GENOTYPING OF *Mycobacterium tuberculosis* (MTB) USING MICROBEAD-BASED SPOLIGOTYPING AND 24-LOCI MIRU-VNTR TO COMPLEMENT TUBERCULOSIS PREVENTION AND MANAGEMENT IN SABAH, EAST MALAYSIA

DAWN CARMEL PAUL

DOCTOR OF PHILOSOPHY

(MEDICAL SCIENCE)

M. KANDIAH FACULTY OF MEDICINE AND HEALTH SCIENCES

UNIVERSITI TUNKU ABDUL RAHMAN

MAY 2024

GENOTYPING OF *Mycobacterium tuberculosis* (MTB) USING MICROBEAD-BASED SPOLIGOTYPING AND 24-LOCI MIRU-VNTR TO COMPLEMENT TUBERCULOSIS PREVENTION AND MANAGEMENT IN SABAH, EAST MALAYSIA

By

DAWN CARMEL PAUL

A thesis submitted to the Department of Pre-clinical Sciences,

M. Kandiah Faculty of Medicine and Health Sciences,

Universiti Tunku Abdul Rahman,

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (Medical Science)

MAY 2024

Dedicated to

my late paternal grandparents Flora Cyprianus, and Liberatus John

&

my late maternal grandparents Robert Tinus Marikam, and Monica Lim

ABSTRACT

GENOTYPING OF *Mycobacterium tuberculosis* (MTB) USING MICROBEAD-BASED SPOLIGOTYPING AND 24-LOCI MIRU-VNTR TO COMPLEMENT TUBERCULOSIS PREVENTION AND MANAGEMENT IN SABAH, EAST MALAYSIA

DAWN CARMEL PAUL

Sabah, a state in East Malaysia contributes to about 20% of Malaysia's TB cases with a prevalence of 128/100,000 population compared to the national prevalence of 92/100,000 population. Molecular epidemiology can help guide decisions pertaining to TB prevention and management, but to date information on MTB genotyping is still scant in Sabah. This study aims to determine the circulating MTB diversity in the state and how genotyping could complement existing TB prevention and management efforts. 1060 unique archived MTB clinical isolates from 2015 to 2016, and prospective isolates from 2017 to 2018 from the whole state were randomly selected and matched against the TB patient registration information in the MyTB database. Epidemiological data were obtained from MyTB, MKAKK Laboratory records and SIMKA database. DNA extraction was performed in the MKAKK High Containment TB Laboratory. Spoligotyping and resistotyping were done using the microbead-based TB-SPRINT Beamedex assay in UTAR Pathogen Laboratory. SITVIT2, SpolSimilarity Search and SpolLineages Tool were employed to assign MTB spoligopatterns. 1019 isolates were finalised for further analyses from which 102 SITs with 68 existing SITs and 34 newly created SITs (SIT4181 to SIT4215),

and 153 orphan spoligopatterns were detected. Eight newly created SITs from an Unknown lineage could be part of an emerging lineage unique to this region. The lineage distributions are: EAI (n=811), Unknown (n=119), T (n=25), Beijing (n=25), LAM (n=17), Manu (n=9), AFRI (n=4), H (n=4), X (n=3), Turkey (n=1), and Atypical (n=1). Duplex 24-loci MIRU-VNTR were performed on selected clustered and unique spoligotyped isolates (N=593) representing Sabah spoligopatterns. 36 isolates were mixed infections. Cluster analysis was done on 556 strains using *MIRU-VNTRplus*. ETR-B, Mtub2, QUB11b and ETR-A were the most discriminatory loci. Eight clonal complexes and 51 MIT24 profile clusters were discovered. There was no significant correlation between MTB lineages and factors such as age, sex, occupation, nationality and ethnicity. Individual HGDI and clustering rate of spoligotyping and 24-loci MIRU-VNTR were 0.814 and 0.748; and 0.997 and 0.223 respectively. Combination of both spoligotyping and 24-loci MIRU-VNTR rendered an HGDI and clustering rate of 0.998 and 0.223. TB transmission and reactivation rates for conventional contact investigation versus genotyping complemented contact investigation are 6.1% and 93.9% versus 23.6% and 76.4%. Genotyping revealed new possible epidemiological links previously undetected by conventional investigation. Two alternative genotype-complemented approaches to contact investigation were proposed namely Direct Cluster approach, and Prioritised Cluster approach. Both approaches are yet to be tested in the field. However, the potential of these strategies in terms of deepening the understanding of TB transmission dynamics in addition to cost reduction in TB prevention and management activities could be of interest to the stakeholders. MTB drug-resistance is still relatively low in Sabah. Eighty-four

MTB isolates (8.2%) in this study were found to be drug resistant (MDRTB: 6, 0.6%; RRTB: 4, 0.4%; INH-R: 34, 3.3%; and STR-R: 40, 3.9%). The ratio between low level and high level INH resistance (*inhA* prom mutation:*katG* mutation) is about 1:1. Five of six MDRTB cases showed the same mutation: *rpoB* 531_mut_TTG and *inhA*_prom_mut-15_T suggesting primary drug resistance. Recommendation was made to the Sabah State Health department on the feasibility and the potential impact of genotyping on TB prevention and management activities in the State with the hope to facilitate reduction in TB incidence in Sabah. Whole genome sequencing and universal genotyping were not done due to technical issues. Distribution of samples over the years (2015 to 2018) were not equal due to limited availability of MTB isolates.

ACKNOWLEDGEMENTS

All thanks and glory to my *Lord Jesus Christ* for the strength and resolve to complete this study

Highest gratitude and appreciation to

1. My supervisors, Prof Yap Sook Fan, Prof Ngeow Yun Fong, Dr Jiloris Frederick Dony, and Assoc. Prof Richard Avoi for their guidance and support
2. My colleagues in M. Kandiah FMHS - Pathogen Laboratory, especially to Dr Ng Hien Fuh and Madam Sargit Kaur for their guidance and friendship
3. My collaborator and friend, Dr Emilyn Costa Conceição for your knowledge sharing and insight pertaining to MTB genotyping
4. UTAR for the grants UTAR RF 6200/N39, 6200/N43 & 6200/N48
5. Malaysian Ministry of Health for ETHICS APPROVAL: NMRR-16-2780-32850 (IIR)
6. Sabah State Health Department, TB/Leprosy Sector and MKAKK

Thank you ever so much to my mum Betty Robert, and dad Paul Liberatus; my husband Jal Gabriel Gai; My friends and family for your kind support, prayers and thoughtfulness throughout this journey.

No man is an island, and all things are possible for those who believe

MAY 2024

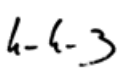
APPROVAL SHEET

This dissertation/thesis entitled "**GENOTYPING OF *Mycobacterium tuberculosis* (MTB) USING MICROBEAD-BASED SPOLIGOTYPING AND 24-LOCI MIRU-VNTR TO COMPLEMENT TUBERCULOSIS PREVENTION AND MANAGEMENT IN SABAH, EAST MALAYSIA"** was prepared by DAWN CARMEL PAUL and submitted as partial fulfilment of the requirements for the degree of Doctor of Philosophy (Medical Science) at Universiti Tunku Abdul Rahman

Approved by:

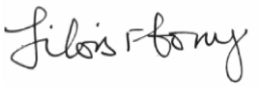
_____

(Prof. Yap Sook Fan) Date: 31 MAY 2024
Senior Professor/Supervisor
Department of Pre-clinical Sciences
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman



_____

(Prof. Ngeow Yun Fong) Date: 31 MAY 2024
Senior Professor/Co-supervisor
Department of Pre-clinical Sciences
Faculty of Medicine and Health Sciences
Universiti Tunku Abdul Rahman



_____

(Dr Jiloris F. Dony) Date: 31 MAY 2024
External Co-supervisor
Kota Kinabalu Public Health Laboratory
Ministry of Health

_____

(Assoc Prof. Dr Richard Avoi) Date: 31 MAY 2024
External Co-supervisor
Faculty of Medicine and Health Sciences
Universiti Malaysia Sabah

M. KANDIAH FACULTY OF MEDICINE AND HEALTH SCIENCES

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 31 MAY 2024

PERMISSION SHEET

It is hereby certified that DAWN CARMEL PAUL (ID No: 16UMD07621) has completed this thesis entitled "**GENOTYPING OF *Mycobacterium tuberculosis* (MTB) USING MICROBEAD-BASED SPOLIGOTYPING AND 24-LOCI MIRU-VNTR TO COMPLEMENT TUBERCULOSIS PREVENTION AND MANAGEMENT IN SABAH, EAST MALAYSIA**" under the supervision of Prof. Yap Sook Fan (Supervisor) from the Department of Pre-clinical Sciences, M. Kandiah Faculty of Medicine and Health Sciences, Prof. Ngeow Yun Fong (Co-Supervisor) from the Department of Pre-clinical Sciences, M. Kandiah Faculty of Medicine and Health Sciences, Dr Jiloris F. Dony (External Supervisor) Director of Kota Kinabalu Public Health Laboratory, and Associate Prof. Dr Richard Avoi (External Supervisor) Senior Lecturer from Faculty of Medicine and Health Sciences, UMS)

I understand that the University will upload a softcopy of my thesis in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

_____

(DAWN CARMEL PAUL)

DECLARATION

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

Signature _____

(DAWN CARMEL PAUL)

Date 31 MAY 2024

TABLE OF CONTENTS

		Page
ABSTRACT		**iv**
ACKNOWLEDGEMENTS		**vii**
APPROVAL SHEET		**viii**
PERMISSION SHEET		**ix**
LIST OF TABLES		**xv**
LIST OF FIGURES		**xviii**
LIST OF APPENDICES		**xxi**
LIST OF ABBREVIATIONS		**xxii**
CHAPTER		
1.0	**INTRODUCTION**	**1**
1.1	Study background	1
1.2	Problem statement	5
1.3	Hypothesis	5
1.4	Study objectives	5
1.4.1	Specific objectives	6
2.0	**LITERATURE REVIEW**	**7**
2.1	Tuberculosis (TB)	7
2.2	*Mycobacterium tuberculosis* (MTB)	11
2.2.1	Origin and strain variation of MTB	12
2.2.2	MTB drug resistance	23
2.2.3	TB relapse and reinfection	31
2.2.4	Mixed infection of MTB	33
2.3	Genotyping and molecular epidemiology of MTB	33
2.3.1	Evolution of genotyping techniques for MTB	35
2.3.2	Spacer oligonucleotide typing (spoligotyping)	40
2.3.3	Mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR)	47
2.3.4	Lineage and sub-lineage designation	51
2.3.5	MTB clustering, clustering rate and Hunter Gaston discriminatory index	57
2.4	TB prevention and management program	59
2.4.1	WHO Stop TB strategy and End TB strategy	62
2.4.2	TB prevention and management program in Sabah	64
2.4.3	TB case contact investigation	67
2.5	Sabah, land below the wind	74
2.5.1	Migrant population in Malaysia	76
2.5.2	TB and the Sabah population	77

			Page
3.0	**MATERIALS AND METHODS**		**83**
	3.1	Conceptual framework	83
	3.2	Registration of study	83
	3.3	Data mining of TB laboratory diagnostics in Sabah, 2012-2018	84
	3.4	Sample collection	84
		3.4.1 Epidemiological data	84
		3.4.2 Isolate selection	85
		3.4.3 AFB microscopy	88
		3.4.4 AFB culture	88
		3.4.5 Isolate identification and drug susceptibility testing	89
		3.4.6 Line probe assay for the detection of MTB / MDRTB (LPA)	89
		3.4.7 DNA extraction	89
	3.5	Genotyping of MTB isolates using two-tiered classical genotyping	90
		3.5.1 TB-SPRINT (TB-SPOL and TB-RINT)	90
		3.5.2 24-loci MIRU-VNTR typing	97
		3.5.3 Lineage designation and diversity analysis	102
	3.6	Genotyping-complemented case investigations MTB isolates using two-tiered classical genotyping	107
	3.7	Statistical analysis	109
4.0	**RESULTS**		**110**
	4.1	General performance of AFB culture and DST in Sabah	110
		4.1.1 Sample population by occupation in Sabah	116
	4.2	Genetic diversity, population structure and clustering of MTB in Sabah	117
		4.2.1 Spoligotype of MTB isolates in Sabah	117
		4.2.2 24-loci MIRU-VNTR typing panel for Sabah MTB population	133
	4.3	MTB clustering and clonal complex clustering	135
		4.3.1 24-loci MIRU VNTR clonal complex cluster	136
		4.3.2 24-loci MIRU VNTR exact clusters MIT24-391 to MIT24-441	141
	4.4	Genotype-complemented TB contact investigation with conventional TB contact investigation	143
		4.4.1 Transmission and reactivation rate	143
		4.4.2 Proposed genotype-complemented contact investigation	146
		4.4.3 Direct cluster approach for TB case-contact investigation	148
		4.4.4 Prioritized cluster approach (PCA) for TB case-contact investigation	156
	4.5	Relationship between MTB sub-lineages and drug-resistance	168
		4.5.1 Mono drug-resistant MTB	170

			Page
	4.5.2	Multi drug-resistant MTB (MDRTB)	177
	4.5.3	Drug-resistant MTB (DRTB) clonal complex	180
4.6		Relationship between sub-lineages and TB relapse cases	183
4.7		Relationship between sub-lineages and mixed infection of MTB	186
4.8		Orphan MTB in Sabah	193
5.0	**DISCUSSIONS**		**197**
5.1		Diversity and distribution of MTB genotypes in Malaysia	197
	5.1.1	Thirty-four newly created MTB SITs	197
	5.1.2	Newly created SITs from Unknown lineage	199
	5.1.3	Differences in lineage profiles in peninsular Malaysia and Sabah	199
	5.1.4	There is no significant association between MTB sub-lineages and ethnicity in Sabah	202
5.2		There is no correlation between MTB sub-lineages and DRTB in Sabah	206
	5.2.1	MTB isoniazid (INH) resistance in Sabah	208
	5.2.2	MTB rifampicin (RIF) resistance prevalence is low in Sabah	209
	5.2.3	Multi drug-resistant TB (MDRTB) prevalence is low in Sabah	210
	5.2.4	Universal genotyping of all patient isolate is important to systematically gauge, and to establish the initial level of MTB drug resistance in a population	212
	5.2.5	Phenotypic (ACM) detected DRTB need to be counter-tested using molecular DST, and vice versa	213
	5.2.6	Challenges in Culture-based MTB DST	215
5.3		Relapse / reinfection cases of MTB in Sabah	217
5.4		Mixed infections of MTB in Sabah	218
5.5		Orphan MTB showed higher proportional sub-lineage variation as compared to total MTB genotypes in Sabah	222
5.6		MTB cluster analysis and genotype-complemented TB contact investigation	223
	5.6.1	Conventional contact investigation needs to be revamped with the use of genotyping	224
	5.6.2	Clustering rate need to be monitored to assess recent transmission, and reactivation of MTB in Sabah	226
	5.6.3	Notification interval should be considered as an Indicator in the TB surveillance system	227
	5.6.4	Implementation of DCA and PCA has the potential of reducing 60% to 80% of the cost allocation for TB case contact investigation	227
	5.6.5	Proposed accessory surveillance modalities for Early detection of TB	229

