, which occurred in 2013 ($p = 0.003$, $RR = 1.79$, Figure 3.6C). Comparing the North and South for association with some risk factors showed association of the North with rural settings ($P=0.0000$), farming (0.0000), contact with cattle (0.0003), compound housing (0.0000) whereas the South was associated with driving as occupation (0.0163) as shown in supplementary data (Table S3.6).

University of Ghana http://ugspace.ug.edu.gh

Table 3.2. Most likely spatial clusters detected in the study area using SaTScan analysis

TB cases	Reference population	Cluster type	Year (s) of observed cluster	Clustered districts	Observed cases	Expected cases	Log Likelihood ratio	Relative risk	p-value	Type of analysis
All TB cases	District population	Most likely	2012 - 2014	AshK, OsuK	213	61.2	123.9	3.99	0.000	Purely spatial
All TB cases	District population	Secondary	2012 - 2014	TamM, MamE	138	67.9	29.8	2.16	0.000	Purely spatial
All TB cases	District population	Most likely	2013	AshK, OsuK	123	29.7	85.1	4.48	0.000	Space-time
All TB cases	District population	Secondary	2013	TamM, MamE	50	20.5	15.4	2.5	0.000	Space-time
MTBss	All cases per district	Most likely	2012 - 2014	MamE, TamM, Ayaw, Okai, Kpes, OsuK	588	398.1	75.0	2.21	0.000	Purely spatial
MAF	All cases per district	Most likely	2012 - 2014	MamE, TamM, Ayaw, Okai, Kpes, OsuK	158	107.8	18.9	2.12	0.000	Purely spatial
L4	All L4 cases in 2012	Most likely	2012 - 2014	Able	285	279.1	0.1	1.03	1.000	Purely spatial
L4	All L4 cases in 2012	Most likely	2012	Able	53	37.1	3.2	1.46	0.204	Space-time
Gh	All L4 cases in 2014	Most likely	2014	MamE, TamM	23	10.8	5.6	2.27	0.013	Space-time
Cam	All L4 cases in 2014	Most likely	2014	Kpes, Okai, Ayaw, TamM, MamE	122	80.0	11.5	1.68	0.000	Space-time
Cam	All L4 cases in 2013	Secondary	2013	AshK, OsuK	52	30.4	6.8	1.79	0.003	Space-time

This table shows the most likely spatial clusters detected from the SaTScan analysis. The TB case in the first column shows the category of TB lineage/sub-lineage to which the spatial or space-time analysis was performed. The districts to which the clusters were observed are shown in column 5 with the respective year of cluster observation shown in column 4. The last column shows the type of cluster analysis performed. *Abbreviations*: TB, tuberculosis; MTBss, *Mycobacterium tuberculosis* sensu stricto; MAF, *Mycobacterium africanum*; L4, Lineage 4; Gh, Ghana genotypes (Ghana sub-lineage); Cam, Cameroon sub-lineage; MamE, Mamprusi East; TamM, Tamale Metropolis; AshK, Ashiedu Keteke; Ayaw, Ayawaso; Able, Ablekuma; OsuK, Osu Klottey; Okai, Okaikoi; Kpes, Kpeshie.

University of Ghana http://ugspace.ug.edu.gh

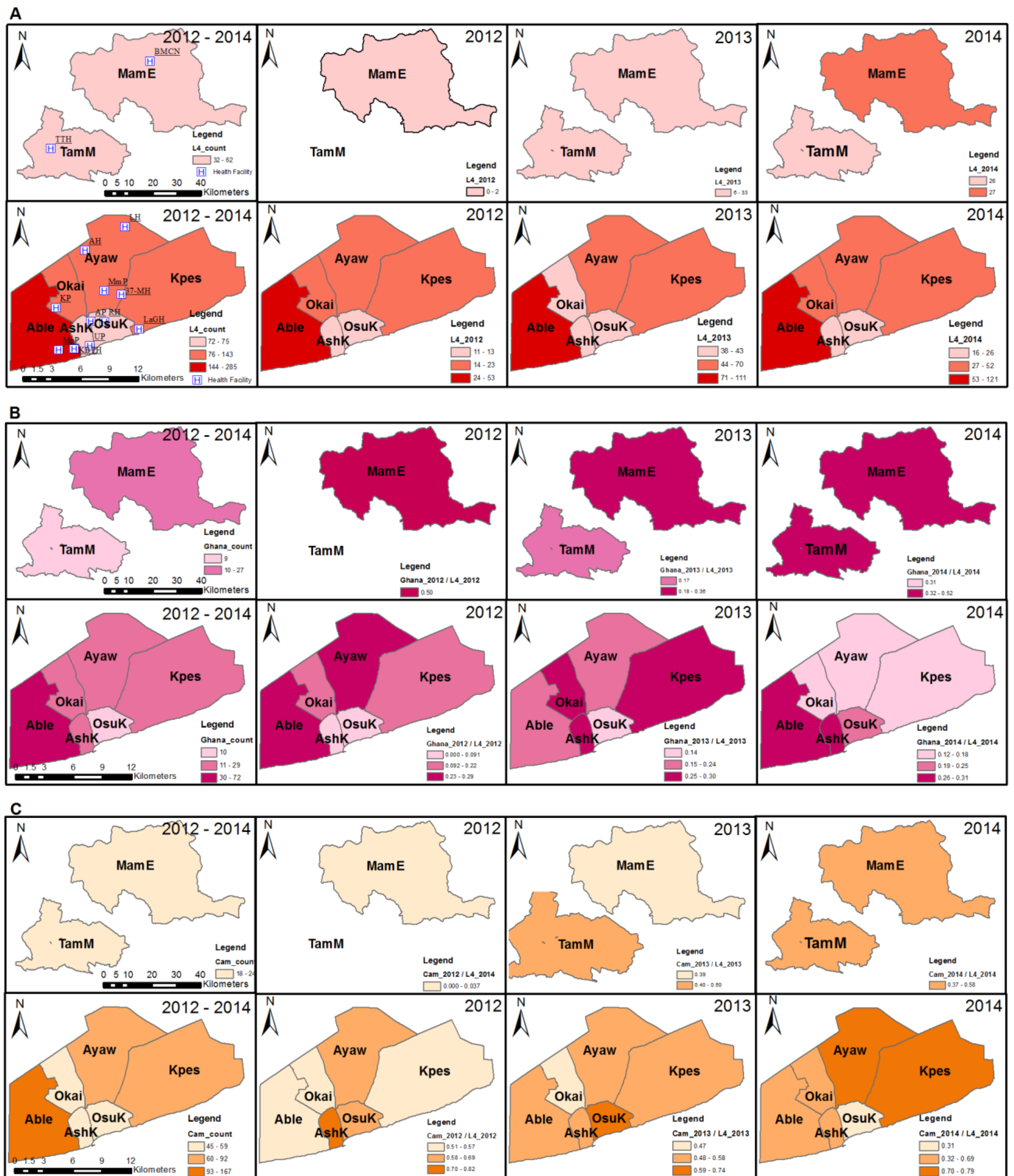


Figure 3.6. Spatial distribution of lineage 4 and major lineage 4 sub-lineages within selected districts (2012 – 2014). The figure shows the (A) Sum and normalized distribution of lineage 4 cases, (B) Sum and normalized distribution of Ghana sub-lineages

University of Ghana http://ugspace.ug.edu.gh

cases, (C) Sum and normalized distribution of Cameroon sub-lineage cases. The total number of cases per year was used as the denominator for normalization. This study recorded no TB cases for TamM in 2012, consequently TamM was excluded from all analysis carried out using 2012 data. This figure was created using the ArcMap program in ArcGIS v. 10.0. Abbreviations: L4, Lineage 4; Ghana, Ghana genotypes (Ghana sub-lineage); Cam, Cameroon sub-lineage; MamE, Mamprusi East district; TamM, Tamale Metropolis; AshK, Ashiedu Keteke; Ayaw, Ayawaso; Able, Ablekuma; OsuK, Osu Klottey; Okai, Okaikoi; Kpes, Kpeshie.

3.5 Discussion and conclusion

Our objective was to analyze in time and space the prevalence of MTBC species and genotypes among isolates obtained from sputum-positive TB cases over an 8-year period in Ghana. A secondary objective was to determine the space clustering of specific genotypes. Our longitudinal analysis indicated that: 1) the Beijing and Ghana genotypes of Lineages 2 and 4, respectively, as well as the animal adapted MTBCs are isolated more often from patients from Northern than Southern Ghana and 2) the proportion of MAF among the isolates over the study period remained fairly constant.

Our first evidence of an association between the north and the Ghana genotype of Lineage 4 as well as the Beijing genotype of Lineage 2 reiterates the phylogeographical nature of the human-adapted MTBCs such that even within a single country there can be variations in the distribution of distinct genotypes within a lineage and specific geographical regions. For example in Senegal, it has been observed that the proportion of *M. africanum* causing TB varies by region (Diop *et al.*, 1976).

University of Ghana http://ugspace.ug.edu.gh

The Ghana genotype clustered in the two districts (MamE and TamM) of the north (Figure 3.6B) whereas the Cameroon sub-lineage clustered in the South as two clusters (Figure 3.6C; Table 3.2). Thus, clustering alone cannot be used as a proxy for active transmission of these genotypes in the specific geographical areas as the resolution of the molecular tool (spoligotyping) used for characterization within this study is not enough to infer on-going transmission because of its low discriminatory power. However, they may be indicative of the areas of origin or introduction of these genotype/spoligotypes into the country. The Cameroon genotype as observed from various molecular studies is the most prevalent genotype causing TB in West Africa and our findings confirm this and show that this genotype may have been introduced into the country through the south. Our findings call for studies to investigate the transmission dynamics of these sub-lineages within the respective geographical areas; this is of public health concern because evidence that MTBC genotypes might influence disease phenotype. For example, an association between Beijing strains (Lineage 2) and drug resistance has often been reported, and we recently showed that the Ghana genotype is also associated with drug resistance in Ghana (manuscript submitted). This means that effective control of TB in the Northern region of Ghana would be challenged with these two genotypes in higher proportions.

We also found the animal-adapted MTBC to be statistically associated with Northern Ghana (p=0.0381, OR=3.73). There were 5 patients among our study population who had direct contact with cattle including 4 butchers and 1 farmer who owns cattle and all 5 were infected with animal strains of the MTBC. This finding supports previous observations that people who are in direct constant contact with cattle and/or their products may be at risk of infection with *M. bovis* (Michel *et al.*, 2010). However, there were some patients from

University of Ghana http://ugspace.ug.edu.gh

whom animal strains were isolated but did not have constant direct contact with cattle or any other farm animal. This finding compares with a similar work done in Mexico where most of the patients from whom animal strains of the MTBC were isolated had no direct contact with livestock but rather had consumed unpasteurized milk products in the past (Laniado-Laborin *et al.*, 2014). Our work design did not include analysis of patients' dietary lifestyle and therefore we have no data on whether those patients are in a similar situation. Nevertheless, this association could be due to the dominance of livestock in the Northern region as compared to the south (GSS, 2012) meaning more possible interaction of humans with animals in the north as compared to the south.

The proportion of MTBss analyzed among our data set remained stable and higher than that of MAF over the eight-year period (Figure 3.2A and Figure 3.3). This result is in contrast with reports from other West African countries, indicating a decline of MAF (Huet *et al.*, 1971; Koro Koro *et al.*, 2013). Various reasons have been suggested to explain the observed decline including non-specificity of biochemical assays used previously. Nevertheless, this study shows that MAF has remained fairly constant over the study period at an average of approximately 20% with significant fluctuations observed within four different periods [(2009/2010, P=0.0422), (2010/2011, P=0.0033), (2011/2012, P=0.0002), (2013/2014, P=0.0014)].

This finding in conjunction with others (Addo *et al.*, 2007; Goyal *et al.*, 1999; Meyer *et al.*, 2008) indicates that the two TB causing pathogens have adapted well within the Ghanaian population. Moreover, two independent studies conducted by our group found a strong association between L5 and an ethnic group in Ghana (Asante-Poku *et al.*, 2016; Asante-Poku *et al.*, 2015).

University of Ghana http://ugspace.ug.edu.gh

CHAPTER 4

4.0 Manuscript 2. Reduced Transmission of *Mycobacterium africanum* Compared to *Mycobacterium tuberculosis* in Urban West Africa.

(Published in International Journal of Infectious Diseases (2018), doi.org/10.1016/j.ijid.2018.05.014)

4.1 Abstract

Objective: Understanding transmission dynamics is useful for tuberculosis (TB) control. We conducted a population-based molecular epidemiological study to understand TB transmission in Ghana.

Methods: *Mycobacterium tuberculosis* complex (MTBC) isolates obtained from prospectively-sampled pulmonary TB patients between July, 2012 and December, 2015 were characterized using spoligotyping and standard 15-loci MIRU-VNTR typing for transmission studies.

Results: Out of 2,309 MTBC isolates, we identified 1,082 (46.9%) unique cases with 1,227 (53.1%) isolates belonging to one of 276 clusters. Recent TB transmission rate was estimated to be 41.2%. Whereas TB strains of lineage-4 belonging to *M. tuberculosis* showed a high recent transmission rate (44.9%), we found reduced recent transmission rates for lineages of *Mycobacterium africanum* (lineage-5 (31.8%); lineage-6 (24.7%)).

Conclusion: Our findings indicate high recent TB transmission suggesting occurrences of unsuspected outbreaks in Ghana. The observed reduced transmission rate of *M. africanum* suggests other factor(s) (host/environmental) may be responsible for its continuous presence in West Africa.

University of Ghana http://ugspace.ug.edu.gh

4.2 Introduction

Tuberculosis (TB) is a global health emergency; in 2016 an estimated 10.4 million people got sick, while 1.7 million died of TB (WHO, 2017). The World Health Organization (WHO) in 1993 declared TB a global health emergency and called for more efforts and resources to fight TB. Due to the largely inefficacy of the bacillus Calmette–Guérin (BCG) vaccine against pulmonary TB in adults, the current TB control strategy relies on case detection and treatment by the Directly Observed Therapy short course (DOTs) strategy. The conventional indicators used for assessing the national control programs under this strategy focus on the proportion of cases that are cured at the end of treatment or whose sputum microscopy becomes negative after the first 2 months of treatment. Such indicators ignore equally important aspects of TB control, which includes the duration of infectivity, the frequency of reactivation, and the risk of progression among the infected contacts and the proportion of TB due to recent transmission.

Understanding transmission dynamics will contribute to knowledge on factors that enhance the spread of the disease, which is useful for developing preventive interventions. Molecular epidemiological studies have been very useful in a number of countries by identifying populations at risk, areas of high transmission as well as providing much understanding on the prevalence of different *Mycobacterium tuberculosis* complex (MTBC) strains with varied virulence and drug resistance rates (Anderson *et al.*, 2014; Malm *et al.*, 2017; Seto *et al.*, 2017; Varghese *et al.*, 2013; Walker *et al.*, 2014; Yang *et al.*, 2017). These studies have shown that the dynamics of TB transmission vary greatly geographically. Even though Africa harbors a large proportion of the global TB cases with current incidence at 254 per 100,000 population (WHO, 2017), population-based

University of Ghana http://ugspace.ug.edu.gh

molecular epidemiological studies needed to understand transmission patterns are rare. The few studies conducted were not population-based and lacked in-depth analysis of the transmission dynamics of MTBC strains belonging to different lineages (Asante-Poku *et al.*, 2016; Glynn *et al.*, 2010; Mulenga *et al.*, 2010).

The molecular typing tools, spacer oligonucleotide typing (spoligotyping) and mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR) typing, have been successfully used for strain differentiation in TB transmission studies due to their combined high discriminatory power and reproducibility, and in combination with epidemiological data, have been remarkably used for the detection of recent TB transmission and outbreaks (Anderson *et al.*, 2014; Barnes & Cave, 2003; Maguire *et al.*, 2002; Surie *et al.*, 2017; Varghese *et al.*, 2013). Currently, due to the high cost and expertise needed for employing whole genome sequencing and analysis, population-based studies may not readily employ its use and considering capacity building in a low resource setting like Ghana, spoligotyping and MIRU-VNTR typing remain good alternatives.

TB in humans is caused mainly by *Mycobacterium tuberculosis* sensu stricto (MTBss) and *M. africanum* (MAF) which are further divided into seven lineages [MTBss: lineage 1-4 and 7 (L1-L4 and L7), MAF: lineages 5 and 6 (L5 and L6)] (Blouin *et al.*, 2012; de Jong, Antonio, *et al.*, 2010); while MTBss is globally distributed MAF is restricted to West Africa where it is responsible for up to 50% of TB cases (Gagneux & Small, 2007). Nevertheless, reports mainly from the Gambia where L6 is prevalent suggest MAF is attenuated compared to MTBss, hence could be outcompeted by MTBss (de Jong, Antonio, *et al.*, 2010; de Jong *et al.*, 2008; Kallenius *et al.*, 1999). However, an 8-year study

University of Ghana http://ugspace.ug.edu.gh

recently conducted in Ghana found the prevalence of MAF to be fairly constant at approximately 20%, indicating that MAF may transmit equally as MTBss (Yeboah-Manu *et al.*, 2016). The objective of this study was to determine the transmission dynamics of TB caused by MTBss and MAF in Ghana.

4.3 Methods

4.3.1 Study design and population

This study was a population-based prospective study in which sputum samples were collected from consecutive clinically diagnosed pulmonary TB patients reporting to 12 selected health facilities within an urban setting [Accra Metropolitan Area (AMA)] and a rural setting East Mamprusi district (MamE) (Figure S4.1). The study was conducted from July, 2012 to December, 2015. A pulmonary TB case was defined as an individual with both clinical and bacteriologically confirmed case of TB. Detailed demographic and epidemiological data were obtained from consented participants.

4.3.2 Mycobacterial isolation, species identification and drug susceptibility testing

The sputum samples were decontaminated and cultured on Lowenstein-Jensen media to obtain mycobacterial isolates which were confirmed as MTBC by detecting the MTBC-specific insertion sequence IS*6110* using PCR (Yeboah-Manu *et al.*, 2001). *In vitro* drug susceptibility to isoniazid and rifampicin were determined using either the microplate alamar blue cell viability assay as described elsewhere (Otchere *et al.*, 2016), and/or the Geno-type MTBDR*plus* (Hain lifescience), according to the manufacturer's protocol (Barnard *et al.*, 2008).

University of Ghana http://ugspace.ug.edu.gh

4.3.3 Lineage and strain classification

Lineage and strain classification of the MTBC was achieved in a step wise manner using large sequence polymorphism typing identifying regions of difference 4, 9, 12, 702, and 711 (de Jong, Antonio*, et al.*, 2010; Gagneux & Small, 2007), single nucleotide polymorphism typing, spoligotyping (Kamerbeek *et al.*, 1997), and MIRU-VNTR typing (Supply *et al.*, 2006). For MIRU-VNTR typing, we first used a customized set of 8-MIRU loci as described by Asante-Poku *et al*. (Asante-Poku, Nyaho*, et al.*, 2014) and resolved clustered cases by analyzing the remaining 7 loci of the standard MIRU-15 loci set (Supply *et al.*, 2006). All assays were well controlled with PCR amplifications and pre-PCR procedures conducted in physically separated compartments to avoid laboratory cross contamination. The presence of more than one allelic repeat number (multiple allele) for any given locus is suggestive of laboratory cross contamination, multiple strain infection or microevolution of a single strain. To prevent bias resulting from cross contamination and multiple strain infection, isolates with multiple allele at more than one MIRU loci (described as *untypeable*) were excluded from further analysis. Isolates with only one multiple allele at any given locus were however included due to the possibility of microevolution.

The spoligotyping patterns and assigned shared type numbers obtained were defined according to SITVITWEB database (http://www.pasteur-guadeloupe.fr:8081/SITVIT_ONLINE/) while sub-lineages were assigned based on the MIRU-VNTRplus database (http://www.miru-vntrplus.org) (Allix-Beguec, Harmsen*, et al.*, 2008). Strains with no lineage nomenclature data were further identified using the TB-lineage database (Shabbeer *et al.*, 2012) or otherwise regarded as orphan strains. A strain was

University of Ghana http://ugspace.ug.edu.gh

defined as an MTBC isolate with a unique molecular signature thus, a unique spoligotype pattern and/or a unique MIRU-VNTR allelic pattern for the number of investigated MIRU loci.

4.3.4 Clustering analysis and risk factor assessment

Clustering analysis was performed using the categorical parameter and the UPGMA coefficient from a constructed phylogenetic tree using the online MIRU-VNTR tool. Clustering analysis was based on the assumption that, strains with the same DNA fingerprint may be epidemiologically linked and associated with recent TB transmission (Hall, 1996). A Cluster was defined as two or more isolates (same strain) that share an indistinguishable spoligotype and 15-loci MIRU-VNTR allelic pattern but allowing for one missing allelic data at any one of the *difficult-to-amplify* MIRU loci (VNTR 2163, 3690 and 4156). We also defined the size of a cluster using the total number of isolates in the cluster into categories of small (2 isolates), medium (3 – 5 isolates), large (6 – 20 isolates) and, very large (>20 isolates).

The recent transmission rate was estimated using the n–1 formulae (Glynn *et al.*, 1999);

$$\frac{(nc-c)}{n}$$

Where; *nc* is the total number of clustered cases, *c* is the number of clusters, and *n* is the total number of cases in the sample.

We included only one strain per participant in our analysis and excluded follow-up cases. Clustering analysis was stratified first by location and then by MTBC lineage. The spatial distribution and clustering among all the observed Spoligo/MIRU strain types were studied

University of Ghana http://ugspace.ug.edu.gh

by constructing a minimum spanning tree (MST) with Bionumerics software (Applied Maths, Sint-Marteen-Latem, Belgium).

4.3.5 Data management and analysis

Both molecular and epidemiological data were analyzed. Epidemiological data retrieved from all participants with positive MTBC cultures were included in the analysis while excluding data from those with no growth, contaminated cultures and isolated non-tuberculous mycobacterial species. All statistical analyses were carried out using the Stata statistical package version 14.2 (Stata Corp., College Station, TX, USA). The association of specific lineages and/or sub-lineages of the MTBC with time and/or geographical locations were explored using chi-square test and a logistic regression model. For the determination of independent predictive factors for recent TB transmission, a multivariate analysis (forward step-wise approach with a probability entry of 0.1) was conducted using a logistic regression model while estimating the odds ratios (OR). P-values < 0.05 were considered significant.

The manuscript was reported according to the "strengthening the reporting of molecular epidemiology for infectious diseases (STROME-ID)" guidelines (Field *et al.*).

University of Ghana http://ugspace.ug.edu.gh

4.4 Results

4.4.1 Characteristics of study participants

We recruited 3,303 sputum smear positive pulmonary TB cases (382/3,303 (11.6%) from the rural setting and 2,921/3,303 (88.4%) from the urban setting) and obtained 2,604 (78.8%) MTBC isolates (Table S4.1). After excluding 13 *M. bovis* and isolates that were untypeable (described in methods), 2,309/2,604 (88.7%) isolates were included for clustering analysis. The participants comprised 71% (1,631) males and 29% (663) females (15 participants had no record of gender) with a median age of 39 (range: 3 to 91) and 33 (range: 4 to 90), respectively (Figure 4.1; Table S4.1). The male/female ratio observed was comparable to the national average of approximately 2:1. Of the 2,309 participants with MTBC genotyping results, 201 (8.7%) were from the rural setting and 2,108 (91.3%) from the urban setting. Among our study cohort, 7.4% (184/2,482) of participants were previously treated cases including relapse, which is similar to the national value of 7.0% (WHO, 2015b). Seventy-one percent (1,561/2,208) presented with a bacterial burden resulting from sputum smear microscopy of at least 2+ and 33% (544/1,665) admitted having contact with at least one TB patient. In a multivariate logistic regression analysis, we found that male patients were less likely to be infected with a L5 strain (adjusted OR 0.7, 95% CI 0.5 - 0.9) and individuals living in villages were more likely to be infected with a L6 strain (OR 6.6, CI 1.2 – 36.1) (Table S4.2).

University of Ghana http://ugspace.ug.edu.gh

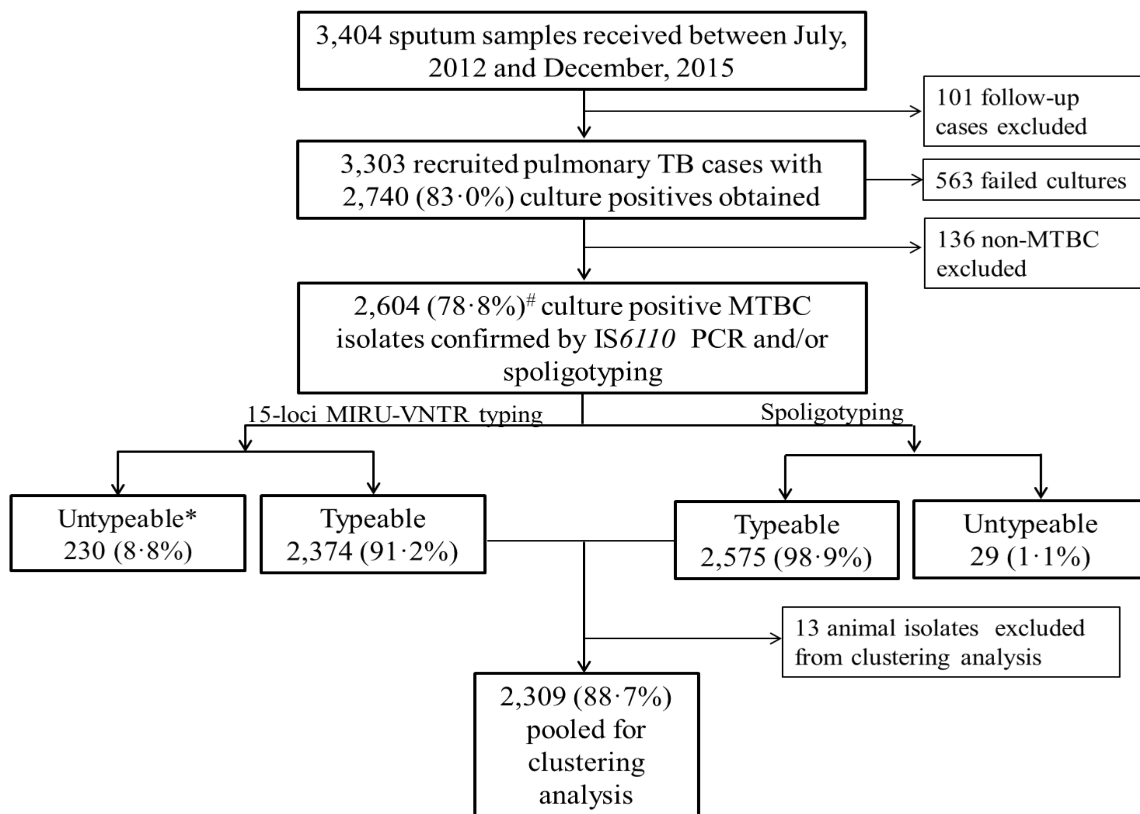


Figure 4.1. Pipeline for recruited participants and culture positive TB cases used for clustering analysis.

*Category described as untypeable for MIRU-VNTR includes isolates with ≥ 2 MIRU-loci un-amplified (164, 71·3%) and isolates with double allele at ≥ 2 MIRU-loci (66, 28·7%). These isolates were described as suspected mix infection or laboratory contamination and hence were removed from further analysis.

Frequency was expressed as the total number of *Mycobacterium tuberculosis* complex (MTBC) isolates obtained

4.4.2 Population structure and recent transmission rate estimation

Among the 2,309 MTBC isolates analyzed for clustering, 1,870 (81.0%) were MTBss and 439 (19.0%) were MAF. Six of the seven human-adapted MTBC lineages were found, with L4, L5 and L6 being most frequent with 1,741 (75.4%), 289 (12.5%), and 150 (6.5%) isolates, respectively (Table 4.1). The relative proportions of the most frequent MTBC lineages remained constant over the entire three and half year study period (ptrend: L4 p=0.72, L5 p=0.84, L6 p=0.25, Figure 4.2).

University of Ghana http://ugspace.ug.edu.gh

Table 4.1. Geographical distribution and population structure of MTBC in Ghana by spoligotyping

	Rural, N (%)	Urban, N (%)	Combined, N (%)*
MTBC isolates	204 (8.8)	2,118 (91.2)	2,322 (100.0)
Species Distribution			
M. tuberculosis	172 (9.2)	1,698 (90.8)	1,870 (80.5)
M. africanum	29 (6.6)	410 (93.4)	439 (18.9)
Animal	3 (23.1)	10 (76.9)	13 (0.6)
Human adapted MTBC lineage distribution			
Lineage_1	4 (10.5)	34 (89.5)	38 (1.6)
Lineage_2	14 (21.5)	51 (78.5)	65 (2.8)
Lineage_3	1 (3.8)	25 (96.2)	26 (1.1)
Lineage_4	153 (8.8)	1,588 (91.2)	1,741 (75.4)
Lineage_5	15 (5.2)	274 (94.8)	289 (12.5)
Lineage_6	14 (9.3)	136 (90.7)	150 (6.5)
Lineage_4 sub-lineage distribution			
Cameroon	77 (7.4)	969 (92.6)	1,046 (60.1)
Ghana	50 (13.3)	326 (86.7)	376 (21.6)
Haarlem	12 (7.7)	144 (92.3)	156 (9.0)
LAM	7 (14.0)	43 (86.0)	50 (2.9)
Uganda	1 (2.5)	39 (97.5)	40 (2.3)
Others (S, U, X, NEW-1)	5 (9.8)	46 (90.2)	51 (2.9)
Not_determined	1 (4.5)	21 (95.5)	22 (1.3)

* Proportion stated here are column wise distributions with respect to the categories of species, lineages or sub-lineages.

MTBC: *Mycobacterium tuberculosis* complex

University of Ghana http://ugspace.ug.edu.gh

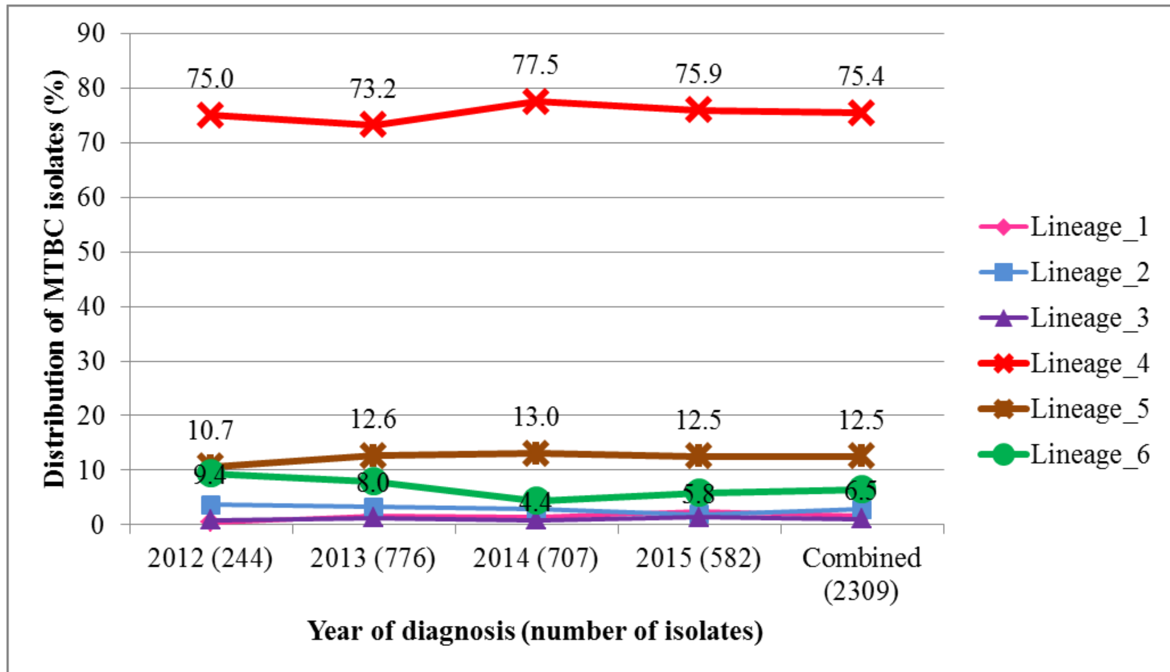


Figure 4.2. Temporal distribution of 2,309 MTBC isolates stratified by lineage.

Lineages have been color coded with the universally accepted color codes for the main MTBC lineages.

Of the 2,309 isolates included for clustering analysis, we identified 1,227 (53.1%) isolates being clustered in 276 different clusters with a mean cluster size of 4 (range: 2 – 35) and 1,082 (46.9%) unique isolates, giving a total of at least 1,358 different MTBC strains circulating within our study population (Table 4.2a). Using the n-1 method, we estimated the overall clustering rate (reflecting recent transmission rate) to be 41.2%. Lineages 2, 4 and 5 contributed high clustering rates of 53.8%, 44.9%, and 31.8%, respectively (Table 4.2a). The Cameroon, Ghana and Haarlem sub-lineages of L4 were the most abundant sub-lineages and, compared to the LAM sub-lineage, contributed significantly to the observed high L4 clustering rate ($p<0.05$) (Figure 4.3). There was no significant difference in the clustering rate between the Cameroon and Ghana sub-lineages ($p=0.57$) (Figure 4.3).

University of Ghana http://ugspace.ug.edu.gh

While we saw no significant difference in the recent transmission rates between members of MAF (L5 and L6, p=0.118), we found that L4 transmitted significantly more (p<0.001) with seven of its clusters having very large cluster sizes (>20 isolates per cluster) made up of the Ghana sub-lineage (4 very large clusters) and Cameroon sub-lineage (3 very large clusters) (Figure 4.3; Figure S4.2). Notwithstanding the lower transmissibility of L5 and L6 compared to L4, we also observed four large clusters for each of these lineages. The urban and rural settings had estimated recent transmission rates of 41.7% and 9.0% respectively.

Table 4.2a. Clustering analysis stratified by lineages and major sub-lineage populations of MTBC

Lineage	Isolates (n)	Clustered cases (c)	Clustered strains (nc)	Single cases (s)	Total strain types (s+c)	Clustering rate[†] (%)
Lineage 1	38	3	7	31	34	10.5
Lineage 2	65	8	43	22	30	53.8
Lineage 3	26	2	4	22	24	7.7
Lineage 4	1,741	201	982	759	960	44.9
Cameroon[#]	1046	123	614	432	555	46.9
Ghana[#]	376	36	206	170	206	45.2
Haarlem[#]	156	23	91	65	88	43.6
LAM [#]	50	6	25	25	31	38.0
Uganda[#]	40	5	16	24	29	27.5
Lineage 5	289	51	143	146	197	31.8
Lineage 6	150	11	48	102	113	24.7
Summary*	2309	276	1227	1082	1358	41.2

[*] Summary was calculated using only the items in cells corresponding to the six main lineages.

[#] Major lineage 4 sub-population

[†] The clustering rate was used for estimating the recent transmission rate

University of Ghana http://ugspace.ug.edu.gh

Table 4.2b. Clustering analysis stratified by study setting and lineages/major sub-lineage populations of MTBC

Lineage	Isolates (n)		Clustered cases (c)		Clustered strains (nc)		Single cases (s)		Total strain types (s+c)		Clustering rate † (%)	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
Lineage 1	34	4	3	0	7	0	27	4	30	4	11.8	0
Lineage 2	51	14	5	1	33	4	18	10	23	11	54.9	21.4
Lineage 3	25	1	2	0	4	0	21	1	23	1	8	0
Lineage 4	1,588	153	183	10	907	25	681	128	864	138	45.6	9.8
Cameroon#	969	77	112	5	575	10	394	67	506	72	47.8	6.5
Ghana#	326	50	32	4	182	12	144	38	176	42	46	16
Haarlem#	144	12	20	1	81	3	63	9	83	10	42.4	16.7
LAM #	43	7	6	0	25	0	18	7	24	7	44.2	0
Uganda#	39	1	5	0	16	0	23	1	28	1	28.2	0
Lineage 5	274	15	49	0	137	0	137	15	186	15	32.1	0
Lineage 6	136	14	10	0	43	0	93	14	103	14	24.3	0
Summary	2108	201	252	11	1131	29	977	172	1229	183	41.7	9

* Summary was calculated using only the items in cells corresponding to the six main lineages.

Major lineage 4 sub-population

† The clustering rate was used for estimating the recent transmission rate

University of Ghana http://ugspace.ug.edu.gh

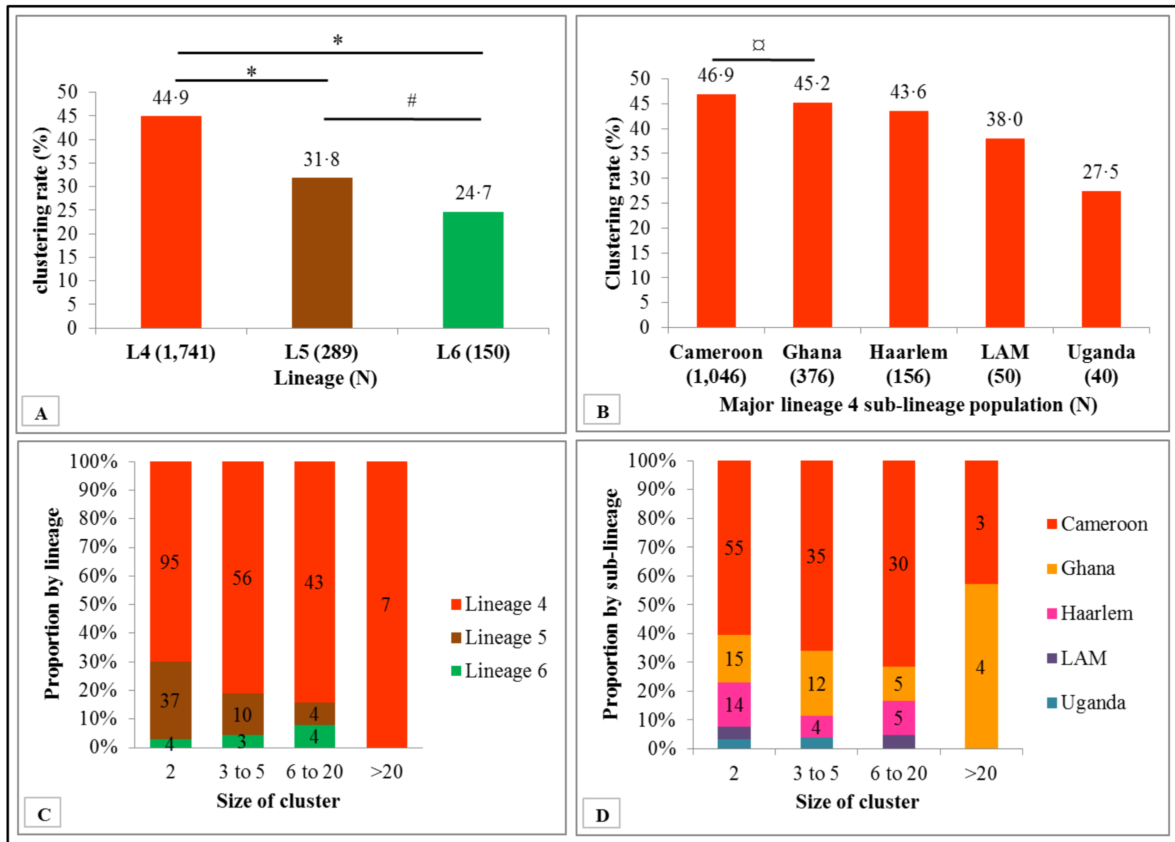


Figure 4.3. Cluster distribution and size stratified by lineage (panel A and C) and sub-lineage (panel B and D).

* p<0·001 # p=0·118 ¤ p=0·565

4.4.3 Exploring the diversity and clustering within the MTBC lineages

Very large molecular clusters (clusters with > 20 isolates: *defined in methods*) were observed for L4 in addition to one strikingly large cluster belonging to the Beijing family of lineage 2 (Figure 4.4; Figure S4.3). Generally, we observed only few multidrug resistant MTBC strains across all the major lineages (Figure S4.4 – S4.6). There was no single large cluster with all isolates being multidrug resistant (Figure S4.4). The spatial distributions of the isolates constituting each cluster stratified by study setting are shown in Figures S4.7 – S4.9.

University of Ghana http://ugspace.ug.edu.gh

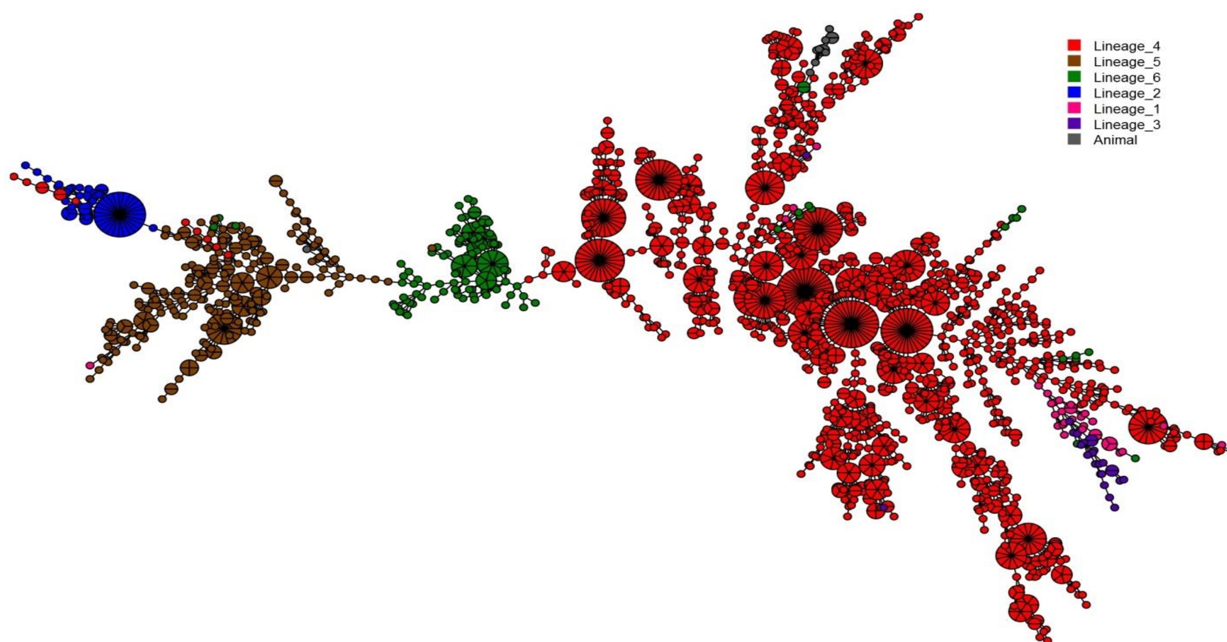


Figure 4.4. Minimum spanning tree (MST) representation of the clustering of 2,322 MTBC isolates from Accra Metropolitan Assembly and East Mamprusi district built with Bionumerics software. The color code reflects the main MTBC lineages 1 to 6 with the size depicting the number of clustered isolates with an identical strain type.

4.4.4 Molecular epidemiology and factors associated with clustering: Logistic regression modeling

Next we looked for risk factors associated with recent TB transmission. We identified a total of 675 individuals belonging to either large (6 – 20 isolates) or very large (>20 isolates) molecular clusters with a combined median cluster size of 14 (range 6 – 35). Majority of the individuals belonging to very large clusters were males with a male to female ratio of approximately 3:1, significantly higher than the 2:1 ratio observed in the general TB patient population (p=0.022). Three large clusters; cluster ID MSC4193, MSC5003.X and MSC4107 with cluster sizes of 9, 7 and 7 respectively, involved only males (Table 4.3).