		Page
5.6.6	Cost considerations for MTB genotyping	231
5.7	Feasibility of application of two-tiered classical genotyping in Sabah	234
5.7.1	Advantages of spoligotyping in TB molecular epidemiology in Sabah	234
5.7.2	Importance of 24-loci MIRU-VNTR in TB case Contact investigation in Sabah	235
5.7.3	The practicability of classical genotyping in TB prevention and management in Sabah	237
5.8	Limitations of research	238
6.0	**CONCLUSIONS**	**239**
6.1	Overall study conclusions	239
6.2	Recommendations to the Sabah State Health Department	243
6.3	Future studies	245
LIST OF REFERENCES		**247**
APPENDICES		**272**
LIST OF PUBLICATIONS AND PAPERS PRESENTED		**282**

LIST OF TABLES

Table		Page
2.1	MTB lineage names	14
2.2	Signature signals of MTB spoligotypes from different lineages / sub-lineages in this study	15
2.3	Genotyping techniques and their limitations	39
3.1	Number of MTB isolates sampled by districts in Sabah (**r**andom selection of samples, with a MyTB registry match)	86
3.2	MTB isolates according to year of isolation and patient country of origin (N=1019)	87
3.3	Optimised duplex MIRU-VNTR PCR cocktail	99
3.4	Primers for duplex 24-loci MIRU-VNTR panel	100
3.5	Correspondence table between amplicon sizes and number of repeats for each VNTR, presented for each duplex	101
4.1	AFB culture performance in Sabah 2014-2016 (by patient)	112
4.2	AFB smear grading in this study	113
4.3	Sample population details (N=1019)	115
4.4	Sub-lineages and SIT of MTB in this study	119
4.5	Predominant MTB SITs in this study	121
4.6	Newly created MTB SITs in Sabah (4181-4216)	122
4.7	New MTB SIT and clonal complex by 24-loci MIRU-VNTR in Sabah, East Malaysia	124
4.8	Possible members of new MTB lineages in Sabah, East Malaysia	125
4.9	Comparison of MIRU VNTR HGDI among different lineages of MTB isolates in Sabah	134

Table		Page
4.10	Clustering analysis of mycobacterial interspersed repetitive units - variable number tandem repeat (MIRU-VNTR) and spoligotyping	137
4.11	Clonal complex clustering pattern of MTB by country of origin (n=556)	140
4.12	TB notification time interval between Case #1 and Case #2 in clonal complex clusters	145
4.13	Direct cluster approach (DCA) and prioritised cluster approach (PCA) - simple heuristic approach for genotype complemented TB contact investigation in this study	147
4.14	Clonal complexes derived from 24-loci MIRU-VNTR DLV2	148
4.15	51 clusters 24-loci MIRU-VNTR MIT24 of MTB in Sabah, n=175	151
4.16	MIT24-397 (24-loci MIRU-VNTR profile 2145243A2923276223342713)	155
4.17	Prioritised cluster approach PCA-B1, step 2 - 24-loci MIRU-VNTR clonal complexes of TB cases for prioritization for contact investigation	157
4.18	Filipino (PHP) – Malaysian (MYS) MTB combined clonal complex	163
4.19	Distribution of drug-resistance by country of origin in Sabah, East Malaysia	169
4.20	MTB isoniazid resistance in Sabah, East Malaysia	173
4.21	MTB rifampicin resistance in Sabah, East Malaysia	178
4.22	Multidrug-resistant MTB (MDRTB) in Sabah, East Malaysia	179
4.23	TB relapse / retreatment cases in Sabah, 2012-2017	184
4.24	MIRU-VNTR loci with multiple allele	188

Table		Page
4.25	Possible transmission pattern for TB cases with mixed infections (1-14)	190
4.26	Mixed infection MTB by nationality in Sabah, East Malaysia	192
4.27	Orphan MTB Isolates by patient nationality in Sabah, East Malaysia	195
5.1	Cost for microbead-based spoligotyping and 24-loci MIRU-VNTR	233

LIST OF FIGURES

Figure		Page
2.1	Various outcomes of MTB infection (Gautam et al., 2023)	10
2.2	DR locus and membrane spoligotyping	43
2.3	MIRU-VNTR loci in MTB H37Rv chromosome	49
2.4	Designation of MIRU-VNTR profile (a) 12-loci MIRU-VNTR; (b) 15-loci MIRU-VNTR and (c) 24-loci MIRU-VNTR	50
2.5	SpolLineages Tool binary spoligo rules (Couvin et al., 2020)	54
2.6	WHO Stop TB Strategy and End TB Strategy	62
2.7	TB prevention and management algorithm in Sabah (Avoi, 2015)	69
2.8	Map showing Sabah (East Malaysia), Philippines (South Philippines) and Indonesia (Kalimantan and Sulawesi) in South East Asia	74
2.9	Map of Sabah showing the 5 divisions and 25 districts	75
3.1	Conceptual framework of this study	83
3.2	Two-tiered genotyping of MTB using spoligotyping and 24-loci MIRU-VNTR, and resistotyping	96
3.3	Examples of spoligotyping results	104
4.1	Stratification of population by occupation	116
4.2	MTB isolates from Unknown lineage were grouped into 8 clonal complexes involving 119 isolates	126
4.3	MST of the newly created MTB SITs from the EAI2-Manila and Unknown lineage	128
4.4	VENN Diagram of MTB SIT distribution among Malaysian, Filipino and Indonesian in Sabah	130

Figure		Page
4.5	Comparison of MTB lineages in Sabah by country of origin – (a) Indonesia, (b) Malaysia, and (c) Philippines	132
4.6	Sub-lineages by spoligotyping (a) Ngeow et al., 2006, MTB isolates from Malaysia (with relatively low representation from Borneo (n=689); (b) Ismail et al., 2014, MTB Isolates from Kuala Lumpur and Kelantan (n=220); (c) This study (n=1019), MTB isolates from Sabah	132
4.7	Eight clonal complexes (CC) of MTB in Sabah (Bionumerics 7.6)	138
4.8	Clonal complex 1 (CC1) distribution by sub-lineage	139
4.9	Distribution of 51 MTB MIT24 exact 24-loci MIRU-VNTR clusters	142
4.10	Figure showing branch 12 (n=274) which is a part of CC1 (a) dendogram of the 593 MTB isolates, (b) branch 12 of CC1 (n=274), (c) detailed members of branch 12 of CC1	149
4.11	Combination of conventional and genotype-complemented TB case contact investigation (MIT24-397)	154
4.12	Prioritised Cluster Approach 1 for genotype complemented contact investigation, (a) DLV2 24-loci MIRU-VNTR Clonal Complexes, (b) Super Clonal Complex of isolates with highest likelihood of transmission	159
4.13	Minimum Spanning Tree of MTBC in Sabah, East Malaysia among(a) Filipinos, CC=5; (b) Indonesian, CC=3; (c) Malaysian, CC=8. Clusters and clonal complexes indicate recent transmission while singletons indicate reactivation	161
4.14	Possible transmission of TB CC2-PHP	162
4.15	Possible transmission of PHP–MAS MTB combined clonal complex (CC PHP-MAS)	164
4.16	Drug resistant TB (DRTB) distribution in Sabah	176
4.17	Minimum spanning tree (MST) of DRTB clonal complexes of MTB isolates in Sabah	182

Figure		Page
4.18	TB relapse cases by nationality and MTB sub-lineage in Sabah, East Malaysia	185
4.19	Distribution of mixed TB infection cases by occupation	188
4.20	Distribution of orphan MTB Spoligotypes in Sabah, East Malaysia	193
4.21	Minimum spanning tree of orphan MTB in Sabah, East Malaysia	195
4.22	Distribution of orphan MTBC spoligotype by nationalities in Sabah, East Malaysia a) Total orphan MTB Spoligotype; b) orphan MTBC spoligotype of Indonesians; c) Orphan MTBC spoligotype of Malaysians, East Malaysia d) Orphan MTBC spoligotype of Filipinos	196

LIST OF APPENDICES

Appendix		Page
A	List of reagents and consumables for spoligotyping and 24-loci MIRU-VNTR	271
B	List of items needed for TB SPRINT	272
C	Correspondence of spacers with the microsphere sets for TB-SPRINT	273
D	Reagent preparation for TB-SPRINT hybridization	274
E	Reagent preparation for buffer Q	275
F	PCR thermocycling profile for duplex 24-loci MIRU-VNTR	276
G	Occupation by nationality in Sabah, East Malaysia	277
H	Distribution of MTB SIT among Malaysian, Filipino and Indonesian in Sabah	278
I	40 GenTypes (Spoligotype+24-loci MIRU-VNTR) clusters of MTB from Sabah, n=129	280

LIST OF ABBREVIATIONS

ACD	Active case detection
ACM	Absolute Concentration Method
AFB	Acid Fast Bacilli
AFRI	*M. africanum* West Africa spoligotyping-based lineage
AI	Artificial Intelligence
All-S	All susceptible / pan susceptible
ASEAN	Association of Southeast Asian Nations
BCG	Bacillus Calmette Guérin
BFT	Beaufort district
CAS	Central Asian Strain MTB spoligotyping-based lineage
CC	Clonal complex
CDC	Centre for Disease Control
CRISPR	Clustered regularly interspaced short palindromic repeats
DCA	Direct cluster approach
DCSM	Drug-containing solid-media
DLV2	Double locus variant
DNA	Deoxyribonucleic acid
DOT	Directly observed treatment
DOTS	Directly observed treatment short course
DR	Direct repeats
DRTB	Drug-resistant tuberculosis
DST	Drug susceptibility test
EAI	East African-Indian
EEI	Enhanced epidemiological investigation
ETR-A	Exact tandem repeat A MIRU-VNTR locus
ETR-B	Exact tandem repeat B MIRU-VNTR locus

FD	Failed drug susceptibility test
FLD	First line drugs
H	Haarlem MTB spoligotyping-based lineage
HGDI	Hunter Gaston discriminatory index
IMR	Institute for Medical Research
ICT	Information and communication technology
INA	Indonesia
INH	Isoniazid
INH-R	Isoniazid resistant
KK	Kota Kinabalu district
KM	Kota Marudu district
LAM	Latin American Mediterranean MTB spoligotyping-based lineage
LDU	Lahad Datu district
LPA	Line Probe Assay
LTBI	Latent TB infection
MANU	MANU MTB spoligotyping-based lineage
MAS	Malaysia
MDRTB	Multidrug resistant tuberculosis
MKAKK	Makmal Kesihatan Awam Kota Kinabalu (Kota Kinabalu Public Health Laboratory)
MGIT	Mycobacteria growth indicator tube
MIRU-VNTR	Mycobacterial interspersed repetitive unit-variable number tandem repeat
MLST	Multi locus sequence typing
MLVA	Multi locus variable number tandem repeat analysis
MIC	Minimal inhibitory concentration
MIT	MIRU international type
MOH	Ministry of Health

MTB	*Mycobacterium tuberculosis*
mtDNA	Mitochondrial DNA
Mtub2	*Mycobacterium tuberculosis* 2 MIRU-VNTR locus
MWM	Microsphere working mix
MyTB	Malaysian TB surveillance database
NGS	Next generation sequencing
NAAT	Nucleic acid amplification test
NJ	Neighbour joining
NTM	Non-tuberculous Mycobacteria
PCA	Prioritised cluster approach
PCD	Passive case detection
PCR	Polymerase chain reaction
PHP	Philippines
PLHIV	People living with HIV
PR1	Treatment Centre 1 (Pusat Rawatan 1)
PR2	Treatment Centre 2 (Pusat Rawatan 2)
PTB	Pulmonary tuberculosis
QUB11b	Queen's University of Belfast 11b MIRU-VNTR locus
rpoB	RNA polymerase B
RIF	Rifampicin
RRDR	Rifampicin resistance determining region
RRTB	Rifampicin resistant TB
SCR	Sputum conversion rate
SCC	Short-course chemotherapy
SDK	Sandakan district
SIMKA	Sistem Informasi Makmal Kesihatan Awam (*Public Health Laboratory Information System*)
SITs	Spoligotype international types

SITVIT2	Spoligotype international type world wide database of genotyping markers for *Mycobacterium tuberculosis*
SLD	Second line drugs
SLV1	Single locus variant
SNP	Single nucleotide polymorphism
SNS	Single nucleotide substitution
SNV	Single nucleotide variation
SPN	Semporna district
STR	Streptomycin
T	T ill-defined MTB spoligotyping-based lineage
TB	Tuberculosis
TbD1	TbD1 locus 2153bp region in ancestral MTB
TBIS	TB information system (MOH Malaysia)
TB-SPRINT	TB-SPOL (spoligotyping) and TB-RINT (resistotyping)
TWU	Tawau district
WGS	Whole genome sequencing
wgMLST	Whole genome multi locus sequence typing
WHO	World Health Organization
WHO/IUATLD	World Health Organization / International Union Against Tuberculosis and Lung Disease
X	X MTB spoligotyping-based lineage
XDRTB	Extensively drug-resistant tuberculosis

CHAPTER 1

INTRODUCTION

1.1 Study background

With a prevalence of 92 cases per 100,000 people in 2019 and an estimated mortality rate of 4 cases per 100,000 people per year, Malaysia is classified as a country with a medium TB burden. Tuberculosis (TB) is still regarded as a serious disease with significant public health implications in Malaysia. On the other hand, Sabah, a state in East Malaysia, makes up about 20% of the total burden of TB in the country, with a prevalence of 128/100,000 and an 8% TB case fatality rate when compared to national data (Goroh et al., 2020; Avoi and Liaw, 2021). Looking at these numbers alone, Sabah would be included in the high TB burden category (WHO, 2023).

In Sabah, laboratory diagnostics for TB are largely microscopy and culture-centric. Solid cultures are performed for more than 80% of the samples and liquid culture for about 20% of the samples (extra-pulmonary and selected cases). TB culture is time consuming. Molecular testing such as line probe assay (Hain Life Science) and nucleic acid amplification tests (NAAT) such as polymerase chain reaction (PCR) MTB detection and GeneXpert® MTB/Rif (Cepheid) which are more costly, have been introduced in the past decade, and used to aid the detection of MTB in select samples i.e., complicated and urgent cases involving mostly extra-pulmonary TB, TB meningitis, HIV, and possible

multi drug resistant TB (MDRTB) cases. The Ministry of Health Clinical practice guidelines (MOH, 2012; MOH, 2016; MOH, 2021), and recently the Sabah TB management guideline (Rosli, 2022) have laid out standard algorithm when requesting for these molecular laboratory tests. MDRTB cases using conventional culture-based drug susceptibility testing (DST) usually has a median turnaround time of 2.7 months and 5.7 months (Paul, 2019). This means that it would usually take weeks or months before meaningful action based on laboratory reports could be taken. Molecular tests are expected to provide rapid results for early appropriate treatment which would result in less treatment failures, reduced disease transmission and ultimately, a decrease in disease incidence. Relatively high cost of tests is usually the impeding factor.

Both Indonesia and the Philippines, which have incidence rates of 301 per 100,000 people and 539 per 100,000 people, respectively, are close neighbours to Sabah in East Malaysia and are considered to have a high TB burden. The Kota Kinabalu district of Sabah saw the highest number of TB cases in the entire state. In Sabah, a significant fraction of TB cases, particularly among migrants, are not discovered until the disease has progressed to a more severe stage and the symptoms are more obvious (Rundi, 2010; Rundi et al., 2011; Goroh et al., 2020a; Goroh et al., 2023; Avoi and Liaw, 2021). The advancement of TB care and disease prevention depend on the early diagnosis of TB cases. Therefore, routine surveillance exercises and systematic screening algorithms need to be adjusted to unique populations' features (WHO, 2015).

Continuous improvements in the TB public health programmes, diagnostic facilities and provision of sufficient drug treatment would further bring down the TB incidence (Dye and Williams, 2010). Thus, it is important to focus on the necessary measures to ensure effective TB control in Sabah (MOH, 2002; Liew et al., 2015; MOH, 2016; Goroh et al., 2020b). To date, routine TB case contact investigation and investigation heavily relies on conventional methods. In this approach, as per recommendation by the Malaysian guidelines (MOH, 2002) contact investigation is done utilising symptom screening, tuberculin skin test (TST) with (or without) a chest radiograph. Sputum collection from household members and other close contacts of people with TB which include social and work place contacts (MOH, 2002; Goroh et al., 2020b). According to the Malaysian National TB Program (NTP) at least 10 contacts for every TB patient need to be identified and screened (MOH, 2022). Conventional contact investigation has its merits, but it can be quite challenging especially when dealing with complicated TB cases in addition to high workload. A genotyping-complemented approach to improve and aid TB case surveillance would prove to be worthwhile. As we move closer to embracing precision medicine, knowledge of the molecular aspects of MTB is no longer a luxury. Molecular epidemiology is vital for optimal day to day function in terms of TB case management, surveillance and outbreak control (Portevin et al., 2011; Merker et al., 2015; Brites and Gagneux, 2017; Bainomugisa et al., 2021). Therefore, determining the MTB lineage profile in Sabah is imperative.

This study attempts to aid TB control efforts by determining and understanding the circulating MTB in Sabah, its transmission dynamics and how

it affects the population. It is also the interest of this study to know how circulating strains of MTB in Sabah compares with the MTB lineage profile in peninsular Malaysia, and the surrounding neighbouring countries from previous studies. Circulating strains may differ according to population composition (Reed et al., 2009). In an earlier study by Ngeow et al. (2006), the Beijing genotype was found in high frequencies in Malaysia. Previous studies also mentioned that the Beijing sub-lineage is more often associated with MTB drug resistance (Parwati et al., 2008; Mokrousov et al., 2012). Thus, it is possible that the drug-resistant TB cases in Malaysia may be associated to the Beijing genotype. It is hoped that the outcome of this study will shed some light on the role of MTB genotyping in complementing existing TB case contact investigation, thus enabling optimal TB surveillance and outbreak control, and in doing so help improve the TB prevention and management (TB control) program as a whole.

1.2 Problem statement

1. The circulating genotypes of MTB in the whole state of Sabah and how they affect the population are still largely unknown.
2. It is unclear if MTB genotypes are more prevalent among specific populations.
3. The proportion of recent transmission versus reactivation among TB in Sabah is unknown.

1.3 Hypothesis

1. The circulating strains of MTB in Sabah differ from those in peninsular Malaysia due to a difference in the host population composition.
2. The Beijing sub-lineage is more often associated with MTB drug-resistance.
3. TB genotyping can complement TB control activities in Sabah especially in TB Contact Investigation.

1.4 Study objectives

This study aims to use genotyping to look into the TB evolutionary development in Sabah and to determine how the application of a two-tiered genotyping scheme of TB spoligotyping and 24-loci MIRU-VNTR could complement TB contact investigation that would eventually lead to improved clinical management and public health control of TB in Sabah.

1.4.1 Specific objectives

1. To determine the genetic diversity, population structure and clustering of *Mycobacterium tuberculosis* in Sabah utilising a two-tiered typing scheme with spoligotyping followed by subtyping of spoligotype clusters with MIRU-VNTR typing.
2. To compare genotype complemented TB contact investigation with conventional TB contact investigation.
3. To study the relationship between spoligotype sub-lineages and drug-resistant and relapse cases.
4. To study the relationship between the predominant MTB strains and orphan genotypes.
5. To study the relationship between spoligotype sub-lineages and mixed infection of MTB.

CHAPTER 2

LITERATURE REVIEW

2.1 Tuberculosis (TB)

‘What's in a name? That which we call a rose by any other name would smell as sweet’ (Shakespeare, 2015). Consumption and White plague, Phthisis and Scrofula, Pott’s disease, Koch disease and Schachepheth (Daniel, 2006; Hussain, 2007) are some of the aliases by which tuberculosis (TB), a disease of old is known. It is still considered an important disease with immense public health impact in many countries in the world. Often touted as a disease of the poor, TB is especially prevalent in low-income countries. With the HIV/AIDS epidemic, and migration of people from developing as well as war-torn countries, developed countries in North America and Europe too are not spared (Gagneux, 2018; Conceição et al., 2021; Chakaya et al., 2022; Mokrousov et al., 2023).

TB is thought to have emerged during the paleolithic era (Bloom et al., 1994). It was suggested by archaeological discoveries in Egyptian mummies that spinal caries existed around 2400 BC (Zink et al., 2004), and discovery in legal text the Babylonian king Hammurabi possibly mentioning about TB 3900 years ago are quite fascinating. In China, TB was first mentioned in a medical tome written by Emperor Shennong (2700 BCE) describing weak consumption or xulao bing, which is believed to be TB (Daniel, 2006). The earliest reported case of paleo-pathologically confirmed TB in East Asia is from Hunan province in

China in a female wet cadaver dated to about 200BC, in the western Han dynasty in the kingdom of Changsha (Daniel, 2006). Between the end of the 19th and the beginning of the 20th century, TB was the main cause of death in Europe and America. Many well-known individuals died of TB, such as Frederick Chopin, Niccolò Paganini, the Brontë sisters, Lord Byron, George Orwell, and Eleanor Roosevelt (Bloom et al., 1994; Daniel, 2006).

Robert Koch announced the discovery of MTB on 24 March, 1882, the bacteria that causes TB. He changed the thinking of TB and infectious disease altogether, establishing the science of microbiology (Bloom et al., 1994; Daniel, 2006). Improved acid-fast bacilli (AFB) staining techniques was developed by Ehrlich and later refined by Ziehl-Neelsen. This staining method is still used today. William Wells in 1930s suggested airborne transmission of TB emphasizing on the role of infectious droplet nuclei and undertook studies of aerial spread, beginning the era of TB control (Bloom et al., 1994; Daniel, 2006).

Within the first 2 to 5 years following the initial infection, active TB cases can typically develop. Over the course of a lifetime, there is a cumulative chance of developing active TB. People who have recently been infected with MTB and those with illnesses that have a poor impact on the immune system are often categorized into two categories when it comes to those who are at risk of developing TB disease (Merker et al., 2017). In general, 5 to 10% of infected individuals who do not receive treatment for latent TB infection (LTBI) will eventually contract TB disease. The chance of contracting TB disease is significantly higher in people with weakened immune systems than in healthy

individuals, particularly in those who are HIV-infected. As clinical signs of active TB are often non-specific, diagnosis may be delayed, resulting in prolonged patient illness and infectiousness. Delays in diagnosis and application of control measures can result in further spread of TB, even among individuals thought to be at low risk (Coitinho et al., 2014).

It is quite remarkable that a disease from antiquity is causing an annual global incidence of 10 million cases. Although notifications of TB infections dipped to 5.8 million in 2020 from 7.1 million in 2019 due to the Covid-19 pandemic (WHO, 2020; WHO, 2021), the numbers went back up to an estimated 10.6 million in 2021 (Chakaya et al., 2022; WHO, 2023). However, it is worth to note that this is not a real drop in TB case numbers, but reflective of undercounting primarily due to health service problems with diagnosis and notification. The World Health Assembly had adopted WHO's End TB Strategy in 2014, and aimed for an 80% decrease in TB incidence by 2030, with the specific goal that include ending the global TB epidemic. 450,000 incident cases of rifampicin-resistant TB (RRTB) were reported in 2021. Eastern Europe, Russia and other countries in central Asia were reported as having among the highest prevalence (>50%) of MDRTB or RRTB among individuals who received treatment (Chakaya et al., 2022).

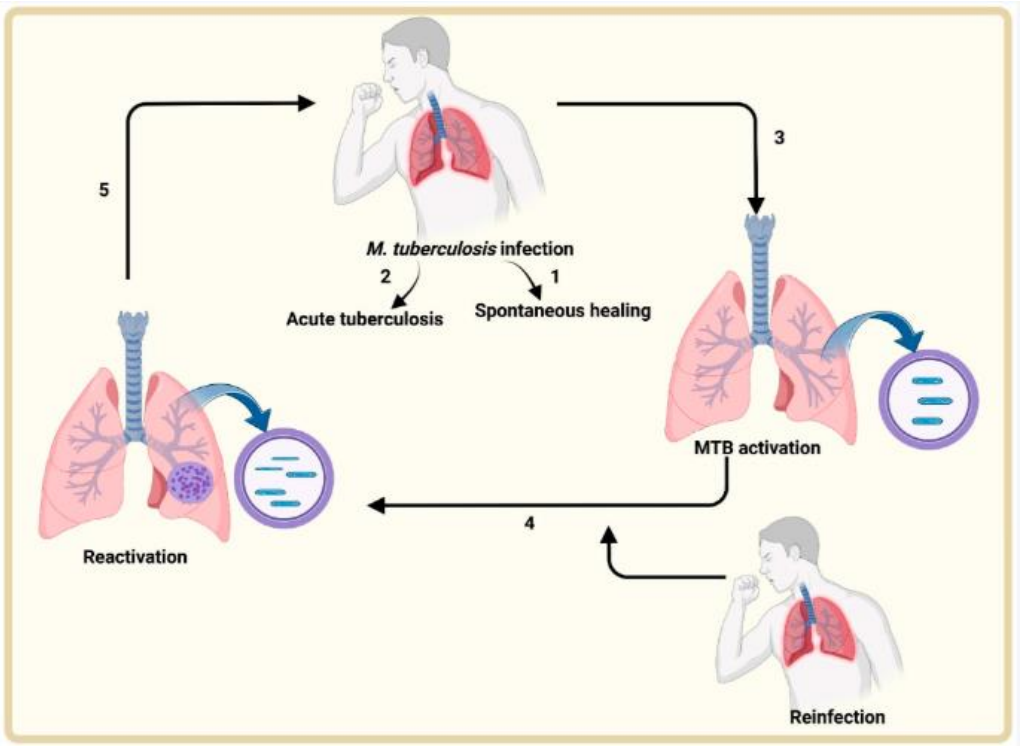


Figure 2.1: Various outcomes of MTB infection (Gautam et al., 2023)

In the TB infection cycle shown in Figure 2.1, when air-borne aerosols encapsulating MTB are inhaled by patients, these MTB will face different outcomes: (1) the pulmonary immune system eliminates MTB; (2) it leads to active TB; (3) the infection progresses to latent TB (LTBI); (4) LTBI is reactivated by an endogenous or exogenous means or both; (5) active MTB infection, progression and transmission (Gautam et al., 2023)

2.2 *Mycobacterium tuberculosis* (MTB)

MTB is classified under the taxonomy as follows:

Kingdom	:	Bacteria
Phylum	:	Actinomycetota
Class	:	Actinomycetia
Order	:	Mycobacteriales
Family	:	Mycobacteriaceae
Genus	:	*Mycobacterium*
Species	:	*M. tuberculosis*

The tubercle bacillus has the remarkable ability to persist in the human host without exhibiting any symptoms. This is known as TB latency. Slow growth, a complex cell envelope, dormancy, intracellular pathogenesis, and genetic homogeneity are some distinguishing characteristics of MTB (Good and Schinnick, 1998; Cirillo et al., 2017; Procop et al., 2020).

The generation time of MTB in growth media or animal hosts is roughly 20 hours. The disease's chronic nature makes therapy difficult and time-consuming (Good and Schinnick, 1998; Cirillo et al., 2017; Procop et al., 2020). By hiding out in an infected tissue, the bacillus continues to exist in a latent stage, causing the metabolic shutdown. The host body becomes more susceptible to the dormant bacteria being reactivated as it ages or as immunity is suppressed. Numerous diseases will start to spread as a result of this. It is yet unknown what molecular processes underlie dormancy and its revival. However, it is believed

that the bacteria are being genetically programmed, whereby the MTB characteristics, and interactions with the host organism are determined by the set of genetic information it possesses (Good and Schinnick, 1998; Cirillo et al., 2017; Procop et al., 2020).

2.2.1 Origin and strain variation of MTB

Genetic diversity is generated within the species through deletion mechanism, duplication and recombination events. Horizontal gene transfer is rare if any in MTB, and this results in clonal evolution of distinct lineages (Nicol and Wilkinson, 2008). This diversity is organized into groups of related strains, families or genotypes that correlate with the geographical origin of the strain (Reed et al., 2009).

Currently, the human-adapted MTB members are classified within 9 phylogenetic lineages (L1-L9): seven major lineages (L1-L7), and two newly discovered lineages (L8 and L9) seemingly restricted to Africa (Ngabonziza, Couvin et al., 2019a; Couvin et al., 2020; Ngabonziza et al., 2020; Coscolla et al., 2021). L2 and L4 are well represented in the taxonomic and phylogeographic evaluations, meanwhile work on L1 and L3 have only been done more recently (Couvin et al., 2019b). Guyeux et al. (2024) recently performed a genome sequencing data analysis from more than 100,000 MTB complex genomes using TB-Annotator software. Through phylogenetic reconstruction, this study

revealed a previously unknown lineage, with the proposed name L10, in central Africa.