University of Ghana http://ugspace.ug.edu.gh

Table 4.3. Characteristics of large molecular clusters resulting from combined 15-MIRU and spoligotyping cluster analysis

Number	Cluster code	Number of cases in cluster	Gender male : female	Median age (IQR)	†Diagnosis lapse (months)	# Same residential district	*Known risk factor (number)	Lineage (sub-lineage)	¤Drug resistance
1	**MSC4063.X**	35	31:4	34 (26 - 44)	40	7/5/5/4/4/5	Smoking (6) other (8)	L4 (Cameroon)	3
2	MSC4060.X	34	24:10	34 (25 – 45)	41	6/4/3/3/3/3	Smoking (6) other (5)	L4 (Cameroon)	4
3	MSC4045.X	30	26:4	40 (29 – 48)	39	7/3/3/3/3	Smoking (5) HIV (4) Other (3)	L4 (Cameroon)	2
4	**MSC2001**	27	22:5	35 (27 – 48)	37	8/5	Smoking (8) HIV (4) Other (2)	L2 (Beijing)	1
5	MSC4031	26	19:7	41 (33 – 52)	36	6/4/3	Smoking (6) HIV (3) Other (1)	L4 (Ghana)	11
6	MSC4110	26	21:5	38 (28 – 51)	39	6/3	Smoking (5) HIV (1) Other (4)	L4 (Ghana)	ND
7	**MSC4095**	24	16:8	35 (24 – 45)	39	7/6	Smoking (5) Other (2)	L4 (Ghana)	9
8	MSC4027	21	16:5	27 (25 – 45)	40	6/3	Smoking (3)	L4 (Ghana)	3
9	MSC4063.3	19	18:1	28 (21 – 45)	41	5/5	Smoking (7)	L4 (Cameroon)	ND
10	**MSC4063.18**	18	10:7	35 (24 – 41)	36	6/3	Smoking (4) HIV (1)	L4 (Cameroon)	ND
11	MSC4013	15	13:2	42 (32 – 55)	32	4/3	Smoking (3) HIV (2) Others (2)	L4 (Haarlem)	2
12	MSC4136	15	13:2	36 (28 – 44)	34	6	Smoking (2) HIV (1) Others (3)	L4 (Haarlem)	ND

University of Ghana http://ugspace.ug.edu.gh

Table 4.3 continued

Number	Cluster code	Number of cases in cluster	Gender male : female	Median age (IQR)	†Diagnosis lapse (months)	# Same residential district	*Known risk factor (number)	Lineage (sub-lineage)	¤Drug resistance
13	MSC4040	14	8:6	31 (27 – 45)	33	3/3	HIV (1) Others (4)	L4 (Cameroon)	1
14	**MSC4069.X**	14	11:3	27 (23 – 38)	34	6/3	Smoking (2) HIV (2) Others (1)	L4 (Cameroon)	ND
15	MSC4073	14	9:5	40 (29 – 47)	24	5/4	Smoking (3)	L4 (Cameroon)	3
16	MSC5002.X	14	7:7	40 (38 – 53)	28	5	HIV (2) Smoking (1)	L5 (West African I)	2
17	MSC4063.2	13	8:4	37 (27- 44)	38	4	Smoking (2) Others (3)	L4 (Cameroon)	ND
18	MSC4068.X	13	9:4	35 (30 – 44)	27	5/3	Smoking (6) HIV (2) Others (2)	L4 (Cameroon)	ND
19	MSC4024	12	6:5	28 (26 – 42)	37	3	Smoking (2) Others (1)	L4 (X3)	4
20	MSC4060.18	12	7:5	35 (32 – 40)	36	3/3	Smoking (2) HIV (2) Others (3)	L4 (Cameroon)	1
21	MSC4063.17	12	7:5	26 (24 – 51)	39	ND	Smoking (4) HIV (1) Others (3)	L4 (Cameroon)	2
22	MSC4138	11	7:4	41 (30 – 48)	28	4	Smoking (4)	L4 (LAM)	ND
23	MSC4069.3	10	5:5	32 (24 – 39)	31	3	Smoking (1) HIV (1)	L4 (Cameroon)	2
24	**MSC4104**	10	7:2	35 (25 – 54)	34	5	Smoking (3) Others (1)	L4 (Ghana)	6
25	MSC6006	10	4:6	41 (35 – 47)	33	5/3	Smoking (1) Others (2)	L6 (West African II)	ND
26	MSC4045.3	9	7:2	43 (32 – 50)	33	3	Smoking (1) Others (1)	L4 (Cameroon)	1

University of Ghana http://ugspace.ug.edu.gh

Table 4.3 continued

Number	Cluster code	Number of cases in cluster	Gender male : female	Median age (IQR)	†Diagnosis lapse (months)	# Same residential district	*Known risk factor (number)	Lineage (sub-lineage)	¤Drug resistance
27	MSC4060.21	9	6:3	32 (26 – 43)	22	ND	HIV (1) Others (2)	L4 (Cameroon)	ND
28	MSC4060.3	9	5:4	32 (25 – 53)	34	3	Smoking (1) Others (2)	L4 (Cameroon)	ND
29	MSC4193	9	9:0	36 (30 – 41)	28	ND	Smoking (6) HIV (1) Others (2)	L4 (Cameroon)	ND
30	MSC4068.3	8	6:2	45 (34 – 54)	27	2	Smoking (2) HIV (1) Others (1)	L4 (Cameroon)	ND
31	MSC4022	7	6:1	50 (46 – 62)	36	4	Smoking (1) Others (1)	L4 (Haarlem)	ND
32	MSC4060.4	7	3:4	34 (30-49)	22	5	Smoking (1)	L4 (Cameroon)	ND
33	MSC4080.13	7	4:3	24 (17 – 50)	20	3	Smoking (1)	L4 (Cameroon)	1
34	MSC4082	7	6:1	35 (28 – 40)	33	3	Smoking (1)	L4 (Ghana)	ND
35	MSC4107	7	7:0	38 (29 – 53)	28	3/3	Smoking (2) Others (1)	L4 (Ghana)	1
36	MSC5003.2	7	4:3	35 (26 – 57)	33	ND	HIV (1)	L5 (West African I)	1
37	MSC5003.X	7	7:0	43 (26 – 66)	34	3	Smoking (2) Others (1)	L5 (West African I)	ND
38	MSC6004	7	5:2	44 (36 – 50)	31	ND	HIV (1) Others (3)	L6 (West African II)	3

† Time lapse (in months) between first diagnosed case and last diagnosed case
# Number of participants with same district of residence. Only >2 individuals in the same residential district were indicated.
"/": was used to separate individuals from different districts
* Others in this category refers to alcohol or substance abuse
¤ Number of participants carrying strains with drug resistance to either isoniazid or rifampicin
L2: lineage 2, L4: lineage 4, L5: lineage 5, L6: lineage 6, ND: none determined, IQR: interquartile range
Cluster codes in **bold** involved evidence of household transmission

University of Ghana http://ugspace.ug.edu.gh

Epidemiological investigations revealed both localized and dispersed recent transmission among the clustered cases with suggested evidence of household transmission in at least six large clusters (MSC4063.X, MSC2001, MSC4095, MSC4063.18, MSC4069.X and MSC4104). Specifically, the same L4 strain (part of cluster MSC4069.X) was found among three individuals belonging to the same household with the oldest person (age 49) reporting having contact with his son who had TB four months prior to his episode (suggestive of household transmission). Majority of the large clusters involved TB strains circulating over almost the entire study period (Figure S4.10). Apart from three Ghana sub-lineage clusters (MSC4104, MSC4031 and MSC4095) and one L6 cluster (MSC6004) with respectively 60% (6/10), 42% (11/26), 38% (9/24) and 43% (3/7) of isolates showing resistance to either rifampicin and/or isoniazid (Table 4.3), we did not observe such high levels of drug resistance in the other large and very large clusters. Only 2% of the isolates belonging to large and very large clusters were MDR-TB strains and this was significantly lower than that for small (2 isolates) and medium (3 – 5 isolates) (4%) clusters (p=0.031).

For the determination of possible factors associated with recent TB transmission, we first performed a general logistic regression model including all MTBC lineages using the event of belonging to a clustered case as the outcome variable and participants' variables as possible predictors (Table 4.4). In a separate logistic regression model, we tested risk factors associated with recent TB transmission stratified independently by L4 and L5 (Table 4.5) excluding L6 due to limited sample size. In the multivariable analysis for the general logistic regression model, we found that harboring either isoniazid or rifampicin resistant MTBC strain (adjusted OR 0.7, 95% CI 0.5 – 0.9) was associated with a lower odds of belonging to a clustered case (Table 4.4). All other factors such as education

University of Ghana http://ugspace.ug.edu.gh

status, occupation, income level, ethnicity, religion or HIV status had no association with recent TB transmission.

Table 4.4. Logistic regression analysis of risk factors associated with TB clustering (recent TB transmission)

Variable	MTBC (N =2,309)		Univariate		Multivariate	
	Total TB cases, N (%)	Clustered[#] cases, N (%)	OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
Year diagnosed	2,309 (100)	1,229 (53·2)				
2012	244 (10·6)	147 (60·3)	1·4 (1·0 - 1·8)	0·043	1·3 (0·9 - 1·7)	0·113
2013	776 (33·6)	410 (52·8)	Reference			
2014	707 (30·6)	365 (51·6)	1·0 (0·8 - 1·2)	0·642	0·9 (0·7 - 1·1)	0·203
2015	582 (25·2)	307(52·8)	1·0 (0·8 - 1·2)	0·975	1·0 (0·8 - 1·2)	0·703
Gender	2,294 (99·4)					
Male	1,631 (71·1)	863 (52·9)	1·0 (0·8 - 1·2)	0·685		
Female	663 (28·9)	357 (53·8)	Reference			
Age[¤] (yrs)	2,224 (96·3)					
<15	37 (1·7)	25 (67·6)	1·6 (0·8 - 3·3)	0·183	1·6 (0·8 - 3·2)	0·221
15 - 29	639 (28·7)	360 (56·3)	Reference			
30 - 39	570 (25·6)	307 (53·9)	0·9 (0·7 - 1·1)	0·387	0·9 (0·7 - 1·2)	0·688
40 - 59	778 (35·0)	398 (51·2)	0·8 (0·7 - 1·0)	0·052	0·9 (0·7 - 1·1)	0·241
>59	200 (9·0)	97 (48·5)	0·7 (0·5 - 1·0)	0·053	0·9 (0·6 - 1·1)	0·211
Nationality	1,781 (77·1)					
Ghanaian	1,714 (96·2)	932 (54·4)	Reference			
Others	67 (3·8)	38 (56·7)	1·1 (0·7 - 1·8)	0·706		
Locality	2,309 (100)	1,229 (53·2)				
Rural	201 (8·7)	74 (36·8)	Reference			
Urban	2,108 (91·3)	1,155 (54·8)	2·1 (1·5 - 2·8)	<0·001		
Residence classification	1,642 (71·1)					
Village	69 (4·2)	27 (39·1)	0·5 (0·3 - 0·8)	0·007		
Town	182 (11·1)	96 (52·7)	0·9 (0·6 - 1·2)	0·415		
City residential area	52 (3·2)	27 (51·9)	0·8 (0·5 - 1·5)	0·564		
City suburb	1,136 (69·2)	636 (56·0)	Reference			
City Slum	203 (12·4)	112 (55·2)	1·0 (0·7 - 1·3)	0·83		
Residential district	1,538 (66·6)					
Ablekuma	545 (35·4)	298 (54·7)	Reference			
Ashiedu Keteke	170 (11·1)	100 (58·8)	1·2 (0·8 - 1·7)	0·343		
Ayawaso	220 (14·3)	124 (56·4)	1·1 (0·8 to 1·5)	0·672		
Kpeshie	224(14·6)	121 (54·0)	1·0 (0·7 - 1·3)	0·867		
Mamprusi East	70 (4·6)	22 (31·4)	0·4 (0·2 - 0·6)	<0·001		
Okaikoi	176 (11·4)	98 (55·7)	1·0 (0·7 to 1·5)	0·816		
Osu Klottey	133 (8·6)	78 (58·7)	1·2 (0·8 - 1·7)	0·409		

University of Ghana http://ugspace.ug.edu.gh

Variable	MTBC (N =2,309)		Univariate		Multivariate	
	Total TB cases, N (%)	Clustered[#] cases, N (%)	OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
Household type	1,624 (70·3)					
Self-contain	412 (25·4)	221 (53·6)	1·0 (0·8 - 1·2)	0·797		
Compound house	1,212 (74·6)	659 (54·4)	Reference			
Education	1,748 (75·7)					
Primary	222 (12·7)	125 (56·3)	1·1 (0·8 - 1·5)	0·637		
Middle/JHS	637 (36·4)	347 (54·5)	Reference			
Secondary	429 (24·5)	232 (54·1)	1·0 (0·8 - 1·3)	0·899		
Tertiary	190 (10·9)	110 (57·9)	1·1 (0·8 - 1·6)	0·405		
No Education	270 (15·4)	141 (52·2)	0·9 (0·7 - 1·2)	0·534		
Occupation	1,722 (74·6)					
Unemployed	390 (22·6)	208 (53·3)	0·9 (0·7 - 1·1)	0·423		
Unskilled	951 (55·2)	530 (55·7)	Reference			
Skilled	381 (22·1)	198 (52·0)	0·9 (0·7 - 1·1)	0·213		
Monthly income (GH¢)	1,622 (70·2)					
None	371 (22·9)	213 (57·4)	Reference			
<301	807 (49·7)	438 (54·3)	0·9 (0·7 - 1·1)	0·315		
301 - 1,000	407 (25·1)	218 (53·6)	0·8 (0·6 - 1·1)	0·281		
>1,000	37 (2·3)	15 (40·5)	0·5 (0·3 - 1·0)	0·052		
Religion	1,771 (76·7)					
Christian	1,361 (76·9)	739 (54·3)	Reference			
Islam	302 (17·0)	161 (53·3)	1·0 (0·7 - 1·2)	0·755		
Others	26 (1·5)	14 (53·9)	1·0 (0·4 - 2·1)	0·963		
Not religious	82 (4·6)	49 (59·7)	1·2 (0·8 - 2·0	0·366		
Ethnicity	1,760 (76·4)					
Akan	570 (32·3)	309 (54·2)	Reference			
Ewe	259 (14·7)	143 (55·2)	1·0 (0·8 - 1·4)	0·788		
Ga/Adangbe	544 (30·8)	310 (57·0)	1·1 (0·9 - 1·4)	0·352		
Others	392 (22·2)	196 (50·0)	0·8 (0·6 - 1·1)	0·199		
Marital Status	1,758 (76·1)					
Single	766 (43·6)	431 (56·3)	Reference			
Married	742 (42·2)	395 (53·2)	0·9 (0·7 - 1·1)	0·237		
Divorced	167 (9·5)	99 (59·3)	1·1 (0·8 - 1·6)	0·476		
Widowed	83 (4·7)	35 (42·2)	0·6 (0·3 - 0·9)	0·015		
Smear positivity	2,208 (95·6)					
Scanty 1 - 9	173 (7·8)	96 (55·5)	1·1 (0·8 - 1·5)	0·714		
1+	474 (21·5)	237 (50·0)	0·9 (0·7 - 1·1)	0·151		
2+	546 (24·7)	294 (53·9)	1·0 (0·8 - 1·2)	0·957		
3+	1,015 (46·0)	548 (54·0)	Reference			
Previous TB treatment	1,737 (75·2)					
Yes	291 (16·8)	153 (52·6)	0·9 (0·7 - 1·2)	0·535		
No	1,446 (83·2)	789 (54·6)	Reference			
Risk of TB contact						
Closefriend/Household	1,665 (72·1)					
No contact	1,121 (67·3)	594 (53·0)	Reference			
1 contact	212 (12·7)	118 (55·7)	1·1 (0·8 - 1·5)	0·475		
2-5 contacts	309 (18·6)	179 (57·9)	1·2 (0·9 - 1·6)	0·123		
6-10 contacts	23 (1·4)	15 (65·2)	1·7 (0·7 - 4·0)	0·249		

University of Ghana http://ugspace.ug.edu.gh

Variable	MTBC (N =2,309)		Univariate		Multivariate	
	Total TB cases, N (%)	Clustered[#] cases, N (%)	OR (95% CI)	p-value	Adjusted OR (95% CI)	p-value
Imprisonment	1,660 (71·9)					
Yes	97 (5·8)	56 (57·7)	1·1 (0·8 - 1·7)	0·513		
No	1,563 (94·2)	849 (54·3)	Reference			
Health/Lab worker	1,661 (71·9)					
Yes	47 (2·8)	25 (53·2)	0·9 (0·5 - 1·7)	0·85		
No	1,614 (97·2)	881 (54·6)	Reference			
Immunosuppressive condition	1,695 (73·4)					
Any	893 (52·7)	488 (54·6)	1·0 (0·9 - 1·2)	0·747		
None	802 (47·3)	432 (53·9)	Reference			
Diabetes Mellitus	534 (23·1)					
Yes	104 (19·5)	54 (51·9)	1·0 (0·7 - 1·5)	0·957		
No	430 (80·5)	222 (51·6)	Reference			
HIV status	1,166 (50·5)					
Yes	144 (12·3)	82 (56·9)	1·1 (0·8 - 1·6)	0·481		
No	1,022 (87·7)	550 (53·8)	Reference			
Smoking	1,518 (65·7)					
Yes	434 (28·6)	237 (54·6)	1·0 (0·8 - 1·2)	0·949		
No	1,084 (71·4)	590 (54·4)	Reference			
Substance abuse (excluding alcohol)	1,401 (60·7)					
Yes	140 (10·0)	84 (60·0)	1·3 (0·9 - 1·8)	0·172		
No	1,261 (90·0)	680 (53·9)	Reference			
Substance abuse (including alcohol)	1,474 (63·8)					
Yes	460 (31·2)	250 (54·3)	1·0 (0·8 - 1·3)	0·858		
No	1,014 (68·8)	546 (53·8)	Reference			
Lineage	2,309 (100)					
Lineage 1	38 (1·7)	7 (18·4)	0·2 (0·08 - 0·4)	<0·001	0·13 (0·05 - 0·36)	<0·001
Lineage 2	65 (2·8)	43 (66·2)	1·5 (0·9 - 2·5)	0·126	1·5 (0·9 - 2·5)	0·155
Lineage 3	26 (1·1)	4 (15·4)	0·1 (0·05 - 0·4)	<0·001	0·15 (0·05 - 0·45)	0·001
Lineage 4	1,741 (75·4)	984 (56·5)	Reference			
Lineage 5	289 (12·5)	143 (49·5)	0·8 (0·6 - 1·0)	0·026	0·7 (0·6 - 0·9)	0·032
Lineage 6	150 (6·5)	48 (32·0)	0·4 (0·3 - 0·5)	<0·001	0·3 (0·2 - 0·5)	<0·001
Lineage 4 sub-lineage						
Cameroon	1,046 (60·1)	616 (58·9)	Reference			
Ghana	376 (21·6)	206 (54·8)	0·8 (0·7 - 1·1)	0·167		
Harleem	156 (9·0)	91(58·3)	1·0 (0·7 - 1·4)	0·895		
LAM	50 (2·9)	25 (50·0)	0·7 (0·4 - 1·2)	0·215		
Uganda	40 (2·3)	16 (40·0)	0·5 (0·2 - 0·9)	0·02		
Others	51 (2·9)	26 (51·0)	0·7 (0·4 - 1·3)	0·265		
Not determined	22 (1·3)	4 (18·2)	0·2 (0·1 - 0·5)	0·001		
Drug resistance	2,300 (99·6)					
Any	313 (13·6)	138 (44·1)	0·6 (0·5 - 0·8)	<0·001	0·7 (0·5 - 0·9)	0·002
None	1,987 (86·4)	1,090 (54·9)	Reference			

University of Ghana http://ugspace.ug.edu.gh

Variable	MTBC (N =2,309)		Univariate		Multivariate	
	Total TB cases, N (%)	**Clustered[#] cases, N (%)**	**OR (95% CI)**	**p-value**	**Adjusted OR (95% CI)**	**p-value**
Isoniazid mono-resistant	2,300 (99·6)					
Yes	295 (12·8)	129 (43·7)	0·6 (0·5 - 0·8)	<0·001		
No	2,005 (87·2)	1,099 (54·8)	Reference			
Multidrug resistant (MDR)	2,300 (99·6)					
Yes	81 (3·5)	35 (43·2)	0·7 (0·4 - 1·0)	0·063		
No	2,219 (96·5)	1,193 (53·8)	Reference			
Cluster size (n)	1,227 (53·1)					
Small (2)	290 (23·6)					
Medium (3 - 5)	262 (21·4)					
Large (6 - 20)	452 (36·8)					
Very large (>20)	223 (18·2)					

For the multivariate model, we included only variables with p<0.1 and with at least 90% of available data. We however excluded "locality" due to reduced sample size from the rural setting. We excluded residence classification, marital status, isoniazid mono-resistance and MDR due to collinearity with other variables in the model.

*p<0.05; **p<0.001

[#] A Cluster was defined as two or more isolates (same strain) that share an indistinguishable spoligotype and 15-loci MIRU-VNTR allelic pattern but allowing for one missing allelic data at any one of the difficult-to-amplify MIRU loci.

[¤] We found a significant decreasing trend in the probability of belonging to a clustered case with increasing age category (p=0.004).

MTBC: *Mycobacterium tuberculosis* complex, OR: odds ratio

University of Ghana http://ugspace.ug.edu.gh

Table 4.5: Risk Factors associated with TB clustering: logistic regression analysis stratified by lineage

Variables	lineage 4 (N=1,741)		Univariate	Multivariate		lineage 5 (N=289)		Univariate	
	TB cases, N (%)	clustered[#] cases, N (%)	OR (95% CI)	Adjusted OR (95% CI)	p-value	TB cases, N (%)	clustered[#] cases, N (%)	OR (95% CI)	p-value
Year diagnosed	1,741 (100)					289 (100)			
2012	183 (10·5)	120 (65·6)	1·5 (1·1 - 2·1)*	1·4 (1·0 - 2·1)	0·062	26 (9·0)	14 (53·8)	1·2 (0·5 - 2·9)	0·659
2013	568 (32·6)	318 (56·0)	Reference			98 (33·9)	48 (49·0)	Reference	
2014	548 (31·5)	300 (54·7)	1·0 (0·8 - 1·2)	1·0 (0·7 - 1·3)	0·847	92 (31·8)	43 (46·7)	0·9 (0·5 - 1·6)	0·757
2015	442 (25·4)	244 (55·2)	1·0 (0·8 - 1·2)	1·0 (0·7 - 1·3)	0·955	73 (25·3)	38 (52·1)	1·2 (0·6 - 2·1)	0·691
Age (yrs)	1,672					283			
<15	27 (1·6)	20 (74·1)	2·1 (0·9 - 5·0)			5 (1·8)	3 (60·0)		
15 - 29	497 (29·7)	289 (58·2)	Reference			78 (27·6)	42 (53·8)		
30 - 39	432 (25·8)	252 (58·3)	1·0 (0·8 - 1·3)			68 (24·0)	31 (45·6)		
40 - 59	580 (34·7)	315 (54·3)	0·9 (0·7 - 1·1)			94 (33·2)	48 (51·1)		
>59	136 (8·1)	71 (52·2)	0·8 (0·5 - 1·2)			38 (13·4)	16 (42·1)		
Locality	1,741 (100)					289 (100)			
Rural	153 (8·8)	59 (38·6)	Reference			15 (5·2)	4 (26·7)	Reference	
Urban	1,588 (91·2)	923 (58·1)	2·2 (1·6 - 3·1)**			274 (94·8)	139 (50·7)	2·8 (0·9 - 9·1)	0·081
Residential district	1,165					189			
Ablekuma	412 (35·4)	237 (57·5)	Reference			77 (40·7)	39 (50·7)	Reference	
Ashiedu Keteke	132 (11·3)	81 (61·4)	1·2 (0·8 - 1·8)			13 (6·9)	5 (38·5)	0·6 (0·2 - 2·0)	0·419
Ayawaso	178 (15·3)	111 (62·4)	1·2 (0·8 - 1·8)			21 (11·1)	7 (33·3)	0·5 (0·2 - 1·4)	0·163
Kpeshie	166 (14·2)	88 (53·0)	0·8 (0·6 - 1·2)			37 (19·6)	25 (67·6)	2·0 (0·9 - 4·6)	0·091
Mamprusi East	56 (4·8)	19 (33·9)	0·4 (0·2 - 0·7)*			4 (2·1)	1 (25·0)	0·32 (0·03 - 3·26)	0·339
Okaikoi	134 (11·5)	80 (59·7)	1·1 (0·7 - 1·6)			24 (12·7)	12 (50·0)	1·0 (0·4 - 2·4)	0·956
Osu Klottey	87 (7·5)	58 (66·7)	1·5 (0·9 - 2·4)			13 (6·9)	5 (38·5)	0·6 (0·2 - 2·0)	0·419
Monthly income (GH¢)	1,222								
None	275 (22·5)	167 (60·7)	Reference						
<301	605 (49·5)	351 (58·0)	0·9 (0·7 - 1·2)						

University of Ghana http://ugspace.ug.edu.gh

Variables	lineage 4 (N=1,741)		Univariate	Multivariate		lineage 5 (N=289)		Univariate	
	TB cases, N (%)	clustered# cases, N (%)	OR (95% CI)	Adjusted OR (95% CI)	p-value	TB cases, N (%)	clustered# cases, N (%)	OR (95% CI)	p-value
301 - 1,000	314 (25·7)	184 (58·6)	0·9 (0·7 - 1·3)						
>1,000	28 (2·3)	11 (39·3)	0·4 (0·2 - 0·9)*						
Marital Status	1,322								
Single	591 (44·7)	355 (60·1)	Reference						
Married	549 (41·5)	312 (56·8)	0·9 (0·7 - 1·1)	0·9 (0·7 - 1·2)	0·589				
Divorced	124 (9·4)	78 (62·9)	1·1 (0·8 - 1·7)	1·1 (0·7 - 1·7)	0·543				
Widowed	58 (4·4)	24 (41·4)	0·5 (0·3 - 0·8)*	0·5 (0·3 - 0·8)	0·011				
Lineage 4 sub-lineage									
Cameroon	1046 (60·1)	614 (58·7)							
Ghana	376 (21·6)	206 (54·8)	0·9 (0·7 - 1·1)	0·9 (0·7 - 1·2)	0·403				
Harleem	156 (9·0)	91 (58·3)	1·0 (0·7 - 1·4)	1·0 (0·7 - 1·5)	0·87				
LAM	50 (2·9)	25 (50·0)	0·7 (0·4 - 1·2)	0·7 (0·4 - 1·4)	0·354				
Uganda	40 (2·3)	16 (40·0)	0·5 (0·2 - 0·9)*	0·4 (0·2 - 0·8)	0·013				
Others	51 (2·9)	26 (51·0)	0·7 (0·4 - 1·3)	0·8 (0·4 - 1·6)	0·558				
Not determined	22 (1·3)	4 (18·2)	0·2 (0·1 - 0·5)*	0·10 (0·03 - 0·35)	<0·001				
Drug resistance	1,736								
Any	241 (13·9)	114 (47·3)	0·7 (0·5 - 0·9)*	0·7 (0·5 - 1·0)	0·059				
None	1,495 (86·1)	867 (58·0)	Reference						

We included in this analysis, only variables with p<0.1 from the general logistic regression model in Table 4.4

For the multivariate model, we included only variables with p<0.1 and with at least 90% of available data.

*p<0.05; **p<0.001

[#] A Cluster was defined as two or more isolates (same strain) that share an indistinguishable spoligotype and 15-loci MIRU-VNTR allelic pattern but allowing for one missing allelic data at any one of the difficult-to-amplify MIRU loci.

University of Ghana http://ugspace.ug.edu.gh

Finally, using adjusted predictions, we found that, the probability of belonging to a clustered case decreased with age and increased with the number of TB contacts (Figure 4.5). In a separate logistic regression analysis, including age as a continuous variable with belonging to a clustered case the outcome variable, we found that, each year increase in age was significantly associated with approximately 1% (CI: 0.13 – 2.00%) decrease in the odds of a TB patient being part of a recent transmission event (p=0.007).

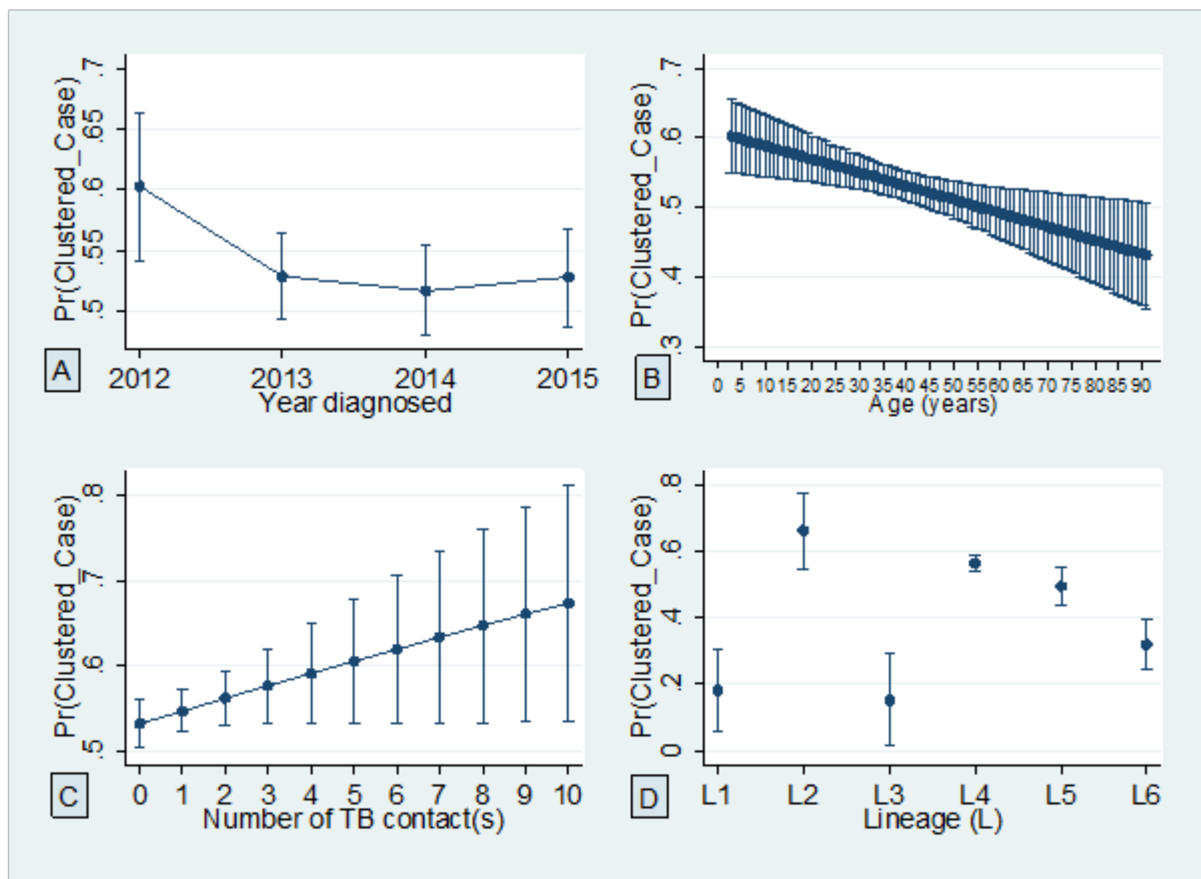


Figure 4.5. Adjusted predictions of the probability of belonging to a clustered case with 95% confidence interval at the year of diagnosis (A), while ageing (B), considering the number of close TB contact(s) (C) and of circulating MTBC lineages (D).

University of Ghana http://ugspace.ug.edu.gh

4.5 Discussion and Conclusion

The aims of this study were to conduct a population-based prospective molecular epidemiological study to analyze the transmission dynamics of MTBC strains circulating in Ghana and to identify risk factors associated with recent TB transmission.

We obtained a high MTBC isolate recovery rate of 78.8% higher than that reported in similar studies (Hamblion *et al.*, 2016; Mears *et al.*, 2015) and this strenghtens the power of the sample size to make assessments of TB transmission rate in Ghana. Our study identified a high TB clustering (recent TB transmission) rate of 41.2% which is quite alarming with the urban and rural areas having estimated rates of 41.7% and 9.0% respectively (Table 4.2b). These findings call for intensifying community outreach programs to encourage early case reporting and infection control. Moreover, our analysis predicted the probability of clustering to generally increase with increase in the number of TB contacts (Figure 4.5). This means that a susceptible individual is likely to have TB and be involved in a recently transmitted event as the number of TB contacts increases.

Within our study population, we found no association of recent TB transmission with education status, occupation, income level, ethnicity, religion or HIV status. However, we observed that individuals below the age of 30 years were associated with recent TB transmission, and this is similar to observations made elsewhere (Hamblion *et al.*, 2016; Vluggen *et al.*, 2017). Also from this study, we observed that each year increase in age is associated with an approximately 1% (CI: 0.13 – 2.00, p=0.007) decrease in the odds of a TB patient being part of a recent transmission event implying that, compared to younger individuals, older individuals are more likely to get active TB disease by reactivation of latent TB infection rather than through a recent transmission event (Hamblion *et al.*, 2016).

University of Ghana http://ugspace.ug.edu.gh

This finding puts age as a risk factor for recent TB transmission in Ghana. This finding was however largely driven by L4 and L5 since separate analysis was not valid for L6 judging from the small sample size. Also, we identified that the male/female ratio among very large clusters was significantly higher than that observed in the general TB patient population (p=0.022). This finding, together with the observation that some large clusters involved only males, also puts males as having a higher risk of recent TB transmission compared to females suggesting that males may engage in certain social activities that predisposes them to belonging to a recent transmission event.

We observed a lower rate of MDR-TB among large clustered cases compared to the general population (2% vs 4%, p=0.031) indicating a low MDR-TB transmissibility within our study population. This finding further suggests that, majority of drug resistant-TB cases in Ghana acquired the drug resistance during treatment which indicate poor patient compliance (Danso *et al.*, 2015). We also found that compared to drug (isoniazid and/or rifampicin) sensitive MTBC strains, it is unlikely to find MTBC strains with either isoniazid and/or rifampicin resistance involved in a recent transmission event (adjusted OR 0.7, 95% CI 0.5 – 0.9).

Within our study setting, we observed a reduced transmission of MAF (L5: 31.8%, L6: 24.7%) compared to MTBss L4 (44.9%). The high recent transmission rate observed for L4 was driven by both the Cameroon and Ghana sub-lineages with no difference in their transmissibility, hence putting these sub-lineages as very important pathogens. The high recent transmission of the Ghana sub-lineage coupled with recently reported association with drug resistance (Otchere *et al.*, 2016) is of public health importance and hence calls

University of Ghana http://ugspace.ug.edu.gh

for the national tuberculosis control program to support peripheral diagnostic laboratories with facilities to accurately detect and help control the spread of the Ghana sub-lineage.

The higher recent transmission rate for L4 compared to L5 and L6 may not necessarily imply the outcompeting of L5 and L6 by L4 as their relative proportions still remained constant over the entire study period (Figure 4.2) and also based on previous reports (Yeboah-Manu *et al.*, 2016). Despite the low transmissibility of MAF, the observed stable relative proportion over the entire study period may be because the pathogen has adapted to infecting specific host populations (possibly due to unidentified host genetic or environmental factors peculiar to some West African inhabitants) and hence enabling the maintenance of a stable prevalence over time. Using adjusted predictions for the probability of clustering, we found that MAF L5 may still have the propensity to transmit as equal as lineage 4 (Figure 4.5) not forgetting the confounding effect of a higher diversity in spoligotype pattern of L5 compared to L4 and hence reducing the clustering of the former (Asante-Poku *et al.*, 2016). Compared to L4, we found a significant association of L6 with individuals living in villages (OR 6.6, $p<0.05$; Table S4.2). The low recent TB transmission in the villages coupled with an association of L6 could be the reason why we observed low frequencies of L6 strains within our study setting.

Our report could be limited by the possibility of underestimating recent transmission rate resulting from misclassifying strains as unique if they were actually clustered outside of our restricted geographic sampling site and sampling period. We however took some measures to address underestimation of recent TB transmission by recruiting up to 90% of the diagnosed TB cases spanning a three and a half year period. In addition, we also understand the possibility of overestimating recent TB transmission rates considering that

University of Ghana http://ugspace.ug.edu.gh

the basis of our clustering analysis was done using a combined 15 loci MIRU-VNTR typing and spoligotyping whereas WGS could have offered a better resolution of strains. Overall, our findings indicate high recent TB transmission suggesting occurrences of unsuspected outbreaks and recommends intensifying community education to improve early case reporting and infection control.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER 5

5.0 Manuscript 3. Whole Genome Sequencing and Spatial Analysis Identifies Recent TB Transmission Hotspots in Ghana

5.1 Abstract

Background: Whole genome sequencing (WGS) is progressively being used to investigate the transmission dynamics of members of the *Mycobacterium tuberculosis* complex (MTBC), the causative agent of tuberculosis (TB). We used WGS and analysis to resolve large traditional genotype clusters and explored the spatial distribution of confirmed large recent transmission clusters.

Methods: Bacterial genomes from a total of 452 MTBC isolates belonging to large traditional clusters from a population-based study spanning July 2012 and December 2015 were obtained through short read next generation sequencing using the illumina HiSeq2500 platform. We performed genomic clustering analysis by employing specified R packages using an adopted threshold of 10-SNPs. The spatial distribution of genomic clusters was done using ArcGIS.

Results: Of the 452 traditional genotype clustered cases, 314 (69.5%) were confirmed as clusters with a median cluster size of 7.5 (IQR, 4-12). While the MIRU/VNTR plus spoligotyping estimated recent transmission rate at 41.2%, WGS estimated it to be 24.7% and we confirmed the existence of low recent transmission among *M. africanum* lineages (lineage 5 and 6) compared to *M. tuberculosis* sensu stricto lineage 4. WGS confirmed a wide spread of a Cameroon sub-lineage clone of lineage 4 with a large cluster of 78 isolates predominantly from the Ablekuma sub-district of Accra metropolis. More

University of Ghana http://ugspace.ug.edu.gh

importantly, we identified a recent transmission cluster associated with isoniazid resistance belonging to the Ghana sub-lineage of lineage 4.

Conclusion: WGS was useful in detecting unsuspected outbreaks hence we recommend its use not only as a research tool but as a surveillance tool to aid in providing the necessary guided steps to track, monitor and control TB.

5.2 Introduction

Tuberculosis (TB), a contagious disease of antiquity, affects millions of people and in 1993 the world health organization (WHO) declared it a global health emergency, calling for more resources and studies for effective control. In 2017 alone, the WHO estimated that 10 million new TB cases occurred globally, out of which 1.6 million died making TB the number one infectious disease killer and one of the ten killer diseases (WHO, 2018b). TB in humans is caused mainly by *Mycobacterium tuberculosis* sensu stricto (MTBss) and *Mycobacterium africanum* (MAF), which are further divided into seven lineages (L): MTBss subdivided into L1–L4 and L7, and MAF L5 and L6 (Blouin *et al.*, 2012; de Jong, Antonio*, et al.*, 2010).

The global TB control strategy aims at having a TB-free world by attaining zero deaths, disease and suffering due to TB. (WHO, 2015a). One of the activities to achieve this, is to investigate the transmission dynamics of the disease to understand risk factors leading to occurrence of the disease within distinct population for design of appropriate preventive interventions. The study of the spread of these TB lineages have been made possible through molecular epidemiology (molepi) studies. The traditional molecular biological

University of Ghana http://ugspace.ug.edu.gh

tools used for such studies include IS*6110* DNA fingerprinting, spacer oligonucleotide typing (spoligotyping) and mycobacteria interspersed repetitive-unit – variable-number tandem repeat typing (MIRU-VNTR typing) (Anderson *et al.*, 2014; Barnes & Cave, 2003; Hamblion *et al.*, 2016). These tools have been used extensively in previous studies and found to have varying discriminatory power (Barnes & Cave, 2003; Jagielski *et al.*, 2014; Jamieson *et al.*, 2014; Kanduma *et al.*, 2003; Mendez *et al.*, 2016; Murray & Alland, 2002). However, WGS analysis is considered the ultimate for strain typing and confirmation of strain clusters (Jamieson *et al.*, 2014; Senghore *et al.*, 2017; Walker, Ip*, et al.*, 2013). The established assumption is isolates of similar fingerprint (clustered isolates) may be epidemiologically linked and hence associated with recent transmission.

In a previous population-based study, we used the combined resolution power of MIRU-VNTR typing and spoligotyping (MIRU/Spoligo) for strain differentiation followed with clustering analysis to estimate the extent of recent transmission in Ghana (Asare *et al.*, 2018). We estimated a high recent transmission rate of 41.2% and found 53.1% of all isolates belonging to one of 276 clusters. Yet, it has been indicated that the combined resolution of spoligotyping and MIRU-VNTR may not be enough to distinguish between unrelated strains, depending on the circulating strains for a given geographic setting (Jamieson *et al.*, 2014). Whole genome sequence data does not only give us the power to identify recent TB transmission but also the ability to trace the route/direction of transmission between such epidemiologically linked TB cases (Walker, Ip*, et al.*, 2013). In this current study, we used a WGS approach to further resolve large MIRU/Spoligo defined clusters and explore epidemiological characteristics including spatial distribution of confirmed large clusters.

University of Ghana http://ugspace.ug.edu.gh

5.3 Materials and methods

5.3.1 Study design and population

This study involved a retrospective analysis of selected isolates obtained from a population-based study conducted from July 2012 to December 2015 and sampled within two districts in Ghana; Accra Metropolitan Area (AMA) and East Mamprusi District (MamE). All isolates were obtained from pulmonary TB cases. All participants consented to the study and provided both epidemiological and demographic data. A detailed description of the study locale and participants data are provided elsewhere (Asare *et al.*, 2018; Yeboah-Manu *et al.*, 2016).

5.3.2 Isolate selection, DNA extraction and whole genome sequencing

The isolate collection for the analysis was a convenient sample of all cases belonging to 40 large clusters (cluster size > 5) comprising 473 isolates from our previous study. Only large clusters belonging to the three most dominant MTBC lineages (L4, L5 and L6) in Ghana were considered for analysis. All isolates were previously characterized using standard phenotypic and genotypic techniques with detailed description provided in earlier reports (Asare *et al.*, 2018; Yeboah-Manu *et al.*, 2016). Drug susceptibility to INH and RIF was carried out by the microplate Alamar Blue cell viability assay, as described elsewhere (Franzblau *et al.*, 1998; Otchere *et al.*, 2016) and/or by GenoType MTBDRplus assay (Hain Lifescience), following the manufacturer's protocol (Barnard *et al.*, 2008).

DNA extraction was performed using a modified CTAB protocol as previously described (Otchere *et al.*, 2016). Briefly, to obtain enough intact genomic DNA (gDNA), up to 200 ul of packed bacteria cells (representing an estimated number of 10^{20} cells per mL) harvested from the L-J media was heat inactivated at 80 ºC for 30 minutes in cell lysis

University of Ghana http://ugspace.ug.edu.gh

bugs buffer. The suspension was then treated with 10 mg/ml lysozyme SDS (20%) and proteinase K (10 mg/ml), followed by CTAB to a final concentration of 2% and incubated at 65 º for 10 minutes to complete the cell lysis. Chloroform:isoamyl alcohol (24:1) extraction was done and the resulting DNA was precipitated and purified with isopropanol 70% ice-cold ethanol respectively and eluted in 1X Tris EDTA buffer (TE buffer). All steps that required centrifugation were carried out at 13,000xg. Prior to DNA sequencing, the purity and fragmentation of the extracted gDNA was checked using Nanodrop and 1% agarose gel electrophoresis.

All gDNA samples were outsourced to Swiss Tropical and Public Health Institute for DNA library preparation and NGS. Sequencing libraries were prepared using NEXTERA XT DNA Preparation Kit (Illumina, San Diego, CA, United States) and multiplex paired-end sequenced at the Genomics Facility of the University of Basel using the illumina HiSeq2500 NGS platform (Illumina, San Diego, CA, United States) with raw read sequence lengths of either 101, 125 or126 nucleotides (nt). Raw sequence data will be made available at European Nucleotide Archive.

5.3.3 Whole genome sequence analysis and variance calling

The raw fastq illumina reads were trimmed of illumina adaptor and low quality reads using Trimmomatic v 0.33 with a sliding window of 5:20 (Bolger *et al.*, 2014). We dropped all reads with read length less than 20 nt and employed the mem algorithm in BWA v0.7.13 (Li & Durbin, 2010) to align the filtered reads to a reconstructed MTBC ancestral sequence obtained from a previous report (Comas *et al.*, 2010). The chromosome coordinates and the annotation used was based on the genome of the laboratory reference strain *M. tuberculosis* H37Rv (NC_000962.3). We excluded also duplicated reads after marking

University of Ghana http://ugspace.ug.edu.gh

with the Mark Duplicate module of Picard v2.9.1 (https://github.com/broadinstitute/picard). SNPs were called with mpileup implemented in Samtools v1.2 (Li, 2011) and VarScan v2.4.1 (Koboldt *et al.*, 2012). We used a quality threshold score of 20 for both minimum mapping quality and minimum base quality. Sample genotypes were called using the majority allele (SNPs were considered to have reached fixation within an isolate with a minimum frequency of 90%) in positions supported by at least 7-fold coverage; otherwise we classified them as missing genotypes. On the other hand, the ancestor state was called when the SNP within-isolate frequency was ≤10%. We classified a genome as a possible mixed infection or contamination if it had more than 120 heterogeneous base calls**.** All SNPs were annotated using snpEff v4.11 (Cingolani *et al.*, 2012) with H37Rv reference annotation (NC_000962.3). We excluded genome positions in highly repetitive and variable regions (PE/PPE genes), phages, insertion sequences and regions with at least 50 bp identities to other regions in the genome (Stucki *et al.*, 2016). After all the filtering steps, we also additionally excluded genomes with average coverage lower than 15x. Our final dataset consisted of 452 genomes.

5.3.4 Phylogenetic analysis

All 452 genomes that passed the filtering steps were selected to produce a multifasta alignment file containing only polymorphic sites using customized python scripts. A position was considered polymorphic if at least one genome had a SNP at that position. We excluded genome positions with > 10% missing calls. The GTR-GAMMA model with 1,000 rapid bootstrap inferences followed by a thorough maximum-likelihood search performed in CIPRES (Miller *et al.*, 2010) was used to infer a maximum likelihood

University of Ghana http://ugspace.ug.edu.gh

phylogenetic tree using the MPI parallel version of RaxML v8.2.3 (Stamatakis, 2014) on the multi-fasta alignment file. Each tree constructed was based on the best-scoring maximum likelihood topology and rooted with *M. canettii*. Phylogenetic trees were plotted and annotated using the ggtree package (Yu *et al.*, 2018; Yu *et al.*, 2017) and graphics enhanced using ggplot2 (Wickham, 2016) all implemented in R (R, 2019) (http://cran.r-project.org/). We calculated pairwise SNP distances between sequences using the ape package (Paradis & Schliep, 2019) implemented in R (R, 2019).