There are two nomenclature conventions used for these lineages, based on the method of genotyping and MTB database used (Table 2.1). The well-known genotypes include Beijing, Central-Asian (CAS), East-African Indian (EAI), Haarlem (H), and Latin-American-Mediterranean (LAM) (Gagneux and Small, 2007). The grouping of MTB isolates into the different lineages and sub-lineages based on the key signals of spoligotypes or spoligopatterns are shown in Table 2.2.

Table 2.1: MTB lineage names

Lineage	Genomic-based (WGS)	Spoligotype (SITVIT2)	Distribution
1	Indo-Oceanic or East Asian lineage	East-African-Indian	Predominant in East and Southeast Asia, including China and Vietnam, East Africa
2	East Asian	Beijing	Globally widespread
3	East-African-Indian (EAI)	Central Asian (CAS)	Commonly found in East Africa, the Indian subcontinent, and parts of the Middle East
4	Euro-American	Cameroon, Haarlem (H), Latin-American-Mediterranean (LAM), NEW-1 (formerly named Ural-2), S, T, Turkey, Ural and X	The most widespread lineage, found in Europe, the Americas, and parts of Africa
5	West-Africa 1	AFRI 2 and AFRI 3	Prevalent in Eastern part of West Africa
6	West-Africa 2	AFRI 1	Found primarily in Eastern part of West Africa
7	Ethiopian or Aethiops vetus	Ethiopian	Predominant in Ethiopia and neighbouring regions
8	yet to be named, diverged before hitherto known MTB MRCA	yet to be named (Atypical as yet) presence of spacers 1-5 and 34-36, absence of spacers 6-33, 38-43	Central Africa (recently reported in Rwanda and Uganda)
9	yet to be named, substantial genetic separation with L6 genomes	yet to be named (Atypical as yet) presence of spacer 37-38, 41, 43 missing spacers 13-24,39	Djibouti and Somalia
*10	yet to be named, sister lineage of L6 and L9	yet to be named	Central Africa

Source: Adapted from Couvin et al., (2019a), Couvin et al. (2020), Ngabonziza et al. (2020), Coscolla et al. (2021) and Guyeux et al. (2024). *The asterisk indicate that the lineage name was recently proposed.

Table 2.2: Signature signals of MTB spoligotypes from different lineages / sub-lineages in this study

Lineage	Sub-lineage (SIT)	Absent Spacers	Present Spacers
AFRI	AFRI	8,9,39	40
Beijing	Beijing (1)	1-34	35-43
	Beijing (2610)	1-34, 43	35-42
East-African Indian (EAI)	EAI	34	11-14, 17, 21, 27-28, 33, 38-39
	EAI1-SOM	34, 40	39, 41
	EAI2	3, 20-21, 29-32, 34	2, 4, 7, 33, 35, 38, 43
	EAI2-Manila	20-21, 29, 31-32, 34	2, 4, 19, 22
	EAI3-IND	2, 29-32, 34, 37-38	36
	EAI5	31-32	-
	EAI6-BGD1	23, 29-32, 34	22, 28, 33
Haarlem (H)	H	31, 33-36	32
	H3	31, 33-36	32
Latin American Mediterranean (LAM)	LAM	21-24, 33-36	-
	LAM1	21-24, 33-36	13, 25, 31
	LAM2 (17)	21-24, 33-36	25
	LAM6	21-24, 29, 33-36	28, 30-32
	LAM8 (290)	21-24, 27, 33-36	1, 25-26, 28-31, 42
	LAM9	21-24, 33-36	3, 13, 40
MANU	MANU_Ancestor	-	1-43
	MANU	16-17, 19-37	1-15,18
	MANU1	34	1, 33, 35
	MANU3	34-36	33, 37
T	T	33-36	31
	T1	33-36	31
	T3	13, 33-36	12, 14, 31
	T4-CEU1	19, 23-24, 38-39	1-3, 17-18, 20-22, 25,30-32, 37, 40-43
	T-H37Rv	20-21, 33-36	6, 8, 10, 12, 16-17, 19, 22-23, 25-26, 31-32, 37, 42-43
Turkey	Turkey (Orphan)	20-24, 26-27, 33-36	25, 28, 31
Unknown (Est.)	Unknown	-	-
	(SIT32)	9-38	1-8, 39-43
	(SIT821)	3, 8-38	1-2, 4-7, 39-43
	(SIT1196)	3-35	1-32, 36-43
	(SIT1509)	3, 20-21, 25, 27-43	1-2, 4-19, 22-24, 26
X	X	18, 33-36	19, 25, 31
	X1	18, 33-36	19, 25, 31
	X3	4-12, 18, 33-36	1, 3, 16-17, 19, 24-25, 31

Source: Adapted from Couvin et al. (2020)

Phylogeographical analyses on global MTB strains (Hershberg et al., 2008; Comas et al., 2013; Brites and Gagneux, 2017) have revealed the African continent as being the most likely origin of the MTB as described by the Out-of-Africa theory (Hershberg et al., 2008; Comas et al., 2013). Genetic diversity of the MTB is highest in Africa hence the discovery of many different lineages. The number of lineages decreases with increasing distance from Africa (Comas et al. 2013, Gagneux 2017). However, when we look at the variation within the lineages, similar to the concept of the mitochondrial (mtDNA) sequence variations, the greater the sequence difference compared to the original, the more recent the species (Koleske et al., 2023). Strong phylogeographic structure of human-adapted MTB lineages have led to the hypothesis that particular MTB variants might be locally adapted, with certain degree of affinity towards specific human populations (Coscolla and Gagneux, 2010; Gagneux, 2012). Local adaptation may have taken place whereby given pathogen phenotype has a higher fitness in a host population with which it has co-evolved (sympatric association) than in a different host population (allopatric association) (Gagneux, 2012; Gagneux, 2018). This is exhibited in the different distribution patterns of the MTB strains all over the world.

Among the genetic strains discovered within the MTB, the Beijing family (L2) is possibly the most studied MTB lineage. The strains were isolated and identified in Beijing, China hence the adoption of the name of the city. The signature spoligopattern is a marker that defines the Beijing genotype and distinguishes it from other families within MTB. These strains are endemically prevalent in Southeast Asia, the area of origin and primary dispersal, as well as

in northern Eurasia and South Africa, areas of their secondary dispersal (Mokrousov et al., 2005; Parwati et al., 2008; Mestre et al., 2011). The Beijing strains possess important pathogenic features that are rapidly being disseminated globally, and is thought to be greatly responsible for current global epidemic of the drug-resistant MTB (Mokrousov**,** 2012). Several reports associate Beijing with increased possibility of MDRTB. It may be possible that the global distribution of MDRTB hotspots is in fact driven by local predominance of the Beijing strain (Borrell and Trauner, 2017). Beijing lineage is predominant in Asia, South East Asia and Eastern Europe and is strongly associated with DRTB. However, the exact biological basis is poorly understood.

The genomic Euro-American Lineage (L4) is the most widespread lineage with the most spoligotype-based lineages / sub-lineages such as Haarlem (H), Latin-American-Mediterranean (LAM), T, Turkey and X (Table 2.1). The genotype Haarlem family (H) was designated as such because the first recognized strain was isolated from a patient living in Haarlem, The Netherlands (Kremer et al., 1999; Rastogi and Sola, 2007). The spoligotype is characterized by the absence of the spacer 31, undetectable due to the presence of a second copy of IS6110 which is asymmetrically inserted within the DR locus (Groenen et al., 1993; Filliol et al., 2000; Legrand et al., 2001). The H family is highly prevalent in Northern Europe, Caribbean to a lesser extent and is also prevalent in Central Africa, where it is believed to have been introduced during the European colonization (Filliol et al., 2003).

The Latin-American-Mediterranean (LAM) family is said to have originated in the Western Mediterranean region (Rastogi and Sola, 2007; Mokrousov et al., 2016). Rastogi and Sola (2007) highlighted that paleo-pathological and ancient DNA data support the existence of TB before the arrival of Spanish settlers to Latin America in the 15th century (Arriaza et al., 1995). It is supposedly linked to the Lusitanian-Hispanian colonization of the New World. Conversely, it may have been endemic in Africa and/or in South America, spreading to Europe later (Rastogi and Sola, 2007). The LAM family was defined by the finding of linkage disequilibrium between the absence of spacers 21-24 in the spoligotyping and the presence of an ETR-A allele equal to 2 (Sola et al., 2001). However, this genotype family is more diverse and its study is more complicated than initially thought (Rastogi and Sola, 2007; Mokrousov et al., 2016).

The MANU family was first identified in India (Gutierrez et al., 2006) is included in the ancestral lineage of MTB, whereby the TbD1 MTB-specific region is still intact. It is closely related to the EAI lineage, as they are believed to have been derived from a last common ancestor (Thomas et al., 2011). It is tentatively subdivided into MANU1 (ST100; loss of spacer 34), MANU2 (ST 54; loss of spacers 33, 34) and MANU3 (ST1378; deletion of spacers 34–36) (Thomas et al., 2011).

The Turkey genotype or formerly LAM7-TUR is mainly distributed in areas surrounding Turkey in Western Asia particularly in Turkey, East and Austral Africa, South and North America, as well as the rest of Europe (Western,

Eastern and Northern), but not in Central America, and only sporadically in the Caribbean and the rest of Africa and Asia (Zozio et al., 2005; Couvin et al., 2019a) This distribution may underline a cycle of displacement of populations animated by successive waves of exploration, trade, or colonization (Zozio et al., 2005; Brudey et al., 2006; Couvin et al., 2019a). This genotype has missing spacers 20–24, 26–27 and 33–36, and is more likely to be drug-resistant (Couvin et al., 2019a).

The MTB T family is an 'ill-defined' genetic family which is found globally, and correspond to about 30% of all entries in the database. It includes strains with missing spacers 33-36, and cannot be classified in other groups. Undoubtedly, WGS data will improve the knowledge of the identity of these isolates designated as 'T' by default (Sebban et al., 2002; Filliol et al., 2003; Brudey et al., 2006; Mokrousov, 2021).

The key features of the X family are: i. a low number of IS6110 copies, and ii. the absence of spacer 18 in the spoligotyping (Sebban et al., 2002). One of the first fully sequenced MTB genome, the CDC1551, has both of these characteristics. The X family was also the first group identified in the French Polynesia (Torrea et al., 1995) and Guadeloupe (Sola et al., 1997). The distribution of this family also appears to be linked to Anglo-Saxon countries (Dale et al., 2003). It is also highly prevalent in South Africa and to a lesser extent in the Caribbean, and poorly documented in India. The X family are also found in high numbers in the US and Mexico (Rastogi and Sola, 2007).

In addition to host and environmental determinants, the genetic diversity of the pathogen may also play a role in the outcome of TB infection and disease (Nicol and Wilkinson 2008; Comas and Gagneux, 2009; Coscolla and Gagneux, 2010). Click et al. (2012) suggested that different MTB lineages may have different preference of anatomical sites. Earlier studies suggested that different lineages rendered different disease outcomes, with the modern lineages e.g., Beijing (L2) and LAM (L4) exhibiting greater virulence and success in spreading the disease on a global scale (Brites and Gagneux, 2017). The general signature signals of MTB spoligotypes are shown in Table 2.2. Thus, MTB lineage identification and recognition is important to guide patient treatment and management as different MTB lineages may have different behaviour and treatment response. Lineage identification is also important to enable a deeper understanding of MTB transmission pattern and dynamics especially in TB prevention and management activities.

The virulence of MTB is caused by the accumulation of different clinical phenotypes transmission rate and disease severity. Clinical phenotypes are influenced by cellular and immunological phenotypes while MTB phenotypes are determined by the genotype. Therefore, finding the genotypes responsible for clinical phenotypes would allow discovering MTB virulence factors. Defining the impact of specific bacterial genomic loci on virulence when other bacterial determinants, human and environmental factors are also impacting the phenotype would contribute to a better knowledge of TB virulence and ultimately benefit TB control (Krishnan et al., 2011; Reiling et al., 2013; Broset, 2015).

Recent improvements in the sequencing capabilities of laboratories in high-TB-burden environments have sped up studies on the origins of L1 and L3. This contains an analysis of how migration and dispersal influence the prevalence of MTB in various regions of the world (Freschi et al., 2021). Couvin, et al. (2019b) reported that on the global scale the ancestral strains of MTB seem to be roughly divided into regions that are dominated by EAI or Indo-Oceanic, EAI (L1) lineages versus Central Asian, CAS (L3). The Indian Ocean rim which is the region that accounts for the majority of new TB cases worldwide, is characterized by the prevalence of L1 and L3 (Couvin et al., 2019b).

2.2.1.1 EAI - EAI2-Manila genotype of MTB

The EAI lineage (L1) is prevalent in many South East Asia, some parts of South America, Mongolia Eastern-Coast of Africa, South-India and Southeast Asia (Douglas et al., 2003; Parwati et al., 2008; Chen et al., 2017; Conceição et al., 2017; Chihota et al., 2018; Couvin et al., 2019a)**.** Couvin et al. (2019b) highlighted that the EAI and CAS lineages of MTB mainly infect TB patients in the eastern hemisphere which contains many of the high TB burden countries which include China and India. Information on the distribution of EAI worldwide could enable better understanding of the prevailing TB epidemic in the context of its historical spread and evolutionary features, as well as providing guidance to better treatment and patient-care in countries and regions where these lineages are prevalent. EAI genotypes are prevalent in South-East Asia, India, and East Africa, with low number of IS6110 copies. Beijing strains of

MTB on the other hand have a large number of IS6110 copies. This high copy numbers suggests its potential implication to increased virulence and capacity for rapid dissemination. IS6110 elements presence and distribution in the MTB genome varies between the different strains, contributing to the genetic diversity among different MTB isolates (Shanmugam et al., 2011; Alonso et al., 2013; Koleske et al., 2023). This difference can be utilized in molecular epidemiology studies for genotyping MTB and understanding transmission patterns of TB. IS6110 is often located in the non-coding regions such as the intergenic regions of the MTB genome. It can also be found inserted within the genes themselves. When this happens, gene function may potentially be disrupted. The number of copies of IS6110 varies among different strains (Shanmugam et al., 2011; Alonso et al., 2013; Koleske et al., 2023). A subgroup of these strains harbouring a single copy of IS6110 was shown to be widespread in Malaysia, Tanzania, and Oman (Dale et al., 1999; Rastogi and Sola, 2007). EAI spoligotypes are shown by absence of spacers 29-32, presence of spacer 33, absence of spacer 34, exact tandem repeat A (ETR-A) allele ≥4 (Rastogi and Sola, 2007). MTB lineage 1 (L1) contributes considerably to the disease morbidity.