5.3.5 Cluster definition and analysis

Clustering analysis was based on the assumption that strains with the same DNA fingerprint may be epidemiologically linked and associated with recent TB transmission (Hall, 1996). Only one strain per participant was included in the analysis. A strain was defined as an MTBC isolate with a unique genotypic signature having a unique genomic content with not more than 12 SNPs shared with another. Based on proposed SNP thresholds from various studies, three genomic cluster definition were explored; 1) a cut off at 5 SNPs (Walker, Ip, *et al.*, 2013), 2) a cut off at 10 SNPs (Guerra-Assunção *et al.*, 2015) and 3) a cut off at 12 SNPs (Walker *et al.*, 2014; Yang *et al.*, 2017). Using the multi-fasta file and the cluster package (Maechler *et al.*, 2019) implemented in R (R, 2019), we set the three thresholds of 5, 10 and 12 and generated separate datasets containing list of clusters per SNP threshold specified. We used these set of cluster data to annotate the phylogenetic tree generated and perform further downstream analysis on selected clustered cases after sticking to a specific cluster definition. The size of a cluster was defined using the total number of genomes in the cluster classified into categories of small (2 genomes), medium (3–5 genomes) and large (>5 genomes).

University of Ghana http://ugspace.ug.edu.gh

5.3.6 Recent TB transmission rate estimation

The recent TB transmission rate was estimated using the n – 1 formula described by Glynn et al., 1999 ; $\frac{(nc-c)}{N}$ where nc is the total number of clustered cases, c is the number of clusters, and N is the total number of cases in the sample (Glynn *et al.*, 1999).

Our sample set for this current study constituted a biased collection of only traditional genotype clustered cases. In order to estimate a recent transmission rate from the WGS clustering analysis, we first estimated the number of single cases from the previous population as follows;

> If the previous population size of 2,309 isolates resulted in an estimated traditional genotypic recent transmission rate of 41.2% from 1,227 clustered isolates of 276 clusters (thus, ((1227-276)/2309) = 41.2%), then a selected traditional genotypic clustered cases of 452 strains from 40 clustered cases will have been drawn from a population size of N given by;
>
> $$\frac{(nc-c)}{N} = 41.2\%$$
>
> But nc = 452 and c = 40, hence
>
> $$\frac{(452-40)}{N} = \frac{41.2}{100}$$

Computing the above gives N = 1,000 isolates.

University of Ghana http://ugspace.ug.edu.gh

5.3.7 Mutation rates and within-host micro-evolution of longitudinal isolates

In addition to the clustered cases, we whole genome sequenced three randomly selected MTBC isolate pairs obtained from three longitudinal cases. Isolates from these three cases belong to the three most dominant lineage/sub-lineages found in Ghana being Cameroon, Ghana and MAF West African 1. These isolates were investigated for within-host micro-evolution. Here, we calculated pairwise SNP distances between each pair of sequence from the same case using Mega v10.0.5 (Kumar *et al.*, 2018).

5.3.8 Data management and analysis

We included in our analysis both molecular and epidemiological data. Analysis that involved statistical inferences was carried out using the Stata statistical package version 14.2 (Stata Corp., College Station, TX, USA). The chi squared test or the Fischer's exact test was used for significant test where appropriate. p-values of <0.05 were considered significant. The GIS co-ordinates of the participants' self-reported district of residence was used to construct a spatial representation of the MTBC isolates using the ArcMap employed in ArcGIS (Economic and Social Research Institute, version 10.1) (ESRI, 2010). The GIS coordinate information was combined with the genomic, epidemiological and other demographic data to analyze for risk factors for clustering.

This study is reported according to the Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID) guidelines (Field *et al.*, 2014).

University of Ghana http://ugspace.ug.edu.gh

5.4 Results

5.4.1 Characteristics of the study population

Out of the 473 isolates, each from a single case, 452 (95.6%) had good sequences and hence was used for downstream analysis (Figure 5.1). Of the passed genomes, 71.4% (319/447) were isolates from males and 28.6% (128/447) from females and the median age was 35 years (range 27 to 45 years). Five participants had no record of gender. A large proportion of the participants (73.6%, 315/428) had a smear grade of at least 2+. Smear grade of 1+ and scanty was found in 18.0% (77/428) and 8.4% (36/428) respectively.

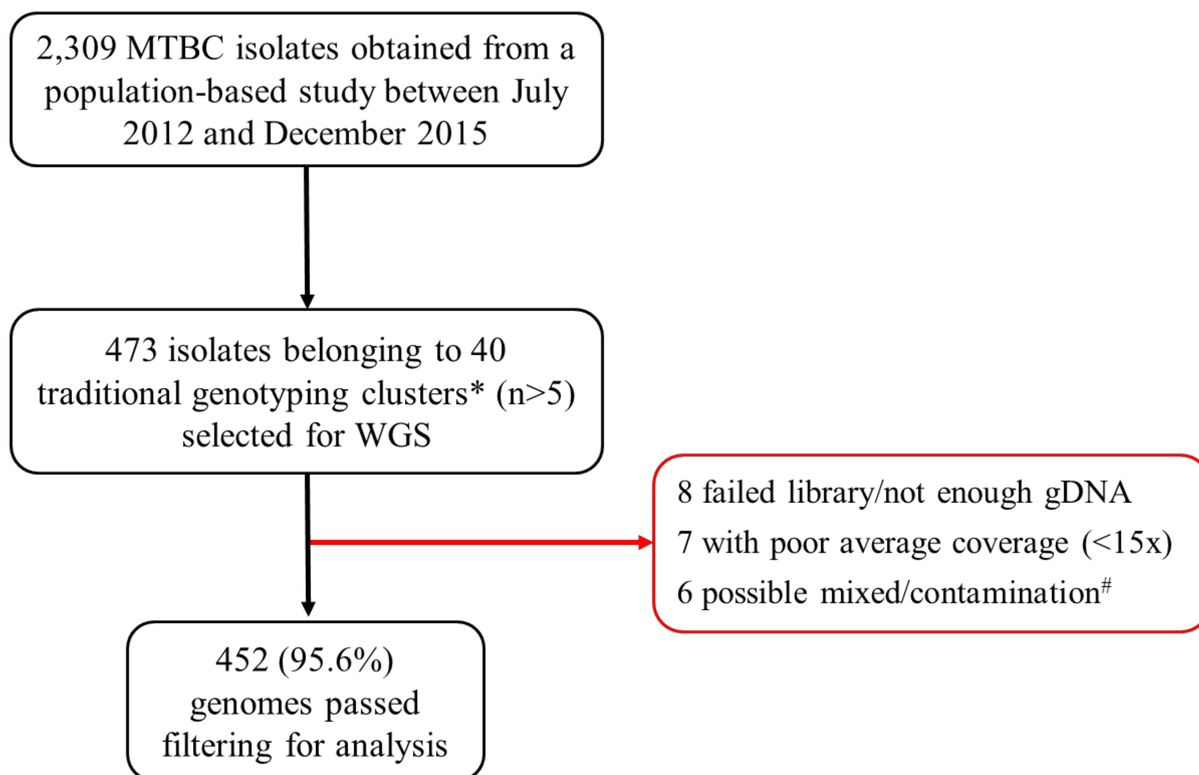


Figure 5.1. Pipeline for sample selection and whole genome sequencing.

*We defined traditional genotyping clusters as previously described clusters identified using mycobacterial interspersed repetitive-unit-variable-number tandem-repeat analysis and spoligotyping tools

#Genomes with heterozygous SNPs greater than 120 were classified as possible mixed infection or contamination and hence removed from further analysis

University of Ghana http://ugspace.ug.edu.gh

5.4.2 Single nucleotide polymorphism threshold for clustering analysis

Three longitudinal TB cases were randomly chosen for within host micro-evolution analysis. The three cases had two isolates each, belonging to the Cameroon (FU080), Ghana (FU049) and MAF West African 1 (FU031) genotypes. All three cases received the same set of anti-TB drugs (isoniazid, rifampicin, ethambutol and pyrazinamide). Case FU080 was a male, 39 years of age, diagnosed with a smear grade of 2+ and the follow-up sample was taken at month five (153 days) of treatment (Figure 5.2). Case FU049 was a female, 33 years of age and diagnosed with a smear grade of 3+ but sputum smear grade of scanty 3 at 49 days of follow-up. Case FU031 was a male, 51 years of age and diagnosed with a smear grade of 3+ and had a smear grade of 1+ at 175 days of follow-up. The SNP distances between each genome pair is shown in Figure 5.2. On average, there were 1.3 (4/3) SNPs accrued in 126 days ((153+49+175)/3). This implies that within three years it is possible to accrue approximately 11 SNPs if all things be equal. Consequently, our analysis and inferences were based on a 10 SNP cut off.

The traditional genotype clusters were found to form close to distinct monophyletic clades upon reconstructing the phylogenetic tree using WGS data (supplementary Figure S5.1). Some monophyletic clades however consisted of more than one large traditional genotype cluster. All 452 genomes were broadly grouped into the three main phylogenetic lineages found in Ghana (lineages 4, 5 and 6) (Figure 5.3, supplementary Figures S5.2 and S5.3). Whereas no genomic cluster was observed for lineage 5, we observed three small clusters for lineage 6 and several clusters for lineage 4. We identified 67 clusters with a median cluster size of 7.5 (range 4 to 12) and total clustered cases of 314 individuals. Eight large clusters were observed with the largest cluster consisting of 78 genomes (Figure S5.3). All

University of Ghana http://ugspace.ug.edu.gh

the large clusters were supported with a bootstrap value of 100. The estimated clustering rate (recent transmission rate) was 24.7% (Figure 5.4). In addition to the SNP threshold at 10, we explored also SNP thresholds at 5 and 12 (supplementary Figures S5.2 and S5.3).

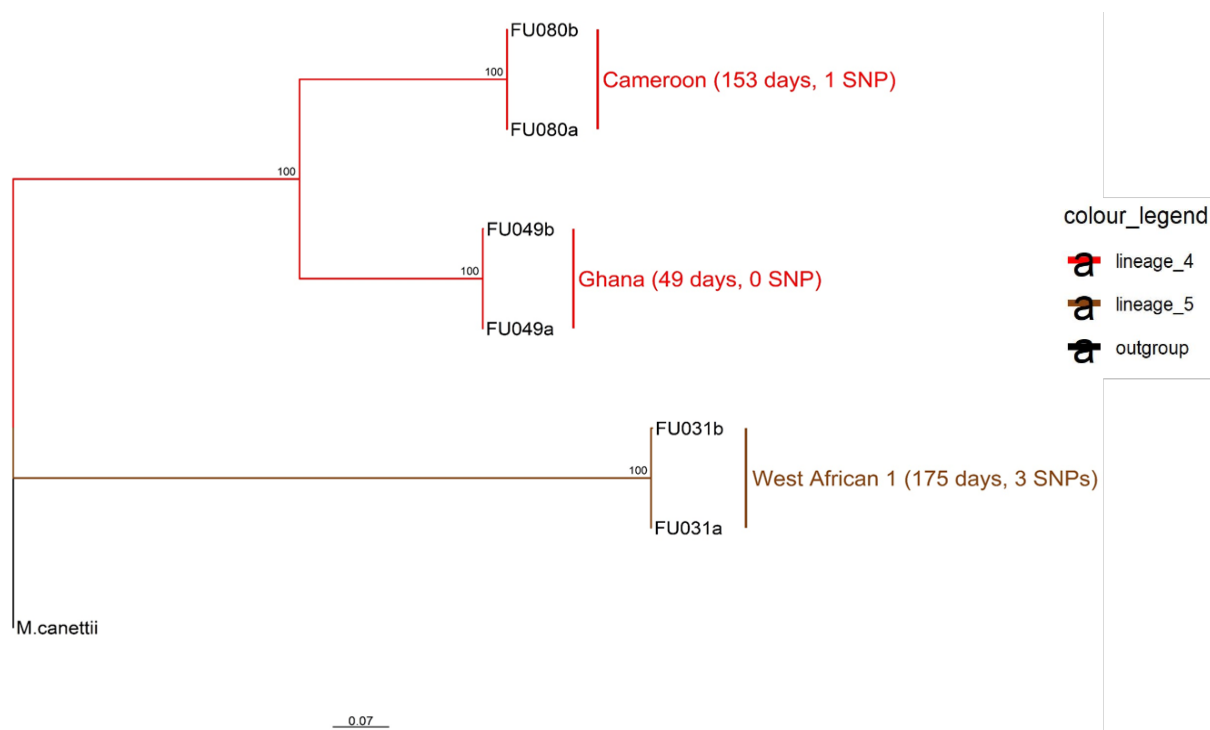


Figure 5.2. Phylogenetic relationship of 6 *M. tuberculosis* complex longitudinal isolates obtained from three participants. The two major branches constitute the top two MTBC lineages found in Ghana and color coded with the universally accepted Gagneux-defined lineages as red for lineage 4 and brown for lineage 5. Lineage 4 is further split into the dominant Cameroon and Ghana sub-lineage. The sub-lineages, days between sampling for each participant as well as the number of SNPs observed between sample pairs are annotated on the tree. This analysis involved 7 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 2,104 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018). Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*.

University of Ghana http://ugspace.ug.edu.gh

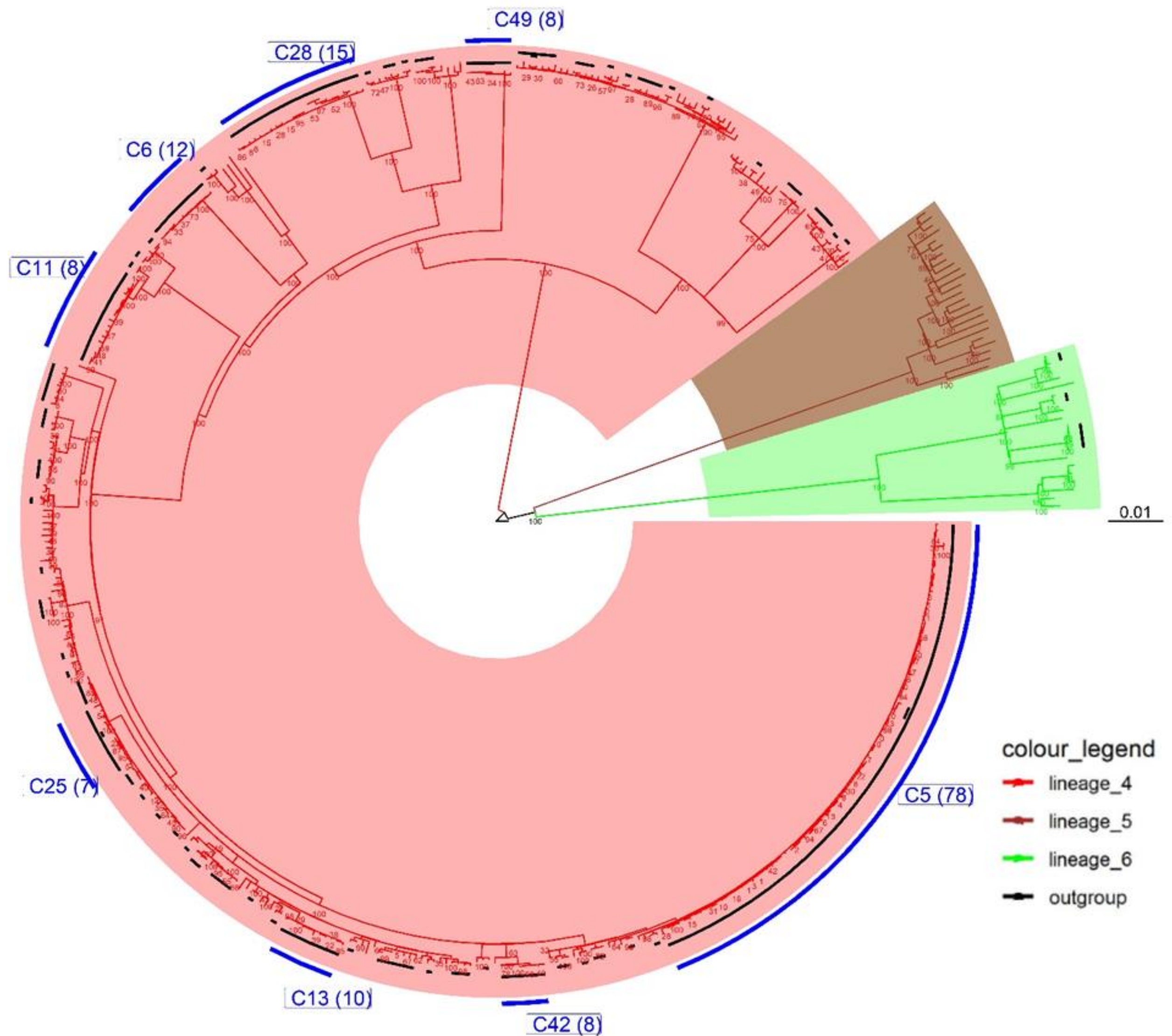
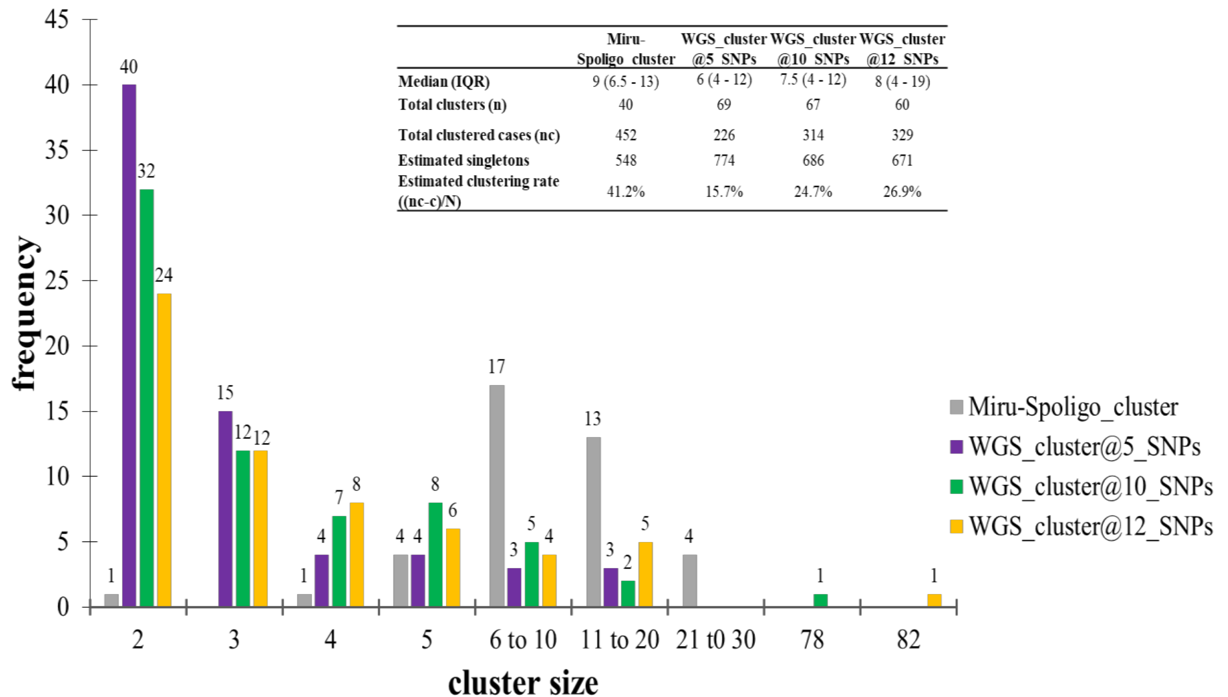


Figure 5.3. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 10-SNPs. The tree was built with an alignment file containing 11,041 variable positions. Black bars plotted on the tips of the branches indicate the clustered cases at the defined threshold of 10-SNPs. Blue bars represent large clusters (cluster size >5) with the number of clustered cases indicated in brackets. The three major branches constitute the three main MTBC lineages found in Ghana and color coded with the universally accepted Gagneux-defined lineages as red for lineage 4, brown for lineage 5 and green for lineage 6. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*.

University of Ghana http://ugspace.ug.edu.gh



	Miru-Spoligo_cluster	WGS_cluster @5_SNPs	WGS_cluster @10_SNPs	WGS_cluster @12_SNPs
Median (IQR)	9 (6.5 - 13)	6 (4 - 12)	7.5 (4 - 12)	8 (4 - 19)
Total clusters (n)	40	69	67	60
Total clustered cases (nc)	452	226	314	329
Estimated singletons	548	774	686	671
Estimated clustering rate ((nc-c)/N)	41.2%	15.7%	24.7%	26.9%

Figure 5.4. Clustering analysis stratified by cluster definition. The estimated population size (N=1,000) used for estimating the clustering rate are explained in methods. Singletons were calculated from the estimated population size of 1,000 individuals. IQR: interquartile range.

5.4.3 Recent transmission hotspots and characteristics of large clusters

A total of 146 genomes constituting eight large clusters (cluster size > 5) duly annotated (Table 5.1, Figure 5.2 and 5.5) were observed from the clustering analysis following a cut off at 10 SNPs. The smallest of these large clusters had a cluster size of 7 (Whole genome sequence cluster 25; WGSC-25) with the largest cluster having a cluster size of 78 (WGSC-5). The number of genomes in this large cluster (WGSC-5) forms a quarter of all clustered cases (78/314=25%). All the large clusters belonged to lineage 4 with Cameroon sub-lineage predominating (WGSC-5, WGSC-13, WGSC-25 and WGSC-42) followed by

University of Ghana http://ugspace.ug.edu.gh

the Ghana sub-lineage (WGSC-11 and WGSC-28). The two remaining large clusters belonged to the Haarlem (WGSC-6) and LAM (WGSC-49) sub-lineages. Not more than 138 variable SNPS were observed between these large clusters. The median of median pairwise SNP distance between these large clusters was 5 SNPs. With the exception of 7 isolates, all remaining 139/146 isolates were sensitive to both isoniazid and rifampicin. Interestingly, 5/7 INH resistant isolates belonged to the same cluster (WGSC-11) and were from individuals residing in the same district (Ayawaso district) (Figures 5.5 and 5.6). Only two isolates belonging to WGSC-5 were resistant to INH. The ratio of male to female among the clustered cases was confirmed to be significantly higher (3:1, 231/81) compared to the general population (2:1) and twice as much among large clusters (4:1, 115/29) ($p<0.05$). Two participants had no record of gender. Two large clusters (WGSC-25 and WGSC-42) were made up of only males.

All cases belonging to large clusters spanned the entire 3.5 years sampling period and were distributed among 20 districts/sub-districts but generally clustered within Accra metropolis (Figure 5.5 and 5.6) which is made up of 6 sub-districts. Most of the large clusters exhibited a geographic distribution even though not exclusive. For example, whereas hotspot for WGSC-11 and WGSC-42 was the Ayawaso sub-district, WGSC-13 and WGSC-5 were found mostly from the Ablekuma sub-district, the main identified hotspot of recent transmission (Figure 5.7 and supplementary Table S5.1).

5.4.3.1 Socio-demographic characteristics of individuals infected with a strain from the largest cluster (WGSC-5)

This largest transmitting cluster made up of 78 cases exhibited an interesting geographical distribution. Except for two cases from Northern Ghana, all 76/78 cases in this cluster were

University of Ghana http://ugspace.ug.edu.gh

from Southern Ghana of which 19 were found in Ablekuma sub-district (Figures 5.5, 5.6b and 5.7). The two cases from Northern Ghana shared no SNP difference between them. One case had no record of residential location. There were 59 males and 17 females with a median age of 34 (IQR, 24 - 43). Two participants had no record of gender. A greater proportion (77.8%, 42/54) of individuals responded living in compound houses at city suburb (66.1%, 37/56) with an average monthly income of not more than 300 Ghanaian cedis (92.8%, 52/56) or 60 USD in its equivalence. The median number of individuals living in a giving household was 12 (IQR, 5 - 20). On average, there were more unskilled laborer's (60.7%, 34/56) than skilled laborer's (16.1%, 9/56) with the remaining 23.2% (13/56) being unemployed including students.

Table 5.1. Characteristics and risk factor analysis of large genomic clusters resulting from a threshold of 10 SNPs

Number	WGS cluster code	Number of cases in cluster	Number of variable fixed SNPs	Median pairwise SNP (IQR)	Lineage (sub-lineage*)	Lineage classification by stucki/coll	¤Any drug resistance	Gender male: female	Median age (IQR)
1	WGSC-5	78	138	7 (6 - 9)	L4 (Cameroon)	L4.6.2/L4.6.2.2	2	59:17	34 (24 - 43)
2	WGSC-28	15	39	7 (5 - 7)	L4 (Ghana)	L4.10/L4.8	ND	11:4	39 (32 - 51)
3	WGSC-6	12	18	5 (3- 6)	L4 (Haarlem)	NA/L4.6	ND	11;1	38 (28 - 48)
4	WGSC-13	10	23	5 (5 - 6)	L4 (Cameroon)	L4.6.2/L4.6.2.2	ND	8:2	42.5 (32 - 49)
5	WGSC-11	8	22	8 (6.5 - 8.5)	L4 (Ghana)	NA/L4.6.2	5	5:3	32.5 (28 - 41.5)
6	WGSC-42	8	6	3 (1 - 3.5)	L4 (Cameroon)	L4.6.2/L4.6.2.2	ND	8:0	25.5 (22.5 - 28.5)
7	WGSC-49	8	13	4 (1 - 7.5)	L4 (LAM)	L4.3/L4.3.1	ND	6:2	42 (32 - 54)
8	WGSC-25	7	16	4 (3 - 9)	L4 (Cameroon)	L4.6.2/L4.6.2.2	ND	7:0	39 (28 - 50)

WGS, Whole genome sequence; L4, lineage 4; ND, none determined; IQR, interquartile range.
* sub-lineage defined using spoligotyping
¤ Number of participants carrying strains with drug resistance to either isoniazid or rifampicin

University of Ghana http://ugspace.ug.edu.gh

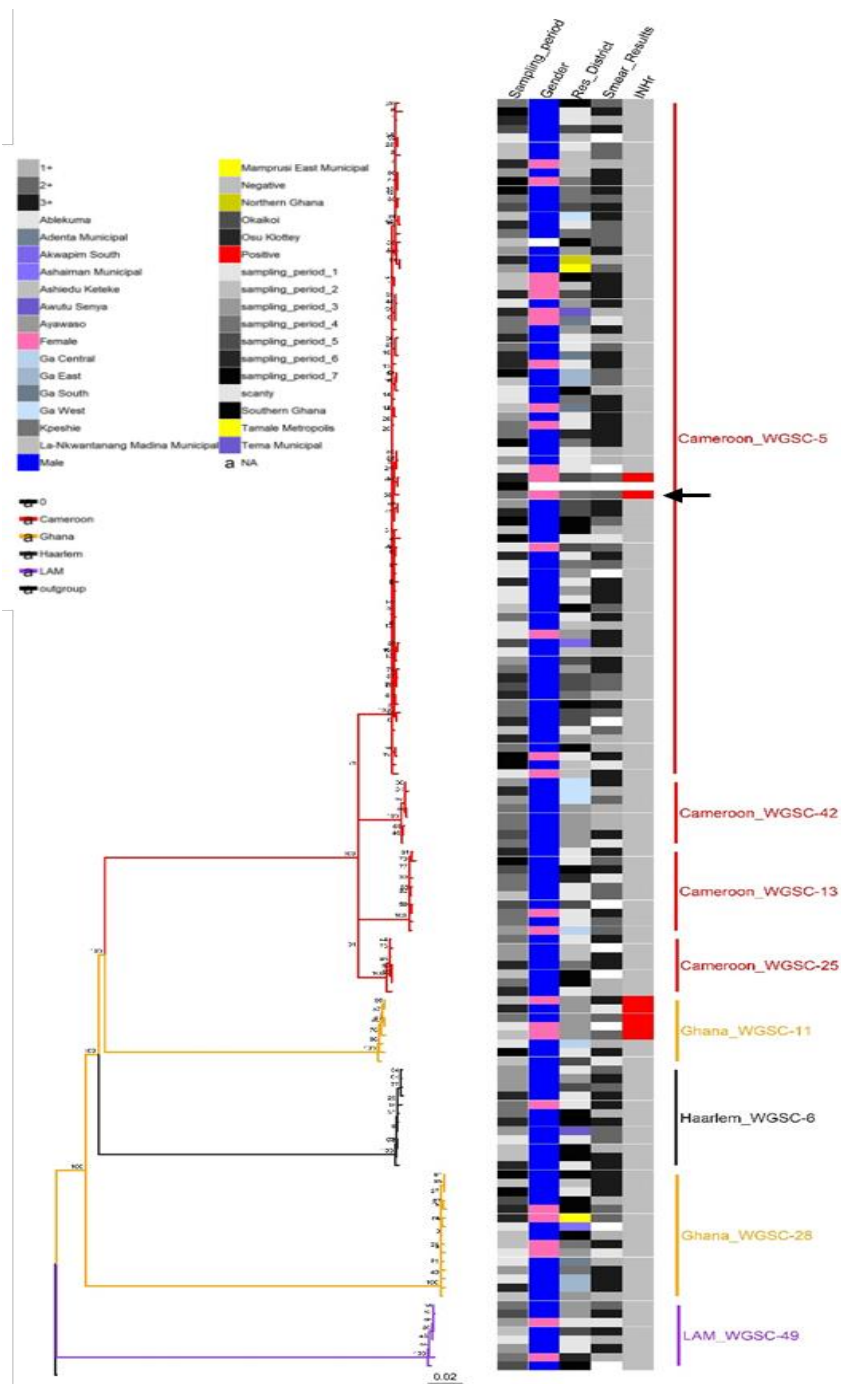


Figure 5.5. Phylogenetic reconstruction of 146 *M. tuberculosis* complex isolates showing characteristics of the 8 identified large clusters as defined by a threshold of 10-SNPs. The heat map shows some characteristics of the clustered cases including sampling period (column 1), gender (column 2), residential district (column 3), smear results (column 4) and drug resistance status to isoniazid (column 5). There was only one rifampicin resistant isolate (black arrow). The color codes are defined in the key. All cases belong to lineage 4. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*.

University of Ghana http://ugspace.ug.edu.gh

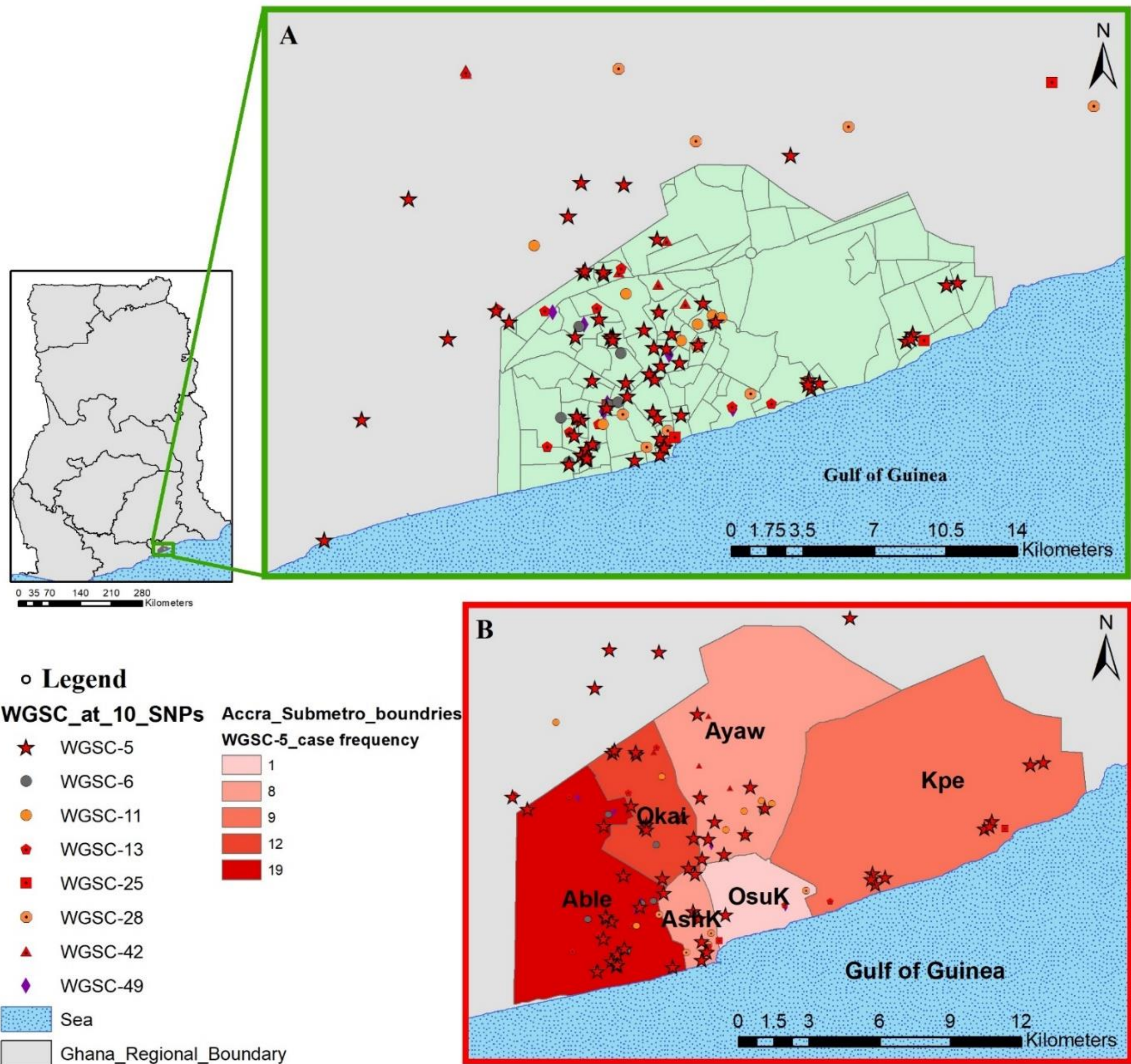


Figure 5.6. Spatial distribution of pulmonary tuberculosis cases belonging to large clusters in and around Accra metropolis. The geographical distribution of cases from all 8 large clusters are shown in 'A' and the relative distribution of cases from the largest cluster highlighted in 'B' stratified by Accra metropolis sub-district. Abbreviations; Able: Ablekuma sub-district, Okai: Okaikoi sub-distict, AshK: AShiedu Keteke sub-district, OsuK: Osu Klottey sub-district, Ayaw: Ayawaso sub-district, Kpe: Kpeshie sub-district.

University of Ghana http://ugspace.ug.edu.gh

	Frequency stratified by cluster							
Residential district	**WGSC_5**	**WGSC_6**	**WGSC_11**	**WGSC_13**	**WGSC_25**	**WGSC_28**	**WGSC_42**	**WGSC_49**
Ablekuma	19	5	1	6	2	1		3
Adenta Municipal						1		
Akwapim South	1							
Ash Keteke	8				2	1		
Ashaiman Municipal						1		
Awutu Senya	1							
Ayawaso	8	1	5			2	5	2
Ga Central			1	1				
Ga East	2					2		
Ga South	3							
Ga West	1						3	
Kpeshei	9				1	2		
La-Nkwantanang Madina	1							
Mamprusi East	1							
Northern Ghana	1							
Okaikoi	12	1	1	1				1
Osu Klottey	1			1				1
Southern Ghana	10	4		1	2	4		1
Tamale Metropolis						1		
Tema Municipal		1						
Total	78	12	8	10	7	15	8	8

Figure 5.7. Frequency and relative geographical distribution of cases in all 8 large clusters identified by a threshold of 10-SNPs.

5.5 Discussion

In this study, our main goal was to use a WGS approach to resolve large traditional genotype clusters (MIRU/Spoligo defined clusters) and explore some epidemiological characteristics including spatial distribution of confirmed large clusters. Major findings from our analysis indicate that: 1) estimated recent TB transmission rate using WGS at a SNP threshold of 10 remains high at 24.7%, 2) there is wide spread of a clone of the Cameroon sub-lineage of lineage 4 with an ongoing transmission at hotspots mostly found within the Ablekuma sub-district of the Accra metropolis.

WGS was first used in 2011 to delineate two unrelated transmission events among a cohort of drug users with identical MIRU-VNTR profiles from Vancouver and ever since it has

University of Ghana http://ugspace.ug.edu.gh

been used in some large studies to understand TB transmission dynamics (Gardy *et al.*, 2011; Lalor *et al.*, 2018; Walker *et al.*, 2017; Walker, Monk*, et al.*, 2013). Despite the continuous progress and decreasing costs of WGS-based typing, there are some important pertaining challenges such as the lack of standardization of WGS analysis pipelines and genomic distances (SNP distance) for defining clusters (Merker *et al.*, 2017). A first step in analyzing WGS data for transmission studies is usually to define SNP threshold to identify cluster and the assumption is that, isolates from cases separated by SNPs less than or equal to the specified threshold are epidemiologically linked (Stimson *et al.*, 2019). The mutation rate established from within-host microevolution analysis using the main lineage/sub-lineage population suggested a cut-off at 11 SNPs, will be adequate to define a cluster. Consequently, we chose a SNP threshold of 10 for our analysis. This chosen threshold is ideal as other similarly high TB transmission settings like Malawi, have used the same threshold to infer recent transmission (Guerra-Assunção *et al.*, 2015). The sampling size used for the within-host micro-evolution was small, we therefore recommend that future studies should confirm this with a larger sample size covering all lineages.

We previously estimated the recent transmission index to be 41.2%, using MIRU/VNTR and spoligotyping which is higher than the current estimate of 24.7% using WGS analysis. This reduced rate was anticipated as the discriminatory power of WGS analysis is higher and other studies have identified reduced rate after using WGS (Stucki *et al.*, 2015). Nevertheless, the 24.7% estimated recent transmission rate is high like 30% from a similarly high transmission setting like Malawi (Guerra-Assunção *et al.*, 2015; Yates *et al.*, 2016) and predicts the occurrence of undetected recent transmission of large clusters. Out

University of Ghana http://ugspace.ug.edu.gh

of the 40 large MIRU/VNTR plus spoligotyping clusters, WGS analysis split most of these large clusters into small and medium size clusters except few large clusters (Figure 5.4). With the exception of three clusters of lineage 6, all the remaining 64 clusters were lineage 4 and no cluster for lineage 5. This finding confirms our previous report of reduced recent transmission of MAF lineages (L5 and L6) compared to MTBss and has stressed the need for studies to investigate the continuous prevalence of MAF in West-Africa. The observation of nearly distinct monophyletic clades from the reconstructed phylogenetic tree implies that traditional genotyping may still be useful as initial screening tools to help reduce the huge cost of WGS of all isolates especially in large-size population-based studies.

We did not identify a cluster consisting of multidrug resistant strains within our study population, confirming our previous report of the unlikeliness of a drug resistant TB strain to be involved in a recent transmission event (Asare *et al.*, 2018). This observation may be due to the low proportion (2 - 4%) of MDR among MTBC isolates in Ghana (Asare *et al.*, 2018; Otchere *et al.*, 2016) or this probably may be due to the reduced fitness cost associated to resistance conferring mutations (Gagneux, 2009; Melnyk *et al.*, 2015). Moreover, only 7/146 cases belonging to large genomic clusters were resistant to INH. Interestingly, 5/7 INH resistant isolates belonged to the same Ghana sub-lineage cluster (WGSC-11). The Ghana sub-lineage has previously been associated with drug resistance (Otchere *et al.*, 2016; Yeboah-Manu *et al.*, 2016). Though the size of the cluster is not very large (cluster size of 8), this is nonetheless worrying since recent transmission of such drug resistant clone may pose a great challenge to TB control in the sub-region. Until recently, drug resistant clones were thought to be less fit and less likely to transmit from person to

University of Ghana http://ugspace.ug.edu.gh

person; however recent studies have documented evidence of transmission even though not involving large clusters (Arandjelovic *et al.*, 2019; Coscolla *et al.*, 2015; Walker *et al.*, 2018). There is therefore the need to identify and control such difficult to treat drug resistant clones to stop their spread.

Our population-based study included two distant regions in Ghana; the Northern region (in Northern Ghana) and the Greater Accra region (in Southern Ghana). With the exception of three cases from Northern Ghana (Figure 5.7, supplementary Table S5.1) all 146 cases belonging to large genomic clusters were found in Southern Ghana. Two cases from Northern Ghana were found within the largest cluster (WGSC-5, Figure 5.5). These cases were however very closely related sharing the same most recent common ancestral node and in fact no SNP difference between them suggesting direct person to person transmission. A careful examination of their demographic data also showed that indeed, these two individuals have the same family name and most probably come from the same family. We show that the clustering of TB cases in Ablekuma sub-district observed in our previous study (Yeboah-Manu *et al.*, 2016) was most probably due to recent TB transmission. This is not surprising as Ablekuma is the most densely populated of the six sub-districts (GSS, 2014). Our analysis suggests that there may be super-spreaders in Ablekuma and probably Okaikoi sub-districts which recorded the second highest numbers (19 and 12 respectively, Figure 5.6 and 5.7) that belonged to the largest cluster (WGSC-5). Majority of the individuals in this high transmitting cluster were found to inhabit the city suburbs in densely populated compound houses. Their low-income status combined with over-crowding may be driving factors for the ongoing transmission in this hotspot. A high smear grade of over 70% of cases being at least 2+ signifies that, these individuals are

University of Ghana http://ugspace.ug.edu.gh

likely to have been actively transmitting the pathogen prior to diagnosis in their homes and neighboring communities indicating that other individuals may have been infected.

The goal of universal screening is what most TB control programs are geared towards especially detecting MDR cases; Our study has identified hotspots not only for recent TB transmission of drug sensitive strains but also spread of INH resistant strain. We encourage more similar studies as it can identify geographical zones of highest need to support the NTP with a targeted and guided approach to controlling TB. Case search approaches targeted at high risk areas may be more effective in TB control (Zelner *et al.*, 2016). We have shown that application of WGS in a molecular epidemiology study has aided the recognition of specific *M. tuberculosis* strains (e.g., cluster WGSC-11 associated with drug resistance) which can be predictive of INH drug-resistant TB in the Ghanaian contexts that could and can help provide indications of the TB case source similar to TB strains elsewhere (Varghese *et al.*, 2013). Also, we have been able to identify hotspots of recent TB transmission within the Accra Metropolitan Area, hence, we recommend that an urgent action to curtail the continual spread of the pathogen.

University of Ghana http://ugspace.ug.edu.gh

CHAPTER 6

6.0 Manuscript 4. Whole Genome Sequencing and Epidemiological Analysis Identifies High Relapse Among Individuals with Recurring Tuberculosis in Ghana

6.1 Abstract

This study aimed to investigate the cause of individuals presenting with recurring tuberculosis (*rc*TB) and to confirm transmission among epidemiologically linked (same household) TB cases within a prospective population-based study between July 2012 and December 2015. Isolates were characterized by IS*6110* PCR, spoligotyping, mycobacterial interspersed repetitive unit - variable number of tandem repeat (MIRU-VNTR) typing and whole genome sequencing (WGS), followed by phylogenetic analysis to assess strain relatedness. In all, 99 MTBC isolates obtained from 47 paired TB cases were analyzed; 36 cases from *rc*TB episodes and 11 from epidemiologically linked cases. The Cameroon sub-lineage was the dominant sub-lineage in both cohorts followed by the Ghana sub-lineage. A greater proportion (94.4%, 34/36) of individuals with *rc*TB episodes were males and 58.6% (21/36) had TB recurrence within 12 months post treatment. Only 19.4% (7/36) of participants with *rc*TB harbored a strain with isoniazid (INH) resistance at baseline of which 29% (2/7) were additionally resistant to rifampicin. However, 27.8% (10/36) harbored an INH resistant strains upon recurring of which 40% (4/10) were MDR-TB strains. Recurrent TB was attributed to relapse (same strain) in 75.0% (27/36) of participants with 25.0% (9/36) attributed to re-infection. The relapse cases were separated by not more than 7 SNPs. Epidemiologically linked TB cases were likely the results of recent TB transmission within the house or from neighboring recent transmission events.

University of Ghana http://ugspace.ug.edu.gh

Finally, we show that though WGS is ideal, 15-locus MIRU-VNTR typing is enough to predict both the cause of TB recurrence and household related transmission at a high agreement compared to WGS (84.6%, Kappa=0.7702, p<0.001). Our findings indicate that unresolved previous infection due to inadequate treatment may be the cause rcTB which can contribute to the development of drug resistant strains.

6.2 Introduction

Tuberculosis (TB) remains a major global public health threat (WHO, 2018b) as such more efforts are needed to deal with this global menace. In addition to prompt diagnosis and provision of appropriate treatment, TB control requires understanding of the transmission dynamics and factors that influence occurrence of the disease among different populations to guide implementation of control strategies. Despite the effective use of the combination therapy employed in the directly observed treatment short course (DOTs) regimen since 1993 (Alipanah *et al.*, 2018; Frieden & Sbarbaro, 2007), some previously treated patients still turn up with secondary cases of the disease, here referred to as recurring TB. This recurring TB falls into one of two categories; either being a relapse case (endogenous reactivation of a previous infection) or a re-infection case (exogenous re-infection). The former occurs in patients who may not have been adequately treated during their primary disease. Situations where the patient was adequately treated but still presents with a second episode implies that the current episode is due to re-infection especially among HIV co-infected patients and those living in high TB burdened countries (Narayanan *et al.*, 2010; Verver *et al.*, 2005). Whereas reinfection has been identified as the principal cause of recurring TB in high TB burden areas (Parvaresh *et al.*, 2018; van Rie *et al.*, 1999) relapse

University of Ghana http://ugspace.ug.edu.gh

may be the likely result of a poor prognosis and an unsuccessful treatment (Alipanah *et al.*, 2018; McIvor *et al.*, 2017) and has been associated with drug resistance (Yang *et al.*, 2017).