Douglas et al. (2003) designated the Manila clade which is also known as EAI2-Manila. of MTB based on their work on 48 MTB strains obtained from TB patients in Manila city, Philippines. Utilising three classical genotyping methods, Douglas et al. (2003) compared the DNA profile against international databases and confirmed the novelty of these isolates. The signature spoligotype of the absence of 8 spacers in 41isolates, and additional 3 or 4 deletions to the right of spacer 24 in the other 3 strains deletions were adequate to classify them

in a new family. In a WGS analysis of 178 isolates from the Philippines National Drug Resistance Survey, Phelan et al. (2019) reported that the majority (80.3%) belonged to the lineage 1 Manila clade, with the minority belonging to lineages 4 (European-American; n = 33) and 2 (East Asian; n = 2). Sixty-eight mutations known to be associated with 10 anti-TB drug resistance were identified in the Philippine strains, and all have been observed in other populations.

2.2.2 MTB drug resistance

Emergence of drug-resistant TB especially MDRTB and XDRTB is a major public health problem, revealing weaknesses in TB management and diagnostic services. Reliable drug susceptibility testing and ability to rapidly detect drug-resistance, especially MDRTB is imperative for effective patient management and TB control (Kam and Yip, 2001; Rashidi Ali et al., 2015).

Drug resistance is caused by mutations in chromosomal genes, primarily the accumulation of point mutations and indels in genes coding for drug targets or drug converting enzymes (Borrell and Gagneux, 2011; Balaban, 2011; Levin-Reisman et al., 2019). MTB relies heavily on two types of resistance mechanisms namely target modification e.g., loss of binding affinity, and loss of prodrug activation. Drug efflux and transcriptional modulation also occur as resistance mechanism but at much lower frequencies as shown by low contributions to resistance. Efflux pumps may be an important aspect to consider for nonspecific resistance in the future (Borrell and Trauner, 2017).

MTB drug-resistant strains develop mainly by selecting genetic mutants resulting from suboptimal treatment and poor patient adherence to the treatment regime (Kam and Yip, 2001; Zhang and Yew, 2015). Recent evidence shows that pharmacokinetic-pharmacodynamic variability scenarios related to the induction of the mycobacterial drug efflux pump may also facilitate the development of genetic mutations in MTB. The development of drug resistance as a result of mutations in drug resistance genes in MTB may incur a cost in terms of fitness and virulence of the organism. Acquired resistance can also be compounded by transmitted resistance (Borrell and Gagneux, 2011; Coscolla and Gagneux, 2014; Zhang and Yew, 2015).

Both *in vitro* and *in vivo* tests may be used to confirm the MTB's fitness. The frequency of these mutants in the population can also be used to gauge fitness. To find out if a strain's transmissibility has been impacted by the development of drug resistance, its distribution in a population can be compared to that of susceptible strains. Another indicator of fitness is the size of the clusters (van Doorn and Dieckmann, 2006). It can be assumed that the strains are successful and have not lost much ground if specific mutations that give resistance are common in the population. Large clusters, however, are not usually a sign of improved fitness in the MTB strain. In countries with high TB/MDRTB burdens, accurate estimations of fitness can be especially important (Knight et al., 2018).

Drug resistance in MTB is either acquired or primary. When a patient with a DRTB is diagnosed but has never had anti-TB therapy, primary resistance

occurs. The spread of drug-resistant MTB isolates from one patient to another causes the acquisition of primary resistance. Therefore, the amount of primary resistance in a community serves as a proxy for community transmission (Yin et al., 2016). Ineffective anti-TB medication can emerge from a variety of man-made reasons, including poor patient adherence to treatment, unsuitable prescriptions, and faulty medical care. This is known as acquired resistance. The degree of acquired resistance is an indicator of how well the ongoing TB control effort is working (Cohn et al., 1997; Dean et al., 2017). However, the two proxies "resistance among new cases" and "resistance among previously treated cases" have now been adopted to replace the terms acquired and primary resistance respectively (Espinal et al., 2001).

2.2.2.1 MTB isoniazid mono-resistance

Isoniazid (INH) was introduced in 1952 as an anti-TB agent. It is the backbone of TB treatment alongside rifampicin. INH is also known as Isonicotinic acid hydrazide. It is a pro-drug that needs to be activated by the catalase/peroxidase enzyme KatG, encoded by the *katG* gene, to exert its effect. It is capable of attacking multiple targets in MTB, but the primary one is by inhibiting the synthesis of mycolic acids, thereby inhibiting cell wall production in MTB. Although simple in its structure, resistance to this drug has been associated with mutations in several genes, such as *ahpC, inhA, kasA,* and *katG*.

The two main molecular mechanisms of isoniazid resistance are associated with gene mutations in *katG* and *inhA*. Numerous studies have found mutations in these two genes as the most commonly associated with isoniazid resistance. High level isoniazid resistance is usually associated with *katG* gene mutation; while low level isoniazid resistance is usually associated with *inhA* promoter gene mutation (Brossier et al., 2006; Gagneux, 2009). Among these, the most prevalent gene mutation has been identified as S315T in *katG* (serine to threonine at codon 315) accounting for 50–95% of isoniazid resistant (INH-R) clinical isolates, resulting in an isoniazid product deficient in forming the isoniazid-NAD adduct needed to exert its antimicrobial activity. This however, differs with geographic location (Musser et al., 1996). This mutation has been consistently associated with high-level resistance (MIC > 1 µg/mL) to isoniazid and occurs more frequently in MDR strains (Palomino and Martin, 2014; Zhang and Yew, 2015; Borrell and Trauner, 2017). *katG* S315 mutations usually do not completely eliminate catalase activity, and such strains may still retain fitness and virulence, which may explain its frequent occurrence among clinical isolates (Ramaswamy and Musser, 1998; Slayden and Barry, 2000).

The second most common mutation occurs in the promoter region of *inhA*. The most prevalent mutation which is found at position −15C/T causes the *inhA* to overexpress, and the increased amount of the enoyl-ACP reductase enzyme neutralizes the effects of INH. This mutation (-15C/T) is also usually linked to low-level resistance to INH. Mutations in *inhA* do not only cause resistance to INH but also affects the drug ethionamide (Ramaswamy and Musser, 1998; Palomino and Martin, 2014; Zhang and Yew, 2015; Borrell and

Trauner, 2017). A recent study found that a mutation in the *inhA* regulatory region together with a mutation in the *inhA* coding region produced high-level isoniazid resistance and also cross-resistance to ethionamide (Borrell and Trauner, 2017).

2.2.2.2 MTB rifampicin mono-resistance

Rifampicin (RIF) is a rifamycin derivative, bactericidal drug introduced in 1972 as an anti-tuberculosis (anti-TB) agent. Together with INH, it constitutes the basis of the first line multidrug treatment regimen for TB. RIF is active against growing and slow metabolizing bacilli. The mode of action of RIF in MTB is by binding to the β-subunit of the RNA polymerase, inhibiting the elongation of messenger RNA. The majority of RIF resistant (RRTB) clinical isolates of MTB harbour mutations in the *rpoB* gene that codes for the β-subunit of the RNA polymerase. As a result of this, conformational changes occur, decreasing the affinity for the drug and results in the development of resistance (Palomino and Martin, 2014; Zhang and Yew, 2015) .

In about 96% of MTB isolates resistant to rifampicin, there are mutations in the hot-spot 81-bp region of *rpoB* gene, spanning codons 507–533, encoding 27 amino acids, also known as the rifampicin resistance-determining region, RRDR (Ramaswamy and Musser, 1998; Palomino and Martin, 2014). Mutations in codons 516, 526 and 531 are the most commonly associated mutations with rifampicin resistance in the majority of studies. Some reports noted in lower

frequencies the occurrence of mutations outside of the hot-spot region of *rpoB* (Palomino and Martin, 2014). Mutations in some codons (e.g., 518 or 529) have been associated with low-level resistance to RIF but still susceptible to other rifamycins, such as rifabutin or rifalazil. This is especially important for TB patients on antiretroviral therapy since rifabutin is a less effective inducer of the cytochrome P450 CYP3A oxidative enzyme (Palomino and Martin, 2014). However, not all mutations in *rpoB* are associated with RIF resistance. For example, mutations at E510H, L511P, D516Y, N518D, H526N and L533P are not associated with RIF resistance and are found in RIF-susceptible strains (Cirillo et al., 2017). Monoresistance to RIF is rare and almost all RRTB strains are also resistant to other drugs, especially to INH. Thus, RIF resistance is considered as a surrogate marker for MDRTB (Brandis et al., 2012). RRTB has been well characterized to be the result of point mutations or small deletions in the RNA polymerase gene (*rpoB*) (Ramaswamy and Musser, 1998). RRTB were reported to occur at much higher frequencies for the Beijing genotype strain compared to the EAI genotype strain (Cirillo et al., 2017).

2.2.2.3 Multidrug-resistant TB (MDRTB)

Multidrug resistance in MTB (MDRTB) is defined as resistance to at least the combination rifampicin (RIF) and isoniazid (INH), caused by the sequential accumulation of mutations in the genes encoding the targets of RIF and INH (Kam and Yip, 2001; Sougakoff, 2011). Extensively drug resistant TB (XDRTB) is defined as ‘MDRTB/RRTB that is resistant to a fluoroquinolone

(levofloxacin or moxifloxacin) and at least one additional Group A drug which is either bedaquiline or linezolid, or both'. The Group A drugs are currently levofloxacin or moxifloxacin, bedaquiline and linezolid (WHO, 2021). The emergence of MDRTB and XDRTB is a major public health problem that threatens TB care and control in many countries. This reveals weaknesses in TB management and diagnostic services. First line TB treatment is very effective, but it cannot be said so for the second line treatment. Drug resistance among retreatment cases is believed to reflect accumulated treatment errors. The rate of resistance to drug combinations is also obviously higher in retreatment MDRTB cases, than in new cases (Borrell and Trauner, 2017; Cirillo et al., 2017).

DRTB and MDRTB could also be caused by primary drug resistance as opposed to acquired drug resistance. In some parts of the world, high rates of MDRTB have been observed not only among previously treated patients, due to poor case management, but also among new cases, due to transmission in the community (Kam and Yip, 2001). In many countries, widespread use of the standard short-course chemotherapy has led to an increasing incidence of RRTB. This created a reservoir of MDRTB that could be disastrous in a densely populated city. Strong laboratory support for the early detection of resistance and to provide guidance in the choice of second-line anti-TB drugs for the treatment of MDRTB, is essential for an effective TB control program. Ng et al. (2013) reported the draft genome of the first XDRTB strain isolated in Malaysia and examined the gene polymorphism related to its extensive resistance, to provide a better understanding of drug resistance patterns and mechanisms in local strains of MTB.

2.2.2.4 Fitness Cost

The capacity of a bacteria to survive, reproduce, and propagate among susceptible hosts may be used to gauge its level of success (Shorten, 2011; Knight et al., 2018; Gagneux et al., 2013). It needs to be assumed that the rate of division is crucial to the success of the organisms in order to do this. The time taken for a bacterium to divide, or generation time, may therefore be taken as a measure of fitness (Shorten, 2011; Knight et al., 2018; Gagneux et al., 2013). The size of clusters can be used to estimate fitness as well (Devaux et al., 2009; Gagneux, 2009; Borrell and Gagneux, 2011; Shorten, 2011). MTB's fitness and drug resistance may also be influenced by the phylogeographic background of the bacterium (Devaux et al., 2009; Gagneux, 2009; Borrell and Gagneux, 2011; Shorten, 2011).

Resistance mutations inflict different levels of fitness cost in absence of the drug, with some showing mild or no detrimental effects. In MTB, such "low-cost mutations" also account for the majority of the resistance detected in clinical isolates. Thus, strains that harbour low-cost mutations are positively selected among DRTB strains either within the bacterial population infecting a host and/or as a result of variable transmission efficiency. High-cost mutations on the other hand, refers to gene mutations that result in significant detrimental effects to MTB often affecting its ability to compete with other bacteria or survive in certain environments. In addition, there may also be compensatory mutations

that are able to offset damaging effects caused by the high-cost mutation. An example of this is the findings by De Vos et al. (2013) whereby the putative compensatory mutations in the genes rpoA and rpoC (encoding the α and β' subunits of RNA polymerase) were able to alleviate the fitness cost incurred by rifampicin resistance-conferring mutations by *rpoB* in the RRTB strain. These mutations have also been associated with higher transmissibility in some settings (Brandis et al., 2012; Palomino and Martin, 2014). Compensatory evolution then seems to be a key driver ensuring that fitness is restored in strains with multiple costly DRTB mutations (Comas et al., 2012; De Vos et al., 2013).

2.2.3 TB relapse and reinfection

A relapse case occurs when a person who has completed TB treatment, and declared cured of TB subsequently develops active TB again due to the reactivation of the same MTB strain that caused the initial infection. The original strain of bacteria that caused the initial infection remained inactive or dormant in the body, usually in the lungs (Huyen et al., 2013; McIvor et al., 2017; Shen et al., 2017). When conditions are favourable, such as when the host immune system weakens, MTB becomes active again. This usually happens within a relatively short period after treatment completion, averagely within one to two years, although it can occur later (Huyen et al., 2013; McIvor et al., 2017; Shen et al., 2017).

TB reinfection on the other hand, occurs when a person who has previously undergone TB treatment and declared cured becomes infected with a new strain of MTB that is different from the original strain that caused the initial infection (McIvor et al., 2017; Shen et al., 2017). In a reinfection case, the patient's immune system successfully cleared the first infection, but is then subsequently exposed to MTB again, leading to a new infection with a different strain of the bacteria (McIvor et al., 2017; Shen et al., 2017).

Recurrent TB may be brought on by a recent reinfection or by the reactivation of an endogenous initial infection (relapse) (Van Rie et al., 2005). Therefore, genotyping by fingerprinting MTB isolates of the original and recurring episodes, acts as a definitive way to distinguish these two occurrences from one another. The recurrent event is regarded as a reactivation if the paired isolates from the primary and recurrent episodes of one patient have identical DNA fingerprints (or nearly identical, with one or two band differences) (typically MIRU-VNTR); otherwise, if the paired isolates have different DNA fingerprints, the recurrent event is regarded as reinfection. Reinfection is now believed to be a factor in TB recurrence even in regions with low incidence rates (Bandera et al., 2001; Garcia de Viedma et al., 2002). For the purpose of calculating the treatment failure rate and transmission level, a precise distinction between reinfection and reactivation must be made.