Differentiating between the two causes of recurring TB has traditionally been done using genotyping tools including IS*6110*-RFLP, mycobacterial interspersed repetitive-unit variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping. Nevertheless, whole genome sequencing (WGS) is increasingly being used in recent times. Whereas the traditional methods have been used extensively for determination of strain relatedness (Velayutham *et al.*, 2018; Zong *et al.*, 2018) WGS, which is considered the ultimate tool for strain differentiation however has not been explored much probably due to cost and technological demands (Jagielski *et al.*, 2016; Jagielski *et al.*, 2014). Whole genome sequence data does not only give us the power to identify recent TB transmission between individuals of the same households but also the ability to trace the route/direction of transmission between such epidemiologically linked TB cases (Walker, Ip*, et al.*, 2013).

Based on direct smear microscopy, TB treatment success rate in Ghana is estimated at 85% (Amo-Adjei & Awusabo-Asare, 2013; WHO, 2018c); we therefore hypothesize that recurring TB episodes in Ghana are driven mainly by re-infection but not relapse due to unsuccessful treatment of previous episodes. The aim of this study was to identify and delineate the occurrence of recurring TB episodes and to confirm transmission among epidemiologically linked cases in distinct Ghanaian communities using whole genome sequencing.

University of Ghana http://ugspace.ug.edu.gh

6.3 Materials and methods

6.3.1 Ethical considerations

The Scientific and Technical Committee and the Institutional Review Board of the Noguchi Memorial Institute for Medical Research, University of Ghana (FWA00001824) reviewed and approved all protocols and procedures for this study. The objectives and procedures of the study were made clear to each participant using a printed consent form and verbally communicated in a language familiar to the participant. Methods for sputum sampling conformed to WHO guidelines (two sputa per patient) and patients' identity was protected. At the time of the study, drug susceptibility results were made available to clinicians upon request.

6.3.2 Study design and population

The study was a retrospective analysis of isolates and data obtained from a population-based study that recruited more than 90% of TB cases from the Accra Metropolitan Area (urban/south) and the Mamprusi East (rural/north) for more than three years (Figure 6.1) (Asante-Poku *et al.*, 2016; Asare *et al.*, 2018; Yeboah-Manu *et al.*, 2016). All individuals who presented with more than one TB episode (> 6 months between visits) following treatment (regardless of the treatment outcome) were included in this study, here referred to as recurring TB cases. We also additionally included all cases belonging to the same household (epidemiologically linked) for households with more than one TB case; herein referred to as household related TB cases. Clinical and socio-demographic characteristics of the cases were obtained by reviewing medical records and using a detailed questionnaire after obtaining written informed consent from each participant. We obtained permission from guardians of participants below the age of 18 years. Participants' information

University of Ghana http://ugspace.ug.edu.gh

obtained included age, sex, ethnicity, marital status, occupation, clinical history including co-morbid illness with HIV or diabetes, TB contacts, sputum smear grading among others. All participants were treated following the DOTs regimen using a cocktail of four first line antibiotics; isoniazid (INH), rifampicin (RIF), ethambutol (EMB) and pyrazinamide (PZA).

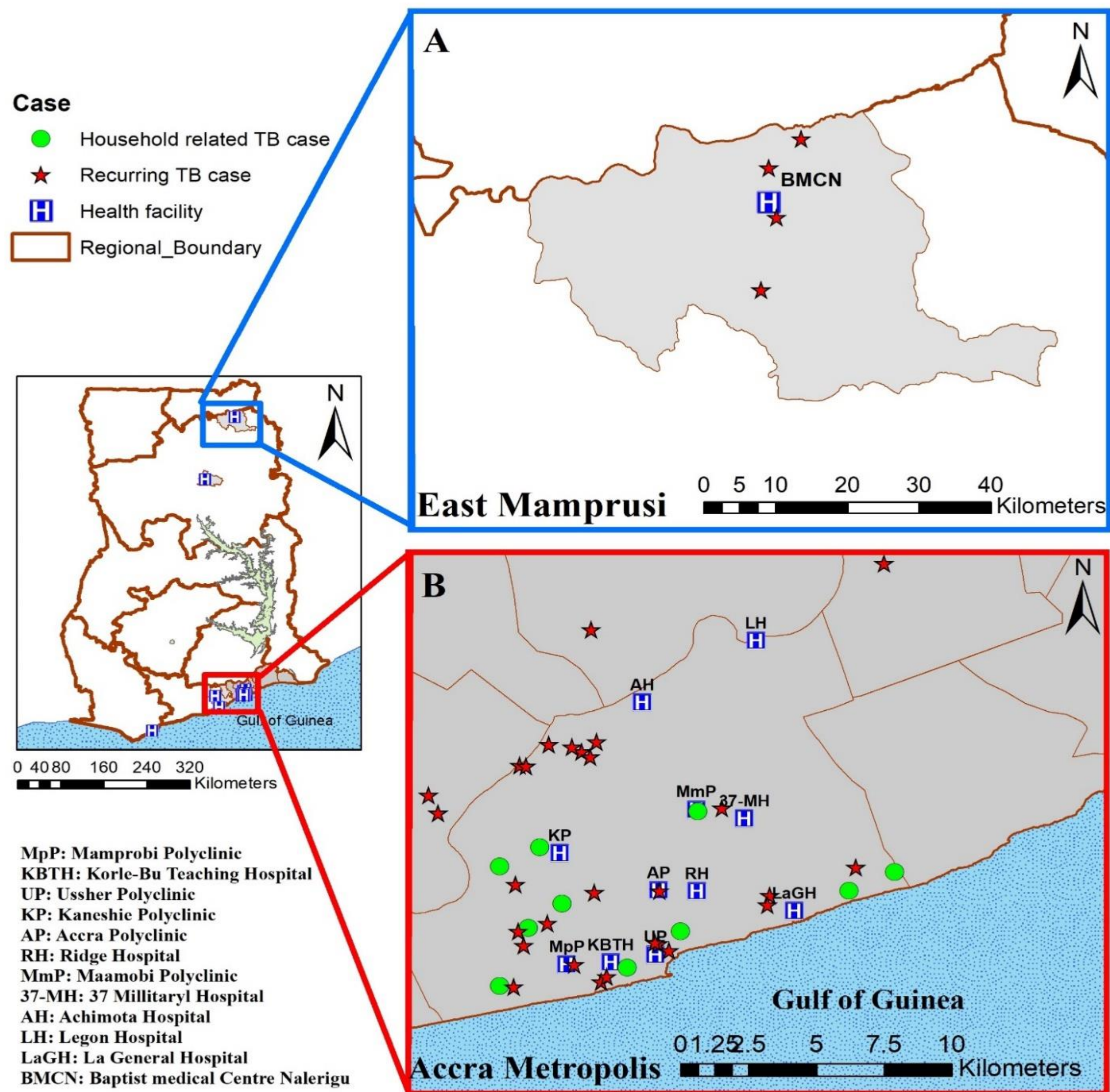


Figure 6.1. Relative geographical location of 13 health facilities and distribution of recurring and household related TB cases. Participant recruitment was carried out in two areas, the East Mamprusi District (A) and Accra Metropolitan Area (B).

University of Ghana http://ugspace.ug.edu.gh

6.3.3 Mycobacterial isolation, characterization and DNA extraction

Mycobacterial isolates were obtained from sputum samples cultured on Lowenstein Jensen media following a 5% oxalic acid treatment and incubation at 37 ºC. After verification of acid fast bacilli through smear microscopy, each isolate was further characterized using standard procedures including PCR detection of the insertion sequence IS*6110,* single nucleotide and large sequence polymorphism typing (SNP/LSP), spoligotyping and as previously described (Yeboah-Manu *et al.*, 2016; Yeboah-Manu *et al.*, 2001). Whole genomic DNA extraction using the CTAB protocol was performed towards whole genome sequencing as reported (Otchere *et al.*, 2016). The only addition to the previous protocol was that, to obtain enough intact genomic DNA (gDNA), harvested mycobacterial cells were heat inactivated at 80 ºC for 30 minutes (instead of the previous 95 ºC for 1 hour) in cell lysis buffer.

6.3.4 Conventional strain typing and drug susceptibility testing

Conventional strain typing was performed using the standard 15-MIRU loci set (Supply *et al.*, 2006) as previously described (Asare *et al.*, 2018). Briefly, we performed a stepwise typing by first interrogating a customized 8 MIRU loci set (Asante-Poku, Nyaho*, et al.*, 2014) following which clustered strains were further delineated at the remaining 7/15 MIRU loci. PCR products were examined on standard 2% agarose gel electrophoresis with each run well controlled. Two or more strains were defined as being the same if they had identical allelic repeat numbers at all 15 MIRU loci and described as similar if they varied at a maximum of one locus. *In vitro* drug susceptibility to INH and RIF was determined using either the microplate Alamar Blue cell viability assay, as reported (Franzblau *et al.*, 1998; Otchere *et al.*, 2016), and/or the GenoType MTBDRplus assay (Hain Lifescience),

University of Ghana http://ugspace.ug.edu.gh

following the manufacturer's protocol (Barnard *et al.*, 2008). In addition to the *in vitro* tests, *in silico* analyses were performed using the TBprofiler package (Coll *et al.*, 2015) to detect mutations associated with drug resistance and also report on other mutations present in drug resistant associated genes.

6.3.5 Whole genome sequencing and analysis

Genomic DNA samples were outsourced to the genomics facility of the University of Basel for whole genome sequencing by the illumina platform after library preparation using the NEXTERA XT DNA Preparation Kit (Illumina, San Diego, CA, United States). Multiplexed paired-end sequencing was performed using the HiSeq2500 (Illumina, San Diego, CA, United States). The sequence reads were quality controlled and analyzed using previously described customized procedures (Brites *et al.*, 2018; Otchere *et al.*, 2018). Briefly, reads that were cleaned and filtered using Trimmomatic v 0.33 (Bolger *et al.*, 2014) and Mark Duplicate module of Picard v2.9.1 (https://github.com/broadinstitute/picard) respectively, were aligned to the reconstructed MTBC ancestral sequence (Comas *et al.*, 2010) using the mem algorithm in BWA v0.7.13 (Li & Durbin, 2010). Samtools v1.2 (Li, 2011) and VarScan v2.4.1 (Koboldt *et al.*, 2012) were used for calling SNPs using the mpileup algorithm and quality threshold score of 20 for minimum mapping and base quality. SnpEff v4.11 was used for annotating SNPs (Cingolani *et al.*, 2012). Genome positions in highly repetitive and variable regions (PE/PPE genes), phages, insertion sequences and regions with at least 50 bp identities to other regions in the genome were excluded (Stucki *et al.*, 2016). Genomes with average coverage less than 15X were excluded from further analysis.

University of Ghana http://ugspace.ug.edu.gh

6.3.6 Phylogenetic reconstruction

Using customized python scripts, we produced multifasta alignment files separately for genomes from recurring TB cases and household related cases. Only polymorphic positions were included for phylogenetic reconstruction analysis after excluding genomic positions with > 10% missing calls. The GTR-GAMMA model with 1,000 rapid bootstrap inferences followed by a thorough maximum-likelihood search performed in CIPRES (Miller *et al.*, 2010) was used to infer a maximum likelihood phylogenetic tree using RaxML v8.2.3 (Stamatakis, 2014). All phylogenetic trees were reconstructed and annotated using the ggtree package in R (Yu *et al.*, 2018; Yu *et al.*, 2017) and graphics enhanced using ggplot2 (Wickham, 2016) also implemented in R (R, 2019) (http://cran.r-project.org/). Pairwise SNP distances were calculated between each pair of genomes from the same patient using Mega v10.0.5 (Kumar *et al.*, 2018).

6.3.7 Case definitions

A case was defined as relapse when MTBC isolates recovered from both episodes had a maximum of one allelic difference in their MIRU profile and 10 or less SNPs difference between their respective genomes (Guerra-Assuncao *et al.*, 2015; Walker, Ip*, et al.*, 2013). Conversely, we defined re-infection when there was more than one allelic repeat difference in MIRU profiles and greater than 50 SNPs between genomes.

6.3.8 Data analysis and epidemiology

Data obtained using the structured questionnaire were double examined for completeness and consistency and entered in Microsoft Access. All statistical analysis was performed using the Stata statistical package version 14.2 (Stata Corp., College Station, TX, USA) and run with significance level pegged at $p < 0.05$ using Fischer's exact test. We used the

University of Ghana http://ugspace.ug.edu.gh

Kappa test to test for concordance between typing methods. The ArcMap tool employed in ArcGIS (Economic and Social Research Institute, version 10.1) (ESRI, 2010) was used for constructing maps.

The study is reported according to the Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID) guidelines (Field *et al.*, 2014).

6.4 Results

We included in our analysis 99 MTBC isolates from 47 epidemiologically linked TB cases of which 36 cases (75 isolates) had recurring TB and 11 (26 isolates) were involved in suspected household related transmission (Figure 6.2). One individual (2 isolates) involved in a household related TB case also had recurring TB hence was included in both analyses. Fifteen isolates were dropped from WGS analysis due to poor gDNA yield (7), poor average genome coverage (6), failed library (1) or possible mixed infection (1). All the 99 isolates (47 cases) were included for downstream analysis (Figure 6.2).

University of Ghana http://ugspace.ug.edu.gh

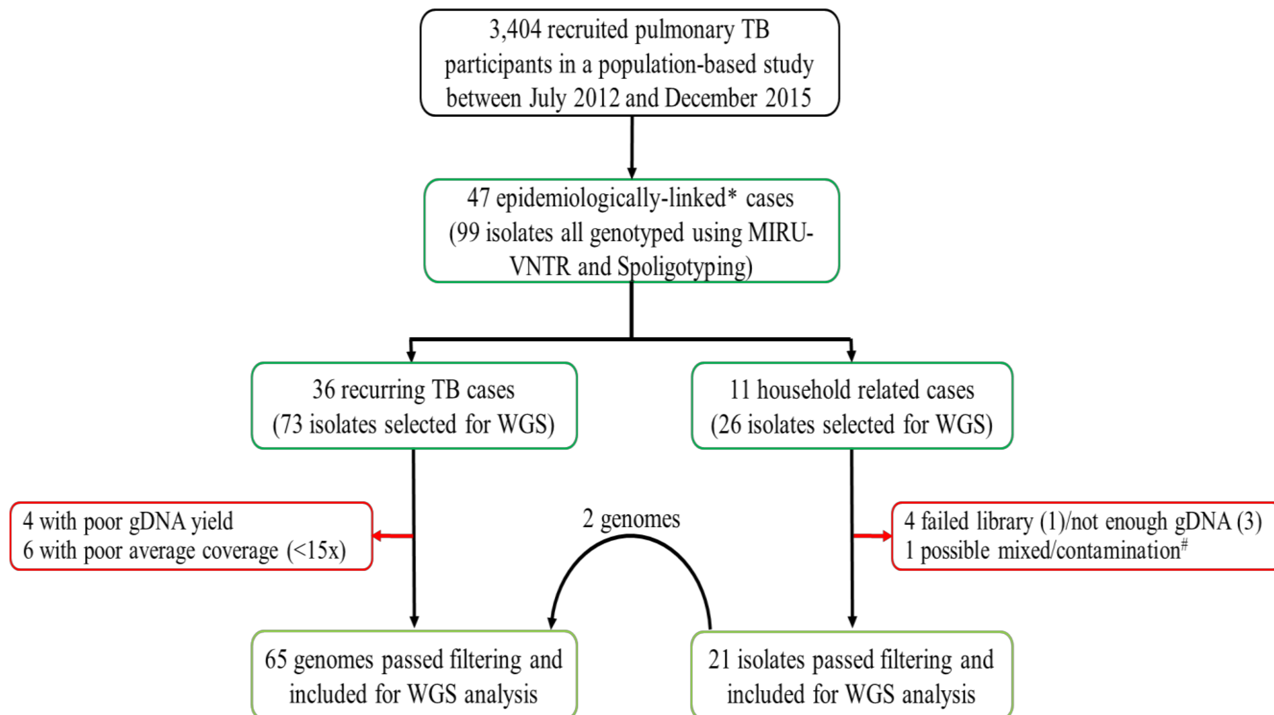


Figure 6.2. Pipeline for sample selection and genotyping.

*Isolates were classified as epidemiologically linked if they came from the same household or same patient (longitudinal isolates)

#Genomes with heterozygous SNPs greater than 120 were classified as possible mixed infection or contamination and hence removed from further analysis

†Excluded due to unavailable genome for isolate pairs

MIRU-VNTR: mycobacteria interspersed repetitive-unit variable-number tandem-repeat typing, WGS: whole genome sequencing, gDNA: genomic DNA

University of Ghana http://ugspace.ug.edu.gh

6.4.1 Characteristics of individuals with recurring TB episodes

A greater proportion (94.4%, 34/36) of individuals with recurring TB episodes were males. Overall, their ages ranged between 22 years and 75 years with a median age of 39.5 years (IQR, 32 – 53). Most participants were ≤ 40 years (19/36, 52.8%) and 21/36 (58.6%) participants had TB recurrence within ≤12 months post treatment. All 36 recurrent TB cases had either been cured (11), completed treatment (4) defaulted (3) or there was no data available (18) on outcomes of the previous episode in their respective health facilities. Twenty-four (66.7%) participants had a smear grade greater than 1+ at baseline. Of the participants tested for co-infections, 8.3% (2/24) and 36.8% (7/22) were respectively HIV and diabetes mellitus positive. Only 19.4% (7/36) of participants harbored a strain with INH resistance at baseline of which 29% (2/7) were additionally resistant to RIF and are referred to as multi-drug resistant TB (MDR-TB) strains. Among the recurring episodes, 27.8% (10/36) harbored INH resistant strains of which 40% (4/10) were MDR-TB strains (Table 6.1). Two of the 4 MDR-TB cases had recurrence with different strains whereas the remaining 2 were the same strains identified from the previous episode. One other participant (RL025) with MDR-TB was found to also be resistant to PZA and EMB; thus, this strain was resistant to all the first line anti-TB drugs. Two of the 3 extra primary cases (RL009 and RL019) which had INH resistant strains during the secondary episode were the same strain found in the previous episode. Majority of participants harbored MTBC lineages belonging to MTBss lineage 2 or 4 (77.8%, 28/36) with the Cameroon sub-lineage causing half (14/28) of such cases in both episodes followed by the Ghana sub-lineage (17.9%, 5/28).

University of Ghana http://ugspace.ug.edu.gh

Table 6.1. Characteristics of cases with recurring TB episodes

Participant ID	Gender	Age	Smear grade		Duration between follow-up (days)	MTBC lineage/Spoligotype		Strain similarity*	Outcome of previous treatment	Conventional DR 1°		Conventional DR 2°/3°		Genotypic DR (mutation†) 1°		Genotypic DR (mutation†) 2°/3°		Pairwise SNP distance	WGS status
			1°	2°/3°		1°	2°/3°			INH	RIF	INH	RIF	INH	RIF	INH	RIF		
RL001	Male	26	3+	3+	397	L4/Cameroon	L4/Cameroon	Same	cured	S	S	S	S	S	S	S	S	0	Available
RL002	Male	48	2+	2+	189	L4/Haarlem	L4/Haarlem	Same	NA	S	S	S	S	S	S	S	S	1	Available
RL003	Male	27	3+	1+	707	L6/West African 2	L4/Cameroon	Different	Completed	S	S	S	S	S	S	S	S	1848	Available
RL004	Male	34	1+	scanty	624	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	NA	NA	S	S	NA	Available for 1
RL005	Male	52	1+	3+	196	L5/West African 1	L5/West African 1	Different	cured	R	S	R	R	S	S	R (-15C>T/fabG1/Rv1483)	S	334	Available
RL006	Male	22	1+	3+	161	L4/Ghana	L6/West African 2	Different	Completed	S	S	S	S	S	S	NA	NA	NA	Available for 1
RL007	Female	25	3+	2+	603	L5/West African 1	L5/West African 1	Different	defaulted	R	S	R	S	S	S	R (-15C>T/fabG1/Rv1483)	R (L452P/rpoB/Rv0667)	243	Available
RL008	Male	40	3+	1+	476	L4/Ghana	L4/Cameroon	Different	NA	S	S	S	S	S	S	S	S	571	Available
RL009	Male	NA	1+	3+	279	L4/Cameroon	L4/Cameroon	Same	NA	S	S	R	S	S	S	S	S	0	Available
RL010	Female	24	2+	1+	570	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	7	Available
RL011	Male	53	1+	1 /3+	231/532	L4/Cameroon	L4/Cameroon	Same	defaulted	S	S	S	S	S	S	S	S	0/0/0	Available
RL012	Male	70	1+	3+	524	L2/Beijing	L2/Beijing	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL013	Male	55	1+	NA/2+	164/408	L4/Cameroon	L4/Cameroon	Same	Completed	S	S	S	S	S	S	S	S	0	Available for 2
RL014	Male	35	3+	3+	484	L4/Cameroon	L4/Cameroon	Same	defaulted	S	S	S	S	S	S	S	S	0	Available
RL015	Male	58	3+	3+	237	L4/Cameroon	L4/Ghana	Different	NA	S	S	S	S	S	S	S	S	655	Available

University of Ghana http://ugspace.ug.edu.gh

Participant ID	Gender	Age	Smear grade 1°	Smear grade 2°/3°	Duration between follow-up (days)	MTBC lineage/Spoligotype 1°	MTBC lineage/Spoligotype 2°/3°	Strain similarity*	Outcome of previous treatment	Conventional DR 1° INH	Conventional DR 1° RIF	Conventional DR 2°/3° INH	Conventional DR 2°/3° RIF	Genotypic DR (mutation†) 1° INH	Genotypic DR (mutation†) 1° RIF	Genotypic DR (mutation†) 2°/3° INH	Genotypic DR (mutation†) 2°/3° RIF	Pairwise SNP distance	WGS status
RL016	Male	31	1+	3+	700	L6/West African 2	L6/West African 2	Same	cured	R	S	R	S	NA	NA	R (S315T/KatG/ Rv1908c)	R(L452P/rpoB/Rv0667)	NA	Available for 1
RL017	Male	75	1+	scanty	252	L4/Ghana	L4/Ghana	Same	cured	S	S	S	S	NA	NA	NA	NA	NA	NA
RL018	Male	52	1+	NA	519	L4/Haarlem	L4/Ghana	Different	Completed	R	S	R	R	NA	NA	NA	NA	NA	NA
RL019	Male	68	3+	scanty	595	L4/Ghana	L4/Ghana	Same	cured	S	S	R	S	S	S	NA	NA	NA	Available for 1
RL020	Male	43	3+	3+	263	L4/Haarlem	L4/Haarlem	Same	cured	S	S	S	S	S	S	S	S	0	Available
RL021	Male	52	3+	1+	413	L4/Cameroon	L4/Cameroon	Same	cured	S	S	S	S	NA	NA	S	S	NA	Available for 1
RL022	Male	37	3+	3+/3+	350/476	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	0/0/0	Available
RL023	Male	53	2+	2+	419	L4/Ghana	L4/Ghana	Same	cured	R	S	R	S	R (S315T/KatG/ Rv1908c)	S	R (S315T/KatG/ Rv1908c)	S	0	Available
RL024	Male	32	1+	2+	196	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL025	Male	27	2+	scanty	271	L4/Ghana	L4/Ghana	Same	NA	R	R	R	R	R (S315T/KatG/ Rv1908c)	R (D435V/rpoB/ Rv0667)	R (S315T/KatG/ Rv1908c)	R (D435V/rpoB/Rv0667)	2	Available
RL026	Male	34	2+	2+	329	L4/Cameroon	L4/Cameroon	Different	cured	S	S	R	S	S	R (S450L,rpoB, Rv0667)	R (-15C>T/fabG1/ Rv1483)	S	62	Available
RL027	Male	39	1+	NA	314	L5/West African 1	L5/West African 1	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL028	Male	59	1+	NA	371	L5/West African 1	L5/West African 1	Same	cured	R	R	R	R	R (S315T/KatG/ Rv1908c)	R (Q432E/rpoB/ Rv0667)	R (S315T/KatG/ Rv1908c)	R (Q432E/rpoB/Rv0667)	0	Available
RL029	Male	70	3+	NA	383	L4/Ghana	L4/Ghana	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL030	Male	34	2+	NA	232	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	0	Available

University of Ghana http://ugspace.ug.edu.gh

Participant ID	Gender	Age	Smear grade		Duration between follow-up (days)	MTBC lineage/Spoligotype		Strain similarity*	Outcome of previous treatment	Conventional DR 1º		Conventional DR 2º/3º		Genotypic DR (mutation†) 1º		Genotypic DR (mutation†) 2º/3º		Pairwise SNP distance	WGS status
			1º	2º/3º		1º	2º/3º			INH	RIF	INH	RIF	INH	RIF	INH	RIF		
RL031	Male	32	3+	NA	352	L5/West African 1	L5/West African 1	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL032	Male	37	3+	NA	288	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL033	Male	68	2+	NA	338	L4/Cameroon	L4/Cameroon	Same	NA	S	S	S	S	S	S	S	S	5	Available
RL034	Male	25	3+	NA	268	L4/Ghana	L5/West African 1	Different	NA	S	S	S	S	S	S	S	S	1749	Available
RL035	Male	36	3+	NA	252	L2/Beijing	L2/Beijing	Same	NA	S	S	S	S	S	S	S	S	0	Available
RL036	Male	50	2+	2+	237	L4/Haarlem	L4Haarlem	Same	cured	S	S	S	S	S	S	S	S	0	Available

*Strain similarity was assessed by tradition genotyping and was defined as being same if the isolate pair shared not more than one loci difference between them and had the same antibiotic resistant pattern

†SNPs in coding regions are annotated using the reference amino acid, codon number and alternative amino acid. SNPs in non-coding regions (i.e. RNA genes and intergenic regions) are annotated using the reference nucleotide, gene coordinate and alternative nucleotide.

We obtained three longitudinal isolates from each of recurrent cases RL011, RL013 and

NA: not available, WGS: whole genome sequence, SNP: single nucleotide polymorphism, INH: isoniazid, RIF: rifampicin, S: Sensitive to specified antibiotic, R: Resistant to specified antibiotic

University of Ghana http://ugspace.ug.edu.gh

6.4.2 Whole genome sequence analysis identifies high relapse rate among recurring TB episodes

Sixty-five of seventy-five (86.7%) isolates obtained from the 36 recurrent TB cases had whole genome sequences available for analysis. Ten genomes were not available due to poor gDNA yield for sequencing (4) or poor average coverage of sequence reads (6).

Based on our definitions for relapse and re-infection and available WGS data, recurrent tuberculosis was attributed to relapse in 61.1% (22/36) of participants with 19.4% (7/36) attributed to reinfection. Of the remaining 7 cases with no WGS data, 5/7 were identified as same and 2/7 as different by MIRU-VNTR typing. Hence overall, there were 75.0% (27/36) relapse cases and 25.0% (9/36) re-infection cases. Of the 22 WGS confirmed relapse cases, 18 (81.8%) had no SNP (0 SNP) distance between their isolate pairs with the remaining 4 having 1, 2, 5 and 7 SNPs respectively separating their pairs of isolates (Table 6.1, Figure 6.3). Also, of the 7 recurring TB cases classified by WGS as re-infection, 6 (85.7%) had > 100 SNPs separating their isolate pairs with the remaining one case having 62 SNP differences. Of the 9 re-infection cases, 1 participant who was both HIV and diabetes mellitus positive was initially infected with a MAF strain (RL003) and subsequently infected with an MTBss strain. Another case (RL006) was also initially infected with an MTBss strain but later got infected with a MAF strain. All remaining 7/9 re-infection cases were infected with the different strains belonging to the same MTBC lineage. The Cameroon sub-lineage was the most common sub-lineage associated with relapse (13/27, 48.1%), followed by the Ghana sub-lineage (5/27, 18.5%) and then the MAF West African I lineage (3/27, 11.1%).

University of Ghana http://ugspace.ug.edu.gh

Phylogenetic reconstruction showed that generally, relapse cases involving MTBss are closely related compared to MAF cases. Four of the relapse cases (RL022, RL013, RL032 and RL010) were closely related with an overall average pairwise distance of 16.0 SNPs seperating each case (Figure 6.3). The only two beijing strains (RL012 and RL035) involved in relapse were also very closely related with only 4.3 SNPs sepearating each other (Figure 6.3).

Statistical analysis comparing relaspe versus re-infection cases showed no significant difference in all characteristics analyzed (Table 6.2). For instance, males with recurring TB were equaly as likely to be relapse cases or re-infected cases just like females ($p=0.443$). We also found no association of the cause of recurrence with age, marital status, formal education, other co-infections or sputum smear grade ($p>0.05$).

Apart from 5 cases, all drug resistance profiles identified using either phenotypically or by MTBDRplus were identical to that suggested by the WGS analysis for DR mutations. At least 12 isolates had one form of drug resistance either only INH (7/12) resistance or MDR (5/12). Majority of these resistant isolates were found to belong to the MAF lineage (6/12; 5 L5 and 1 L6) followed by the Ghana sub-lineage (4/12), with the remaining 2 resistant isolates being Cameroon sub-lineage. In addition to INH and RIF resistance, we found that some of the isolates also harbored resistance conferring mutations to other antibiotics including streptomycin (RL005b, RL026), Pyrazinamide (RL016b, RL025, RL026 and RL028) and ethambutol (RL006a, RL016b and RL025). Isolates from participant RL005 had discrepant drug resistance profile; whereas the MTBDRplus suggested INH resistance (katG MUT 1 Present) for the primary case (RL005a), no resistance associated mutation was found in our WGS analysis (Table 6.1). Even though we observed both INH and RIF

University of Ghana http://ugspace.ug.edu.gh

resistance in the isolate from the secondary episode (rpoBMUT2B present, katG WT absent, katG MUT1 present) using MTBDR plus, our WGS analysis only confirmed the presence of INH resistance. RL007 also had discrepant drug resistance profile results. Except for recurrent cases with different strains in both episodes, all other recurrent cases had the same set of non-resistant associated mutations in the list of resistant genes investigated.

University of Ghana http://ugspace.ug.edu.gh

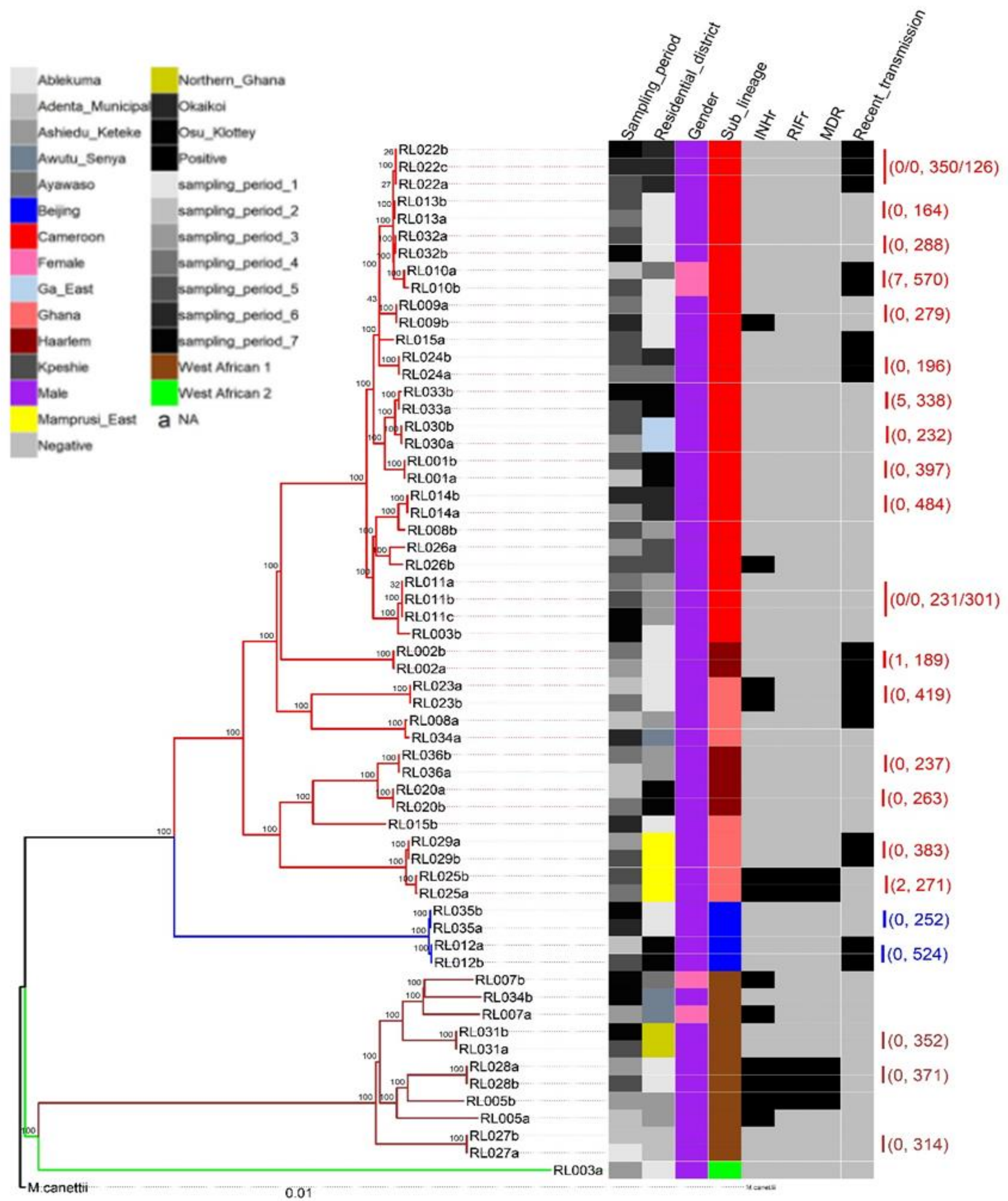


Figure 6.3. Phylogenetic reconstruction showing the genomic relationship between 60 *M. tuberculosis* complex isolates from 29 recurring TB cases. The tree was built with an alignment file containing 6,485 variable positions. The heat map shows some characteristics of the cases including sampling period (column 1), residential district (column 2), gender (column 3), sub-lineage (column 4), isoniazid resistance status (column

University of Ghana http://ugspace.ug.edu.gh

5), rifampicin resistance status (column 6), multi-drug resistance status (column 7) and whether that strain was involved in a recent transmission event (column 8). The color codes for the heat map are defined in the key. The four major branches constitute four MTBC lineages and color coded with the universally accepted Gagneux-defined lineages as red for lineage 4, blue for lineage 2, brown for lineage 5 and green for lineage 6. Data to the right of the heat map shows the number of SNP differences and number of days (in brackets) separating each isolate pair for only relapse cases. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*.

Table 6.2. Demographic and clinical characteristics of participants with recurring TB episodes

Characteristics	Relapse (N=27) n(%)	Reinfection (N=9) n(%)	Total (N=36) n(%)	p-value
Gender (n=36)				
Male	26 (96.3%)	8 (88.9%)	34 (94.4%)	0.443
Female	1 (3.7%)	1 (11.1%)	2 (5.6%)	
Age Category (n=36)				
≤40	13 (48.1%)	6 (66.7%)	19 (52.8%)	0.451
>40	14 (51.9%)	3 (33.3%)	17 (47.2%)	
Time to recurring episode (n=36)				
≤12 months	16 (59.3%)	5 (55.6%)	21 (58.3%)	1.000
>12 months	11 (40.7%)	4 (44.4%)	15 (41.7%)	
Marital Status (n=33)				
Single	6 (24.0%)	3 (37.5%)	9 (27.7%)	0.521
Married	15 (60.0%)	5 (62.5%)	20 (60.6%)	
Others	3 (16.0%)	0 (0.0%)	4 (12.1%)	
Formal Education (at least				

University of Ghana http://ugspace.ug.edu.gh

Characteristics	Relapse (N=27) n(%)	Reinfection (N=9) n(%)	Total (N=36) n(%)	p-value
lower grade) (n=34)				
Yes	20 (80.0%)	8 (88.9%)	28 (82.4%)	1.000
No	5 (20.0%)	1 (11.1%)	6 (17.6%)	
HIV positive (n=24)				
Yes	1 (5.9%)	1 (14.3%)	2 (8.3%)	0.683
No	16 (94.1%)	6 (85.7%)	22 (81.7%)	
Diabetes mellitus (n=19)				
Yes	5 (35.7%)	2 (40.0%)	7 (36.8%)	1.000
No	9 (64.3%)	3 (60.0%)	12 (63.2%)	
Current smoker (n=33)				
Yes	12 I48.0%)	3 (37.5%)	15 (45.4%)	0.699
No	13 (52.0%)	5 (62.5%)	18 (54.5%)	
Sputum smear grade (n=36)				
≤1+	9 (33.3%)	3 (33.3%)	12 (33.3%)	1.000
>1+	18 (66.7%)	6 (66.7%)	24 (66.7%)	
Isoniazid resistant strain (n=36)				
Yes	4 (14.8%)	3 (33.3%)	7 (19.4%)	0.333
No	23 (85.2%)	6 (66.7%)	29 (80.6%)	
Infecting MTBC lineage (n=36)				
MTBss L2 and L4	23 (85.2%)	5 (55.6%)	28 (77.8%)	0.086
MAF L5 and L6	4 (14.8%)	4 (44.4%)	8 (22.2%)	

6.4.3 Characteristics of individuals belonging to household related TB transmission cases

In all, a total of 26 isolates from 11 households were analyzed for household related transmission. Majority of the cases were male (17/26, 65.4%) with median age of 40 (IQR, 26 – 49) and the remaining 9/26 (34.6%) were female with median age of 27 (IQR, 20 –

University of Ghana http://ugspace.ug.edu.gh

30). Overall, the age ranged between 13 and 54 years with a median age of 28.5 years (IQR, 22 - 47). Nineteen (73.1%) of the 26 participants had a smear grade greater than 1+. None of the participants tested for co-infections, was positive for HIV, and only one was positive for diabetes mellitus. Of the respondents, 26.7% (4/11) were smokers.

Twenty-one of 26 (70.8%) isolates obtained from the 11 household related TB cases had whole genome sequences available for analysis. Five genomes were excluded, three due to poor gDNA yield for sequencing, one due to failed library and one due to possible mixed infection or contamination (Figure 6.2).

Based on our cut off for identifying strain relatedness, 8/11 (72.7%) households had individuals infected with the same strain with 2 household having individuals infected with different MTBC strains. The two participants in the remaining household (FT014) shared the same MIRU-VNTR allelic pattern but were separated by a genomic distance of 35 SNPs. All participant harbored MTBss lineages with the predominant sub-lineage being Cameroon (46.2%, 12/26) followed by Ghana (19.2%, 5/26) and then Beijing (15.4%, 4/26). In addition to these lineage 2 and 4 strains, one other participant harbored a lineage 3 strain (FT006a, Delhi/Cas).

For all cases with available WGS data, drug resistance profile was same for both phenotypic and WGS mutation-predicted resistance. Only 11.5% (3/26) of participants harbored a strain with INH resistance and no participant was harboring a strain with RIF resistance (Table 6.3). Two of the only 3 INH resistant cases were females (Figure 6.4) harboring a Ghana sub-lineage with the remaining one being a male. In addition to INH and RIF resistance, we found a few isolates also harboring resistance associated mutations

University of Ghana http://ugspace.ug.edu.gh

to other anti-TB drugs including streptomycin (FT007, FT014a), ethambutol (FT001 and FT016) and capreomycin (FT001 and FT016)

Table 6.3. Characteristics of cases involved in household related transmission

Participant ID	Gender	Age (yrs)	Smear grade	MTBC lineage/Spoligotype	Strain similarity		Drug resistance			Pairwise SNP difference	WGS status
					Traditional genotyping*	WGS	INH	RIF	Mutation identified[†]		
FT001a	Male	26	3+	Ghana	Same	Same	S	S		0	Available
FT001b	Female	20	2+	Ghana	Same	Same	S	S			Available
FT002a	Female	28	3+	Cameroon	Similar	NA	S	S			NA
FT002b	Male	29	scanty	Cameroon	Similar	NA	S	S			Available
FT003a	Male	19	3+	Cameroon	Same	Same	S	S		0	Available
FT003b	Female	50	3+	Cameroon	Same	Same	S	S			Available
FT004a	Male	15	3+	Cameroon	Same	Same	S	S		2	Available
FT004b	Male	49	1+	Cameroon	Same	Same	S	S			Available
FT004c	Male	27	2+	Cameroon	Same	NA	S	S			NA
FT006a	Male	NA	3+	Delhi/CAS	Different	NA	R	S	-8T>A,fabG1, Rv1483		Available
FT006b	Male	13	3+	Cameroon	Different	NA	S	S			NA
FT007a	Female	45	2+	Beijing	Similar	Same	S	S		0	Available
FT007b	Female	16	3+	Beijing	Similar	Same	S	S			Available
FT012a	Male	49	2+	Haarlem	Same	Same	S	S		0/0/0	Available
FT012b	Male	49	scanty	Haarlem	Same	Same	S	S			Available
FT012c	Male	50	2+	Haarlem	Similar	Same	S	S			Available
FT012d	Male	54	3+	S	Different	NA	S	S			NA
FT013a	Male	40	3+	Cameroon	Same	Same	S	S		0	Available
FT013b	Male		3+	Cameroon	Same	Same	S	S			Available
FT014a	Male	28	3+	Beijing	Same	Similar	S	S		35	Available
FT014b	Male	40	1+	Beijing	Same	Similar	S	S			Available
FT015a	Female	15	scanty	Cameroon	Same	Same	S	S		0	Available
FT015b	Male	44	1+	Cameroon	Same	Same	S	S			Available
FT016a	Female	27	scanty	Ghana	Same	Same	R	S	-15C>T,fabGI,Rv1483	4	Available
FT016b	Female	30	3+	Ghana	Same	Same	R	S	-15C>T,fabGI,Rv1483		Available
FT016c	Female	24	2+	Ghana	Same	NA	S	S			NA

*Traditional genotyping clusters are defined as previously described clusters using mycobacterial interspersed repetitive-unit-variable number tandem-repeat analysis and spoligotyping genotyping tools

[†]SNPs in coding regions are annotated using the reference amino acid, codon number and alternative amino acid. SNPs in non-coding regions (i.e. RNA genes and intergenic regions) are annotated using the reference nucleotide, gene coordinate and alternative nucleotide.

There were 4 participants in household FT012 and 3 participants in FT004 and FT016

NA: not available, WGS: whole genome sequence, SNP: single nucleotide polymorphism, INH: isoniazid, RIF: rifampicin, S: sensitive to specified antibiotic, R: resistant to specified antibiotic

University of Ghana http://ugspace.ug.edu.gh

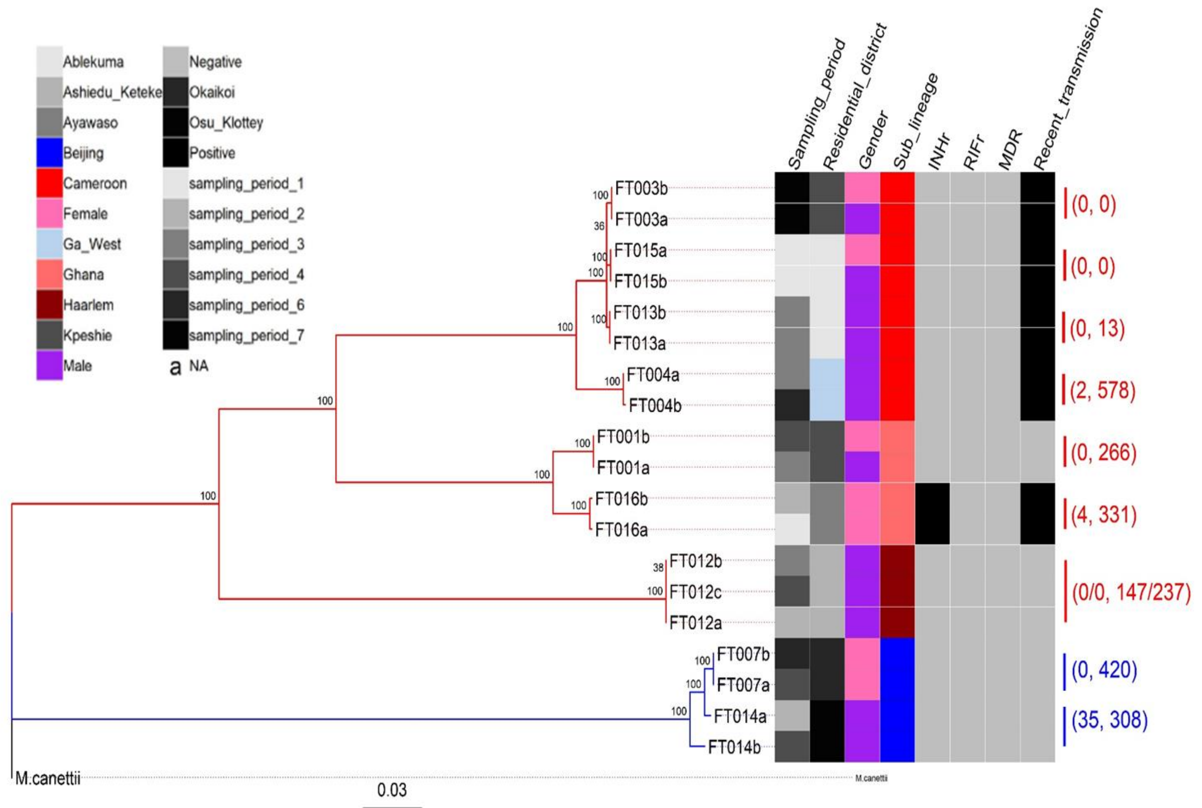


Figure 6.4. Phylogenetic reconstruction showing the genomic relationship between 19 *M. tuberculosis* complex isolates from 9 household transmission groups. The tree was built with an alignment file containing 2,022 variable positions. The heat map shows some characteristics of the cases including sampling period (column 1), residential district (column 2), gender (column 3), sub-lineage (column 4), isoniazid resistance status (column 5), rifampicin resistance status (column 6), multi-drug resistance status (column 7) and involvement of strain in a recent transmission event (column 8). The color codes for the heat map are defined in the key. The two major branches constitute two MTBC lineages and color coded with the universally accepted Gagneux-defined lineages as red for blue for lineage 2, lineage 4. Data to the right of the heat map shows the number of SNP differences and number of days (in brackets) separating each isolate pair. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*.

University of Ghana http://ugspace.ug.edu.gh

6.4.4 Evidence of household recent transmission

Eleven of the 26 participants had reported contact with more than one TB patient; 9/26 from a family member or individual living in the same house who has been coughing within 1 year, 2/26 from close work mates. FT004, a Cameroon sub-lineage case involved three male participants in the same household. Participant FT004a aged 15 years was the first case to be identified reportedly coming into contact with a TB patient at his workplace 1 year previously and shared 2 SNP distance with the next case (FT004b) being his father aged 49 years. The two SNP distances were accumulated within a period of 578 days (Figure 6.4 and 6.5). The third case (FT004c) had no WGS data but had the same MIRU-VNTR allelic pattern as the two previous cases. FT004b and FT004c shard the same room in the house. Participant FT007b, a 16-year-old girl had also been in contact with her mum (FT007a, Beijing sub-lineage, aged 45 years) within 420 days and shared no SNP with her.

6.4.5 15-locus MIRU-VNTR typing is enough for predicting the cause of recurring TB and identifying suspected household related TB transmission

MIRU-VNTR analysis revealed that, 75.6% (34/45) paired isolates with MIRU-VNTR allelic information had at most one loci variance between isolate pairs (Figure 6.6). There was a high concordance (94.6%, Kappa=0.7702, $p<0.001$) and positive correlation ($R^2 = 0.817$, $p<0.001$, Figure 6.7) between 15-locus MIRU-VNTR typing and WGS typing. Apart from one isolate pair (FT014) that MIRU-VNTR predicted to be the same case but WGS analysis found 35 SNPs, all other isolate pairs that were predicted to be same or similar were also same by WGS with not more than 7 SNPs separating each isolate pair. All remaining cases that MIRU-15 typing predicted as different strains were also confirmed as different strains by WGS with more than 62 SNPs separating them.

University of Ghana http://ugspace.ug.edu.gh

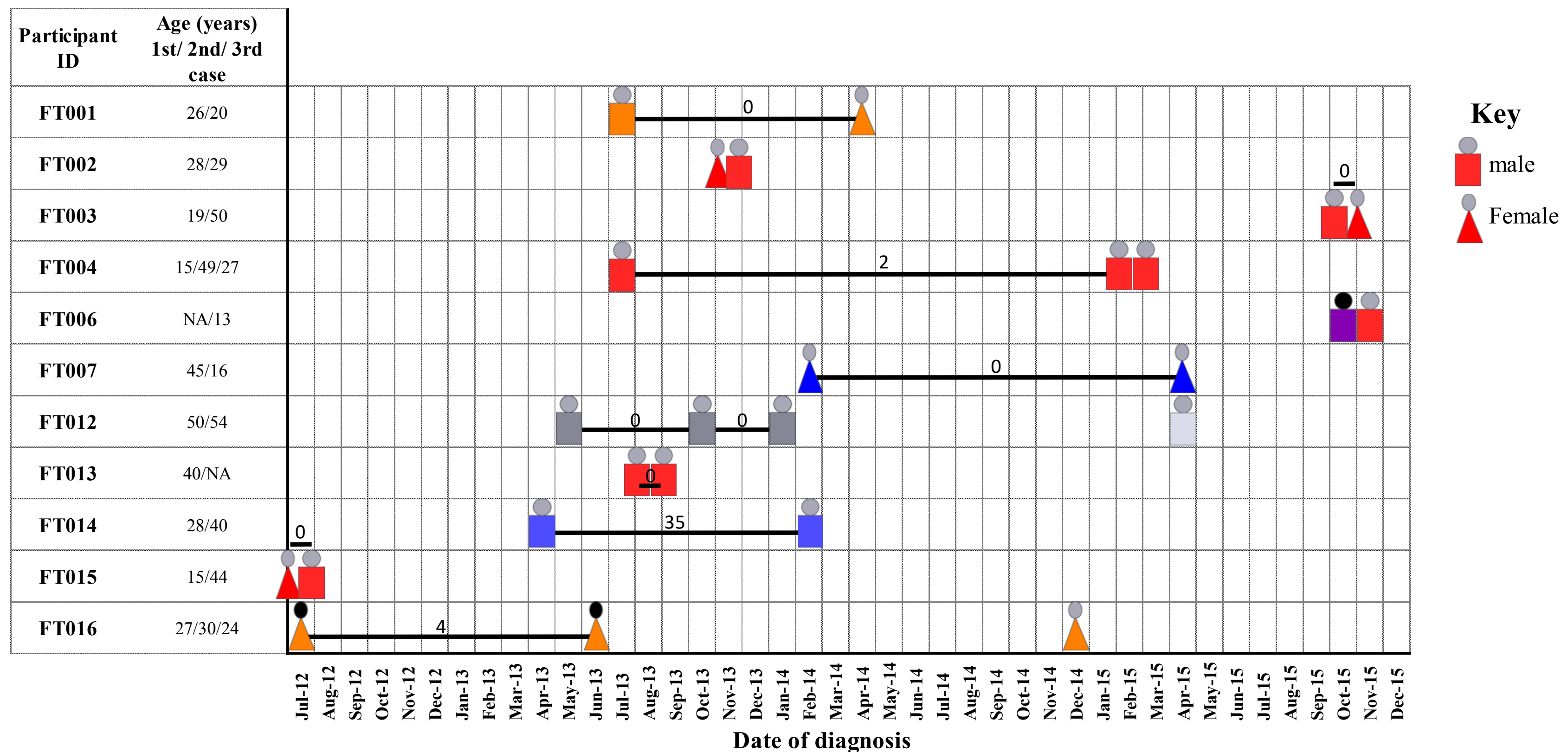


Figure 6.5. Time till event of household related tuberculosis cases. The time of diagnosis for each household related TB case is shown. The color codes represent the various human adapted *M. tuberculosis* complex *sub*-lineages; red for 'Cameroon', orange for 'Ghana', blue for 'Beijing', Grey for 'Haarlem' and light grey for 'S' sub-lineages. For household related cases that are likely the results of the same or similar MTBC strain, the SNP distances between each pair have been indicated on the bars.

University of Ghana http://ugspace.ug.edu.gh

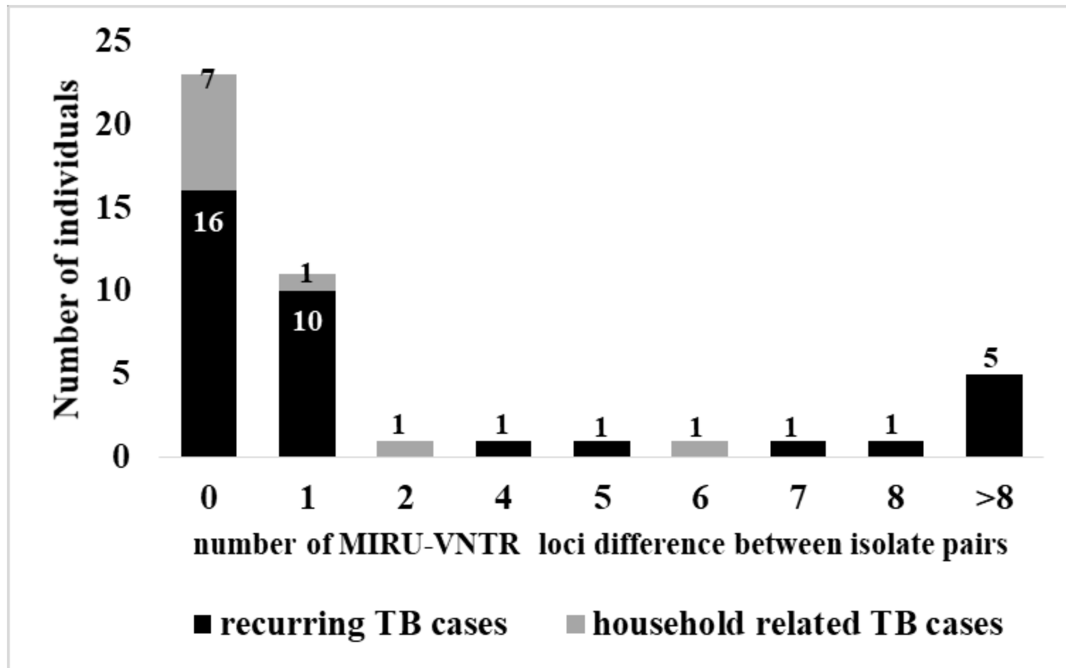


Figure 6.6. Number of MIRU-VNTR allelic variations between identified recurrent TB cases (black bars) and household related cases (grey bars).

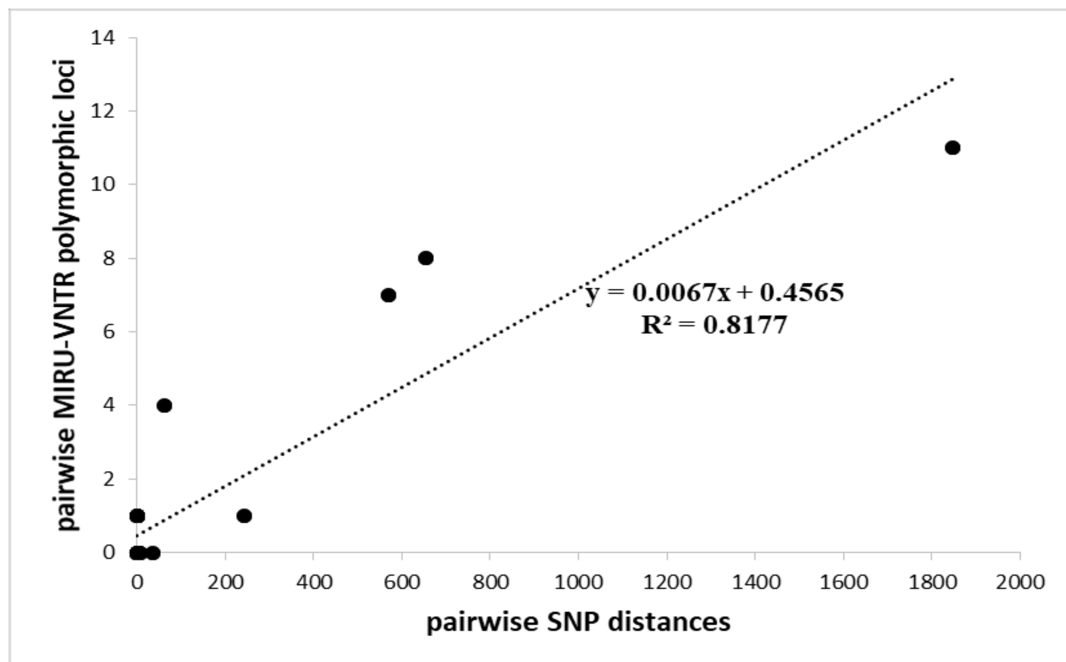


Figure 6.7. Concordance between MIRU-VNTR allelic polymorphism and WGS SNP distances.

University of Ghana http://ugspace.ug.edu.gh

6.5 Discussion

The goals of this study were to delineate the occurrence of recurring TB and to describe household related transmissions among individuals with pulmonary TB reporting to selected health facilities in Ghana by combining WGS and standard molecular epidemiological tools. In addition, we investigated the occurrence of drug resistant TB strains among the study participants. Key findings from our analysis indicate that: 1) Majority (75.0%, 27/36) of recurring pulmonary TB cases are as a result of relapse of previous infection probably due to inadequate treatment 2) Household related TB cases are likely the results of recent TB transmission within the house or from neighboring recent transmission events, 3) Though WGS is ideal, 15-locus MIRU-VNTR typing is enough to predict the cause of TB recurrence and additionally able to predict household related recent TB transmission.

For the first time in Ghana, through WGS analysis, we have shown evidence of both endogenous reactivation (relapse) and exogenous re-infection (re-infection) among individuals with recurring TB episode and found that a majority (75.0%, 27/36) of recurring pulmonary TB cases are as a result of relapse of previous infection but not re-infection. Other similar studies (Varghese *et al.*, 2013) have used only traditional genotyping tools but here we performed both traditional MIRU-VNTR typing and the gold standard WGS to investigate strain relatedness. One participant who was involved in a household (FT012) related transmission also had recurring TB (RL036) and so two genomes were shared between the cohort of recurring TB cases and household related cases. We observed some discrepancies in DR status of the isolates by comparing the conventional tools to WGS-predicted-resistance. This could be as a result of possible

University of Ghana http://ugspace.ug.edu.gh

mixed infection or inability to obtain difficult-to-isolate DR clones upon repeated cultures and hence not represented in WGS data.

Studies conducted in India by Velayutham and collegues have reported that up to 55% of patients had TB recurrence within 3 months post-treatment and a total of 77% occurred within 6 months (Velayutham *et al.*, 2018). Here, we found that majority (58.3%) of TB recurrences occur within the first year post treatment similar to observations made elsewhere (Thomas *et al.*, 2005; Velayutham *et al.*, 2018). It is generally assumed that in high endemic regions, recurrent TB cases should more likely be the results of reinfection rather than relapse (Parvaresh *et al.*, 2018; van Rie *et al.*, 1999). However, studies conducted in endemic regions like India and China have attributed up to 93% of recurrent TB cases to relapse and also found male gender to be a risk factor for TB recurrence (Velayutham *et al.*, 2018; Zong *et al.*, 2018). These studies however, used MIRU-VNTR as the genotyping tools for discriminating between strains and this might have resulted in overestimation of the true incidence of relapse among recurrent cases. These observations were however not different from our finding that up to 75% of recurrent TB cases were as a result of relapses. Our finding is strengthened by the use of WGS for our investigation, but we did not find male gender to be significantly associated with recurrence of TB, probably due to our small sample size. Some of the recurring TB cases (RL022, RL010, RL024, RL002, RL023, RL029 and RL012) harbored strains that were found circulating in large recent transmission clusters (chapter 5). This may suggest a possible re-infection provided they were totally cured of the previous infection and that they were somehow related by geographic location to those transmission clusters. However, we are confident that at least the 20/27 remaining cases classified as relapse are true relapse cases as they do

University of Ghana http://ugspace.ug.edu.gh

not cluster with any other recent transmission clusters previously defined from the same population.

Predominance of relapse over re-infection indicates high-quality public health practices and a low risk of local transmission. Even though the high rate of relapse may be a good sign of low recent transmission in the community, it is still not the best as relapse cases have been associated with development of MDR (Alipanah *et al.*, 2018; Yang *et al.*, 2017). Within our population-based study we identified that 7.4% (184/2482) of participants were previously treated cases (Asare *et al.*, 2018) similar to the national average of 7% (WHO, 2015d). Our observation that 75% of these recurring TB cases are due to relapse implies that, overall, 5.3% ((75/100) x 7) of recurring TB in Ghana can be attributed to relapse. To help reduce the occurrence of relapse cases and improve treatment outcome, some measures that can be put in place include; using adherence interventions such as patient education and counseling, psychological interventions, incentives and enablers and digital health technologies (Alipanah *et al.*, 2018; Amo-Adjei & Awusabo-Asare, 2013). These measures are necessary as drug resistance strains can emerge due to lack of treatment adherence.

In Ghana, there has been reported increase in treatment success from 44% in 1997 to 87% in 2013 with current rates estimated at 85% (Amo-Adjei & Awusabo-Asare, 2013; WHO, 2018c). If treatment success is that high, then we expected that individuals with recurring TB will likely be due to re-infection with new strain and not relapse, but rather we observed the contrary. Consequently we intended to explore more into characterizing the drug resistance profiles of isolates from our study cohort. Interestingly, of the 4 relapse cases (RL005, RL018, RL025 and RL028) with MDR-TB strains during their secondary

University of Ghana http://ugspace.ug.edu.gh

episode, 2 were confirmed cured after their first episode; one participant (RL028) had an MDR-TB strain even during the first episode but was declared cured and left to go back to the community. This individual came back after a year with the same strain (no MIRU loci variance with 0 SNP separating the isolate pair) with a possibility that he might have spread the MDR bacilli to other people since they have the propensity to transmit (Cohen *et al.*, 2019; Lalor *et al.*, 2018). With a good control system, public health measures including contact tracing of such an individual are needed to control the spread of such notorious MDR-TB strain. Another interesting finding of this study was that two relapse cases (RL009 and RL019, Table 6.1) with the same strain at both episodes had INH resistant strains only during the secondary episode suggesting an acquisition of drug resistance within the host.

In addition to recurring TB cases, our analysis of household related TB cases allowed us to identify 8/11 (72.7%) households being involved in recent TB transmission. We acknowledge that for household related transmission studies, the classical approach will be to adopt a contact tracing method. Though we did not use this approach, we took advantage of recruiting participants from the same household and analyzing them since our study time period was large enough to capture such cases. With our current approach, we were able to show that majority of household related TB cases are due to recent transmission of the same strain and greater than 80 % of the first identified case in each household had a smear grade of >1+ implying high infectiousness. We confirmed a household transmission of one INH resistant strain (FT016, Figure 6.6) which shared only 4 SNPs between the isolate pair within a period of one year. This same strain has been found circulating in a large recent transmission cluster within the same population

University of Ghana http://ugspace.ug.edu.gh

(unpublished data). Even though we did not identify any MDR strain, these strains have the potential to evolve into MDR strains especially since some were identified in the recurrent cohort. This is worrying as MDR transmission can particularly hinder the control of TB both locally and internationally (Cohen *et al.*, 2019; Lalor *et al.*, 2018). We therefore recommend that detailed contact tracing, compound house screening and follow-up study be employed to help identify household related cases early enough to aid in control of spread of the disease.

Some research groups have wondered if 15-locus MIRU-VNTR is enough to study strain relationships (Gibson *et al.*, 2005; Kozinska & Augustynowicz-Kopec, 2016). Here, we show that though WGS is ideal, 15-locus MIRU-VNTR typing is enough to predict the cause of TB recurrence and household related transmission. Hence, we recommend the use of 15-locus MIRU-VNTR typing as an initial screening tool in resource-limited settings. This can help improve TB control measures as it will enable the early identification of the source of infection and likely inform treatment selection based on previous antimicrobial susceptibility history of that strain.

To the best of our knowledge, this is the first extensive report of analysis of recurring TB cases as well as household related TB transmission using WGS in Ghana and West Africa. We were however limited in our study design as we had only three and half years for recruitment and this meant that we did not obtain isolates from recurring TB cases who showed up later and also we did not have isolates for recurring cases to whom their primary cases occurred prior to our time of sampling. As a result, we had lower than expected number of cases.

University of Ghana http://ugspace.ug.edu.gh

6.6 Conclusion

It is possible to monitor recurring TB cases and follow-up household related transmission in a resource-limited setting. We recommend local control programs to invest more resources into such studies as they have been proven to provide vital findings that positively influence TB control (Velayutham *et al.*, 2018).

University of Ghana http://ugspace.ug.edu.gh

7.0 GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

7.1 General Discussion

Tuberculosis (TB), a disease of antiquity still infects millions of people annually. In 1993 the world health organization (WHO) declared it a global health emergency, calling for more resources and studies to aid control this global menace (WHO, 2018b). The End TB strategy, unanimously endorsed by all member countries of the WHO and the United Nations (UN) in May 2014, outlined three pillars to aid global TB control. These included; 1) an integrated, patient-centered care and prevention, 2) bold policies and supportive systems and 3) intensified research and innovation (WHO, 2015a). The set targets are to reduce the absolute numbers of TB deaths and TB incidence by 90% and 80% respectively by 2030 and 95% and 90% respectively by 2035 (WHO, 2018b). An objective under the research pillar, is to investigate the transmission dynamics of the disease using molecular epidemiology (molepi) to understand factors leading to occurrence of the disease within distinct population. Consequently, this work sought to use molepi study to determine the transmission pattern and dynamics of *Mycobacterium tuberculosis* complex (MTBC) strains in two areas (Accra Metropolitan Area (AMA) and East Mamprusi District) in Ghana to complement conventional national TB control program (NTP) efforts with the main objectives to 1) explore the spatio-temporal distribution of the different circulating MTBC genotypes, 2) study the transmission dynamics of TB and identify associated risk factors including drug resistance as well as demography and 3) investigate the occurrence of relapse and re-infection among individuals with recurring TB episodes, in the two districts.

University of Ghana http://ugspace.ug.edu.gh

7.1.1 Phylogeography of *Mycobacterium tuberculosis* complex: Implications for TB control

Tuberculosis is caused by nine closely related species collectively known as the MTBC of which *M. tuberculosis* sensu stricto (MTBss) and *M. africanum* (MAF) are the two main species known to cause disease in humans. These human adapted species are sub-divided into seven phylogenetic lineages; lineage (L) 1 – 4 and 7 are subtypes of MTBss, while L5 and L6 are MAF subtypes (Blouin *et al.*, 2012; Gagneux & Small, 2007). The seven lineages exhibit a phylogeographical structure, which indicate that distinct lineages infect specific human populations, with MAF lineages intriguingly restricted to West-Africa (Blouin *et al.*, 2012; Gagneux & Small, 2007). As has previously been shown (Addo *et al.*, 2007; Asante-Poku, Aning, *et al.*, 2014; Asante-Poku *et al.*, 2015; Yeboah-Manu *et al.*, 2012; Yeboah-Manu *et al.*, 2011), with the exception of L7, Ghana houses all remaining 6 human adapted lineages which has been confirmed by this study. This makes Ghana one of the few countries globally with the highest diverse TB strains and provides an opportunity to make varying investigations. The West African region is the only region globally with all of the 6/7 main human TB lineages for reasons not understood (Gehre *et al.*, 2016). Together with Ghana, some other countries that have all the six major human lineages in appreciable numbers are Sierra Leone, Ivory Coast and Benin (de Jong, Antonio, *et al.*, 2010; Winglee *et al.*, 2016). Overall, using standard genotyping tools, we identified 79% of TB in Ghana to be caused by MTBss, 20% by MAF and 1% by animal strains (mainly *M. bovis*). This finding supports the prevalence of all two West African restricted MAF lineages in the sub-region (de Jong, Antonio, *et al.*, 2010; Gagneux & Small, 2007); showing that MAF is still a very important pathogen hence advocating that more studies to

University of Ghana http://ugspace.ug.edu.gh

understand the biology and genomic variations compared to MTBss lineages needs to be considered in the design of control tools. For instance, a study by Ofori-Anyinam and colleagues found that some *mpt*64-based diagnostic kits had reduced detection for MAF L6 and so some L6 TB cases in the Gambia were misclassified as non-tuberculosis mycobacteria (NTM). This is crucial as it has consequences on treatment regimen (Ofori-Anyinam *et al.*, 2016). Most NTMs are intrinsically resistant to the standard TB first line drug, thus NTM infected cases are put on other treatment regimen for instance; azithromycin, ethambutol and rifampicin are for treating *M. avium* complex infections and amikacin, cefoxitin, imipenem and tigecycline for treating *M. abscessus* infection. Also, a study by Otchere and colleagues conducted within the same study population, found that whereas L6 is significantly diverse compared to L4, L5 was the most conserved and additionally found lineage specificity of T cell antigenic variation among these three MTBC lineages (Otchere *et al.*, 2018). These observations have implication for TB control as most vaccines are formulated based on the biology of L4 - hence the need to include the diversity of L5 and L6 in vaccine formulations.

This study was conducted in two varying geographical settings; one setting in Southern Ghana predominantly urban (Accra Metropolitan Area) and the other in Northern Ghana predominantly rural (East Mamprusi district). This also afforded us the opportunity to compare characteristics across these varying geographical settings. Even though the lineage distribution was similar in both study areas, we found a significant association ($p<0.05$) of animal strains (*M. bovis*), Beijing lineage and the Ghana sub-lineage with Northern Ghana whereas the Cameroon sub-lineage (though not statistically significant) was more predominant in Southern Ghana. There are more livestock as well as livestock

University of Ghana http://ugspace.ug.edu.gh

farmers in northern Ghana (GSS, 2012, 2013c) resulting in increased human-livestock contact compared to southern Ghana. It was therefore not surprising to find the association of animal adapted strains with northern Ghana. In fact, 5 individuals infected with animal strains were either cattle farmers or direct cattle handlers supporting the observation that direct constant contact with cattle and/or their products may be a risk factor of *M. bovis* infection (Michel *et al.*, 2010). The dominance of both the Beijing strains and the Ghana sub-lineage in the Northern Region of Ghana may also pose a challenge to TB control in the region since various studies including this current one have identified both strains to be associated with drug resistance (Hu *et al.*, 2011; Otchere *et al.*, 2016; Yeboah-Manu *et al.*, 2016). This calls for the NTP to support peripheral diagnostic laboratories with facilities to accurately detect and help control the spread of drug resistant strains. The observed varying distribution of distinct genotypes even within the same country reiterates the phylogeographic nature of the human adapted MTBC (Gagneux & Small, 2007). Though we observed some geographical clustering of MTBC genotypes in southern Ghana, we were unable to tell (based on the poor strain resolution of spoligotyping) whether the observed clustering was as a result of recent TB transmission of these specific genotypes or chance reactivation of latent TB infection (LTBI) of the same genotype at the same time and geographical location. Consequently, as discussed in the next sub-section, we explored better strain resolving tools to confirm the observed clusters.

Whereas some TB strains do cause disease in a large group of individuals due to their large transmissibility (referred to as generalist), others are geographically restricted probably due to their low transmissibility (referred to as specialists) (Stucki *et al.*, 2016). Whereas the Haarlem, LAM and PGG3 sub-lineages of L4 have been described by Stucki and colleges

University of Ghana http://ugspace.ug.edu.gh

as being generalist the Ghana, Cameroon and Uganda sub-lineages where described as specialist (Stucki *et al.*, 2016). Other MTBC populations including L5 and L6 have been described as specialists due to their geographic restriction and poor transmissibility (de Jong, Antonio*, et al.*, 2010; Gehre *et al.*, 2013; Gehre *et al.*, 2016). Within the L4, we identified the Cameroon sub-lineage (mainly SIT61) as the most dominant sub-lineage that was similar to strain compositions in other West African countries (Affolabi *et al.*, 2009; Chihota *et al.*, 2018; Koro Koro *et al.*, 2013) followed by the Ghana sub-lineage. In addition to the Ghana sub-lineage being associated with drug resistance (Otchere *et al.*, 2016) as previously discussed, the Cameroon sub-lineage has also been found to have a high propensity to develop drug resistance (Koro Koro *et al.*, 2013; Lawson *et al.*, 2012). This poses a great challenge to TB control in Ghana as both specialist strains are endemic. The dominance of the Cameroon sub-lineage suggests it is an important cause of TB and may likely be the most virulent strain in the sub-region (Assam Assam *et al.*, 2013). Socio-economic factors could account for the predominance of the Cameroon sub-lineage in some West Africa countries especially since there is free flow of individuals across borders and hence increased cross-border transmissibility. Active case detection followed up with treatment at the borders will aid to reduce transmission of this successful sub-lineage.

7.1.2 Transmissibility and virulence of members of MTBC: Need for host-pathogen co-evolution studies

Work done mainly in the Gambia, suggests the attenuation of MAF indicating that MTBss has a competitive advantage over MAF (de Jong, Adetifa*, et al.*, 2010; de Jong, Antonio*, et al.*, 2010). While both transmitted equally (shown by the tuberculin skin test conversion of TB contacts), they found that the rate of progression to disease was slower in MAF

University of Ghana http://ugspace.ug.edu.gh

infected contacts. Using combined spoligotyping and 15-locus mycobacterial interspersed repetitive-unit variable-number tandem repeat (MIRU-VNTR) typing as well as whole genome sequence (WGS), we identified (Asare *et al.*, 2018) and confirmed a reduced recent transmission of MAF lineages compared to MTBss lineages (Nebenzahl-Guimaraes *et al.*, 2015). MAF lineages have been suggested to be opportunistic pathogens associated with the aged and immunocompromised individuals (de Jong *et al.*, 2005). Within our study population, we have identified L6 as being associated with both HIV (Asante-Poku *et al.*, 2016) and diabetes mellitus (Asante-Poku *et al.*, 2019) co-morbidity.

Three hypothesis have been proposed to explain the restricted prevalence of MAF in West-Africa: 1) Reduced transmissibility (Nebenzahl-Guimaraes *et al.*, 2015) and virulence (Bold *et al.*, 2012; de Jong *et al.*, 2008) of MAF lineages suggests that MAF will be outcompeted by MTBss leading to reduced incidence of MAF lineages and a decline with time 2) The probable adaptation of MAF to distinct West African human populations has contributed to its restricted prevalence (Asante-Poku *et al.*, 2016; Asante-Poku *et al.*, 2015) and 3) MAF might be zoonotic with an animal reservoir restricted to West Africa. Despite our observation of a reduced transmission of MAF, which tends to support the first hypothesis, our temporal analysis showed persistence of MAF (average of approximately 20%) over an 8-year period (Yeboah-Manu *et al.*, 2016). The observed persistence of MAF within our study population and in West African human population suggests the association of West African human genotype to MAF hence supporting the second hypothesis.

Data from epidemiological studies suggest that the different MTBC phylogenetic lineages exhibit a phylogeographic structure and might have adapted to different human

University of Ghana http://ugspace.ug.edu.gh

populations. Studies in San Francisco supports this phenomenon of MTBC lineages transmitting among their sympatric host populations (Gagneux *et al.*, 2006). This shows the possibility of a host-pathogen-interaction existing between human genetic diversity and MTBC variation (Gagneux, 2012). The human adapted MTBC are obligate human pathogens with no known animal reservoir (de Jong, Antonio*, et al.*, 2010); as such, they depend on their host to transmit from person to person. This is a principal characteristic of host-pathogen co-evolution which has contributed to some of the characteristics of the host-pathogen interactions in human TB. Such investigations in other fields like malaria have led to the association of a particular specie of *Plasmodium, P. vivax*, to distinct human populations with the Duffy blood group (Langhi & Bordin, 2006), thus individuals with Duffy-negative phenotype are resistant to *P. vivax* invasion. There is therefore an urgent need for host-pathogen co-evolution investigations to throw more light on understanding MAF endemicity in West Africa in order to guide control interventions.

7.1.3 Molecular Epidemiology and TB control

Within this PhD, I explored two genotyping tools; the 15-locus MIRU-VNTR and WGS typing tools to study the epidemiology of TB in Ghana. These tools have been widely used in developed countries to study recent TB transmission as well as monitoring and evaluating TB control programs (Gardy *et al.*, 2011; Jamieson *et al.*, 2014; Lalor *et al.*, 2018; Walker *et al.*, 2017; Walker, Monk*, et al.*, 2013). Such population-based molepi studies are necessary to provide information to guide targeted efforts of TB control but are lacking in sub-Saharan Africa which is the most burdened and has the highest disease prevalence and human population ratio. The molepi studies can predict how well TB is being controlled in a given setting. For instance; a high recent transmission rate signifies

University of Ghana http://ugspace.ug.edu.gh

increased spread of the pathogen hence poor control efforts. Moreover, predominance of relapse over re-infection among recurring TB cases may indicate high-quality public health practices and a low risk of recent transmission, on the other hand may also indicate poor treatment monitoring.

Combining the two tools the estimated recent transmission rate was 24.7%, which is high (Guerra-Assunção *et al.*, 2015; Yates *et al.*, 2016) and predicted the occurrence of undetected recent transmission clusters. At the sub-lineage level, we identified the Cameroon and Ghana sub-lineages to be very important pathogens driving recent TB transmission in Ghana. The high transmissibility of the Cameroon sub-lineage probably accounts for its high dominance in the sub-region and calls for further investigation into its virulence to help control it.

The increase in rural-urban migration in search for 'greener pastures' has led to overcrowding of the urban sectors. Judging from the mode of transmission of the causative pathogen being mainly by inhalation of aerosolized droplet nuclei containing the bacilli, it suggests that overcrowded population provides an ideal means for increased transmission through frequent person-to-person contact with an infected person. This is more likely evident in congested areas including overpopulated compound houses. Our observation of high recent transmission in urban settings compared to rural settings confirms this phenomenon and identifies urbanization as an important risk factor in Ghana. High rates of TB have been reported in urban areas due to overcrowding of human populations throughout the world (Corbett *et al.*, 2009; Lonnroth *et al.*, 2009; Wood *et al.*, 2010) thereby raising concerns of possible future urban TB epidemics. Analysis of isolates from the urban setting, we were able to identify a large recent TB transmitting cluster made up

University of Ghana http://ugspace.ug.edu.gh

of the most dominant Cameroon sub-lineage of L4. We showed that the clustering of TB cases in Ablekuma sub-district of AMA observed in our previous report (Yeboah-Manu *et al.*, 2016) was most probably due to recent TB transmission. This is not surprising as Ablekuma is the most densely populated of the six sub-districts within AMA (GSS, 2014). Our analysis suggests that there may be super-spreaders in Ablekuma and probably Okaikoi sub-districts which recorded the second highest prevalence that belonged to the largest recent transmitting cluster. Majority of the individuals in this high transmitting cluster were found to inhabit the city suburbs in highly populated compound houses. Their low-income status combined with over-crowding may be driving factors for the ongoing transmission in this hotspot. A high smear grade of over 70% of cases being at least 2+ signifies delay reporting and these individuals are likely to have been actively transmitting the pathogen prior to diagnosis. We therefore advocate for intensifying community education and outreach programs to improve early case reporting and infection control. The active case search approaches should be targeted at high risk areas.

The quantity of infectious particles and duration of exposure to an active TB patient also directly correlates with the risk of MTBC infection. In epidemiological investigations, household contacts and close neighbors are needed to be frequently screened as they are particularly exposed to viable pathogen (Kozinska & Augustynowicz-Kopec, 2016). Our analysis confirmed recent TB transmission among individuals belonging to the same household. We were able to show that majority of household related TB cases are due to recent transmission of the same strain and greater than 80 % of the first identified case in each household had a smear grade of >1+ implying high infectiousness. More importantly, we confirmed a household transmission of one INH resistant strain which shared only 4

University of Ghana http://ugspace.ug.edu.gh

SNPs between the isolate pair within a period of one year. This same strain was found circulating in a large recent transmission cluster within the same community. Also, another household related case involved three participants with the index case (who was likely infected from his workplace) spreading to his father and brother. Our finding supports a study in Zimbabwe which found that living in a home with two or more people per room was a significant risk factor for TB (Corbett *et al.*, 2009). Our study also supports a similar one in England which confirms that greater than 50% of household related cases are due to recent TB transmission (Lalor *et al.*, 2017). Public health actions needed to control such cases include extended contact tracing, venue screening and TB awareness-raising (Black *et al.*, 2018).

We identified age (p=0.007) and gender (p=0.022) to be significant risk factors for recent TB transmission but found no significant association of recent TB transmission with education status, occupation, income level, ethnicity, religion, or HIV status. Individuals below the age of 30 years as well as males were associated with recent TB transmission similar to observations made elsewhere (Hamblion *et al.*, 2016; Vluggen *et al.*, 2017). We observed that each year increase in age was associated with an approximately 1% decrease in the odds of a TB patient being part of a recent transmission event, implying that compared to younger individuals, older individuals are more likely to get active TB disease by reactivation of latent TB infection rather than through a recent transmission event. This is expected as older individuals are unlikely to involve in social activities that will expose them to recent infections resulting in disease. Global reports indicate more males have TB compared to females; However, the cause of this bias is uncertain. Our observation of males being more likely to be involved in a recent TB transmission event may be due to

University of Ghana http://ugspace.ug.edu.gh

one of several factors including socio-epidemiological factors (under-reporting of females probably due to stigma), behavioral factors (more social contacts; alcohol consumption; smoking) or perhaps biological factors (the role of sex hormones to modulate the immune responses necessary for resistance to TB) (Guerra-Silveira & Abad-Franch, 2013; Nhamoyebonde & Leslie, 2014).

In addition to these risk factors, we identified that it was unlikely (adjusted OR 0.7, 95% CI 0.5–0.9) for a drug resistant strain to be involved in recent transmission; we detected low to no recent transmission of multidrug-resistant TB suggesting that the majority of drug-resistant TB cases in Ghana acquired the drug resistance during treatment, which indicates poor patient compliance. This observation may be due to the low proportion (2 - 4%) of MDR among MTBC isolates in Ghana (Asare *et al.*, 2018; Otchere *et al.*, 2016) or probably due to the reduced fitness cost associated to resistance conferring mutations (Gagneux, 2009; Melnyk *et al.*, 2015). Until recently, drug resistant clones were thought to be less fit and less likely to transmit from person to person. However, recent studies have documented evidence of transmission even though not involving large clusters (Arandjelovic *et al.*, 2019; Coscolla *et al.*, 2015; Walker *et al.*, 2018). There is therefore the need to identify and control such difficult to treat drug resistant clones (especially when they are not transmitting actively) to stop their spread.

Molecular epidemiology is very helpful for tracking TB strains among recurring TB patients and provide indications of the cause of secondary case source, thus differentiating between exogenous re-infection (re-infection with new strain) as opposed to endogenous re-infections (relapse due to inadequate treatment) (Varghese *et al.*, 2013), for appropriate treatment, evaluation of program performance and epidemiology (Barnes & Cave, 2003;

University of Ghana http://ugspace.ug.edu.gh

van Soolingen *et al.*, 2003). Outcome of TB treatment is predominantly based on follow up sputum microscopy at the end of treatment, although sputum-smear microscopy has a low sensitivity. Based on direct smear microscopy, TB treatment success rate in Ghana is estimated at 85% (Amo-Adjei & Awusabo-Asare, 2013; WHO, 2018c); For the first time in Ghana, this study through comprehensive investigation and by employing WGS analysis, has provided substantial evidence of both relapse and re-infection among individuals with recurring TB episode. The majority (75.0%, 27/36) of recurring pulmonary TB cases were as a result of relapse of previous infection but not re-infection, similar to studies by Bryant *et al.,* that observed 70% relapse (Bryant *et al.*, 2013). Other studies (Varghese *et al.*, 2013) have used only traditional genotyping tools but we employed both traditional MIRU-VNTR typing and WGS to investigate strain relatedness. We are confident that at least 74% (20/27) of cases classified as relapse are true relapse cases as they do not cluster with any other recent transmission clusters previously defined from the same population. Predominance of relapse over re-infection indicates inadequate treatment and such cases have been associated with development of MDR (Alipanah *et al.*, 2018; Yang *et al.*, 2017). This implies that some individuals that are declared treatment successful may still harbor metabolically active bacilli. This is probably due to the poor sensitivity of smear microscopy used to monitor treatment. There is the need to adopt better alternative treatment monitoring strategy instead of relying mainly on smear microscopy to declare treatment success. Research on the development of biomarkers for monitoring of TB treatment may be a good area to invest resources into identifying useful biomarkers.

University of Ghana http://ugspace.ug.edu.gh

Overall, a greater proportion (58.3%) of the recurring TB cases occurred within a year post treatment, a duration that has been reported in similar studies (Thomas *et al.*, 2005; Velayutham *et al.*, 2018). Generally, it is assumed that in high endemic regions, recurrent TB cases should more likely be the result of re-infection rather than relapse (Parvaresh *et al.*, 2018; van Rie *et al.*, 1999). Our unexpected higher rate of relapse even in a high endemic setting has also been observed in other studies conducted in endemic regions like India and China where up to 93% of recurrent TB cases have been attributed to relapse and not re-infection (Velayutham *et al.*, 2018; Zong *et al.*, 2018).

7.1.4 Tools for studying TB transmission dynamics

Whole genome sequencing of MTBC has rapidly progressed from a research tool to a clinical application for diagnosing, managing and public health surveillance of TB. As an alternative, most studies to date have employed spoligotyping, IS*6110*-RFLP or MIRU-VNTR analysis. However, the lower cost with fairly good resolution methods (IS*6110*-RFLP and MIRU-VNTR) are fortunately powerful enough to effectively distinguish between MTBC strains associated with relapse/re-infection (G. M. Shen *et al.*, 2006; X. Shen *et al.*, 2017) or epidemiologically linked (household related) recent transmission (Lalor *et al.*, 2017; Sia *et al.*, 2013) for a given patient or closely related patients respectively.

Some research groups have wondered if 15-locus MIRU-VNTR is enough to study strain relationship (Gibson *et al.*, 2005; Kozinska & Augustynowicz-Kopec, 2016). Here we show that though WGS has a higher resolution power, 15-locus MIRU-VNTR typing plus spoligotyping is enough to predict the cause of TB recurrence and additionally able to predict household related recent transmission. Using 15-locus MIRU-VNTR analysis, we

University of Ghana http://ugspace.ug.edu.gh

showed a high concordance (94.6%, Kappa=0.7702, p<0.001) and positive correlation ($R^2 = 0.817$, p<0.001) with WGS typing using paired isolates from recurring and household related TB cases. Thus 15-locus MIRU-VNTR typing is enough to predict the cause of recurrent TB as well as identify household related recent TB transmission. This can help improve TB control measures in resource-limited settings as it will enable the early identification of the source of infection and likely inform treatment selection based on previous antimicrobial susceptibility history of that strain. Similar studies by Bryant *et al.,* identified only 6 out of 33 relapse cases to be misclassified by MIRUVNTR (Bryant *et al.*, 2013). To further reduce the total cost of the 15-locus MIRU-VNTR, we performed the analysis in a more cost-effective way by first interrogating a customized 8-locus set (Asante-Poku, Nyaho*, et al.*, 2014) and further differentiating resulting clustered cases using the remaining 7/15-loci.

After previously estimating the recent transmission index to be 41.2% using combined MIRU-VNTR and spoligotyping we found that the estimated index using WGS analysis was 24.7%. This reduced rate was anticipated as the discriminatory power of WGS analysis is higher (Walker, Ip*, et al.*, 2013) and other studies have also identified reduced rate after using WGS (Stucki *et al.*, 2015). Apart from being able to predict relapse, re-infection or household related recent transmission, we explored the ability of MIRU-VNTR combined with spoligotyping to define true monophyletic clades. Mapping the strain differentiation of combined MIRU-VNTR and spoligotyping with that of WGS, we observed nearly distinct monophyletic clades from a reconstructed phylogenetic tree using WGS implying that traditional genotyping may still be useful as initial screening tools to

University of Ghana http://ugspace.ug.edu.gh

help reduce the huge cost of WGS of all isolates especially in large-size population-based studies.

7.1.5 Limitations

1. I did not collect GPS coordinate of the residence of the study participants, so each participant residential district was generated using his or her residential address from the questionnaire. This might not be accurate and so future studies should be done using GPS coordinates taken from each individual participant's actual residence.
2. This report was strengthened by recruiting up to 90% of all diagnosed TB cases through a population based study spanning 3.5 years; however the study could be limited by the possibility of an underestimation of the recent transmission rate resulting from the misclassification of strains as unique if they were clustered outside of the restricted geographic sampling site and sampling period.

University of Ghana http://ugspace.ug.edu.gh

7.2 Conclusions

This PhD work was carried out in Ghana which is one of the few countries with the most diverse MTBC lineages. The work was mainly carried out at the Noguchi Memorial Institute for Medical Research under the College of Health Sciences, University of Ghana. However, other aspects were conducted at the Swiss Tropical and Public Health Institute, University of Basel, Switzerland. The study, by far, represents the most comprehensive investigation into the transmission dynamics of MTBC in urban West Africa, Ghana and calls for replicative studies in other distinct populations to aid in intervention development. The most important scientific achievements were:

1. Using standard genotyping tools, this study demonstrated that overall, 79% of TB in Ghana is caused by MTBss, 20% by MAF and 1% by animal strains. Unlike what has been found in some West African countries, we did not observe a decline in MAF prevalence in Ghana.
2. The Beijing and Ghana genotypes of lineages 2 and 4, respectively, as well as the animal adapted MTBCs were isolated more often from patients from Northern than Southern Ghana.
3. For the first time using WGS, we confirm high recent TB transmission within the population driven largely by MTBss sub-lineages Cameroon and Ghana. Observed reduced recent transmission of MAF suggests other factor(s) (host/environmental) may be responsible for its continuous presence in West Africa.
4. There is a wide spread of a clone of the Cameroon sub-lineage of lineage 4 with an ongoing transmission at hotspots mostly found within the Ablekuma sub-district of the Accra metropolitan area.

University of Ghana http://ugspace.ug.edu.gh

5. Risk factor analysis using logistic regression modeling, identified age and gender as significant risk factors for recent TB transmission in Ghana; individuals with age < 30 years as well as males have a higher risk of recent TB transmission. Compared to younger individuals, older individuals are more likely to get active TB disease by reactivation of latent TB infection rather than through a recent transmission event.
6. There is low to no recent transmission of multidrug-resistant TB suggesting that most drug-resistant TB cases in Ghana acquired the drug resistance during treatment, which may indicate poor patient compliance and speaks for improvement in procedures for monitoring treatment.
7. WGS was useful in detecting unsuspected outbreaks hence we recommend its use not only as a research tool but as a surveillance tool to aid in providing the necessary guided steps to track, monitor and control TB.
8. For the first time in Ghana, using WGS, we have shown that unresolved previous infection due to inadequate treatment is largely the cause of recurring pulmonary TB and also show that household related TB cases are likely the results of recent TB transmission within the house or from neighboring recent transmission events.
9. Though WGS is ideal, 15-locus MIRU-VNTR typing is enough to predict both the cause of TB recurrence and household related transmission at a high agreement compared to WGS (84.6%, Kappa=0.7702, $p<0.001$). Thus, it is possible to monitor recurring TB cases and follow-up household related transmission in a resource-limited setting.