2.2.4 Mixed infection of MTB

Mixed infections of TB, refers to a state where an individual is infected with multiple strains of MTB. The introduction of molecular techniques enables the extensive study of TB infection (Huyen et al., 2012). In this study, mixed infection of MTB is defined as presence of ≥2 alleles in the MIRU-VNTR loci (de Vries et al., 2009). Although the relevance of mixed infections for the patient and for TB control is not completely clear, understanding these factors is essential for guiding TB prevention and management efforts and developing effective strategies for prevention, diagnosis, and treatment (Stavrum et al., 2009).

2.3 Genotyping and molecular epidemiology of MTB

TB is an infectious disease with complex epidemiology. Understanding the biology and genomics of MTB may be the needed approach as to how the enormous disease burden can be tackled. Traditional epidemiology is able to identify cases of TB that are related in time and space, but this is often insufficient.

Genotyping MTB enables scientists to study evolutionary relationships in addition to the routes of transmission of the organism in between hosts. It could effectively guide outbreak investigation, define transmission dynamics, determine the re-infection or reactivation status of cases provided if universal

genotyping was done, confirm mixed infections and aid disease surveillance at a global scale through the clustering analyses of MTB isolates. The combination of MTB genotypes and epidemiological data available in the database (Niemann et al., 2016; Merker et al., 2017; Meehan et al., 2018) could allow research on current trends of disease within a given population, help identify persons with TB disease involved in the same chain of recent transmission, and to distinguish between persons whose TB disease is the result of past infection, or newly acquired infection (CDC 2006).

Large-scale genotyping is also needed to get better insights into the biological and geographical diversity of the pathogen and its association with host populations, understanding the origin of MTB, and retracing the evolutionary history and the demography of the MTB, and some of its major clones in association with MDRTB resistance (Gagneux et al., 2006). MTB genotyping has been extensively used for investigating clusters of MTB, MDRTB and XDRTB infections, in order to identify factors involved in the transmission of these strains. The knowledge gained may be useful for the formulation of effective control programmes to limit the spread of MTB, especially resistant strains.

Genotyping is important in the clinical and public health setting (Shorten, 2011). TB programs will be able to evaluate completeness of routine contact investigations and progress toward TB elimination by monitoring surrogate measures of recent TB transmission. Prospective, real-time genotyping of MTB isolates can therefore benefit both the laboratory, and the clinical teams. The

real-time investigation of suspected and known outbreaks can be used to rationally guide contact investigation and early detection and intervention, especially in the case of drug resistant strains (Allix-Béguec et al., 2008; Niemann et al., 2016; Meehan et al., 2018).

2.3.1 Evolution of genotyping techniques for MTB

Within two decades, significant efforts have been made evidenced by developments of rapid molecular diagnostics methods in order to understand strain-level genetic diversity in MTB and its geographic distribution (Niemann et al., 2016; Merker et al., 2017). However, these methods have been slow to be implemented worldwide, notably in Asia. A robust classification of MTB strains into evolutionarily meaningful sub-lineages is important for taxonomic purposes and because sub-lineages can differ in virulence or antibiotic resistance (Borrell and Gagneux, 2011; Brites and Gagneux, 2015; Cirillo et al., 2017).

Molecular epidemiology is now an essential tool in determining transmission patterns of MTB and other etiologic agents. Identification of MTB strains utilising genotyping methods enables the clustering of isolates which can be used to complement traditional contact investigation and allows for better understanding of transmission dynamics. Although whole genome sequencing (WGS) is fast becoming the tool of choice for most researchers, basic genotyping such as spoligotyping is still a smart choice due to its ease of handling, lower cost and straightforward result interpretation.

Genotyping methods used for other organisms such as multi-locus sequence typing (MLST) and pulse-field gel electrophoresis (PFGE) could not be employed for MTB as they require substantial chromosomal variation and diversity of housekeeping genes for molecular discrimination at the strain level to be able to work. Thus, development of alternative genotyping method for MTB had to be developed (Table 2.3).

Based on two neutral single nucleotide substitutions (SNSs) in the antibiotic resistance genes *katG* (codon 463) and gyrA (codon 95), MTB was originally divided into three major genetic groupings in 1997. Since then, numerous researches have tried to classify data at a better resolution by employing massive genomic deletions, spoligotyping, and SNSs (Merker et al., 2015; Niemann et al., 2016; Merker et al., 2017). Single nucleotide polymorphisms (SNP) were used for more phylogenetic application especially for low level genetic diversity and high degree of clonality. Sreevatsan et al. (1997) reported that the rate of mutation in MTB is 1 SNP per 2000bp. Major impediment of these methods was that they could only be used for samples in retrospect, to guide investigations. It is slow and labour-intensive.

A number of genotyping technologies have been developed throughout the years to study these genetic differences based on genomic deletions: such as detection of Region of Difference (RD) and Large Sequence polymorphisms (LSP). The discovery of CRISPR has led to the establishment of Spoligotyping. Multi Locus VNTR Analysis (MLVA), Multi Locus Sequence Typing (MLST) and Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem

Repeat (MIRU-VNTR) were developed based on the presence of minisatellites throughout the genome. These techniques have enabled more advanced studies in prokaryotes especially in MTB. These tests are commonly used for ease of handling, lower cost and straightforward result interpretation. Short turnaround time, simple nomenclature system, MIRU-VNTR and spoligotyping replaced IS6110 DNA fingerprinting as the classical genotyping gold standard. For a method to work well in TB control, it needs to work fast and give fast information for action to be taken.

Sougakoff (2011) and Hatherell et al. (2016) studied the molecular epidemiology of MTB in the context of MDRTB/XDRTB spread and general MTB transmission respectively. They compared WGS with classical genotyping namely RFLP analysis of IS6110, spoligotyping and MIRU-VNTR typing of MTB isolates as well as phylogeographical analysis of strains. They found WGS to be useful for the confirmation of MTB transmission, as well as the study of TB recurrence, 'within host diversity' and drug resistance. They also showed that MTB genomic diversity plays a significant role in the inference of transmission and that WGS and conventional typing are able to identify superinfections, implying limited cross immunity between strains of MTB.

With the continuous progress and decreasing costs of next-generation sequencing (NGS) technologies, typing based on WGS is now increasingly performed for near complete exploitation of the available genetic information. Standardization of WGS analysis pipeline and database is being refined, for sharing at the global level. Classical genotyping tests are commonly used for

ease of handling, lower cost and straightforward result interpretation. However, WGS is the ultimate test for the most comprehensive genetic information. Selection of methods to be used depends on the research question and its level of complexity.

Table 2.3: Genotyping techniques and their limitations

NO	TYPING METHODS	PRINCIPLE	ADVANTAGES	DISADVANTAGES
1	Large sequence polymorphism (LSP)	Based on the presence or absence of specific segments of DNA	Robust marker for phylogenetic classification	Not useful to track specific strains in the community
2	Restriction fragment length polymorphism (RFLP-IS6110)	Based on the variability in the number of copies of IS6110 and the molecular weights of DNA fragments in which the insertions are found	Number of copies and positions in genome may vary from isolate to isolate, high discriminatory power, widely used	Limited discriminatory power in isolates with ≥IS6110 bands, lengthy process, reproducibility problems in isolates representing large amounts of IS6110 copies difficult to compare results between laboratories
3	Polymorphic GC-rich repetitive sequence (PGRS-RFLP)	Based on the variability in the number and location of the PGRS regions	Higher discrimination power in isolates with six or l ess IS6110 copies	Limited data using this technique, requiring large amounts of high-quality DNA, lower discrimination power in isolates with multiple copies, lengthy process, difficult to analyse, inter-lab comparison not possible
4	Spacer oligonucleotide typing (Spoligotyping)	Based on polymorphism in the direct repeat locus which is a member of the CRISPRs	Typing isolates with less than six IS6110 and it has shown a 98% identification specificity and 96% sensitivity in clinical samples, PCR-based, data in an exchangeable format, highly reproducible	It shows a 63% sensitivity and 49% specificity as compared with the standard RFLP-IS6110 method, limited discriminatory power
5	Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR)	Based on polymorphisms of MIRU loci Set of 15 loci used for molecular epidemiology studies; Set of 24 for phylogenetic studies	Results in a code system, data in an exchangeable format, useful for identification of polyclonal infecti ons (similar discrimination power to RFLP IS6110) more specific than spoligotyping when IS6110 copies is <six; High discriminatory power.	65% specificity and a 35% sensitivity as compared with the RFLP-IS6110 in molecular studies; Set of 12 loci less discriminatory than IS6110 RFLP. Electrophoresis method, determination of band size is less reproducible than the high-cost sequencer-based method
6	Single Nucleotide Polymorphism (SNP)	Based on detection of specific nucleotide changes, which can provide valuable information about MTB	Allows multiple polymorphisms analysis in a short time, rendering results in a binary format, and easy storage and inter-lab comparison; high resolution and discriminatory power	Similar results to RFLP-IS6110 and spoligotyping, limited discriminatory power compared to MIRU-VNTR and WGS. Expensive, need specialized technology and software
7	Whole Genome Sequencing (WGS)	Based on the analysis of the whole-genome sequence results	Gold standard for phylogenetic classification	Expensive, need for specialized technology and software

Source: Adapted from Kanduma et al. (2003), Rozo-Anaya et al. (2010), Kato-Maeda et al. (2011)

Spoligotyping and 24-loci MIRU-VNTR were selected as the two-tiered genotyping method in this study to complement TB prevention and management in Sabah because they are relatively cheaper than other molecular testing, and relatively simple to perform in the laboratory especially for high sample numbers, as well as having reasonably good discriminatory power (Allix-Béguec et al., 2008; Weniger et al., 2010; Gomgnimbou et al., 2013).

2.3.2 Spacer oligonucleotide typing (spoligotyping)

The clustered regularly interspaced short palindromic repeats (CRISPR) loci genetic family is found in around 40% of Eubacteria and in all Archaea. CRISPR loci are believed to have a functional role against bacteriophages. These loci also contain remnants and 'memories' of previous encounters with bacteriophages at the evolutionary genetics level. The polymorphisms in the CRISPR loci makes them good epidemiological markers (Zhang et al., 2010).

In MTB, the CRISPR loci are repeated patterns composed of succession of conserved sequences called Direct Repeats (DR) constituting 36 nucleotides interspersed by variable individual sequences single "spacer" sequences ranging from 25 to 45 nucleotide. The DR locus spans up to 5 kb and represents 0.1% of the MTB genome. This was the basis for the development of a genotyping method known as spacer-oligonucleotide typing or spoligotyping which was standardised by Kamerbeek et al. (1997). Spoligotyping has since been mainly

used for investigating clusters of MTB, MDRTB and XDRTB infections. Analysis of presence or absence of each of these spacers aided the understanding of the MTB diversity. The spoligotype is a visual description of the CRISPR loci that is present in MTB. The 43-spacers format is the standard utilised version of the test. This spoligotype, which can be easily obtained either *in vitro* or *in silico*, allows us to have a summary information of lineage or even antibiotic resistance when known to be associated to a particular cluster at a lower cost (Kamerbeek et al., 1997; Cowan et al., 2004; Zhang et al., 2010).

2.3.2.1 Conventional spoligotyping

Conventional spoligotyping, comprises a PCR-based reverse line blot hybridization method which assesses the genetic diversity of the direct repeat (DR) locus (Kamerbeek et al., 1997). The spoligotyping assay is performed on a membrane. In the conventional format, each of the 43 spacers produces either a dark band (indicating the presence of the spacer) or no band (indicating the spacer's absence). As Figure 2.2 shows, the spoligotyping assay for each MTB isolate produces a series of bands, much like a barcode. MTB is differentiated based on the presence or absence of 43 unique spacers in the direct repeat (DR) region in the MTB genome. It consists of minisatellite alternating exact direct repeats and variable spacers. Different MTB strains have various complements of the 43 spacers, and these different complements form the basis of the assay (Kamerbeek et al., 1997). Spoligotyping determines the presence or absence of spacers flanking the DRs by amplifying the DR locus by PCR

with one biotin-labeled primer, followed by hybridization of the PCR product with 43 oligonucleotides that are immobilized on a membrane. These oligonucleotides are designed from known spacer sequences of MTB strain H37Rv, and *M. bovis* BCG strain P3. One of the most recognisable spoligotype patterns is that of SIT1/Beijing (binary code: 000-000-000-000-000-000-000-000-000-000-000-011-111-111-1; octal code 000000000003771) as shown in Figure 2.2: spoligotype 6,12 and 37 correspond to Beijing genotype as spacers 1 to 34 are absent, and only nine spacers, 35 to 43 are present. Hybridization signals is detected by using the enhanced chemiluminescence direct nucleic acid labelling and detection system (Amersham Life Science). The method was described in detail by (Kamerbeek et al., 1997). Conventional spoligotyping is a low-cost test but it requires long hands-on-time. Variable membrane quality, contaminations due to carry over, mismatching of oligonucleotide probes due to production error or systematic error during read / interpretation and weak signals are some of the factors that could cause erroneous results.

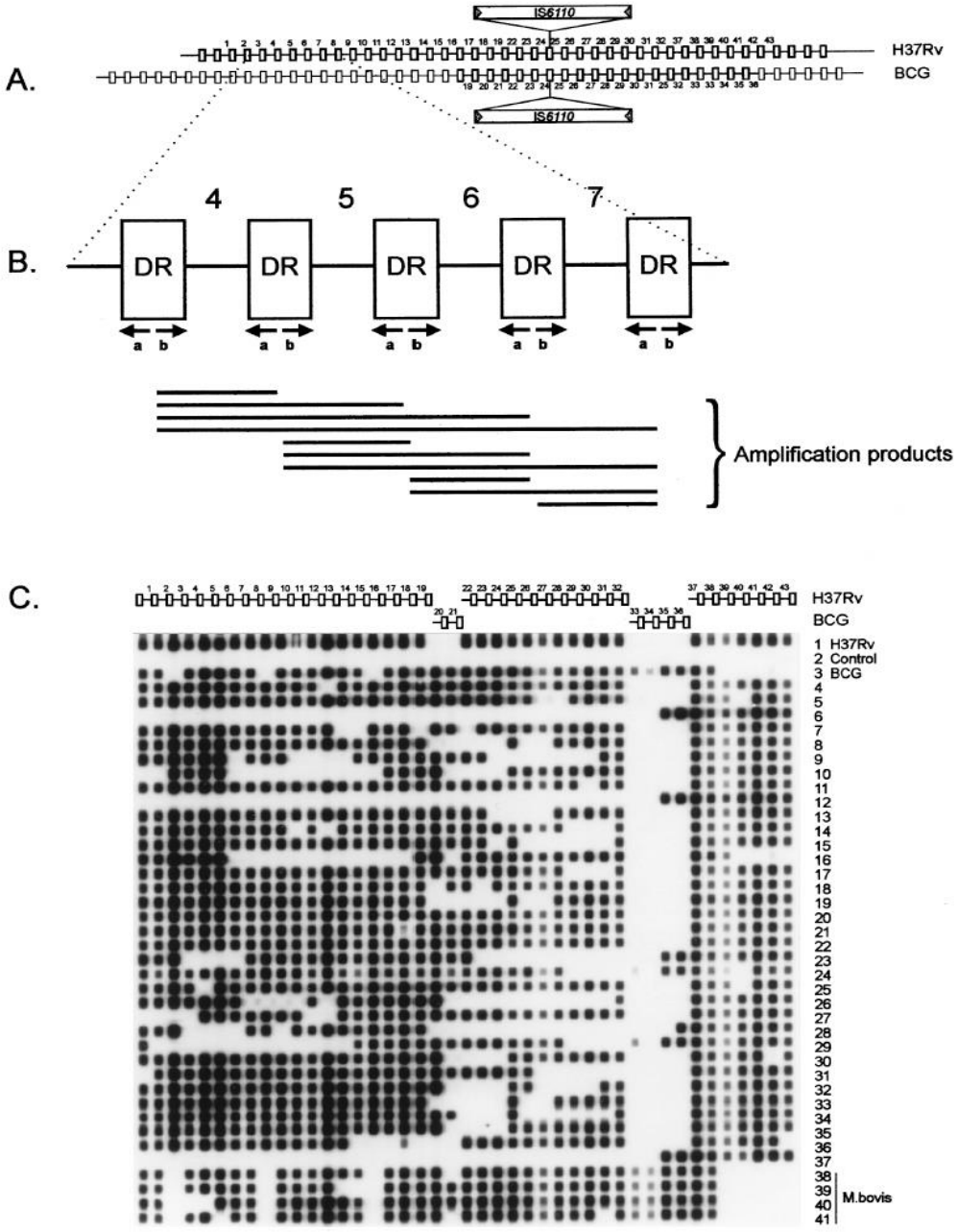


Figure 2.2: DR locus and membrane spoligotyping: A. Structure of DR locus in the mycobacterial genome, B. Principle of DR amplification of the DR region by PCR. Any DR in the DR region may serve as target for primers; C. Hybridization patterns (spoligotype) of amplified mycobacterial DNAs of 35 MTB and 5 *M. bovis* strains. Spoligotype 6,12 and 37 correspond to Beijing genotype (only spacers 35-43 are present). Spoligopatterns obtained with detection of hybridisation signals with an ECL detection system followed by autoradiography of the membrane
Source: Kamerbeek et al. (1997)

2.3.2.2 Spoligotyping using Luminex technology (TB-SPRINT)

A Luminex assay is a type of immunoassay that accurately measures multiple analytes in one sample. The Luminex® xMAP® technology is a bead-based immunoassay that allows for multiplex detection of up to a maximum of 100 analytes simultaneously, depending on the machine that is being used (MagPix® analysed 50 analytes at a time). Colour-coded microspheres, or beads, are internally dyed with different proportions of red and infrared fluorophores that correspond to a distinct spectral signature, or bead region. The quantification of multiple biomarkers in a single sample provides critical information about biological processes and diseases (Sola et al., 2016).

Cowan et al. (2004) successfully transferred the spoligotyping technology principle of interrogating the reference spacer sequences from the 43-spacers spoligotyping from a membrane-based format onto the Luminex multianalyte profiling system (multiplex microbead-based suspension array format), enabling faster and more automated analysis. Zhang et al. (2010) evaluated improvements of spoligotyping by the addition of a panel of 25 extra spacers creating a 68 spacer spoligotyping panel. Zhang et al. (2010) confirmed the high sensitivity and reproducibility of the classical technique using the 43-spacer panel and obtained perfect agreement between the membrane-based and the microbead-based techniques. Zhang et al. (2010) also confirmed that the addition of the 25 spacers to the standard 43 spacer spoligotyping is significant for the study the EAI clade that is prevalent in South-East Asia, but not so much for the Beijing family.

The TB-SPRINT is a 59-plex microbead-based assay developed by Beamedex, Orsay France utilising the Luminex technology (Luminex Corp., Austin, TX) (Gomgnimbou et al., 2012; Gomgnimbou et al., 2013; Sola et al., 2016). It is able to perform spoligotyping (TB-SPOL) as well as resistotyping (TB-RINT), and it is a good alternative method to replace conventional membrane-based spoligotyping. Microbead-based assay is said to be more efficient due to the increased surface for reactions to take place. The microbeads in the Luminex assay have different concentrations of two dyes (Infrared dye and red dye-bead ID). Each bead is coupled with a specific probe to capture a specific target. Biotin is added to all targets during PCR. Biotin attracts Streptavidin-conjugated Phycoerythrin (SAPE) during hybridization and SAPE is then detected by the green LED. Multiplexing enables detection of up to 100 analytes/assay. However, the maximum detection is 50 analytes/assay in the MagPix® (Gomgnimbou et al., 2012; Gomgnimbou et al., 2013; Sola et al., 2016). The microbead-based format allows a digital-numerical data output less prone to interpretation errors. This output is also more amenable to being managed through large international databases that can be curated for genotyping errors (Zhang et al., 2010).

2.3.2.3 Resistotyping and drug susceptibility test (DST)

The mechanisms of MTB drug resistance are well known and are reported to be involved with mutations in specific genes. Rifampicin (RIF) resistance involves mutations in the *rpoB* gene in the 81-bp rifampicin

resistance-determining region [RRDR]), and isoniazid (INH) resistance is associated with mutations in *katG* (codon 315), as well as *inhA* (positions 15 and 8 in the *inhA* promoter sequence), and other genes. Our knowledge of the genetic basis of MTB drug resistance and the use of nucleic acid amplification tests (NAATs) have allowed development of commercial PCR-based tests for resistance detection, such as the INNO-LiPA Rif-TB, Genotype MTBDRplus, GeneXpert® MTB/RIF and TB-RINT assays. DNA sequencing remains the standard method although other research and diagnostic methods have been and are being developed (Gomgnimbou et al., 2012; Gomgnimbou et al., 2013).

Implementation of molecular DST would improve epidemiological surveillance, and provide preliminary drug susceptibility profiles before phenotypic results are available, thus improving the management of TB patients and preventing further spread of drug-resistant MTB strains (Molina-Moya et al., 2017). The multiplexing capacity of this molecular DST method has the potential to simultaneously target a well-defined set of mutations associated with drug resistance, increasing the sensitivity of molecular resistance detection. Molecular DST cannot be fully implemented in Sabah as yet largely due to cost constraints, but it needs to be used to complement phenotypic DST, especially to rule out the possibility of drug resistance of the drug being considered for treatment.

2.3.3 Mycobacterial interspersed repetitive unit-variable number of tandem repeat (MIRU-VNTR)

Mycobacterial interspersed repetitive units-variable number of tandem repeats (MIRU-VNTR) typing is an important genotyping method, as it allows high-throughput, discriminatory and reproducible analysis of clinical isolates. Due to its portable data format, MIRU-VNTR typing has the potential to be a versatile tool for individual strain identification based on large reference databases (Mazars et al., 2001; Supply et al., 2001; Supply, 2005). The 24-loci MIRU-VNTR panel used in this study was developed and proposed by (Supply et al., 2006). The basis of MIRU-VNTR depends on the (variable) number of tandem repeat elements called mycobacterial interspersed repetitive units (MIRU). Four additional hypervariable were suggested by Allix-Béguec et al. (2014) for sub-typing of Beijing MTB, as the 24-loci MIRU VNTR was less discriminatory for Beijing Genotypes. Sun et al. (2004) used MIRU-VNTR to examine genetic diversity of MTB in Singapore.

MIRU-VNTR queries multiple genomic loci, especially in intergenic regions, scattered all throughout the bacterial chromosome which contain VNTR (Figure 2.3). Repeat unit sizes range from about 50 to 110 bp; the loci thus show a similar structure as minisatellite regions in higher eukaryote genomes (Supply et al., 2000). The number of repeat copies per locus varies among strains, and the use of several such loci allows sufficient inter-strain differentiation (Supply et al., 2001; Supply, 2005). The MIRU-VNTR profiles are presented as multidigit numerical codes ranging from 12–24 characters in

similar fashion to our identity cards ("complex haplotypes"), each digit representing the copy number or number of repeats in a single locus (Figure 2.4). The evolution of VNTR loci appears as a whole neutral, and these loci are independent from each other and apparently from the human host genome (Supply et al., 2001; Supply, 2005). This method depends on PCR efficiency, specifically the quantity of repeats which is based on the size of the amplified product.

Slow mutation rate for their repeats as shown by mathematical modelling makes them suitable for inferring long-term evolutionary histories of MTB genotypes. Even though the utility of VNTR loci as a whole (and when used alone) for inferring global phylogeny for MTB species is uncertain, nevertheless they are good markers for dealing with particular clonal lineages within the species (Mokrousov et al., 2004; Jiao et al., 2008; Allix-Béguec et al., 2014). Shorter turnaround and simple numerical nomenclature system, pushed forward mycobacterial interspersed repetitive unit–variable-number tandem repeat (MIRU-VNTR) typing, based on 24 standardized plus four hypervariable loci, optionally combined with spoligotyping, to replace IS*6110* DNA, the former gold standard of fingerprinting over the last decade among classical strain typing methods for many applications. VNTRs have become essential in forensic crime investigations, to study genetic diversity (DNA fingerprinting) and breeding patterns in animals (Marwal and Gaur, 2020).

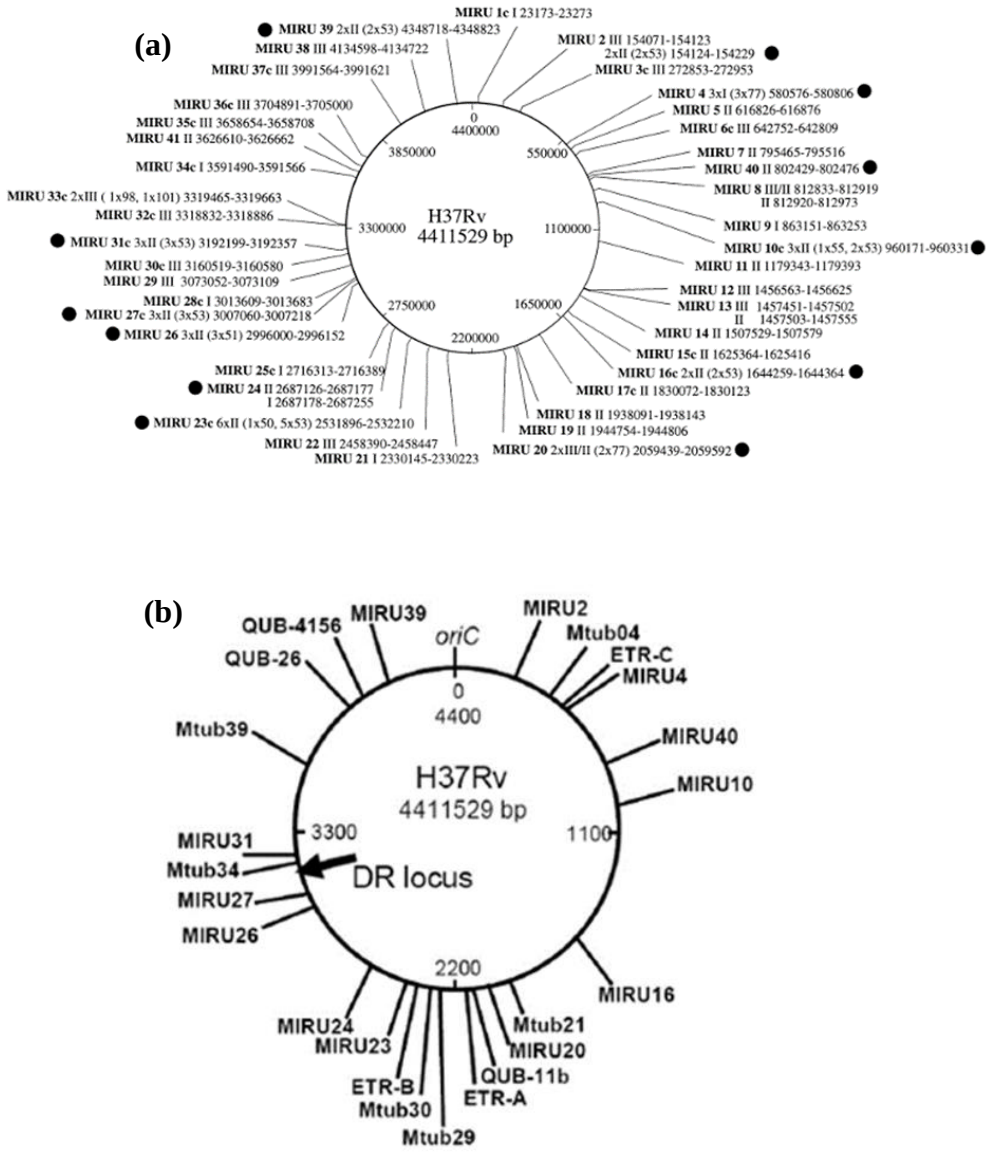


Figure 2.3: MIRU-VNTR Loci in MTB H37Rv chromosome (a) position of the 41 MIRU loci on the MTB H37Rv chromosome (Supply et al., 2000); (b) position of the 24 MIRU loci used in this study (Violeta et al., 2009)

There are 41 MIRU loci in MTB H37Rv chromosome as shown in Figure 2.3 (a), and the Arabic number in bold specify the MIRU locus respective numbers (Supply et al., 2000). Figure 2.3 (b) shows the 24 MIRU loci that are usually used for phylogenetic studies. 15-loci MIRU-VNTR is usually used for epidemiological studies (Allix-Béguec et al., 2008; Weniger et al., 2010).