University of Ghana http://ugspace.ug.edu.gh

7.3 Recommendations

1. Community education and outreach programs should be intensified to improve early case reporting and infection control.
   - ➢ Case search approaches should be targeted at high risk areas
2. There should be periodic screening of both livestock farmers and direct livestock handlers for early detection and treatment of cases.
3. There should be periodic epidemiological investigation and screening of household contacts and close neighbors as they may be particularly exposed to viable pathogen.
   - ➢ Through comprehensive contact tracing, compound house screening and follow-up study.
4. There is the need to adopt good measures to help reduce the occurrence of relapse cases and improve treatment outcome through;
   - ➢ using adherence interventions such as patient education and counseling
   - ➢ psychological interventions
   - ➢ incentives and enablers
   - ➢ digital health technologies
   - ➢ better methods of declaring treatment success by including culture.

   These measures are necessary as drug resistance strains can emerge due to lack of treatment adherence.
5. We recommend the use of the low cost 15-locus MIRU-VNTR typing for initial screening to identify possible sources of recurring TB and household related transmissions.

University of Ghana http://ugspace.ug.edu.gh

- We recommend local control programs to invest more resources into such studies as they have been proven to provide vital findings that positively influence TB control (Velayutham et al., 2018).

6. Due to the high recent transmission of the Ghana sub-lineage coupled with recently reported association with drug resistance, the national tuberculosis control program should support peripheral diagnostic laboratories with facilities to accurately detect and help control the spread of the Ghana sub-lineage.
7. There is a great need for more resources to be devoted to host-pathogen interaction studies to understand why MAF lineages are endemic in West Africa.

University of Ghana http://ugspace.ug.edu.gh

References

Aboagye, S. Y., Danso, E., Ampah, K. A., Nakobu, Z., Asare, P., Otchere, I. D., . . . Yeboah-Manu, D. (2016). Isolation of Nontuberculous Mycobacteria from the Environment of Buruli Ulcer Endemic Communities in Ghana. *Appl Environ Microbiol*. doi:10.1128/aem.01002-16

Addo, K., Owusu-Darko, K., Yeboah-Manu, D., Caulley, P., Minamikawa, M., Bonsu, F., . . . Ofori-Adjei, D. (2007). Mycobacterial species causing pulmonary tuberculosis at the korle bu teaching hospital, accra, ghana. *Ghana Medical Journal, 41*(2), 52-57.

Aeras. (2014). *TB Vaccine Research and Development: A Business Case for Investment*. Retrieved from Rockville, Aeras: http://bit.ly/1EodJBj

Afaghi-Gharamaleki, A., Moaddab, S., Darbouy, M., Ansarin, K., & Hanifian, S. (2017). Determining the Risk of Intra-Community Transmission of Tuberculosis in the Northwest of Iran Through 15 Loci Miru-Vntr Typing. *European Journal of Microbiology & Immunology, 7*(1), 46-54. doi:10.1556/1886.2016.00033

Affolabi, D., Anyo, G., Faihun, F., Sanoussi, N., Shamputa, I. C., Rigouts, L., . . . Portaels, F. (2009). First molecular epidemiological study of tuberculosis in Benin. *Int J Tuberc Lung Dis, 13*(3), 317-322.

Ahmad, S. (2011). Pathogenesis, immunology, and diagnosis of latent *Mycobacterium tuberculosis* infection. *Clin Dev Immunol, 2011*, 814943. doi:10.1155/2011/814943

Alene, K. A., Yi, H., Viney, K., McBryde, E. S., Yang, K., Bai, L., . . . Xu, Z. (2017). Treatment outcomes of patients with multidrug-resistant and extensively drug resistant tuberculosis in Hunan Province, China. *BMC Infectious Diseases, 17*(1), 573-573. doi:10.1186/s12879-017-2662-8

Alexander, D. C., Guthrie, J. L., Pyskir, D., Maki, A., Kurepina, N., Kreiswirth, B. N., . . . Jamieson, F. (2009). *Mycobacterium tuberculosis* in Ontario, Canada: Insights from IS6110 restriction fragment length polymorphism and mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping. *Journal of Clinical Microbiology, 47*(8), 2651-2654. doi:10.1128/JCM.01946-08

Alipanah, N., Jarlsberg, L., Miller, C., Linh, N. N., Falzon, D., Jaramillo, E., & Nahid, P. (2018). Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies. *PLoS Medicine, 15*(7), e1002595. doi:10.1371/journal.pmed.1002595

University of Ghana http://ugspace.ug.edu.gh

Allix-Beguec, C., Harmsen, D., Weniger, T., Supply, P., & Niemann, S. (2008). Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *Journal of Clinical Microbiology, 46*(8), 2692-2699. doi:10.1128/JCM.00540-08

Allix-Beguec, C., Supply, P., Wanlin, M., Bifani, P., & Fauville-Dufaux, M. (2008). Standardised PCR-based molecular epidemiology of tuberculosis. *European Respiratory Journal, 31*(5), 1077-1084. doi:10.1183/09031936.00053307

Allix-Béguec, C., Wahl, C., Hanekom, M., Nikolayevskyy, V., Drobniewski, F., Maeda, S., . . . Supply, P. (2014). Proposal of a Consensus Set of Hypervariable Mycobacterial Interspersed Repetitive-Unit–Variable-Number Tandem-Repeat Loci for Subtyping of *Mycobacterium tuberculosis* Beijing Isolates. *Journal of Clinical Microbiology, 52*(1), 164-172. doi:10.1128/JCM.02519-13

Allix, C., Supply, P., & Fauville-Dufaux, M. (2004). Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis. *Clinical Infectious Diseases, 39*(6), 783-789. doi:10.1086/423383

Alonso-Rodriguez, N., Martinez-Lirola, M., Herranz, M., Sanchez-Benitez, M., Barroso, P., group, I.-T., . . . Garcia de Viedma, D. (2008). Evaluation of the new advanced 15-loci MIRU-VNTR genotyping tool in *Mycobacterium tuberculosis* molecular epidemiology studies. *BMC Microbiology, 8*, 34. doi:10.1186/1471-2180-8-34

Amo-Adjei, J., & Awusabo-Asare, K. (2013). Reflections on tuberculosis diagnosis and treatment outcomes in Ghana. *Arch Public Health, 71*(1), 22. doi:10.1186/2049-3258-71-22

Anderson, L. F., Tamne, S., Brown, T., Watson, J. P., Mullarkey, C., Zenner, D., & Abubakar, I. (2014). Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. *Lancet Infectious Diseases, 14*(5), 406-415. doi:10.1016/S1473-3099(14)70022-2

Ano, H., Matsumoto, T., Yoshida, H., Nagai, M., Tamura, Y., Nishimori, K., . . . Tsuyuguchi, I. (2006). [Molecular epidemiology of tuberculosis by the use of IS6110 restriction fragment length polymorphism: a study from 2001 to 2003]. *Kekkaku, 81*(4), 321-328.

University of Ghana http://ugspace.ug.edu.gh

Aranaz, A., Liebana, E., Gomez-Mampaso, E., Galan, J. C., Cousins, D., Ortega, A., . . . Dominguez, L. (1999). *Mycobacterium tuberculosis* subsp. caprae subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. *International Journal of Systematic Bacteriology, 49 Pt 3*, 1263-1273. doi:10.1099/00207713-49-3-1263

Arandjelovic, I., Merker, M., Richter, E., Kohl, T. A., Savic, B., Soldatovic, I., . . . Niemann, S. (2019). Longitudinal Outbreak of Multidrug-Resistant Tuberculosis in a Hospital Setting, Serbia. *Emerging Infectious Diseases, 25*(3), 555-558. doi:10.3201/eid2503.181220

Asante-Poku, A., Aning, K. G., Boi-Kikimoto, B., & Yeboah-Manu, D. (2014). Prevalence of bovine tuberculosis in a dairy cattle farm and a research farm in Ghana. *Onderstepoort Journal of Veterinary Research, 81*(2), E1-6. doi:10.4102/ojvr.v81i2.716

Asante-Poku, A., Asare, P., Baddoo, N. A., Forson, A., Klevor, P., Otchere, I. D., . . . Yeboah-Manu, D. (2019). TB-diabetes co-morbidity in Ghana: The importance of *Mycobacterium africanum* infection. *PloS One, 14*(2), e0211822. doi:10.1371/journal.pone.0211822

Asante-Poku, A., Nyaho, M. S., Borrell, S., Comas, I., Gagneux, S., & Yeboah-Manu, D. (2014). Evaluation of customised lineage-specific sets of MIRU-VNTR loci for genotyping *Mycobacterium tuberculosis* complex isolates in Ghana. *PloS One, 9*(3), e92675. doi:10.1371/journal.pone.0092675

Asante-Poku, A., Otchere, I. D., Osei-Wusu, S., Sarpong, E., Baddoo, A., Forson, A., . . . Yeboah-Manu, D. (2016). Molecular epidemiology of *Mycobacterium africanum* in Ghana. *BMC Infectious Diseases, 16*, 385. doi:10.1186/s12879-016-1725-6

Asante-Poku, A., Yeboah-Manu, D., Otchere, I. D., Aboagye, S. Y., Stucki, D., Hattendorf, J., . . . Gagneux, S. (2015). *Mycobacterium africanum* is associated with patient ethnicity in Ghana. *PLoS Neglected Tropical Diseases, 9*(1), e3370. doi:10.1371/journal.pntd.0003370

Asare, P., Asante-Poku, A., Prah, D. A., Borrell, S., Osei-Wusu, S., Otchere, I. D., . . . Yeboah-Manu, D. (2018). Reduced transmission of *Mycobacterium africanum* compared to *Mycobacterium tuberculosis* in urban West Africa. *International Journal of Infectious Diseases*. doi:10.1016/j.ijid.2018.05.014

University of Ghana http://ugspace.ug.edu.gh

Assam Assam, J. P., Penlap Beng, V., Cho-Ngwa, F., Toukam, M., Ngoh, A. A., Kitavi, M., . . . Titanji, V. P. (2013). *Mycobacterium tuberculosis* is the causative agent of tuberculosis in the southern ecological zones of Cameroon, as shown by genetic analysis. *BMC Infectious Diseases, 13*, 431. doi:10.1186/1471-2334-13-431

Balasubramanian, V., Wiegeshaus, E. H., Taylor, B. T., & Smith, D. W. (1994). Pathogenesis of tuberculosis: pathway to apical localization. *Tubercle and Lung Disease, 75*(3), 168-178. doi:10.1016/0962-8479(94)90002-7

Barnard, M., Albert, H., Coetzee, G., O'Brien, R., & Bosman, M. E. (2008). Rapid molecular screening for multidrug-resistant tuberculosis in a high-volume public health laboratory in South Africa. *Am J Respir Crit Care Med, 177*(7), 787-792. doi:10.1164/rccm.200709-1436OC

Barnes, P. F., & Cave, M. D. (2003). Molecular Epidemiology of Tuberculosis. *New England Journal of Medicine, 349*(12), 1149-1156. doi:doi:10.1056/NEJMra021964

Bastida, Z. A., Hernández Tellez, M., Bustamante Montes, L. P., Medina Torres, I., Jaramillo Paniagua, J. N., Mendoza Martínez, G. D., & Ramírez Durán, N. (2012). Spatial and Temporal Distribution of Tuberculosis in the State of Mexico, Mexico. *The Scientific World Journal, 2012*, 570278. doi:10.1100/2012/570278

Benjak, A., Sala, C., & Hartkoorn, R. C. (2015). Whole-genome sequencing for comparative genomics and de novo genome assembly. *Methods in Molecular Biology, 1285*, 1-16. doi:10.1007/978-1-4939-2450-9_1

Black, A. T., Hamblion, E. L., Buttivant, H., Anderson, S. R., Stone, M., Casali, N., . . . Booth, L. (2018). Tracking and responding to an outbreak of tuberculosis using MIRU-VNTR genotyping and whole genome sequencing as epidemiological tools. *J Public Health (Oxf), 40*(2), e66-e73. doi:10.1093/pubmed/fdx075

Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., Dehan, C., . . . Vergnaud, G. (2012). Significance of the identification in the Horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. *PloS One, 7*(12), e52841. doi:10.1371/journal.pone.0052841

Bodnar, K. A., Serbina, N. V., & Flynn, J. L. (2001). Fate of *Mycobacterium tuberculosis* within murine dendritic cells. *Infect Immun, 69*(2), 800-809. doi:10.1128/iai.69.2.800-809.2001

University of Ghana http://ugspace.ug.edu.gh

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. *Journal of Infectious Diseases, 205*(6), 984-990. doi:10.1093/infdis/jir883

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., . . . Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature, 514*(7523), 494-497. doi:10.1038/nature13591

Brites, D., Loiseau, C., Menardo, F., Borrell, S., Boniotti, M. B., Warren, R., . . . Gagneux, S. (2018). A New Phylogenetic Framework for the Animal-Adapted *Mycobacterium tuberculosis* Complex. *Frontiers in Microbiology, 9*, 2820. doi:10.3389/fmicb.2018.02820

Brosch, R., Gordon, S. V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., . . . Cole, S. T. (2002). A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proceedings of the National Academy of Sciences of the United States of America, 99*(6), 3684-3689. doi:10.1073/pnas.052548299

Bryant, J. M., Harris, S. R., Parkhill, J., Dawson, R., Diacon, A. H., van Helden, P., . . . Bentley, S. D. (2013). Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir Med, 1*(10), 786-792. doi:10.1016/S2213-2600(13)70231-5

Chihota, V. N., Niehaus, A., Streicher, E. M., Wang, X., Sampson, S. L., Mason, P., . . . Warren, R. M. (2018). Geospatial distribution of *Mycobacterium tuberculosis* genotypes in Africa. *PloS One, 13*(8), e0200632-e0200632. doi:10.1371/journal.pone.0200632

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin), 6*(2), 80-92. doi:10.4161/fly.19695

Cohen, K. A., Manson, A. L., Abeel, T., Desjardins, C. A., Chapman, S. B., Hoffner, S., . . . Earl, A. M. (2019). Extensive global movement of multidrug-resistant *M. tuberculosis* strains revealed by whole-genome analysis. *Thorax*. doi:10.1136/thoraxjnl-2018-211616

University of Ghana http://ugspace.ug.edu.gh

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., . . . Barrell, B. G. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature, 393*(6685), 537-544. doi:10.1038/31159

Coll, F., McNerney, R., Preston, M. D., Guerra-Assuncao, J. A., Warry, A., Hill-Cawthorne, G., . . . Clark, T. G. (2015). Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Medicine, 7*(1), 51. doi:10.1186/s13073-015-0164-0

Comas, I., Chakravartti, J., Small, P. M., Galagan, J., Niemann, S., Kremer, K., . . . Gagneux, S. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nature Genetics, 42*(6), 498-503. doi:10.1038/ng.590

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. doi:10.1371/journal.pone.0007815

Corbett, E. L., Bandason, T., Cheung, Y. B., Makamure, B., Dauya, E., Munyati, S. S., . . . Mason, P. R. (2009). Prevalent infectious tuberculosis in Harare, Zimbabwe: burden, risk factors and implications for control. *Int J Tuberc Lung Dis, 13*(10), 1231-1237.

Coscolla, M., Barry, P. M., Oeltmann, J. E., Koshinsky, H., Shaw, T., Cilnis, M., . . . Metcalfe, J. Z. (2015). Genomic epidemiology of multidrug-resistant *Mycobacterium tuberculosis* during transcontinental spread. *Journal of Infectious Diseases, 212*(2), 302-310. doi:10.1093/infdis/jiv025

Cousins, D. V., Bastida, R., Cataldi, A., Quse, V., Redrobe, S., Dow, S., . . . Bernardelli, A. (2003). Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol, 53*(Pt 5), 1305-1314. doi:10.1099/ijs.0.02401-0

Crampin, A. C., Glynn, J. R., Traore, H., Yates, M. D., Mwaungulu, L., Mwenebabu, M., . . . Fine, P. E. M. (2006). Tuberculosis Transmission Attributable to Close Contacts and HIV Status, Malawi. *Emerging Infectious Diseases, 12*(5), 729-735. doi:10.3201/eid1205.050789

Dangisso, M. H., Datiko, D. G., & Lindtjorn, B. (2015). Spatio-temporal analysis of smear-positive tuberculosis in the Sidama Zone, southern Ethiopia. *PloS One, 10*(6), e0126369. doi:10.1371/journal.pone.0126369

University of Ghana http://ugspace.ug.edu.gh

Danso, E., Addo, I. Y., & Ampomah, I. G. (2015). Patients' Compliance with Tuberculosis Medication in Ghana: Evidence from a Periurban Community. *Advances in Public Health, 2015*, 6. doi:10.1155/2015/948487

Davenne, T., & McShane, H. (2016). Why don't we have an effective tuberculosis vaccine yet? *Expert review of vaccines, 15*(8), 1009-1013. doi:10.1586/14760584.2016.1170599

de Boer, A. S., Borgdorff, M. W., de Haas, P. E., Nagelkerke, N. J., van Embden, J. D., & van Soolingen, D. (1999). Analysis of rate of change of IS6110 RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. *Journal of Infectious Diseases, 180*(4), 1238-1244. doi:10.1086/314979

de Jong, B. C., Adetifa, I., Walther, B., Hill, P. C., Antonio, M., Ota, M., & Adegbola, R. A. (2010). Differences between TB cases infected with *M. africanum*, West-African type 2, relative to Euro-American *M. tuberculosis*- an update. *FEMS immunology and medical microbiology, 58*(1), 102-105. doi:10.1111/j.1574-695X.2009.00628.x

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. doi:10.1371/journal.pntd.0000744

de Jong, B. C., Hill, P. C., Aiken, A., Awine, T., Antonio, M., Adetifa, I. M., . . . Adegbola, R. A. (2008). Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in The Gambia. *Journal of Infectious Diseases, 198*(7), 1037-1043. doi:10.1086/591504

de Jong, B. C., Hill, P. C., Brookes, R. H., Gagneux, S., Jeffries, D. J., Otu, J. K., . . . Adegbola, R. A. (2006). *Mycobacterium africanum* elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. *Journal of Infectious Diseases, 193*(9), 1279-1286. doi:10.1086/502977

de Jong, B. C., Hill, P. C., Brookes, R. H., Otu, J. K., Peterson, K. L., Small, P. M., & Adegbola, R. A. (2005). *Mycobacterium africanum*: a new opportunistic pathogen in HIV infection? *AIDS, 19*(15), 1714-1715.

University of Ghana http://ugspace.ug.edu.gh

Debebe, T., Admassu, A., Mamo, G., & Ameni, G. (2014). Molecular characterization of *Mycobacterium tuberculosis* isolated from pulmonary tuberculosis patients in Felege Hiwot Referral Hospital, northwest Ethiopia. *Journal of Microbiology, Immunology and Infection, 47*(4), 333-338. doi:https://doi.org/10.1016/j.jmii.2013.03.012

Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., . . . Rastogi, N. (2012). SITVITWEB – A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infection, Genetics and Evolution, 12*(4), 755-766. doi:http://doi.org/10.1016/j.meegid.2012.02.004

Diop, S., de Medeiros, D., de Medeiros, G., Baylet, R., & Sankale, M. (1976). [Incidence and geographic distribution of *Mycobacterium africanum* in Senegal]. *Bulletin de la Société Médicale d'Afrique Noire de Langue Française, 21*(1), 50-56.

Dosso, M., Bonard, D., Msellati, P., Bamba, A., Doulhourou, C., Vincent, V., . . . Coulibaly, I. M. (1999). Primary resistance to antituberculosis drugs: a national survey conducted in Cote d'Ivoire in 1995-1996. Ivoirian Study Group on Tuberculosis Resistance. *Int J Tuberc Lung Dis, 3*(9), 805-809.

Eisenach, K. D., Crawford, J. T., & Bates, J. H. (1988). Repetitive DNA sequences as probes for *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology, 26*(11), 2240-2245.

ESRI. (2010). Environmental Systems Resource Institute. Redlands, California.

Esteban, J., & Munoz-Egea, M. C. (2016). *Mycobacterium bovis* and Other Uncommon Members of the *Mycobacterium tuberculosis* Complex. *Microbiol Spectr, 4*(6). doi:10.1128/microbiolspec.TNMI7-0021-2016

Falkinham, J. O., 3rd. (2015). Environmental sources of nontuberculous mycobacteria. *Clinics in Chest Medicine, 36*(1), 35-41. doi:10.1016/j.ccm.2014.10.003

Field, N., Cohen, T., Struelens, M. J., Palm, D., Cookson, B., Glynn, J. R., . . . Abubakar, I. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *The Lancet Infectious Diseases, 14*(4), 341-352. doi:10.1016/S1473-3099(13)70324-4

Field, N., Cohen, T., Struelens, M. J., Palm, D., Cookson, B., Glynn, J. R., . . . Abubakar, I. (2014). Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infectious Diseases, 14*(4), 341-352. doi:10.1016/S1473-3099(13)70324-4

University of Ghana http://ugspace.ug.edu.gh

Franzblau, S. G., Witzig, R. S., McLaughlin, J. C., Torres, P., Madico, G., Hernandez, A., . . . Gilman, R. H. (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.

Frieden, T. R., & Sbarbaro, J. A. (2007). Promoting adherence to treatment for tuberculosis: the importance of direct observation. *World Hospitals and Health Services, 43*(2), 30-33.

Frota, C. C., Hunt, D. M., Buxton, R. S., Rickman, L., Hinds, J., Kremer, K., . . . Colston, M. J. (2004). Genome structure in the vole bacillus, *Mycobacterium microti*, a member of the *Mycobacterium tuberculosis* complex with a low virulence for humans. *Microbiology, 150*(Pt 5), 1519-1527. doi:10.1099/mic.0.26660-0

Gagneux, S. (2009). Fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection, 15*, 66-68. doi:https://doi.org/10.1111/j.1469-0691.2008.02685.x

Gagneux, S. (2012). Host-pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 367*(1590), 850-859. doi:10.1098/rstb.2011.0316

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B. C., Narayanan, S., . . . Small, P. M. (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. doi:10.1073/pnas.0511240103

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. doi:10.1016/s1473-3099(07)70108-1

Gardy, J. L., Johnston, J. C., Ho Sui, S. J., Cook, V. J., Shah, L., Brodkin, E., . . . Tang, P. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine, 364*(8), 730-739. doi:10.1056/NEJMoa1003176

Garnier, T., Eiglmeier, K., Camus, J. C., Medina, N., Mansoor, H., Pryor, M., . . . Hewinson, R. G. (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. doi:10.1073/pnas.1130426100

University of Ghana http://ugspace.ug.edu.gh

Gauthier, M., Bidault, F., Mosnier, A., Bablishvili, N., Tukvadze, N., Somphavong, S., . . . Berland, J. L. (2015). High-throughput mycobacterial interspersed repetitive-unit-variable-number tandem-repeat genotyping for *Mycobacterium tuberculosis* epidemiological studies. *Journal of Clinical Microbiology, 53*(2), 498-503. doi:10.1128/JCM.01611-14

Gehre, F., Antonio, M., Faihun, F., Odoun, M., Uwizeye, C., de Rijk, P., . . . Affolabi, D. (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. doi:10.1371/journal.pone.0077000

Gehre, F., Kumar, S., Kendall, L., Ejo, M., Secka, O., Ofori-Anyinam, B., . . . de Jong, B. C. (2016). A Mycobacterial Perspective on Tuberculosis in West Africa: Significant Geographical Variation of *M. africanum* and Other *M. tuberculosis* Complex Lineages. *PLoS Neglected Tropical Diseases, 10*(3), e0004408. doi:10.1371/journal.pntd.0004408

GHS. (2015). Ghana health Service (GHS) Annual Report for 2014.

Gibson, A., Brown, T., Baker, L., & Drobniewski, F. (2005). Can 15-locus mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis provide insight into the evolution of *Mycobacterium tuberculosis*? *Appl Environ Microbiol, 71*(12), 8207-8213. doi:10.1128/AEM.71.12.8207-8213.2005

Glynn, J. R., Alghamdi, S., Mallard, K., McNerney, R., Ndlovu, R., Munthali, L., . . . Crampin, A. C. (2010). Changes in *Mycobacterium tuberculosis* genotype families over 20 years in a population-based study in Northern Malawi. *PloS One, 5*(8), e12259. doi:10.1371/journal.pone.0012259

Glynn, J. R., Vynnycky, E., & Fine, P. E. (1999). Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. *American Journal of Epidemiology, 149*(4), 366-371.

Gorge, O., Lopez, S., Hilaire, V., Lisanti, O., Ramisse, V., & Vergnaud, G. (2008). Selection and validation of a multilocus variable-number tandem-repeat analysis panel for typing Shigella spp. *Journal of Clinical Microbiology, 46*(3), 1026-1036. doi:10.1128/jcm.02027-07

University of Ghana http://ugspace.ug.edu.gh

Goyal, M., Lawn, S., Afful, B., Acheampong, J. W., Griffin, G., & Shaw, R. (1999). Spoligotyping in molecular epidemiology of tuberculosis in Ghana. *Journal of Infection, 38*(3), 171-175.

Groenen, P. M., Bunschoten, A. E., van Soolingen, D., & van Embden, J. D. (1993). Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular Microbiology, 10*(5), 1057-1065.

GSS. (2012). Ghana Statistical Service, National Analytical Report. 2010 Population and Housing Census. . Retrieved 31/01/2016 http://www.statsghana.gov.gh/publications.html

GSS. (2013a). Ghana Statistical Service, Central Region Analytical Report, 2010 Population and Housing Census. Retrieved 31/01/2016 http://www.statsghana.gov.gh/publications.html

GSS. (2013b). Ghana Statistical Service, Greater Accra Regional Analytical Report, 2010 Population and Housing Census. Retrieved 31/01/2016 http://www.statsghana.gov.gh/publications.html

GSS. (2013c). Ghana Statistical Service, Northern Region Analytical Report, 2010 Population and Housing Census. Retrieved 31/01/2016 http://www.statsghana.gov.gh/publications.html

GSS. (2014). Ghana Statistical Service, District Analytical Report. 2010 Population and Housing Census. Accra Metropolitan. Retrieved 31/01/2016 http://www.statsghana.gov.gh/publications.html

Guerra-Assunção, J. A., Crampin, A. C., Houben, R., Mzembe, T., Mallard, K., Coll, F., . . . Glynn, J. R. (2015). Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife, 4*, e05166. doi:10.7554/eLife.05166

Guerra-Assuncao, J. A., Houben, R. M., Crampin, A. C., Mzembe, T., Mallard, K., Coll, F., . . . Glynn, J. R. (2015). Recurrence due to relapse or reinfection with *Mycobacterium tuberculosis*: a whole-genome sequencing approach in a large, population-based cohort with a high HIV infection prevalence and active follow-up. *Journal of Infectious Diseases, 211*(7), 1154-1163. doi:10.1093/infdis/jiu574

Guerra-Silveira, F., & Abad-Franch, F. (2013). Sex bias in infectious disease epidemiology: patterns and processes. *PloS One, 8*(4), e62390. doi:10.1371/journal.pone.0062390

University of Ghana http://ugspace.ug.edu.gh

Gutierrez, M., Samper, S., Jimenez, M. S., van Embden, J. D., Marin, J. F., & Martin, C. (1997). Identification by spoligotyping of a caprine genotype in *Mycobacterium bovis* strains causing human tuberculosis. *Journal of Clinical Microbiology, 35*(12), 3328-3330.

Hall, A. (1996). What is molecular epidemiology? *Trop Med Int Health, 1*(4), 407-408.

Hamblion, E. L., Le Menach, A., Anderson, L. F., Lalor, M. K., Brown, T., Abubakar, I., . . . Anderson, S. R. (2016). Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax, 71*(8), 749-756. doi:10.1136/thoraxjnl-2014-206608

Hershberg, R., Lipatov, M., Small, P. M., Sheffer, H., Niemann, S., Homolka, S., . . . Gagneux, S. (2008). High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biology, 6*(12), e311. doi:10.1371/journal.pbio.0060311

Horstkotte, M. A., Sobottka, I., Schewe, C. K., Schafer, P., Laufs, R., Rusch-Gerdes, S., & Niemann, S. (2001). *Mycobacterium microti* llama-type infection presenting as pulmonary tuberculosis in a human immunodeficiency virus-positive patient. *Journal of Clinical Microbiology, 39*(1), 406-407. doi:10.1128/JCM.39.1.406-407.2001

Hu, Y., Ma, X., Graviss, E. A., Wang, W., Jiang, W., & Xu, B. (2011). A major subgroup of Beijing family *Mycobacterium tuberculosis* is associated with multidrug resistance and increased transmissibility. *Epidemiol Infect, 139*(1), 130-138. doi:10.1017/S0950268810000890

Huard, R. C., Lazzarini, L. C., Butler, W. R., van Soolingen, D., & Ho, J. L. (2003). PCR-based method to differentiate the subspecies of the *Mycobacterium tuberculosis* complex on the basis of genomic deletions. *Journal of Clinical Microbiology, 41*(4), 1637-1650. doi:10.1128/jcm.41.4.1637-1650.2003

Huet, M., Rist, N., Boube, G., & Potier, D. (1971). [Bacteriological study of tuberculosis in Cameroon]. *Revue de Tuberculose et de Pneumologie, 35*(4), 413-426.

Hunter, P. R., & Gaston, M. A. (1988). Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology, 26*(11), 2465-2466.

University of Ghana http://ugspace.ug.edu.gh

Jagielski, T., Minias, A., van Ingen, J., Rastogi, N., Brzostek, A., Żaczek, A., & Dziadek, J. (2016). Methodological and Clinical Aspects of the Molecular Epidemiology of *Mycobacterium tuberculosis* and Other Mycobacteria. *Clinical Microbiology Reviews, 29*(2), 239-290. doi:10.1128/CMR.00055-15

Jagielski, T., van Ingen, J., Rastogi, N., Dziadek, J., Mazur, P. K., & Bielecki, J. (2014). Current methods in the molecular typing of *Mycobacterium tuberculosis* and other mycobacteria. *Biomed Res Int, 2014*, 645802. doi:10.1155/2014/645802

Jakko van, I., Zeaur, R., Arnout, M., Martin, J. B., Roxane, S., Roland, B., & Dick van, S. (2012). Characterization of *Mycobacterium orygis* as *M. tuberculosis* Complex Subspecies. *Emerging Infectious Disease journal, 18*(4), 653. doi:10.3201/eid1804.110888

Jamieson, F. B., Teatero, S., Guthrie, J. L., Neemuchwala, A., Fittipaldi, N., & Mehaffy, C. (2014). Whole-genome sequencing of the *Mycobacterium tuberculosis* Manila sublineage results in less clustering and better resolution than mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing and spoligotyping. *Journal of Clinical Microbiology, 52*(10), 3795-3798. doi:10.1128/JCM.01726-14

Kallenius, G., Koivula, T., Ghebremichael, S., Hoffner, S. E., Norberg, R., Svensson, E., . . . Svenson, S. B. (1999). Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. *Journal of Clinical Microbiology, 37*(12), 3872-3878.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., . . . van Embden, J. (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology, 35*(4), 907-914.

Kammerer, J. S., Shang, N., Althomsons, S. P., Haddad, M. B., Grant, J., & Navin, T. R. (2013). Using statistical methods and genotyping to detect tuberculosis outbreaks. *Int J Health Geogr, 12*, 15. doi:10.1186/1476-072x-12-15

Kanduma, E., McHugh, T. D., & Gillespie, S. H. (2003). Molecular methods for *Mycobacterium tuberculosis* strain typing: a users guide. *Journal of Applied Microbiology, 94*(5), 781-791.

Kang, D. D., Lin, Y., Moreno, J. R., Randall, T. D., & Khader, S. A. (2011). Profiling early lung immune responses in the mouse model of tuberculosis. *PloS One, 6*(1), e16161. doi:10.1371/journal.pone.0016161

University of Ghana http://ugspace.ug.edu.gh

Klevytska, A. M., Price, L. B., Schupp, J. M., Worsham, P. L., Wong, J., & Keim, P. (2001). Identification and characterization of variable-number tandem repeats in the Yersinia pestis genome. *Journal of Clinical Microbiology, 39*(9), 3179-3185.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., . . . Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research, 22*(3), 568-576. doi:10.1101/gr.129684.111

Koro Koro, F., Kamdem Simo, Y., 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. doi:10.1128/jcm.01196-12

Kozinska, M., & Augustynowicz-Kopec, E. (2016). The incidence of tuberculosis transmission among family members and outside households. *Pneumonologia i Alergologia Polska, 84*(5), 271-277. doi:10.5603/PiAP.2016.0034

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics. Theory and Methods, 26*(6), 1481-1496.

Kulldorff, M. (2015). *SaTScanTM User Guide for version 9.4*.

Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine, 14*(8), 799-810.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

Lalor, M. K., Anderson, L. F., Hamblion, E. L., Burkitt, A., Davidson, J. A., Maguire, H., . . . Thomas, H. L. (2017). Recent household transmission of tuberculosis in England, 2010-2012: retrospective national cohort study combining epidemiological and molecular strain typing data. *BMC Medicine, 15*(1), 105. doi:10.1186/s12916-017-0864-y

Lalor, M. K., Casali, N., Walker, T. M., Anderson, L. F., Davidson, J. A., Ratna, N., . . . Abubakar, I. (2018). The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak. *European Respiratory Journal, 51*(6). doi:10.1183/13993003.02313-2017

Langhi, D. M., Jr., & Bordin, J. O. (2006). Duffy blood group and malaria. *Hematology (Amsterdam, Netherlands), 11*(5), 389-398. doi:10.1080/10245330500469841

University of Ghana http://ugspace.ug.edu.gh

Laniado-Laborin, R., Muniz-Salazar, R., Garcia-Ortiz, R. A., Vargas-Ojeda, A. C., Villa-Rosas, C., & Oceguera-Palao, L. (2014). Molecular characterization of *Mycobacterium bovis* isolates from patients with tuberculosis in Baja California, Mexico. *Infection, Genetics and Evolution, 27*, 1-5. doi:10.1016/j.meegid.2014.06.020

Last, J. M. (2001). *A dictionary of epidemiology* (J. M. Last & A. International Epidemiological Eds. 4th ed ed.). New York: Oxford University Press.

Lawson, L., Zhang, J., Gomgnimbou, M. K., Abdurrahman, S. T., Le Moullec, S., Mohamed, F., . . . Sola, C. (2012). A molecular epidemiological and genetic diversity study of tuberculosis in Ibadan, Nnewi and Abuja, Nigeria. *PloS One, 7*(6), e38409. doi:10.1371/journal.pone.0038409

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

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

Lonnroth, K., Jaramillo, E., Williams, B. G., Dye, C., & Raviglione, M. (2009). Drivers of tuberculosis epidemics: the role of risk factors and social determinants. *Soc Sci Med, 68*(12), 2240-2246. doi:10.1016/j.socscimed.2009.03.041

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2019). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.9.

Maguire, H., Dale, J. W., McHugh, T. D., Butcher, P. D., Gillespie, S. H., Costetsos, A., . . . Banerjee, D. K. (2002). Molecular epidemiology of tuberculosis in London 1995-7 showing low rate of active transmission. *Thorax, 57*(7), 617-622.

Malm, S., Linguissi, L. S., Tekwu, E. M., Vouvoungui, J. C., Kohl, T. A., Beckert, P., . . . Niemann, S. (2017). New *Mycobacterium tuberculosis* Complex Sublineage, Brazzaville, Congo. *Emerging Infectious Diseases, 23*(3), 423-429. doi:10.3201/eid2303.160679

Manges, A. R., Tellis, P. A., Vincent, C., Lifeso, K., Geneau, G., Reid-Smith, R. J., & Boerlin, P. (2009). Multi-locus variable number tandem repeat analysis for Escherichia coli causing extraintestinal infections. *Journal of Microbiological Methods, 79*(2), 211-213. doi:10.1016/j.mimet.2009.09.006

University of Ghana http://ugspace.ug.edu.gh

Marsh, J. W., O'Leary, M. M., Shutt, K. A., Pasculle, A. W., Johnson, S., Gerding, D. N., . . . Harrison, L. H. (2006). Multilocus variable-number tandem-repeat analysis for investigation of Clostridium difficile transmission in Hospitals. *Journal of Clinical Microbiology, 44*(7), 2558-2566. doi:10.1128/jcm.02364-05

Masala, S., Molicotti, P., Bua, A., Zumbo, A., Delogu, G., Sechi, L. A., & Zanetti, S. (2010). Molecular characterization of Sardinian *Mycobacterium tuberculosis* isolates by IS6110 restriction fragment length polymorphism, MIRU-VNTR and rep-PCR. *New Microbiologica, 33*(2), 155-162.

McIvor, A., Koornhof, H., & Kana, B. D. (2017). Relapse, re-infection and mixed infections in tuberculosis disease. *Pathog Dis, 75*(3). doi:10.1093/femspd/ftx020

Mears, J., Abubakar, I., Cohen, T., McHugh, T. D., & Sonnenberg, P. (2015). Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): a systematic review. *BMJ Open, 5*(1), e005636. doi:10.1136/bmjopen-2014-005636

Melnyk, A. H., Wong, A., & Kassen, R. (2015). The fitness costs of antibiotic resistance mutations. *Evolutionary Applications, 8*(3), 273-283. doi:10.1111/eva.12196

Mendez, M. V., Leon, C., Escalona, A., Abadia, E., Da Mata, O., de Waard, J., & Takiff, H. E. (2016). [Evaluation of discriminatory power of molecular epidemiology techniques in *Mycobacterium tuberculosis* Venezuelan isolates]. *Investigacion Clinica, 57*(1), 25-37.

Merker, M., Kohl, T. A., Niemann, S., & Supply, P. (2017). The Evolution of Strain Typing in the *Mycobacterium tuberculosis* Complex. *Adv Exp Med Biol, 1019*, 43-78. doi:10.1007/978-3-319-64371-7_3

Meyer, C. G., Scarisbrick, G., Niemann, S., Browne, E. N. L., Chinbuah, M. A., Gyapong, J., . . . Horstmann, R. D. (2008). Pulmonary tuberculosis: Virulence of *Mycobacterium africanum* and relevance in HIV co-infection. *Tuberculosis, 88*(5), 482-489. doi:10.1016/j.tube.2008.05.004

Michel, A. L., Muller, B., & van Helden, P. D. (2010). *Mycobacterium bovis* at the animal-human interface: a problem, or not? *Veterinary Microbiology, 140*(3-4), 371-381. doi:10.1016/j.vetmic.2009.08.029

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010, 14-14 Nov. 2010). *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* Paper presented at the 2010 Gateway Computing Environments Workshop (GCE).

University of Ghana http://ugspace.ug.edu.gh

Mostowy, S., Cousins, D., Brinkman, J., Aranaz, A., & Behr, M. A. (2002). Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. *Journal of Infectious Diseases, 186*(1), 74-80. doi:10.1086/341068

Moström, P., Gordon, M., Sola, C., Ridell, M., & Rastogi, N. (2002). Methods used in the molecular epidemiology of tuberculosis Clinical Microbiology and Infection Volume 8, Issue 11. *Clinical Microbiology and Infection, 8*(11), 694-704. Retrieved from http://onlinelibrary.wiley.com/doi/10.1046/j.1469-0691.2002.00460.x/abstract; http://ac.els-cdn.com/S1198743X14626574/1-s2.0-S1198743X14626574-main.pdf?_tid=fb493748-9064-11e5-acd8-00000aacb35d&acdnat=1448120089_7e194f824b9e87dae90b7639e2823b48

Mulenga, C., Shamputa, I. C., Mwakazanga, D., Kapata, N., Portaels, F., & Rigouts, L. (2010). Diversity of *Mycobacterium tuberculosis* genotypes circulating in Ndola, Zambia. *BMC Infectious Diseases, 10*, 177. doi:10.1186/1471-2334-10-177

Murray, M., & Alland, D. (2002). Methodological problems in the molecular epidemiology of tuberculosis. *American Journal of Epidemiology, 155*(6), 565-571.

Na Songkhla, M., Tantipong, H., Tongsai, S., & Angkasekwinai, N. (2017). Lateral Flow Urine Lipoarabinomannan Assay (LF-LAM) for Diagnosis of Active Tuberculosis in HIV-Infected Adults: a Prospective Cohort Study. *Open Forum Infectious Diseases, 4*(Suppl 1), S620-S620. doi:10.1093/ofid/ofx163.1638

Narayanan, S., Swaminathan, S., Supply, P., Shanmugam, S., Narendran, G., Hari, L., . . . Narayanan, P. R. (2010). Impact of HIV infection on the recurrence of tuberculosis in South India. *Journal of Infectious Diseases, 201*(5), 691-703. doi:10.1086/650528

Ndiaye, B. P., Thienemann, F., Ota, M., Landry, B. S., Camara, M., Dieye, S., . . . investigators, M. A. t. (2015). Safety, immunogenicity, and efficacy of the candidate tuberculosis vaccine MVA85A in healthy adults infected with HIV-1: a randomised, placebo-controlled, phase 2 trial. *Lancet Respir Med, 3*(3), 190-200. doi:10.1016/S2213-2600(15)00037-5

Nebenzahl-Guimaraes, H., Verhagen, L. M., Borgdorff, M. W., & van Soolingen, D. (2015). Transmission and Progression to Disease of *Mycobacterium tuberculosis* Phylogenetic Lineages in The Netherlands. *Journal of Clinical Microbiology, 53*(10), 3264-3271. doi:10.1128/JCM.01370-15

University of Ghana http://ugspace.ug.edu.gh

Nhamoyebonde, S., & Leslie, A. (2014). Biological differences between the sexes and susceptibility to tuberculosis. *Journal of Infectious Diseases, 209 Suppl 3*, S100-106. doi:10.1093/infdis/jiu147

Niemann, S., Richter, E., & Rüsch-Gerdes, S. (1999). Stability of *Mycobacterium tuberculosis* IS6110 restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. *Journal of Clinical Microbiology, 37*(2), 409-412.

Niobe-Eyangoh, S. N., Kuaban, C., Sorlin, P., Cunin, P., Thonnon, J., Sola, C., . . . Gutierrez, M. C. (2003). Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. *Journal of Clinical Microbiology, 41*(6), 2547-2553.

Norheim, G., Seterelv, S., Arnesen, T. M., Mengshoel, A. T., Tonjum, T., Ronning, J. O., & Eldholm, V. (2017). Tuberculosis outbreak in an educational institution in Norway. *Journal of Clinical Microbiology*. doi:10.1128/JCM.01152-16

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 (Edinb), 96*, 107-119. doi:10.1016/j.tube.2015.09.006

Ofori-Anyinam, B., Kanuteh, F., Agbla, S. C., Adetifa, I., Okoi, C., Dolganov, G., . . . Gehre, F. (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. doi:10.1371/journal.pntd.0004801

Osei-Wusu, S., Amo Omari, M., Asante-Poku, A., Darko Otchere, I., Asare, P., Forson, A., . . . Yeboah-Manu, D. (2018). Second-line anti-tuberculosis drug resistance testing in Ghana identifies the first extensively drug-resistant tuberculosis case. *Infection and Drug Resistance, Volume 11*, 239-246. doi:10.2147/IDR.S152720

Osei, E., Der, J., Owusu, R., Kofie, P., & Axame, W. K. (2017). The burden of HIV on Tuberculosis patients in the Volta region of Ghana from 2012 to 2015: implication for Tuberculosis control. *BMC Infectious Diseases, 17*(1), 504. doi:10.1186/s12879-017-2598-z

Otal, I., Martin, C., Vincent-Levy-Frebault, V., Thierry, D., & Gicquel, B. (1991). Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. *Journal of Clinical Microbiology, 29*.

University of Ghana http://ugspace.ug.edu.gh

Otchere, I. D., Asante-Poku, A., Osei-Wusu, S., Baddoo, A., Sarpong, E., Ganiyu, A. H., . . . Yeboah-Manu, D. (2016). Detection and characterization of drug-resistant conferring genes in *Mycobacterium tuberculosis* complex strains: A prospective study in two distant regions of Ghana. *Tuberculosis (Edinb), 99*, 147-154. doi:10.1016/j.tube.2016.05.014

Otchere, I. D., Coscollá, M., Sánchez-Busó, L., Asante-Poku, A., Brites, D., Loiseau, C., . . . Yeboah-Manu, D. (2018). Comparative genomics of *Mycobacterium africanum* Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. *Scientific Reports, 8*(1), 11269. doi:10.1038/s41598-018-29620-2

Pai, M., Behr, M. A., Dowdy, D., Dheda, K., Divangahi, M., Boehme, C. C., . . . Raviglione, M. (2016). Tuberculosis. *Nat Rev Dis Primers, 2*, 16076. doi:10.1038/nrdp.2016.76

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528. doi:10.1093/bioinformatics/bty633

Parida, S. K., Axelsson-Robertson, R., Rao, M. V., Singh, N., Master, I., Lutckii, A., . . . Maeurer, M. (2015). Totally drug-resistant tuberculosis and adjunct therapies. *Journal of Internal Medicine, 277*(4), 388-405. doi:10.1111/joim.12264

Parvaresh, L., Crighton, T., Martinez, E., Bustamante, A., Chen, S., & Sintchenko, V. (2018). Recurrence of tuberculosis in a low-incidence setting: a retrospective cross-sectional study augmented by whole genome sequencing. *BMC Infectious Diseases, 18*(1), 265. doi:10.1186/s12879-018-3164-z

Paulson, T. (2013). Epidemiology: A mortal foe. *Nature, 502*(7470), S2-3. doi:10.1038/502S2a

Pham, T. H., Peter, J., Mello, F. C. Q., Parraga, T., Lan, N. T. N., Nabeta, P., . . . Perkins, M. D. (2018). Performance of the TB-LAMP diagnostic assay in reference laboratories: Results from a multicentre study. *International Journal of Infectious Diseases, 68*, 44-49. doi:10.1016/j.ijid.2018.01.005

Philips, J. A., & Ernst, J. D. (2012). Tuberculosis pathogenesis and immunity. *Annual Review of Pathology, 7*, 353-384. doi:10.1146/annurev-pathol-011811-132458

Purohit, M., & Mustafa, T. (2015). Laboratory Diagnosis of Extra-pulmonary Tuberculosis (EPTB) in Resource-constrained Setting: State of the Art, Challenges and the Need. *Journal of clinical and diagnostic research : JCDR, 9*(4), EE01-EE06. doi:10.7860/JCDR/2015/12422.5792

University of Ghana http://ugspace.ug.edu.gh

R, C. T. (2019). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.

Ramakrishnan, G., Ochoa-Montano, B., Raghavender, U. S., Mudgal, R., Joshi, A. G., Chandra, N. R., . . . Srinivasan, N. (2015). Enriching the annotation of *Mycobacterium tuberculosis* H37Rv proteome using remote homology detection approaches: insights into structure and function. *Tuberculosis (Edinb), 95*(1), 14-25. doi:10.1016/j.tube.2014.10.009

Randremanana, R. V., Sabatier, P., Rakotomanana, F., Randriamanantena, A., & Richard, V. (2009). Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. *Trop Med Int Health, 14*(4), 429-437. doi:10.1111/j.1365-3156.2009.02239.x

Ravansalar, H., Tadayon, K., & Ghazvini, K. (2016). Molecular typing methods used in studies of *Mycobacterium tuberculosis* in Iran: a systematic review. *Iran J Microbiol, 8*(5), 338-346.

Rebuma, F., Stefan, B., Elena, H., Esther, S., Balako, G., Girume, E., . . . Abraham, A. (2013). Mycobacterial Lineages Causing Pulmonary and Extrapulmonary Tuberculosis, Ethiopia. *Emerging Infectious Disease journal, 19*(3), 460. doi:10.3201/eid1903.120256

Roring, S., Scott, A. N., Glyn Hewinson, R., Neill, S. D., & Skuce, R. A. (2004). Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. *Veterinary Microbiology, 101*(1), 65-73. doi:10.1016/j.vetmic.2004.02.013

Sakamoto, K. (2012). The pathology of *Mycobacterium tuberculosis* infection. *Veterinary Pathology, 49*(3), 423-439. doi:10.1177/0300985811429313

Senghore, M., Otu, J., Witney, A., Gehre, F., Doughty, E. L., Kay, G. L., . . . Antonio, M. (2017). Whole-genome sequencing illuminates the evolution and spread of multidrug-resistant tuberculosis in Southwest Nigeria. *PloS One, 12*(9), e0184510. doi:10.1371/journal.pone.0184510

Seraphin, M. N., Lauzardo, M., Doggett, R. T., Zabala, J., Morris, J. G., Jr., & Blackburn, J. K. (2016). Spatiotemporal Clustering of *Mycobacterium tuberculosis* Complex Genotypes in Florida: Genetic Diversity Segregated by Country of Birth. *PloS One, 11*(4), e0153575. doi:10.1371/journal.pone.0153575

University of Ghana http://ugspace.ug.edu.gh

Seto, J., Wada, T., Suzuki, Y., Ikeda, T., Mizuta, K., Yamamoto, T., & Ahiko, T. (2017). *Mycobacterium tuberculosis* Transmission among Elderly Persons, Yamagata Prefecture, Japan, 2009-2015. *Emerging Infectious Diseases, 23*(3), 448-455. doi:10.3201/eid2303.161571

Shabbeer, A., Cowan, L. S., Ozcaglar, C., Rastogi, N., Vandenberg, S. L., Yener, B., & Bennett, K. P. (2012). TB-Lineage: an online tool for classification and analysis of strains of *Mycobacterium tuberculosis* complex. *Infection, Genetics and Evolution, 12*(4), 789-797. doi:10.1016/j.meegid.2012.02.010

Shen, G. M., Xue, Z., Shen, X., Gui, X. H., Wang, Y. F., Sun, B., . . . Gao, Q. (2006). Study on the cause of tuberculosis recurrence by *Mycobacterium tuberculosis* genotyping method. *Zhonghua Jie He He Hu Xi Za Zhi. Chinese Journal of Tuberculosis and Respiratory Diseases, 29*(2), 79-82.

Shen, X., Yang, C., Wu, J., Lin, S., Gao, X., Wu, Z., . . . Gao, Q. (2017). Recurrent tuberculosis in an urban area in China: Relapse or exogenous reinfection? *Tuberculosis (Edinb), 103*, 97-104. doi:10.1016/j.tube.2017.01.007

Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1980). *The Methods and Materials of Demography*: Department of Commerce, Bureau of the Census.

Sia, I. G., Buckwalter, S. P., Doerr, K. A., Lugos, S., Kramer, R., Orillaza-Chi, R., . . . Wengenack, N. L. (2013). Genotypic characteristics of *Mycobacterium tuberculosis* isolated from household contacts of tuberculosis patients in the Philippines. *BMC Infectious Diseases, 13*, 571. doi:10.1186/1471-2334-13-571

Singhal, R., & Myneedu, V. P. (2015). Microscopy as a diagnostic tool in pulmonary tuberculosis. *International Journal of Mycobacteriology, 4*(1), 1-6. doi:https://doi.org/10.1016/j.ijmyco.2014.12.006

Small, P. M., Hopewell, P. C., Singh, S. P., Paz, A., Parsonnet, J., Ruston, D. C., . . . Schoolnik, G. K. (1994). The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *New England Journal of Medicine, 330*(24), 1703-1709. doi:10.1056/nejm199406163302402

Smith, I. (2003). *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. *Clinical Microbiology Reviews, 16*(3), 463-496. doi:10.1128/CMR.16.3.463-496.2003

University of Ghana http://ugspace.ug.edu.gh

Soolingen, D., Borgdorff, M. W., Haas, P. E., Sebek, M. M., Veen, J., Dessens, M., . . . Embden, J. D. (1999). Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *Journal of Infectious Diseases, 180*. doi:10.1086/314930

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England), 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stavrum, R., Mphahlele, M., Ovreas, K., Muthivhi, T., Fourie, P. B., Weyer, K., & Grewal, H. M. (2009). High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *Journal of Clinical Microbiology, 47*(6), 1848-1856. doi:10.1128/JCM.02167-08

Stimson, J., Gardy, J., Mathema, B., Crudu, V., Cohen, T., & Colijn, C. (2019). Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. *Mol Biol Evol, 36*(3), 587-603. doi:10.1093/molbev/msy242

Streit, E., Baboolal, S., Akpaka, P. E., Millet, J., & Rastogi, N. (2015). Finer characterization of *Mycobacterium tuberculosis* using spoligotyping and 15-loci MIRU-VNTRs reveals phylogeographical specificities of isolates circulating in Guyana and Suriname. *Infection, Genetics and Evolution, 30*, 114-119. doi:10.1016/j.meegid.2014.12.015

Stucki, D., Ballif, M., Bodmer, T., Coscolla, M., Maurer, A. M., Droz, S., . . . Fenner, L. (2015). Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. *Journal of Infectious Diseases, 211*(8), 1306-1316. doi:10.1093/infdis/jiu601

Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., . . . Gagneux, S. (2016). *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature Genetics, 48*(12), 1535-1543. doi:10.1038/ng.3704

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., . . . van Soolingen, D. (2006). Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology, 44*(12), 4498-4510. doi:10.1128/JCM.01392-06

University of Ghana http://ugspace.ug.edu.gh

Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D., & Locht, C. (2001). Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *Journal of Clinical Microbiology, 39*(10), 3563-3571. doi:10.1128/JCM.39.10.3563-3571.2001

Supply, P., Magdalena, J., Himpens, S., & Locht, C. (1997). Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Molecular Microbiology, 26*(5), 991-1003.

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.

Surie, D., Fane, O., Finlay, A., Ogopotse, M., Tobias, J. L., Click, E. S., . . . Kopanyo Study, G. (2017). Molecular, Spatial, and Field Epidemiology Suggesting TB Transmission in Community, Not Hospital, Gaborone, Botswana. *Emerging Infectious Diseases, 23*(3), 487-490. doi:10.3201/eid2303.161183

Surikova, O. V., Voitech, D. S., Kuzmicheva, G., Tatkov, S. I., Mokrousov, I. V., Narvskaya, O. V., . . . Filipenko, M. L. (2005). Efficient differentiation of *Mycobacterium tuberculosis* strains of the W-Beijing family from Russia using highly polymorphic VNTR loci. *European Journal of Epidemiology, 20*(11), 963-974. doi:10.1007/s10654-005-3636-5

Taylor, Z., Nolan, C. M., & Blumberg, H. M. (2005). Controlling tuberculosis in the United States. Recommendations from the American Thoracic Society, CDC, and the Infectious Diseases Society of America. *MMWR: Recommendations and Reports, 54*(Rr-12), 1-81.

Thierry, D., Cave, M. D., Eisenach, K. D., Crawford, J. T., Bates, J. H., Gicquel, B., & Guesdon, J. L. (1990). IS6110, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Research, 18*(1), 188.

Thomas, A., Gopi, P. G., Santha, T., Chandrasekaran, V., Subramani, R., Selvakumar, N., . . . Narayanan, P. R. (2005). Predictors of relapse among pulmonary tuberculosis patients treated in a DOTS programme in South India. *Int J Tuberc Lung Dis, 9*(5), 556-561.

University of Ghana http://ugspace.ug.edu.gh

Togun, T. O., Egere, U., Sillah, A. K., Ayorinde, A., Mendy, F., Tientcheu, L., . . . Kampmann, B. (2015). Contribution of Xpert(R) MTB/RIF to the diagnosis of pulmonary tuberculosis among TB-exposed children in The Gambia. *Int J Tuberc Lung Dis, 19*(9), 1091-1097, i-ii. doi:10.5588/ijtld.15.0228

Touray, K., Adetifa, I. M., Jallow, A., Rigby, J., Jeffries, D., Cheung, Y. B., . . . Hill, P. C. (2010). Spatial analysis of tuberculosis in an urban west African setting: is there evidence of clustering? *Trop Med Int Health, 15*(6), 664-672. doi:10.1111/j.1365-3156.2010.02533.x

UNICEF. (2014). *Generation 2030 Africa: child demographics in Africa. Division of Data, Research Policy.* In. Retrieved from http://www.unicef.org/media/files/UNICEF_Africa_Generation_2030_en_11Aug.pdf

Unis, G., Ribeiro, A. W., Esteves, L. S., Spies, F. S., Picon, P. D., Dalla Costa, E. R., & Rossetti, M. L. (2014). Tuberculosis recurrence in a high incidence setting for HIV and tuberculosis in Brazil. *BMC Infectious Diseases, 14*, 548. doi:10.1186/s12879-014-0548-6

van Crevel, R., Ottenhoff, T. H., & van der Meer, J. W. (2002). Innate immunity to *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews, 15*(2), 294-309.

van Ingen, J., Rahim, Z., Mulder, A., Boeree, M. J., Simeone, R., Brosch, R., & van Soolingen, D. (2012). Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. *Emerging Infectious Diseases, 18*(4), 653-655. doi:10.3201/eid1804.110888

van Rie, A., Warren, R., Richardson, M., Victor, T. C., Gie, R. P., Enarson, D. A., . . . van Helden, P. D. (1999). Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment. *New England Journal of Medicine, 341*(16), 1174-1179. doi:10.1056/nejm199910143411602

van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R., & van Embden, J. D. (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.

University of Ghana http://ugspace.ug.edu.gh

van Soolingen, D., Hoogenboezem, T., de Haas, P. E., Hermans, P. W., Koedam, M. A., Teppema, K. S., . . . van Embden, J. D. (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. doi:10.1099/00207713-47-4-1236

van Soolingen, D., Kremer, K., & Vynycky, E. (2003). New perspectives in the molecular epidemiology of tuberculosis. In *Mycobacteria and TB* (Vol. 2, pp. 17-45): Karger Publishers.

Varghese, B., al-Omari, R., Grimshaw, C., & Al-Hajoj, S. (2013). Endogenous reactivation followed by exogenous re-infection with drug resistant strains, a new challenge for tuberculosis control in Saudi Arabia. *Tuberculosis (Edinb), 93*(2), 246-249. doi:10.1016/j.tube.2012.12.001

Velayati, A. A., Farnia, P., & Farahbod, A. M. (2016). Overview of drug-resistant tuberculosis worldwide. *Int J Mycobacteriol, 5 Suppl 1*, S161. doi:10.1016/j.ijmyco.2016.09.066

Velayati, A. A., Farnia, P., & Masjedi, M. R. (2013). The totally drug resistant tuberculosis (TDR-TB). *International Journal of Clinical and Experimental Medicine, 6*(4), 307-309.

Velayutham, B., Chadha, V. K., Singla, N., Narang, P., Gangadhar Rao, V., Nair, S., . . . Natrajan, M. (2018). Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: A multi-centric prospective study. *PloS One, 13*(7), e0200150. doi:10.1371/journal.pone.0200150

Verma, A., Sampla, A. K., & Tyagi, J. S. (1999). *Mycobacterium tuberculosis* rrn promoters: differential usage and growth rate-dependent control. *Journal of Bacteriology, 181*(14), 4326-4333.

Verver, S., Warren, R. M., Beyers, N., Richardson, M., van der Spuy, G. D., Borgdorff, M. W., . . . van Helden, P. D. (2005). Rate of reinfection tuberculosis after successful treatment is higher than rate of new tuberculosis. *Am J Respir Crit Care Med, 171*(12), 1430-1435. doi:10.1164/rccm.200409-1200OC

Vluggen, C., Soetaert, K., Groenen, G., Wanlin, M., Spitaels, M., Arrazola de Onate, W., . . . Mathys, V. (2017). Molecular epidemiology of *Mycobacterium tuberculosis* complex in Brussels, 2010-2013. *PloS One, 12*(2), e0172554. doi:10.1371/journal.pone.0172554

University of Ghana http://ugspace.ug.edu.gh

Walker, T. M., Ip, C. L., Harrell, R. H., Evans, J. T., Kapatai, G., Dedicoat, M. J., . . . Peto, T. E. (2013). Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infectious Diseases, 13*(2), 137-146. doi:10.1016/s1473-3099(12)70277-3

Walker, T. M., Lalor, M. K., Broda, A., Saldana Ortega, L., Morgan, M., Parker, L., . . . Conlon, C. P. (2014). Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med, 2*(4), 285-292. doi:10.1016/s2213-2600(14)70027-x

Walker, T. M., Merker, M., Knoblauch, A. M., Helbling, P., Schoch, O. D., van der Werf, M. J., . . . Keller, P. M. (2018). A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infectious Diseases*. doi:10.1016/s1473-3099(18)30004-5

Walker, T. M., Merker, M., Kohl, T. A., Crook, D. W., Niemann, S., & Peto, T. E. (2017). Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin Microbiol Infect, 23*(3), 161-166. doi:10.1016/j.cmi.2016.10.014

Walker, T. M., Monk, P., Smith, E. G., & Peto, T. E. A. (2013). Contact investigations for outbreaks of *Mycobacterium tuberculosis* advances through whole genome sequencing, Clinical Microbiology and Infection Volume 19, Issue 9. *Clinical Microbiology and Infection, 19*(9), 796-802.

Warren, R. M., van der Spuy, G. D., Richardson, M., Beyers, N., Borgdorff, M. W., Behr, M. A., & van Helden, P. D. (2002). Calculation of the stability of the IS6110 banding pattern in patients with persistent *Mycobacterium tuberculosis* disease. *Journal of Clinical Microbiology, 40*(5), 1705-1708.

Wells, A. Q., & Oxon, D. M. (1937). Tuberculosis in wild voles. *Lancet i*, 1221.

Weniger, T., Krawczyk, J., Supply, P., Harmsen, D., & Niemann, S. (2012). Online tools for polyphasic analysis of *Mycobacterium tuberculosis* complex genotyping data: now and next. *Infection, Genetics and Evolution, 12*(4), 748-754. doi:10.1016/j.meegid.2012.01.021

WHO. (2014). Global tuberculosis report. Retrieved 2015/06/06, from World Health Organization http://www.who.int/tb/publications/global_report/en/

University of Ghana http://ugspace.ug.edu.gh

WHO. (2015a). Gear up to end TB: introducing the end TB strategy. Available from Google Scholar Retrieved 2019/02/02, from World Health Organization https://www.who.int/tb/End_TB_brochure.pdf

WHO. (2015b). Global tuberculosis report. Available from Open WorldCat. (978-92-4-156505-9). Retrieved 14/09/2016, from World Health Organization http://www.who.int/tb/publications/global_report/en/

WHO. (2015c). Guidelines on the management of latent tuberculosis infection. Available from Open WorldCat Retrieved 2018/12/28, from World Health Organization https://www.who.int/tb/publications/ltbi_document_page/en/

WHO. (2015d). Tuberculosis country profile_Ghana. World Health Organization Retrieved 14/09/2016 http://www.who.int/tb/country/data/profiles/en/

WHO. (2017). Global tuberculosis report. Available from Open WorldCat Retrieved 2018/09/26, from World Health Organization http://www.who.int/tb/publications/global_report/en/

WHO. (2018a). BCG vaccines: WHO position paper. *Weekly Epidemiological Record, 93*(8), 73-96.

WHO. (2018b). Global tuberculosis report. Available from Open WorldCat Retrieved 2018/09/26, from World Health Organization http://www.who.int/tb/publications/global_report/en/

WHO. (2018c). Tuberculosis country profiles. Retrieved 2018/09/26, from World Health Organization http://www.who.int/tb/country/data/profiles/en/

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag New York.

Winglee, K., Manson McGuire, A., Maiga, M., Abeel, T., Shea, T., Desjardins, C. A., . . . Bishai, W. R. (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. doi:10.1371/journal.pntd.0004332

Wolinsky, E. (1979). Nontuberculous mycobacteria and associated diseases. *American Review of Respiratory Disease, 119*(1), 107-159. doi:10.1164/arrd.1979.119.1.107

University of Ghana http://ugspace.ug.edu.gh

Wood, R., Liang, H., Wu, H., Middelkoop, K., Oni, T., Rangaka, M. X., . . . Lawn, S. D. (2010). Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. *Int J Tuberc Lung Dis, 14*(4), 406-412.

Yakam, N. A., Noeske, J., Dambach, P., Bowong, S., Fono, L. A., & Ngatchou-Wandji, J. (2014). Spatial analysis of tuberculosis in Douala, Cameroon: clustering and links with socio-economic status. *Int J Tuberc Lung Dis, 18*(3), 292-297. doi:10.5588/ijtld.13.0573

Yang, C., Luo, T., Shen, X., Wu, J., Gan, M., Xu, P., . . . Gao, Q. (2017). Transmission of multidrug-resistant *Mycobacterium tuberculosis* in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *The Lancet Infectious Diseases, 17*(3), 275-284. doi:10.1016/s1473-3099(16)30418-2

Yates, T. A., Khan, P. Y., Knight, G. M., Taylor, J. G., McHugh, T. D., Lipman, M., . . . Abubakar, I. (2016). The transmission of *Mycobacterium tuberculosis* in high burden settings. *Lancet Infectious Diseases, 16*(2), 227-238. doi:10.1016/S1473-3099(15)00499-5

Yeboah-Manu, D., Asante-Poku, A., Ampah, K. A., Kpeli, G., Danso, E., Owusu-Darko, K., & Bonsu, F. A. (2012). Drug Susceptibility Pattern of *Mycobacterium tuberculosis* Isolates From Ghana; Correlation with Clinical Response. *Mycobacterial Diseases, 2*(3), 1-5. doi:10.4172/2161-1068.1000107

Yeboah-Manu, D., Asante-Poku, A., Bodmer, T., Stucki, D., Koram, K., Bonsu, F., . . . Gagneux, S. (2011). Genotypic diversity and drug susceptibility patterns among *M. tuberculosis* complex isolates from South-Western Ghana. *PloS One, 6*(7), e21906. doi:10.1371/journal.pone.0021906

Yeboah-Manu, D., Asare, P., Asante-Poku, A., Otchere, I. D., Osei-Wusu, S., Danso, E., . . . Gagneux, S. (2016). Spatio-Temporal Distribution of *Mycobacterium tuberculosis* Complex Strains in Ghana. *PloS One, 11*(8), e0161892. doi:10.1371/journal.pone.0161892

Yeboah-Manu, D., Bodmer, T., Mensah-Quainoo, E., Owusu, S., Ofori-Adjei, D., & Pluschke, G. (2004). Evaluation of decontamination methods and growth media for primary isolation of *Mycobacterium ulcerans* from surgical specimens. *Journal of Clinical Microbiology, 42*(12), 5875-5876. doi:10.1128/JCM.42.12.5875-5876.2004

University of Ghana http://ugspace.ug.edu.gh

Yeboah-Manu, D., Yates, M. D., & Wilson, S. M. (2001). Application of a simple multiplex PCR to aid in routine work of the mycobacterium reference laboratory. *Journal of Clinical Microbiology, 39*(11), 4166-4168. doi:10.1128/JCM.39.11.4166-4168.2001

Yu, G., Lam, T. T.-Y., Zhu, H., & Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Molecular Biology and Evolution, 35*(12), 3041-3043. doi:10.1093/molbev/msy194

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution, 8*(1), 28-36. doi:10.1111/2041-210x.12628

Zamani, S., Haeili, M., Nasiri, M. J., Imani Fooladi, A. A., Javadpour, S., & Feizabadi, M. M. (2016). Genotyping of *Mycobacterium tuberculosis* Isolates from Hormozgan Province of Iran Based on 15-Locus MIRU-VNTR and Spoligotyping. *Int J Bacteriol, 2016*, 7146470. doi:10.1155/2016/7146470

Zelner, J. L., Murray, M. B., Becerra, M. C., Galea, J., Lecca, L., Calderon, R., . . . Cohen, T. (2016). Identifying Hotspots of Multidrug-Resistant Tuberculosis Transmission Using Spatial and Molecular Genetic Data. *Journal of Infectious Diseases, 213*(2), 287-294. doi:10.1093/infdis/jiv387

Zhao, F., Cheng, S., He, G., Huang, F., Zhang, H., Xu, B., . . . Wang, L. (2013). Space-Time Clustering Characteristics of Tuberculosis in China, 2005-2011. *PloS One, 8*(12), e83605. doi:10.1371/journal.pone.0083605

Zheng, C., Hu, M., & Gao, F. (2017). Diabetes and pulmonary tuberculosis: a global overview with special focus on the situation in Asian countries with high TB-DM burden. *Glob Health Action, 10*(1), 1-11. doi:10.1080/16549716.2016.1264702

Zong, Z., Huo, F., Shi, J., Jing, W., Ma, Y., Liang, Q., . . . Pang, Y. (2018). Relapse Versus Reinfection of Recurrent Tuberculosis Patients in a National Tuberculosis Specialized Hospital in Beijing, China. *Frontiers in Microbiology, 9*, 1858. doi:10.3389/fmicb.2018.01858

Zumla, A., Abubakar, I., Raviglione, M., Hoelscher, M., Ditiu, L., McHugh, T. D., . . . Atun, R. (2012). Drug-resistant tuberculosis--current dilemmas, unanswered questions, challenges, and priority needs. *Journal of Infectious Diseases, 205 Suppl 2*, S228-240. doi:10.1093/infdis/jir858

University of Ghana http://ugspace.ug.edu.gh

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), e1001012-e1001012. doi:10.1371/journal.pmed.1001012

University of Ghana http://ugspace.ug.edu.gh

APPENDICES

APPENDIX 1

Supplementary data to chapter 3

Table S3.1. Districts and population statistics within designated time points within the study period

District/sub-district	Abbreviation	Region	2010	2012†	2013†	2014†
Accra Metropolis	AMA	Greater Accra	1665086	1771589	1827368	1884904
Ablekuma*	Able	Greater Accra (AMA)	679362	722816	745574	769049
Ashiedu Keteke*	AshK	Greater Accra (AMA)	117525	125042	128979	133040
Ayawaso*	Ayaw	Greater Accra (AMA)	396487	421847	435129	448830
Okaikoi*	Okai	Greater Accra (AMA)	349989	372375	384100	396193
Osu Klottey*	OsuK	Greater Accra (AMA)	121723	129509	133586	137792
Kpeshie	Kpes	Greater Accra	411460	437778	451562	465779
Adenta Municipal	AdeM	Greater Accra	78215	83218	85838	88541
La-Nkwantanang Madina Municipal	LaNM	Greater Accra	111926	119085	122835	126702
Ga Central Municipal	GaCM	Greater Accra	117220	124718	128644	132695
Ga East Municipal	GaEM	Greater Accra	147742	157192	162141	167246
Ga South Municipal	GaSM	Greater Accra	411377	437690	451471	465685
Ga West Municipal	GaWM	Greater Accra	219788	233846	241209	248804
Mamprusi East	MamE	Northern	121009	128235	132008	135893
Tamale Metropolis	TamM	Northern	371351	393526	405106	417026
Gomoa East	GomE	Central	207071	220316	227252	234408
Agona West Municipal	AgWM	Central	115358	122737	126601	130587
Ewutu Senya	AwuS	Central	195306	207798	214341	221089

The last population census conducted in Ghana was in 2010. As a results, to obtain the population statistics for the period 2012 to 2014 (columns 5 to 7) we used the exponential growth rate formulae as described in methods. The intercensal growth rates used per region were: Greater Accra (3.1%), Central region (3.1%), Northern region (2.9%).

**sub-districts within AMA; †Projected population from 2010 population census data.*

University of Ghana http://ugspace.ug.edu.gh

Table S3.2. Annual distribution of TB cases used for spatial or space-time analysis (2007 – 2014)

District/sub-district	2007	2008	2009	2010	2011	2012	2013	2014	Total per district
Accra Metropolis	N/A	N/A	N/A	119	141	N/A	N/A	N/A	260
Ablekuma*	N/A	N/A	N/A	7	2	72	174	153	408
Ashiedu Keteke*	N/A	N/A	N/A	5	N/A	18	53	31	107
Ayawaso*	N/A	N/A	N/A	N/A	N/A	32	90	71	193
Okaikoi*	N/A	N/A	N/A	N/A	N/A	30	62	46	138
Osu Klottey*	N/A	N/A	N/A	1	N/A	17	70	24	112
Kpeshie*	N/A	N/A	N/A	38	25	34	86	51	234
Adenta Municipal	N/A	N/A	N/A	N/A	N/A	4	3	3	10
La-Nkwantanang Madina	N/A	N/A	N/A	N/A	N/A	2	12	6	20
Ga Central Municipal	N/A	N/A	N/A	N/A	N/A	4	11	16	31
Ga East Municipal	N/A	N/A	N/A	N/A	N/A	15	23	11	49
Ga South Municipal	N/A	N/A	N/A	N/A	N/A	4	17	18	39
Ga West Municipal	N/A	N/A	N/A	1	N/A	6	19	11	37
Mamprusi East	N/A	N/A	N/A	N/A	N/A	7	42	44	93
Tamale Metropolis	N/A	N/A	N/A	N/A	N/A	N/A	8	37	45
Agona West Municipal	2	57	31	22	6	N/A	N/A	N/A	118
Ewutu Senya	8	59	18	17	N/A	7	35	21	165
Gomoa East	1	13	6	3	N/A	N/A	N/A	N/A	23
Total per year	11	129	55	213	174	252	705	543	2082

The table lists the number of tuberculosis cases sampled within each district/sub-district for only participants with residential status. Periods within the 8 year study period where no sampling was done are marked N/A (not available). The final column and row contains total counts for each district/sub-district and year respectively.

**sub-districts within AMA.*

University of Ghana http://ugspace.ug.edu.gh

Table S3.3. Lineage distribution of isolated MTBC recruited and used for spatial or space-time analysis (2007 – 2014)

	TB Species/Lineages/Sub-lineages								
Districts/ sub-districts	**MTBss**	**Maf**	**L4**	**L5**	**L6**	**Gh**	**Cam**	**Animal**	**Total**
Accra Metropolis (2010/2011)	215	45	207	35	10	0	155	-	260
Ablekuma*	315	90	292	59	31	72	168	3	408
Ashiedu Keteke*	81	26	79	12	14	20	46	-	107
Ayawaso*	149	43	143	26	17	29	92	1	193
Okaikoi*	109	28	105	16	12	25	59	1	138
Osu Klottey*	87	24	72	15	9	10	46	1	112
Kpeshie*	182	51	168	43	8	27	101	1	234
Adenta Municipal	6	4	5	3	1	2	1	-	10
La-Nkwantanang Madina	17	3	17	2	1	3	13	-	20
Ga Central Municipal	27	4	27	2	2	3	18	-	31
Ga East Municipal	38	11	36	6	5	7	24	-	49
Ga South Municipal	31	8	29	6	2	8	17	-	39
Ga West Municipal	31	5	25	3	2	7	15	1	37
Agona West Municipal	105	13	102	10	3	0	63	-	118
Ewutu Senya	127	36	118	26	10	7	77	2	165
Gomoa East	18	5	15	3	2	0	9	-	23
Mamprusi East	79	13	69	7	6	28	27	-	92
Tamale Metropolis	34	10	32	4	6	9	18	1	45
Total	1651	420	1541	279	141	257	949	11	2082

The table lists the distribution of tuberculosis species/major lineages/sub-lineages sampled within each district/sub-district for only participants with well-defined residential status. The final column and row contains total counts for each district/sub-district and TB species/lineage/sub-lineage respectively.

Abbreviations: TB, tuberculosis; MTBss, *Mycobacterium tuberculosis* sensu stricto; MAF, *Mycobacterium africanum*; L4, Lineage 4; L5, Lineage 5; L6, Lineage 6; Gh, Ghana genotypes (Ghana sub-lineage); Cam, Cameroon sub-lineage.

**sub-districts within Accra metropolis.*

University of Ghana http://ugspace.ug.edu.gh

Table S3.4. Default parameters used in SaTScan for clustering analysis

Category	Parameters	Settings
Input	Time precision	Year
	Coordinates	Latitude/Longitude
Data checking	Temporal data check	Check to ensure that all cases and controls are within the specified temporal study period
	Geographical data check	Check to ensure that all observations (cases, controls and populations) are within the specified geographical area
Analysis	Time aggregation units	Year
	Time aggregation length	1
Spatial window	Maximum spatial cluster size	50% of population at risk
Temporal window	Minimum temporal cluster size	1 year
	Maximum temporal cluster size	50% of study period
Inference	Number of replications	999

The table contains a list of the default settings used in performing the clustering analysis using the SaTScan software.

University of Ghana http://ugspace.ug.edu.gh

Table S3.5. Primer Sequence for large sequence polymorphism LSP Assays

Region of Difference	Sequence (5'-3')
RD4-Flanking Forward	CTC GTC GAA GGC CAC TAA AG
RD4-Flanking Reverse	AAG GCG AAC AGA TTC AGC AT
RD9-Flanking Forward	ACT CCC AGC GCT CGG CGG TGA CGG TAT CGT
RD9-Flanking Reverse	ATT CCG TGG GCG CTG CGG CCA ATG TTT GTT
RD12-Flanking Forward	GCC ATC AAC GTC AAG AAC CT
RD12-Flanking Reverse	CGG CCA GGT AAC AAG GAG T
RD702-Flanking Forward	TTC CGA GGA CCC GTT GTT GAG TGC
RD702-Flanking Reverse	GGG CGG GTT GGG TTG CTG GTC
RD711-Flanking Forward	GGC CGC CCT GCT CAA GAA CCT
RD711-Flanking Reverse	CCT AGG CCG GCG ACG AAG TG

The table contains a list of primer sequences used for the LSP assay

University of Ghana http://ugspace.ug.edu.gh

Table S3.6. Comparison of selected risk factors among the two regions

Variables	North	South	p-value	OR	95% CI
Settlement (1938)	***212***	***1726***			
Rural (152)	151 (71.2%)	1 (0.1%)	0.0000	4154.6	$6.9e^{+2}$-$4.5e^{+15}$
Urban (1786)	61(28.8%)	1725 (99.9%)	0.0000	0.0	0.0-$1.4e^{-3}$
Occupation (1468)	***168***	***1300***			
Farmer (108)	81 (48.2%)	27 (2.1%)	0.0000	43.6	26.4-74.2
Driver (156)	9 (5.4%)	147 (11.3%)	0.0163	0.4	0.2-0.9
Constant contact with Cattle (7)	5 (3.0%)	2 (0.2%)	0.0003	19.8	3.2-209.3
Housing (1459)	***169***	***1290***			
Compound (1207)	160 (96.7%)	1047 (81.2%)	0.0000	4.1	2.1-9.3
Self-Contained (252)	9 (3.3%)	243 (18.8%)	0.0000	0.2	0.1-0.5
HIV (732)	***50***	***682***			
Positive (102)	3 (6.0%)	99 (14.5%)	0.1350	0.4	0.1-1.2
Negative (630)	47 (94.0%)	583 (85.5%)	0.1350	2.7	0.8-13.6

*there were 5 butchers (Abattoir workers) and 2 farmers who own cattle

University of Ghana http://ugspace.ug.edu.gh

Table S3.7. Genotyping profile of 2551 MTBC isolates from Ghana

Supplementary table 7: Genotyping profile of 2551 MTBC isolates from Ghana

Specie	SNP	Spoligotyping profile	Sublineage	SIT	No	%
MTBss	L1		EAI	236	6	0.20
MTBss	L1		EAI	340	11	0.40
MTBss	L1		EAI	342	15	0.60
MTBss	L1		EAI	380	2	0.07
MTBss	L1		EAI	Orphan	1	0.04
MTBss	L1		EAI	Orphan	1	0.04
MTBss	L2		Beijing	1	71	3.00
MTBss	L2		Beijing	1184	3	0.12
MTBss	L2		Beijing	941	2	0.07
MTBss	L2		Beijing		1	0.04
MTBss	L3		Dehli_CAS	25	4	0.20
MTBss	L3		Dehli_CAS	129	2	0.07
MTBss	L3		Dehli_CAS	702	3	0.12
MTBss	L3		Dehli_CAS	1199	4	0.20
MTBss	L3		Dehli_CAS		2	0.07
MTBss	L3		Dehli_CAS		3	0.12
MTBss	L3		Dehli_CAS		2	0.07
MTBss	L3		Dehli_CAS		3	0.12
MTBss	L4		Cameroon	57	13	0.50
MTBss	L4		Cameroon	61	811	32.00
MTBss	L4		Cameroon	114	1	0.04
MTBss	L4		Cameroon	403	5	0.20
MTBss	L4		Cameroon	772	92	3.60
MTBss	L4		Cameroon	838	28	1.10
MTBss	L4		Cameroon	1141	14	0.50
MTBss	L4		Cameroon	1580	15	0.60
MTBss	L4		Cameroon		25	1.00
MTBss	L4		Cameroon		35	1.40
MTBss	L4		Cameroon		10	0.40
MTBss	L4		Cameroon		30	1.20
MTBss	L4		Cameroon		20	1.00
MTBss	L4		Cameroon		25	1.00
MTBss	L4		Cameroon		9	0.40
MTBss	L4		Cameroon		18	0.70
MTBss	L4		Ghana	37	2	0.07
MTBss	L4		Ghana	44	2	0.07
MTBss	L4		Ghana	53	269	11.0
MTBss	L4		Ghana	54	3	0.12
MTBss	L4		Ghana	58	1	0.04
MTBss	L4		Ghana	278	1	0.04
MTBss	L4		Ghana	373	1	0.04
MTBss	L4		Ghana	462	3	0.12
MTBss	L4		Ghana	504	28	1.10
MTBss	L4		Ghana	804	1	0.04
MTBss	L4		Ghana	926	1	0.04
MTBss	L4		Ghana	1105	1	0.04
MTBss	L4		Ghana	1196	3	0.12
MTBss	L4		Ghana	1227	3	0.12
MTBss	L4		Ghana	1547	1	0.04
MTBss	L4		Ghana		7	0.30
MTBss	L4		Ghana		3	0.12
MTBss	L4		Ghana	orphan	1	0.04
MTBss	L4		H37Rv		2	0.07
MTBss	L4		Haarlem	36	1	0.04
MTBss	L4		Haarlem	45	3	0.12
MTBss	L4		Haarlem	47	1	0.04
MTBss	L4		Haarlem	50	51	2.00
MTBss	L4		Haarlem	62	1	0.07
MTBss	L4		Haarlem	124	1	0.04
MTBss	L4		Haarlem	144	2	0.04
MTBss	L4		Haarlem	316	8	0.30
MTBss	L4		Haarlem	655	18	0.70
MTBss	L4		Haarlem	775	3	0.12
MTBss	L4		Haarlem	1159	1	0.04
MTBss	L4		Haarlem	1498	16	1.00
MTBss	L4		Haarlem		4	0.20
MTBss	L4		Haarlem		3	0.12
MTBss	L4		Haarlem		3	0.12
MTBss	L4		Haarlem		3	0.12
MTBss	L4		LAM	20	1	0.04
MTBss	L4		LAM	42	26	1.00
MTBss	L4		LAM	535	2	0.07
MTBss	L4		LAM	765	1	0.04
MTBss	L4		LAM		2	0.07
MTBss	L4		LAM	orphan	1	0.04
MTBss	L4		LAM		2	0.07
MTBss	L4		LAM		2	0.07
MTBss	L4		NEW	orphan	1	0.07
MTBss	L4		S		3	0.12

University of Ghana http://ugspace.ug.edu.gh

MTBss	L4		Uganda I	49	2	0.07
MTBss	L4		Uganda I	52	2	0.07
MTBss	L4		Uganda I	78	1	0.07
Specie	**SNP**	**Spoligotyping profile**	**Sublineage**	**SIT**	**No**	**%**
MTBss	L4		Uganda I	848	1	0.04
MTBss	L4		Uganda I	524	1	0.04
MTBss	L4		Uganda I	712	1	0.04
MTBss	L4		Uganda I	1056	1	0.04
MTBss	L4		Uganda I		22	1.00
MTBss	L4		Uganda I		3	0.10
MTBss	L4		Uganda I		4	0.20
MTBss	L4		Uganda II	237	2	0.07
MTBss	L4		Uganda II	92	11	0.40
MTBss	L4		Uganda II	200	5	0.20
MTBss	L4		Uganda II	1178	4	0.20
MTBss	L4		Uganda II		5	0.20
MTBss	L4		Uganda II	orphan	1	0.04
MTBss	L4		Uganda II		6	0.20
MTBss	L4		Uganda II		2	0.07
MTBss	L4		Uganda II		4	0.20
MTBss	L4		Uganda II	orphan	1	0.04
MTBss	L4		Uganda II		3	0.10
MTBss	L4		Uganda II	orphan	1	0.04
Mafric	L5		West Africa I	319	52	2.00
Mafric	L5		West Africa I	320	10	0.40
Mafric	L5		West Africa I	330	11	0.40
Mafric	L5		West Africa I	331	61	2.40
Mafric	L5		West Africa I	438	14	1.00
Mafric	L5		West Africa I	1592	2	0.07
Mafric	L5		West Africa I		50	1.96
Mafric	L5		West Africa I		10	0.40
Mafric	L5		West Africa I		11	0.40
Mafric	L5		West Africa I		11	0.40
Mafric	L5		West Africa I		11	0.40
Mafric	L5		West Africa I		30	1.20
Mafric	L5		West Africa I		1	0.04
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I		1	0.04

Specie	SNP	Spoligotyping profile	Sub lineage	SIT	No	
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		3	0.12
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		4	0.20
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I		1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		5	0.20
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		2	0.07

University of Ghana http://ugspace.ug.edu.gh

Specie	SNP	Spoligotyping profile	Sub lineage	SIT	No	
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		2	0.07
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I		4	0.20
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L5		West Africa I	orphan	1	0.04
Mafric	L6		West Africa II	181	55	2.20
Mafric	L6		West Africa II	326	40	2.00
Mafric	L6		West Africa II	1200	1	0.04
Mafric	L6		West Africa II	1867	1	0.04
Mafric	L6		West Africa II		4	0.20
Mafric	L6		West Africa II		4	0.20
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II		31	1.20
Mafric	L6		West Africa II		3	0.12
Mafric	L6		West Africa II		3	0.12
Mafric	L6		West Africa II	orphan	1	0.10
Mafric	L6		West Africa II		3	0.12
Mafric	L6		West Africa II		3	0.12

Specie	SNP	Spoligotyping profile	Sub lineage	SIT	No	%
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II		8	0.30
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II		7	0.27
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
Mafric	L6		West Africa II	orphan	1	0.04
M. bovis			BOVIS 1_BCG	482	2	0.07
M. bovis			BOVIS 1	1037	5	0.20
M. bovis			BOVIS		3	0.12
M. bovis			BOVIS		2	0.07
M. bovis			BOVIS		1	0.04
M. bovis			BOVIS		2	0.07
M. caprae			caprae		1	0.04

University of Ghana http://ugspace.ug.edu.gh

Table S3.8. Distribution of Species and Lineages of MTBC

Species	Lineages/species	Sub-Lineages	Number of Isolates (%)
MTBSS			2019 (79.1%)
	L1		36 (1.4%)
	L2		77 (3.0%)
	L3		23 (0.9%)
	L4		1883 (74.3%)*
		Cameroon	1151 (61.1%)*
		Ghana	330 (17.5%)*
		Haarlem	119 (6.3%)*
		LAM	39 (2.1%)*
		Uganda I	38 (2.0%)*
		Uganda II	5 (0.3%)*
		New-1	1 (0.1%)*
		S	3 (0.2%)*
		H37Rv-like	2 (0.2%)*
MAF			516 (20.3%)
	L5		338 (13.3%)
	L6		178 (7.0%)
Animal Strains			16 (0.7%)
	M. bovis		15 (0.6%)
	M. caprae		1 (0.0%)

*as proportion of the total number of Lineage 4 strains (1883) and not the total number of MTBC (2551)

University of Ghana http://ugspace.ug.edu.gh

APPENDIX 2

Supplementary data to chapter 4

Table S4.1. Summary of recruited cases

Parameters	**Rural**	**Urban**	**Combined**
Number of pulmonary TB cases recruited	382 (11·6%)	2,921 (88·4%)	3,303
Number of MTBC isolates obtained	254 (9·7%)	2,350 (90·3%)	2,604 (78·8%)*
Number of MTBC isolates passed for clustering analysis	201 (8·7%)	2,108 (91·3%)	2,309 (88·7%)#
Gender	199	2,095	2,294
Male	151 (75·9%)	1,480 (70·6%)	1,631 (71·1%)
Female	48 (24·1%)	615 (29·4%)	663 (28·9%)
Median age (range)	37 (4 to 91)	37 (3 to 90)	37 (3 to 91)
Male	38 (6 to 91)	39 (3 to 89)	39 (3 to 91)
Female	33 (4 to 78)	32 (4 to 90)	33 (4 to 90)

* Frequency expressed as a percentage of the total number of TB cases recruited
# Frequency expressed as the total number of MTBC isolates obtained

University of Ghana http://ugspace.ug.edu.gh

Table S4.2. Distribution of participants' characteristics stratified by the main MTBC lineages found in Ghana: A univariate and multivariate logistic regression analysis

Variables	Distribution of Participants variables				MAF lineage 5 vs MTBss L1-4		MAF lineage 6 vs MTBss L1-4	
	All lineages (N=2,309)	MTBss L1 - 4 (N = 1,870)	MAF L5 (N = 289)	MAF L6 (N = 150)	Univariate	Multivariate	Univariate	Multivariate
	N (%)	N (%)	N (%)	N (%)	OR (95% CI)	adj·OR (95% CI)	OR (95% CI)	adj·OR (95% CI)
Gender	2,294 (99·4)							
Male	1,631 (71·1)	1,348 (72·6)	183 (63·5)	100 (67·6)	0·7 (0·5 - 0·9)*	0·7 (0·5 - 1·0)*	0·8 (0·5 - 1·1)	
Female	663 (28·9)	510 (27·5)	105 (36·5)	48 (32·4)	Reference		Reference	
Age (yrs)	2,224 (96·3)							
<15	37 (1·7)	30 (1·7)	5 (1·8)	2 (1·4)	1·1 (0·4 (3·0)	0·6 (0·2 - 2·3)	1·0 (0·2 - 4·4)	
15 - 29	639 (28·7)	526 (29·3)	78 (27·6)	35 (24·0)	Reference		Reference	
30 - 39	570 (25·6)	465 (25·9)	68 (24·0)	37 (25·3)	1·0 (0·7 - 1·4)	0·8 (0·5 - 1·3)	1·2 (0·7 - 1·9)	
40 - 59	778 (35·0)	626 (34·9)	94 (33·2)	58 (39·7)	1·0 (0·7 - 1·4)	1·1 (0·7 - 1·7)	1·4 (0·9 - 2·1)	
>59	200 (9·0)	148 (8·3)	38 (13·4)	14 (9·6)	1·7 (1·1 - 2·7)*	1·5 (0·7 - 2·9)	1·4 (0·7 - 2·7)	
Nationality	1,781 (77·1)							
Ghanaian	1,714 (96·2)	1,378 (96·0)	227 (97·4)	109 (96·5)	Reference		Reference	
Others	67 (3·8)	57 (4·0)	6 (2·6)	4 (3·5)	0·6 (0·3 - 1·5)		0·9 (0·3 - 2·5)	
Locality	2,309 (100)							
Rural	201 (8·7)	172 (9·2)	15 (5·2)	14 (9·3)	Reference		Reference	
Urban	2,108 (91·3)	1,698 (90·8)	274 (94·8)	136 (90·7)	1·9 (1·1 - 3·2)*		1·0 (0·6 - 1·7)	
Residence classification	1,642 (71·1)							
Village	69 (4·2)	56 (4·2)	4 (1·9)	9 (8·3)	0·4 (0·2 - 1·3)		2·1 (1·0 - 4·4)*	6·6 (1·2 - 36·1)*
Town	182 (11·1)	145 (10·9)	24 (11·6)	13 (11·9)	1·0 (0·6 - 1·7)		1·2 (0·6 - 2·2)	1·6 (0·7 - 3·9)
City residential area	52 (3·2)	39 (2·9)	10 (4·8)	3 (2·8)	1·6 (0·8 - 3·3)		1·0 (0·3 - 3·3)	1·1 (0·2 - 5·2)
City suburb	1,136 (69·2)	920 (69·4)	146 (70·5)	70 (64·2)	Reference		Reference	
City Slum	203 (12·4)	166 (12·5)	23 (11·1)	14 (12·8)	0·9 (0·5 - 1·4)		1·1 (0·6 - 2·0)	0·9 (0·4 - 1·7)
Residential district	1,538 (66·6)							
Ablekuma	545 (35·4)	438 (35·1)	77 (40·7)	30 (29·1)			Reference	
Ashiedu Keteke	170 (11·1)	136 (10·9)	13 (6·9)	21 (20·4)	0·5 (0·3 - 1·0)		2·3 (1·2 - 4·0)*	2·2 (1·1 - 4·4)*
Ayawaso	220 (14·3)	185 (14·8)	21 (11·1)	14 (13·6)	0·6 (0·4 - 1·1)		1·1 (0·6 - 2·1)	0·9 (0·4 - 1·9)
Kpeshie	224(14·6)	179 (14·4)	37 (19·6)	8 (7·8)	1·2 (0·8 - 1·8)		0·6 (0·3 - 1·4)	0·4 (0·1 - 1·2)
Mamprusi East	70 (4·6)	60 (4·8)	4 (2·1)	6 (5·8)	0·4 (0·1 - 1·0)		1·4 (0·6 - 3·6)	0·24 (0·04 - 1·55)
Okaikoi	176 (11·4)	139 (11·2)	24 (12·7)	13 (12·6)	1·0 (0·6 - 1·6)		1·6 (0·7 - 2·7)	1·3 (0·6 - 2·8)
Osu Klottey	133 (8·6)	109 (8·8)	13 (69)	11 (10·7)	0·7 (0·4 - 1·3)		1·5 (0·7 - 3·0)	1·6 (0·7 - 3·5)
Education	1,748 (75·7)							
Primary	222 (12·7)	180 (12·8)	23 (10·1)	19 (17·0)	0·7 (0·4 - 1·1)		1·8 (1·0 - 3·3)*	1·5 (0·7 - 3·1)
Middle/JHS	637 (36·4)	514 (36·5)	93 (41·0)	30 (26·8)	Reference		Reference	
Secondary	429 (24·5)	348 (24·7)	55 (24·2)	26 (23·21)	0·9 (0·6 - 1·3)		1·3 (0·7 - 2·2)	1·2 (0·6 - 2·2)
Tertiary	190 (10·9)	151 (10·7)	27 (11·9)	12 (10·7)	1·0 (0·6 - 1·6)		1·4 (0·7 - 2·7)	1·9 (0·8 - 4·4)
No Education	270 (15·4)	216 (15·3)	29 (12·8)	25 (22·3)	0·7 (0·5 -1·2)		2·0 (1·1 - 3·4)*	1·8 (0·9 - 3·7)
Occupation	1,722 (74·6)							

University of Ghana http://ugspace.ug.edu.gh

Unemployed	390 (22·6)	311 (22·4)	60 (27·0)	19 (16·8)	1·2 (0·9 - 1·7)		0·7 (0·4 - 1·2)	
Unskilled	951 (55·2)	765 (55·2)	118 (53·1)	68 (60·2)	Reference		Reference	
Skilled	381 (22·1)	311 (22·4)	44 (19·8)	23 (23·0)	0·9 (0·6 - 1·3)		0·9 (0·6 - 1·5)	
Religion	1,771 (76·7)							
Christian	1,361 (76·9)	1,081 (75·8)	198 (84·6)	82 (73·9)	Reference		Reference	
Islam	302 (17·0)	254 (17·8)	29 (12·4)	19 (17·1)	0·6 (0·4 - 0·9)*		1·0 (0·6 - 1·6)	
Others	26 (1·5)	20 (1·4)	3 (1·3)	3 (2·7)	0·8 (0·2 - 2·8)		2·0 (0·6 - 6·8)	
Not religious	82 (4·6)	71 (50)	4 (1·7)	7 (6·3)	0·3 (0·1 - 0·8)*		1·3 (0·6 - 2·9)	
Ethnicity	1,760 (76·4)							
Akan	570 (32·3)	458 (32·2)	90 (39·3)	22 (19·6)	Reference		Reference	
Ewe	259 (14·7)	204 (14·3)	35 (15·3)	20 (17·9)	0·9 (0·6 - 1·3)		2·0 (1·1 - 3·8)*	2·3 (1·1 - 5·1)*
Ga/Adangbe	544 (30·8)	437 (30·7)	71 (31·0)	36 (32·1)	0·8 (0·6 - 1·2)		1·7 (1·0 - 3·0)	1·8 (0·9 - 3·6)
Others	392 (22·2)	325 (22·8)	33 (14·4)	34 (30·4)	0·5 (0·3 - 0·8)		2·2 (1·2 - 3·8)*	1·6 (0·7 - 3·4)
Marital Status	1,758 (76·1)							
Single	766 (43·6)	618 (43·6)	101 (44·3)	47 (41·2)	Reference		Reference	
Married	742 (42·2)	599 (42·3)	93 (40·8)	50 (43·9)	0·9 (0·7 - 1·3)		1·1 (0·7 - 1·7)	
Divorced	167 (9·5)	137 (9·7)	19 (8·3)	11 (9·6)	0·8 (0·5 - 1·4)		1·1 (0·5 - 2·1)	
Widowed	83 (4·7)	62 (4·4)	15 (6·6)	6 (5·3)	1·5 (0·8 - 2·7)		1·3 (0·5 - 3·1)	
Smear positivity	2,208 (95·6)							
Scanty 1 - 9	173 (7·8)	147 (8·2)	18 (6·6)	8 (5·6)	0·8 (0·5 - 1·4)		0·6 (0·3 - 1·3)	
1+	474 (21·5)	381 (21·2)	67 (24·6)	26 (18·3)	1·5 (0·9 - 1·6)		0·8 (0·5 - 1·2)	
2+	546 (24·7)	445 (24·8)	65 (23·9)	36 (25·4)	1·0 (0·7 - 1·4)		0·9 (0·6 - 1·4)	
3+	1,015 (46·0)	821 (45·8)	122 (44·9)	72 (50·7)	Reference		Reference	
Previous TB treatment	1,737 (75·2)							
Yes	291 (16·8)	242 (17·2)	30 (13·5)	19 (17·0)	0·7 (0·5 - 1·1)		1·0 (0·6 - 1·6)	
No	1,446 (83·2)	1,161 (82·8)	192 (86·5)	93 (83·0)	Reference		Reference	
Outcome of TB Treatment	151 (51·9)							
Cured	79 (52·3)	69 (54·3)	8 (44·4)	2 (33·3)	Reference		Reference	
Not cured	36 (23·8)	28 (22·1)	6 (33·3)	2 (33·3)	1·8 (0·6 - 5·8)		2·5 (0·3 - 18·4)	
Defaulted	36 (23·8)	30 (23·6)	4 (22·2)	2 (33·3)	1·1 (0·3 - 4·1)		2·3 (0·3 - 17·0)	
Risk of TB contact								
Closefriend/Household	1,665 (72·1)							
Any	544 (32·7)	442 (33·0)	62 (29·1)	40 (35·7)	0·8 (0·6 - 1·1)		1·1 (0·7 - 1·7)	
No	1,121 (67·3)	898 (67·0)	151 (70·9)	72 (64·3)	Reference		Reference	
Closefriend/Household	1,665 (72·1)							
No contact	1,121 (67·3)	898 (67·0)	151 (70·9)	72 (64·3)	Reference		Reference	
1 contact	212 (12·7)	179 (13·4)	20 (9·4)	13 (11·6)	0·7 (0·4 - 1·1)		0·9 (0·5 - 1·7)	
2-5 contacts	309 (18·6)	243 (18·1)	39 (18·3)	27 (24·1)	0·9 (0·6 - 1·4)		1·4 (0·9 - 2·2)	
6-10 contacts	23 (1·4)	20 (1·5)	3 (1·4)	0	0·9 (0·3 - 3·0)			
Imprisonment	1,660 (71·9)							
Yes	97 (5·8)	81 (6·1)	15 (7·1)	1 (0·9)	1·2 (0·7 - 2·1)			
No	1,563 (94·2)	1,257 (93·9)	196 (92·9)	110 (99·1)	Reference			
Health/Lab worker	1,661 (71·9)							

University of Ghana http://ugspace.ug.edu.gh

Yes	47 (2·8)	40 (3·0)	5 (2·4)	2 (1·8)	0·8 (0·3 - 2·0)		0·6 (0·1 - 2·5)	
No	1,614 (97·2)	1,299 (97·0)	206 (97·6)	109 (98·2)	Reference		Reference	
Immunosuppressive condition	1,695 (73·4)							
Any	893 (52·7)	720 (52·9)	111 (49·8)	62 (56·4)	0·9 (0·7 - 1·2)		1·1 (0·8 - 1·7)	
None	802 (47·3)	642 (47·1)	112 (50·2)	48 (43·6)	Reference		Reference	
Diabetes Mellitus	534 (23·1)							
Yes	104 (19·5)	78 (18·1)	18 (24·7)	8 (25·8)	1·5 (0·8 - 2·6)		1·6 (0·7 - 3·6)	
No	430 (80·5)	352 (81·9)	55 (75·3)	23 (74·2)	Reference		Reference	
HIV status	1,166 (50·5)							
Yes	144 (12·3)	117 (12·4)	20 (13·7)	7 (9·2)	1·1 (0·7 - 1·9)		0·7 (0·3 - 1·6)	
No	1,022 (87·7)	827 (87·6)	126 (86·3)	69 (90·8)	Reference		Reference	
Smoking	1,518 (65·7)							
Yes	434 (28·6)	366 (30·1)	37 (18·6)	31 (29·8)	0·5 (0·4 - 0·8)*		1·0 (0·6 - 1·5)	
No	1,084 (71·4)	849 (69·9)	162 (81·4)	73 (70·2)	Reference		Reference	
Substance abuse (including acohol)	1,474 (63·8)							
Yes	460 (31·2)	375 (31·7)	57 (29·5)	28 (28·3)	0·9 (0·6 - 1·3)		0·8 (0·5 - 1·3)	
No	1,014 (68·8)	807 (68·3)	136 (70·5)	71 (71·7)	Reference		Reference	
Drug resistance	2,300 (99·6)							
Any	313 (13·6)	258 (13·8)	37 (12·8)	18 (12·2)	0·9 (0·6 - 1·3)		0·9 (0·5 - 1·4)	
None	1,987 (86·4)	1,606 (86·2)	251 (87·2)	130 (87·8)	Reference		Reference	
Cluster size (number of isolates)	1,227 (53·1)							
Small (2)	290 (23·6)	208 (20·1)	74 (51·7)	8 (16·7)	Reference		Reference	
Medium (3 - 5)	262 (21·4)	216 (20·9)	35 (24·5)	11 (22·9)	0·5 (0·3 - 0·7)*	0·5 (0·3 - 0·7)*	1·3 (0·5 - 3·4)	
Large (6 - 20)	452 (36·8)	389 (37·5)	34 (23·8)	29 (60·4)	0·2 (0·2 - 0·3)**	0·2 (0·1 - 0·4)**	1·9 (0·9 - 4·3)	
Very large (>20)	223 (18·2)	223 (21·5)	0	0				

For the multivariate model, we included only variables with p<0·1

*p<0·05 **p<0·001

University of Ghana http://ugspace.ug.edu.gh

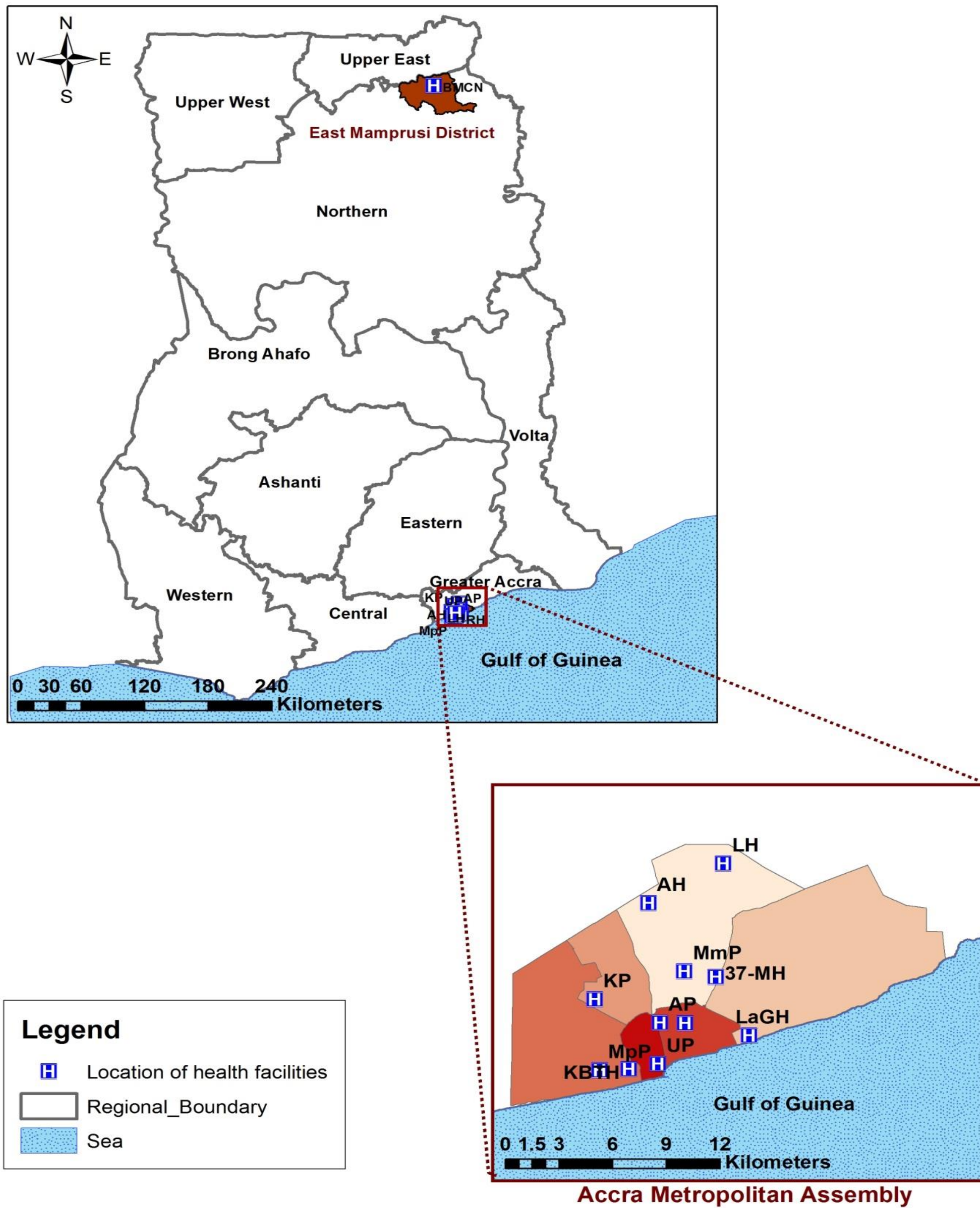


Figure S4.1. Map of study area showing the location of the 12 sampling sites (health facilities)

University of Ghana http://ugspace.ug.edu.gh

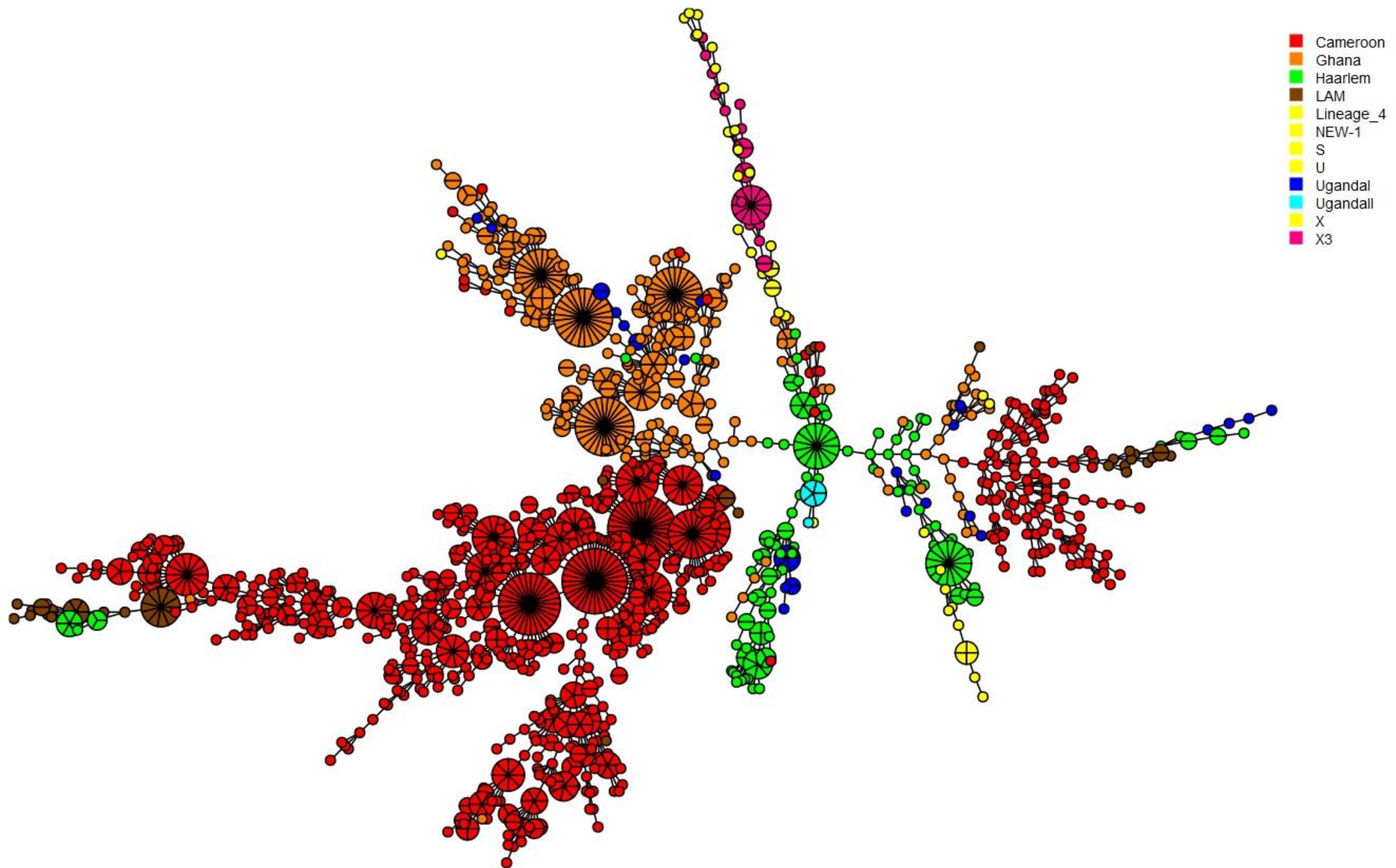


Figure S4.2. Minimum spanning tree (MST) representation of the clustering of sub-lineages of lineage 4 MTBC isolates built with Bionumerics software. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. The color codes indicate the various L4 sub-lineages and have been provided in the figure legend.

University of Ghana http://ugspace.ug.edu.gh

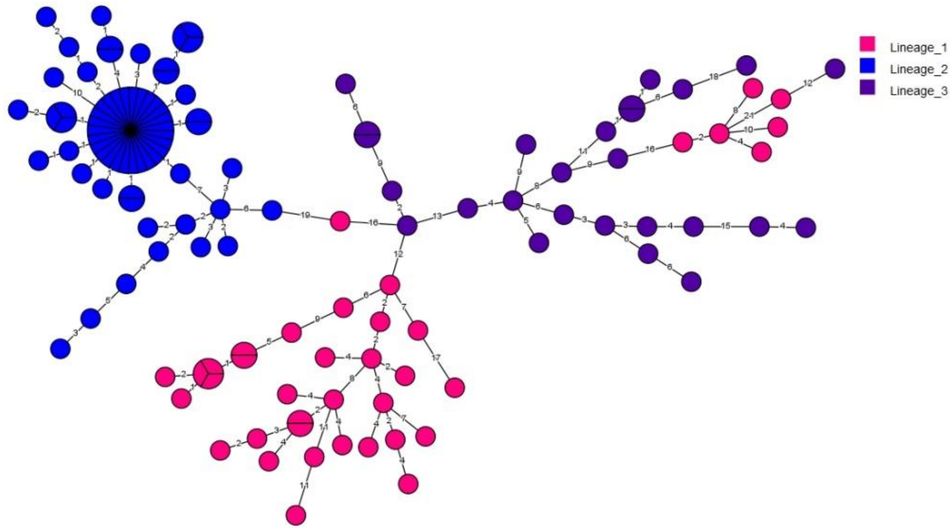


Figure S4.3. MST representation of the clustering of lineage 1 - 3 MTBC isolates built with Bionumerics software. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Digits on the branches indicate number of polymorphic VNTR/Spoligo-spacer between isolates among the linked cases.

Figure S4.4. MST representation of the clustering of lineage 4 MTBC isolates built with Bionumerics software. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Orange circles or sectors indicate individual cases with multidrug resistant (MDR) TB strains.

University of Ghana http://ugspace.ug.edu.gh

Figure S4.5. MST representation of the clustering of lineage 5 MTBC isolates built with Bionumerics software. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Digits on the branches indicate number of polymorphic VNTR/Spoligo-spacer between isolates among the linked cases. Orange circles or sectors indicate individual cases with multidrug resistant (MDR) TB strains.

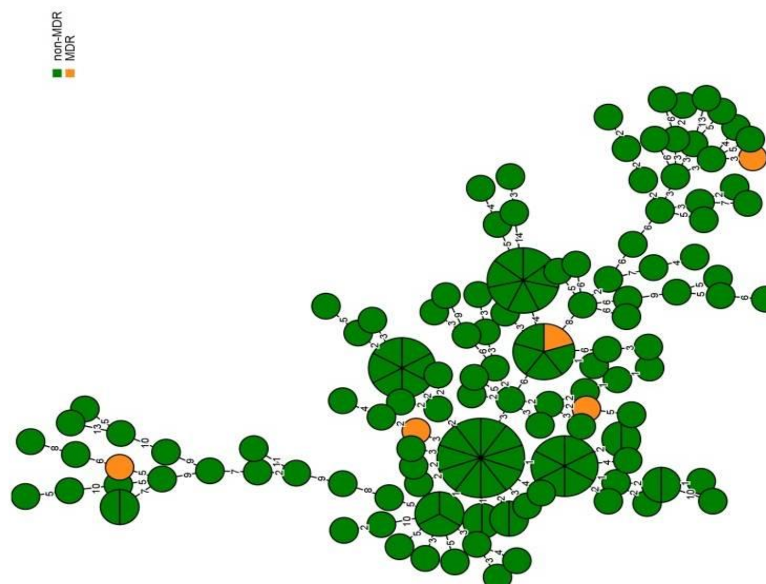


Figure S4.6. MST representation of the clustering of lineage 6 MTBC isolates built with Bionumerics software. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Digits on the branches indicate number of polymorphic VNTR/Spoligo-spacer between isolates among the linked cases. Orange circles or sectors indicate individual cases with multidrug resistant (MDR) TB strains.

University of Ghana http://ugspace.ug.edu.gh

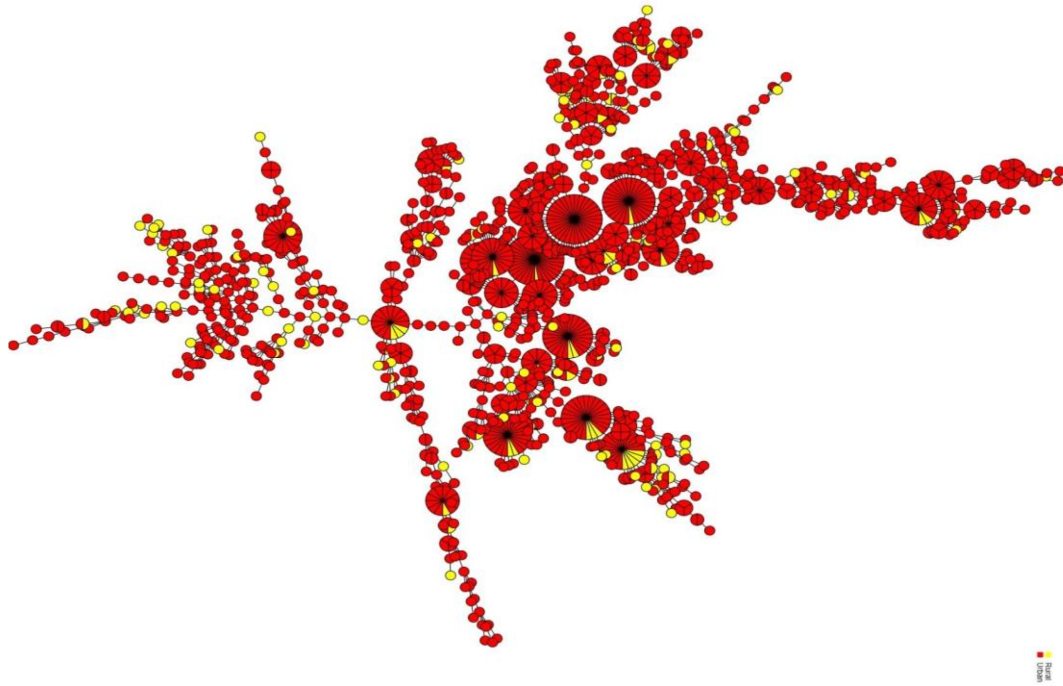


Figure S4.7. MST representation of the clustering of lineage 4 MTBC isolates stratified by study setting. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Yellow circles or sectors indicate individual cases form the rural setting.

Figure S4.8. MST representation of the clustering of lineage 5 MTBC isolates stratified by study setting. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Orange circles or sectors indicate individual cases form the rural setting.

University of Ghana http://ugspace.ug.edu.gh

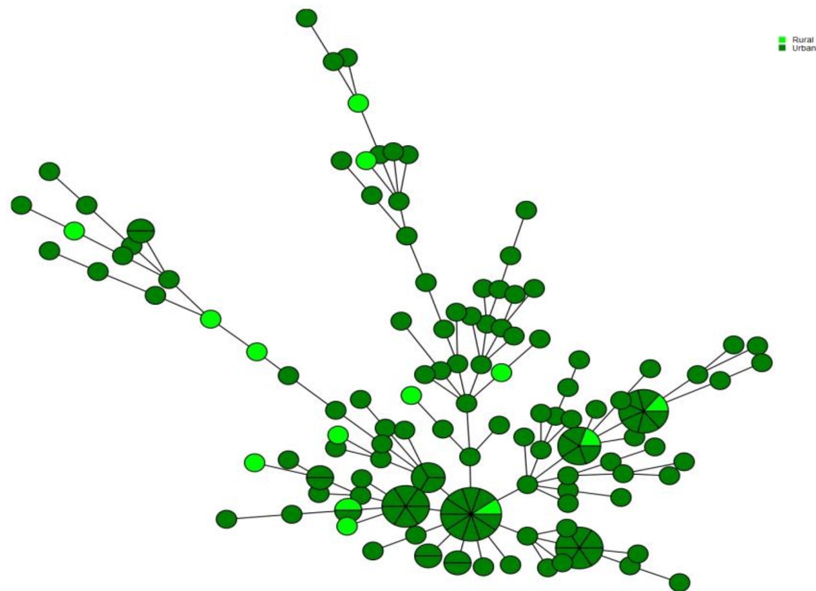


Figure S4.9. MST representation of the clustering of lineage 6 MTBC isolates stratified by study setting. The size of the clusters shows the number of clustered isolates within the cluster with the same strain type. Light green circles or sectors indicate individual cases form the rural setting.

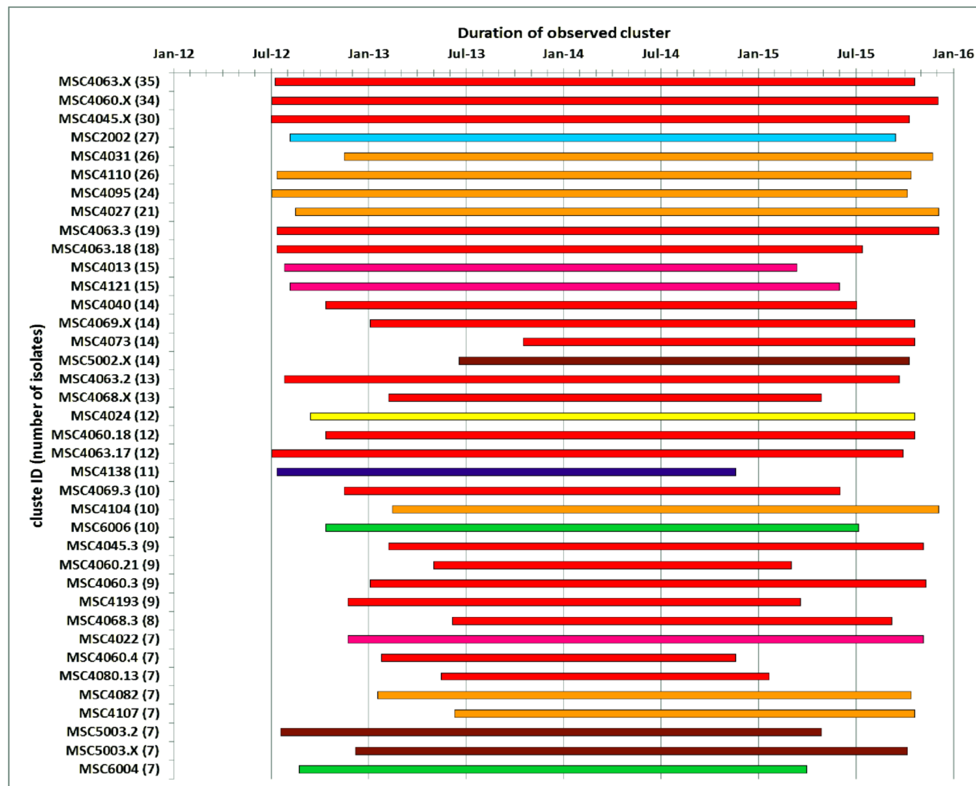


Figure S4.10. Time lapse between first and last diagnosed case within selected large and very large clusters. Lineages 5 and 6 have been color coded with the universally accepted color codes for the main MTBC lineages whereas sub-lineages of lineage 4 have been color coded, Cameroon: red, Ghana: gold, Haarlem: pink, LAM: purple and X3: yellow.

University of Ghana http://ugspace.ug.edu.gh

APPENDIX 3

Supplementary data to chapter 5

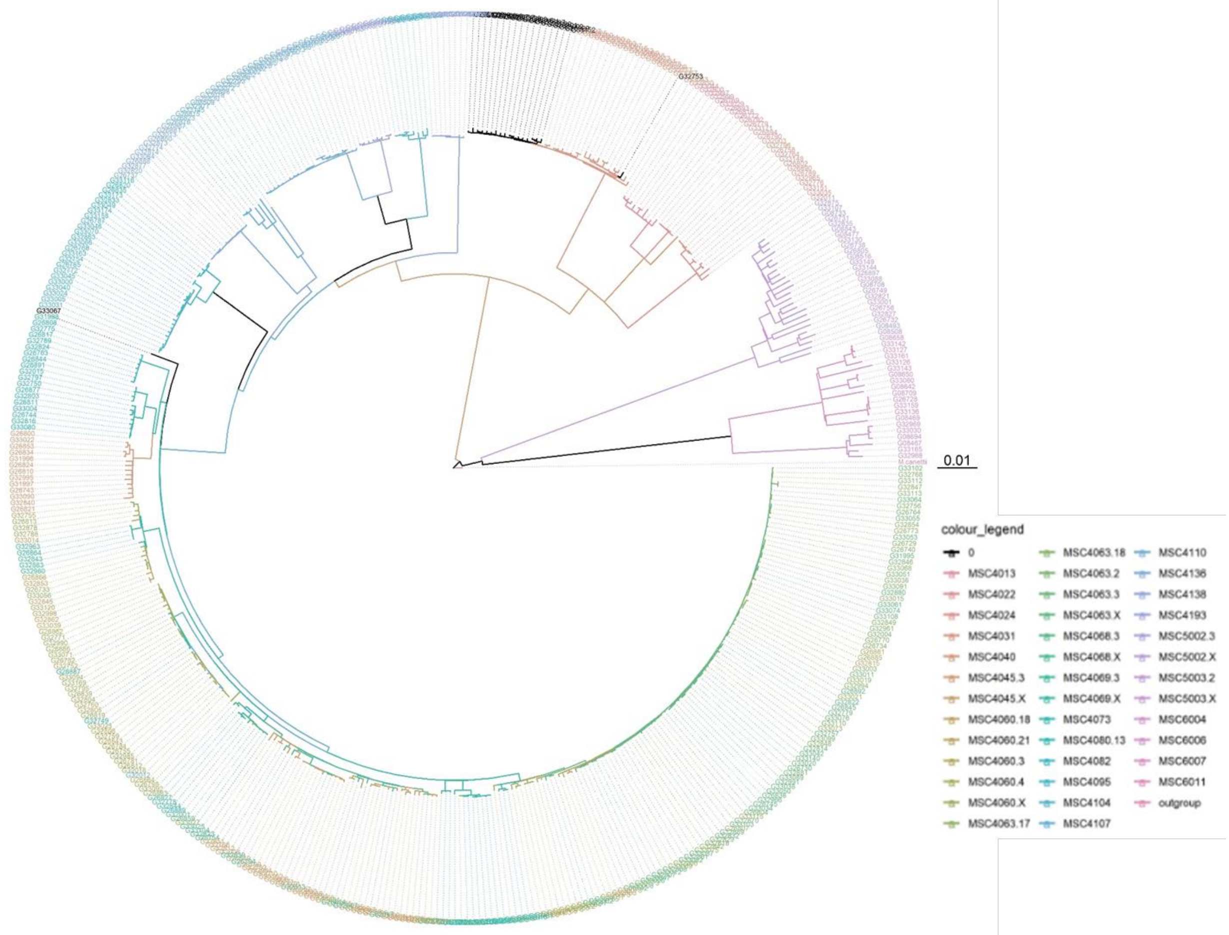


Figure S5.1. Clustering and phylogenomic relationship of 452 *M. tuberculosis* complex isolates. The color codes represent clusters previously defined by combined resolution of MIRU-VNTR typing and spoligotyping, here referred to as traditional genotype clusters. Near distinct monophylectic clades represent the various large traditional genotype clusters. The tree was rooted with *M. canettii*.

University of Ghana http://ugspace.ug.edu.gh

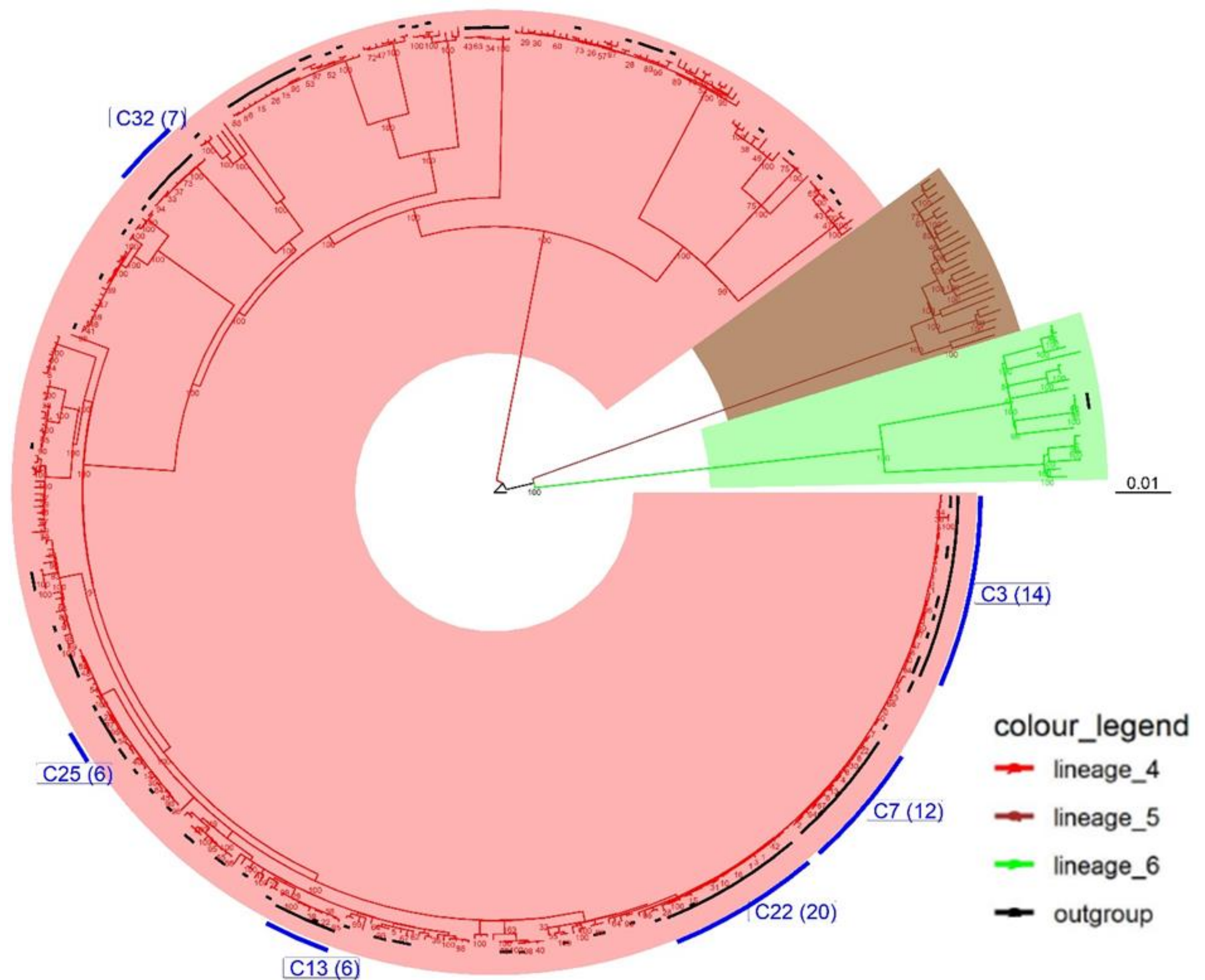


Figure S5.2. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 5-SNPs. The tree was built with an alignment file containing 11,041 variable positions. Black bars plotted on the tips of the branches indicate the clustered cases at the defined threshold of 5-SNPs. Blue bars represent large clusters (cluster size >5) with the number of clustered cases indicated in brackets. The three major branches constitute the three main MTBC lineages found in Ghana and color coded with the universally accepted Gagneux-defined lineages as red for lineage 4, brown for lineage 5 and green for lineage 6. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*. Using a SNP threshold of 5, we identified 69 clusters with a median cluster size of 6 (range 4 to 12) and total clustered cases of 226 individuals. Six large clusters were observed. All the large clusters were supported with a bootstrap value of 100.

University of Ghana http://ugspace.ug.edu.gh

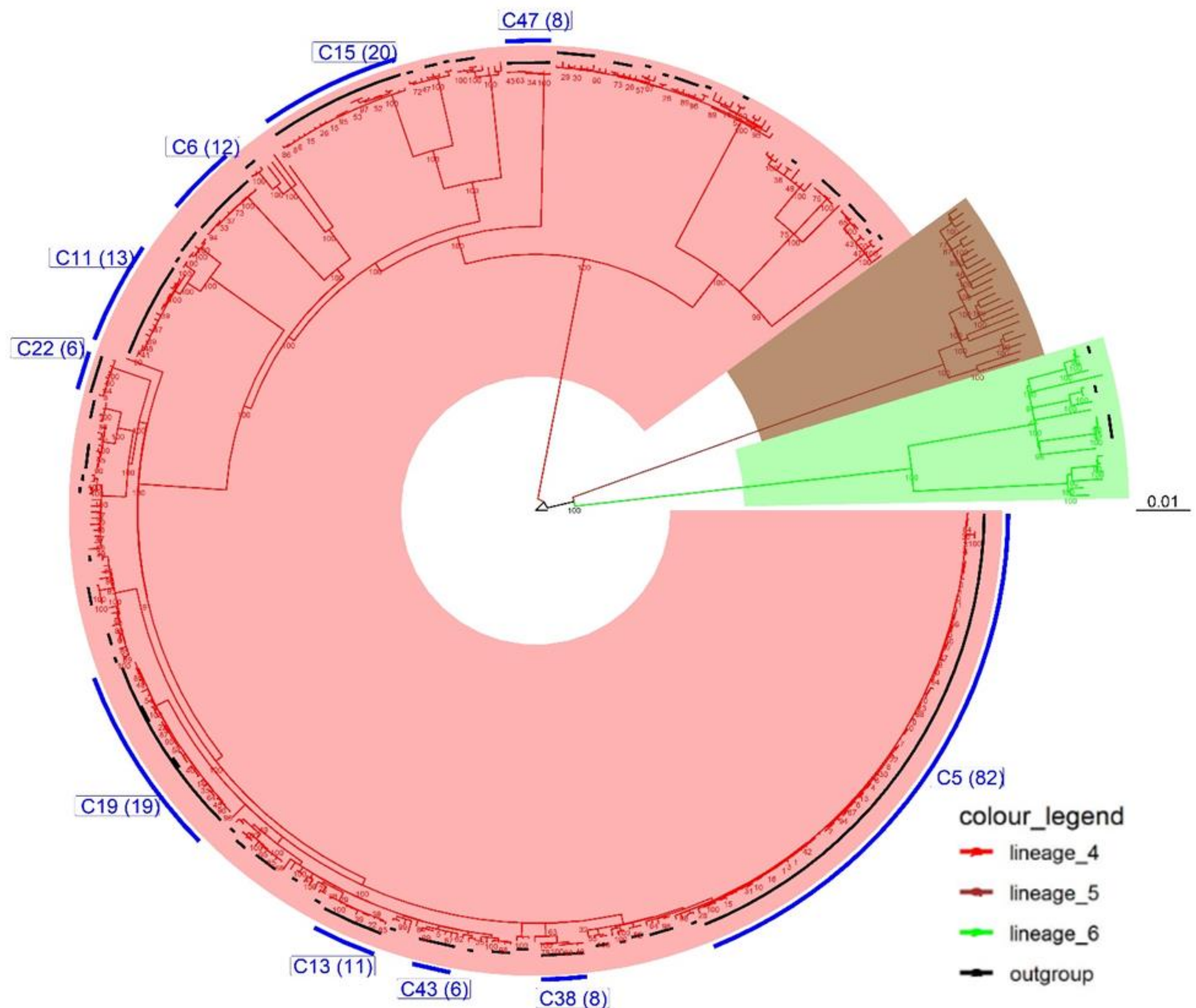


Figure S5.3. Phylogenetic reconstruction of 452 *M. tuberculosis* complex isolates showing clustering at a threshold of 12-SNPs. The tree was built with an alignment file containing 11,041 variable positions. Black bars plotted on the tips of the branches indicate the clustered cases at the defined threshold of 12-SNPs. Blue bars represent large clusters (cluster size >5) with the number of clustered cases indicated in brackets. The three major branches constitute the three main MTBC lineages found in Ghana and color coded with the universally accepted Gagneux-defined lineages as red for lineage 4, brown for lineage 5 and green for lineage 6. Numbers on branches represent bootstrap support after 1,000 rapid bootstrap inferences with the best tree shown. The tree was rooted with *M. canettii*. Using a SNP threshold of 12, we identified 60 clusters with a median cluster size of 6 (range 4 to 12) and total clustered cases of 329 individuals. Ten large clusters were observed with the largest consisting of 82 genomes. All the large clusters were supported with a bootstrap value of 100.

University of Ghana http://ugspace.ug.edu.gh

Table S5.1. Clustering analysis stratified by location of residence for large clusters

Residential district	all large clusters n(%)	WGSC-5 n(%)
Ablekuma	37 (25.52)	19 (24.68)
Adenta Municipal	1 (0.69)	-
Akwapim South	1 (0.69)	1 (1.3)
Ashiedu Keteke	11 (7.59)	8 (10.39)
Ashaiman Municipal	1 (0.69)	-
Awutu Senya	1 (0.69)	1 (1.3)
Ayawaso	23 (15.86)	8 (10.39)
Ga Central	2 (1.38)	-
Ga East	4 (2.76)	2 (2.6)
Ga South	3 (2.07)	3 (3.9)
Ga West	4 (2.76)	1 (1.3)
Kpeshie	12 (8.28)	9 (11.69)
La-Nkwantanang Madina Municipal	1 (0.69)	1 (1.3)
Mamprusi East	1 (0.69)	1 (1.3)
Northern Ghana	1 (0.69)	1 (1.3)
Okaikoi	16 (11.03)	12 (15.58)
Osu Klottey	3 (2.07)	1 (1.3)
Southern Ghana	21 (14.48)	9 (11.69)
Tamale Metropolis	1 (0.69)	-
Tema Municipal	1 (0.69)	-
Total	145 (100)	77 (100)