(a) **12-loci MIRU-VNTR of two different SIT19/EAI2-Manila MTB isolate**

MIT12	SIT / Sublineage	154_MIRU02	580_MIRU04_ETRD	960_MIRU10	1644_MIRU16	2059_MIRU20	2531_MIRU23	2687_MIRU24	2996_MIRU26	3007_MIRU27_QUB5	3192_MIRU31_ETRE	4348_MIRU39	802_MIRU40	12-loci MIRU-VNTR
80	SIT19/EAI2-Manila	2	5	4	3	2	6	2	2	3	4	2	2	254326223422
56	SIT19/EAI2-Manila	2	5	4	3	2	6	2	2	3	4	3	2	254326223432

(b) **15-loci MIRU-VNTR of two different SIT19/EAI2-Manila MTB isolate**

MIT15	SIT / Sublineage	580_MIRU04_ETRD	960_MIRU10	1644_MIRU16	2996_MIRU26	3192_MIRU31_ETRE	802_MIRU40	2165_ETRA	577_ETRC	2163b_QUB11b	4052_QUB26	4156_QUB4156	424_Mtub04	1955_Mtub21	2401_Mtub30	3690_Mtub39	15-loci MIRU-VNTR
1034	SIT19/EAI2-Manila	5	4	3	2	4	2	4	4	9	7	1	1	10	2	2	543242449711A22
1035	SIT19/EAI2-Manila	5	4	3	2	4	2	2	4	9	7	1	1	10	2	2	543242249711A22

(c) **24-loci MIRU-VNTR of two different SIT19/EAI2-Manila MTB isolate**

MIT24	SIT / Sublineage	154_MIRU02	424_Mtub04	577_ETRC	580_MIRU04_ETRD	802_MIRU40	960_MIRU10	1644_MIRU16	1955_Mtub21	2059_MIRU20	2163b_QUB11b	2165_ETRA	2347_Mtub29	2401_Mtub30	2461_ETRB	2531_MIRU23	2687_MIRU24	2996_MIRU26	3007_MIRU27_QUB5	3171_Mtub34	3192_MIRU31_ETRE	3690_Mtub39	4052_QUB26	4156_QUB4156	4348_MIRU39	24-loci MIRU-VNTR
396	SIT19/EAI2-Manila	2	1	4	5	2	4	3	10	2	9	4	3	2	6	6	2	2	3	3	4	2	7	1	2	2145243A2943266223342712
397	SIT19/EAI2-Manila	2	1	4	5	2	4	3	10	2	9	2	3	2	7	6	2	2	3	3	4	2	7	1	3	2145243A2923276223342713

Figure 2.4: Designation of MIRU-VNTR profile (a) 12-loci MIRU-VNTR; (b) 15-loci MIRU-VNTR and (c) 24-loci MIRU-VNTR. Each digit represents the number of repeats detected at the respective MIRU-VNTR locus. For loci with more than nine repeats, letters are used (e.g. A=10 repeats, B=11 repeats etc.)

In the example shown in Figure 2.4, two isolates from the SIT19/EAI2-Manila spoligotype were subjected to 24-loci MIRU-VNTR (Yasmin et al., 2016). The MIRU-VNTR profile from this analysis were recorded as a 12-loci, 15-loci or 24-loci MIRU-VNTR profile. There is a specific sequence to record the MIRU-VNTR profile which must be adhered to in order to allow comparison with other isolates in the different MTB genotype databases e.g. SITVIT2 (Allix-Béguec et al., 2008; Demay et al., 2012; Couvin et al., 2019a).

2.3.4 Lineage and sub-lineage designation

There are a number of available free-access web-based platforms to help assign lineage and or sub-lineage to MTB isolates obtained from different parts of the world. The complementary information available in these different databases can be used to render a more comprehensive and complete outcome or conclusion.

2.3.4.1 SITVIT2

SITVIT2 is a web-based database that is provided, curated and regularly updated by the Institut Pasteur de la Guadeloupe that focuses on genotyping molecular markers for MTB, specifically for spoligotyping-based MTB lineage assignment (Demay et al., 2012; Couvin et al., 2019a). This website provides a user-friendly platform that is equipped with several Bioinformatics tools for scientists to utilise for better monitoring and description the global distribution of MTB strains. SITVIT2 website allows the scientific community to do in-depth studies of the MTB, to obtain exhaustive overview and information pertaining to the molecular epidemiology and drug resistance of MTB based on the collection of 111,635 clinical isolates from 169 countries of patient origin (131 countries of isolation, representing 1032 cities) in its repository (Demay et al., 2012; Couvin et al., 2019a). Despite the increased focus on WGS data alone, SITVIT2 is able to maintain the relevance of classical genotyping data. This is achieved through the establishment of useful links, and allowing correlations to be made between the “older” and “newer” genotyping methodologies, which

provide a more complete picture and knowledge on DNA-based data necessary for a deeper understanding of the global TB molecular epidemiology (Demay et al., 2012; Couvin et al., 2019).

The available molecular markers for study and exploration include Spoligotypes, 5ETRs, 12-loci MIRU-VNTR, 15-loci MIRU-VNTR, and 24-loci MIRU-VNTR markers. Spoligopatterns are determined by the presence or absence of 43 unique spacers in the DR locus of MTB. This pattern can be seen as a 43-digit pattern or under a reduced octal form of only 15 numbers (14 triplet + 1). MIRU-VNTRs markers are variable number of tandem repeats (number of copies). Users of this database can draw some important conclusions on TB geo-epidemiology. The information is accessible through a website (available at: http://www.pasteur-guadeloupe.fr:8081/SITVIT2).

2.3.4.2 *MIRU-VNTRplus*

MIRU-VNTRplus is a web-based platform commonly used to compare and analyse genotyping data in the MIRU-VNTR, spoligotype format and answer questions regarding MTB clustering and genetic relatedness. Queried isolates are compared against a reference database containing collection of 186 strains complete with their 24 MIRU-VNTR loci information and lineage assignments (Allix-Béguec et al., 2008; Weniger et al., 2010). This platform

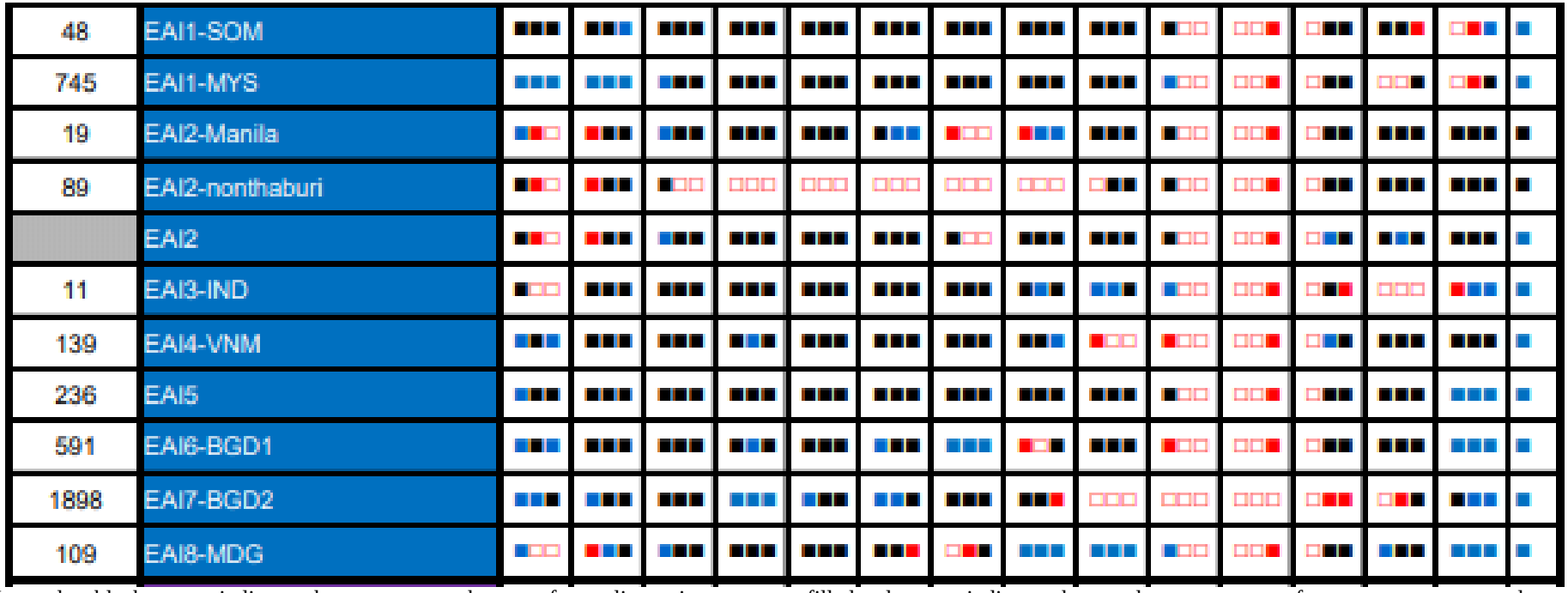
allows users to assign to a strain of interest the lineage of the strain in the database that has a difference of less than 4 loci.

MIRU-VNTRplus has three main functions as follows (Allix-Béguec et al., 2008; Weniger et al., 2010):

i. phylogenetic lineage identification by using a reference database
ii. analysis and visualization of genotyping data
iii. access to MLVA MTB15-9 nomenclature service. Mixed population can result from a true mixed infection or from culture or DNA contamination. In contrast, the occurrence of double allele in a single locus suggests the presence of a given allelic variant within a clonal isolate

2.3.4.3 SpolLineages Tool

SpolLineages tool incorporated some novel methods for fast and precise prediction of MTB spoligotype families. Couvin et al. (2020) developed the 'SpolLineages' software tool which is a novel algorithmic complementary data-driven approach allowing quick and precise prediction of MTB genotypic families from spoligotyping data, using a decision tree (DT), an evolutionary algorithm (EA) or classical binary rules. These approaches are helpful for a better understanding and analysis of genotypic variability and evolution of MTB. The tool can be accessed online at: (https://github.com/dcouvin/SpolLineages).



Legend: a black square indicates the presence or absence of a spoligotyping spacer; a filled red square indicates the mandatory presence of a spacer; an empty red square indicates the mandatory absence of a spacer; a filled blue square represents a high probability of the presence of a spacer; an empty blue square represents a high probability of the absence of a spacer

Figure 2.5: SpolLineages Tool binary spoligo rules (Couvin et al., 2020)

The SpolLineages Tool (Figure 2.5) shows that there are different key signals or spoligotyping patterns among the different sub-lineages, and on this basis consistently identify and group the patterns into the corresponding lineage / sub-lineages using the built-in binary spoligo rules. Isolates that do not fit the spoligo rules are marked as Unknown, and should be tested further using WGS or other suitable genotyping tests to determine their respective lineages / sub-lineages.

2.3.4.4 SPOTCLUST

SPOTCLUST is an approach to advance global studies of MTB genotyping data (Vitol et al., 2006). SPOTCLUST uses mixture models to identify strain families of MTB based on their spoligopatterns. The algorithm incorporates biological information on spoligotype evolution, without attempting to derive the full phylogeny of MTB. SPOTCLUST results both confirm previously defined families of MTB strains and suggest certain new families. This approach can potentially provide a simple first-step tool for epidemiology of TB (Vitol et al., 2006).

2.3.4.5 TBminer

TBminer uses popular machine learning techniques to learn lineage assignments based on either MIRU-VNTR loci, spoligotypes, or both, and

outputs the majority-vote prediction (Azé et al., 2015). TBminer is able to provide both major lineages, as well as its own “consensus” classification.

The present approach of geographical mapping of predominant clinical isolates of tubercle bacilli causing the bulk of the disease both at country and regional level in conjunction with epidemiologic and demographic characteristics allows to shed new light on TB geo-epidemiology in relation with the continued waves of peopling and human migration (Demay et al., 2012; Couvin et al., 2019a).

2.3.5 MTB clustering, clustering rate and Hunter Gaston discriminatory index (HGDI)

Clustering and clonal complex clustering occurs when at least two TB cases have matching or closely matched MTB genotypes to suggest that a recent (active) transmission from a common source, or because of reactivation of a latent infection acquired from the same source in the past has taken place (Allix-Béguec et al., 2008; de Vries et al., 2009; Weniger et al., 2010; García De Viedma and Pérez-Lago, 2018).

The clustering rate refers to the proportion of TB cases that share identical or highly similar MTB genotype within a defined population or setting. It can be used to estimate and assess the extent of recent transmission, as well as to identify possible transmission clusters within the population (Vynnycky et al., 2003; de Vries et al., 2009; Weniger et al., 2010; García De Viedma and Pérez-Lago, 2018; Somphavong et al., 2019). The clustering rate can be calculated using the formula as follows:

$$\mathbf{CR\text{-}[nc\text{-}c]/n}$$

CR: clustering rate; nc:total number of clustered isolates; c: total number of clusters; and n: total number of isolates (Somphavong et al., 2019).

Understanding the clustering rate is crucial for TB prevention and management strategies as it provides insight into the dynamics of TB transmission within communities (Vynnycky et al., 2003; García De Viedma and Pérez-Lago, 2018; Somphavong et al., 2019). Higher clustering rates suggests that a significant proportion of TB cases are linked by recent

transmission within a community, indicating higher rates of ongoing transmission (García De Viedma and Pérez-Lago, 2018). This is usually seen in settings where TB prevention measures are inadequate and may indicate areas or populations where targeted interventions are needed to interrupt the spread of the disease. Although it is not a direct correlation, it is helpful information to guide TB prevention and management activities.

In the context of genotyping-complemented contact investigation, it is important to take into account the sensitivity and specificity, as well as discriminatory power of the genotyping method utilized (Hunter and Gaston, 1988; Navin et al., 2004; CDC, 2022) as it affects the calculation of the clustering rate. Hunter Gaston discriminatory Index (HGDI) is a single numerical index of discrimination commonly used as an estimator of the diversity of molecular markers in bacterial pathogens. It is based on the probability that two unrelated strains from the test population will be placed into different typing groups (Hunter and Gaston, 1988; Hunter, 1990). It is used to validate the hierarchical typing methods, and is most valuable for large collections of distinct strains. A >0.90 index would be desirable for the results to be interpreted confidently (Hunter and Gaston, 1988; Hunter, 1990). HGDI can be calculated by using the Discriminatory Power Calculator which is based on the Simpson's index of diversity (Hunter and Gaston, 1988; Hunter, 1990). (http://insilico.ehu.es/mini_tools/discriminatory_power/). Thus, it is important to complement genotyping data with epidemiological investigations and other surveillance methods to gain a comprehensive understanding of TB transmission dynamics in a particular setting.

2.4 TB prevention and management program

The terminology TB control has been common usage and still in use in Sabah, and MOH Malaysia, but increasingly public health programs internationally have moved away from this and other potentially stigmatizing terms (Zachariah et al., 2012). Thus, taking this into consideration, the terminology 'TB prevention and management' or simply 'TB prevention' will be used in the context of this study to replace the terminology TB control (Zachariah et al., 2012). It is hoped that the use of this phrase will bring the focus on the proactive aspects of addressing TB: emphasizing efforts to prevent disease transmission from occurring, and effectively manage TB when it does (Zachariah et al., 2012). However, the terminology TB control will still be used to refer to TB prevention and management activities in the past.

TB prevention and management plays a vital role in curbing TB transmission through implementation of strategies and approaches. The prevention and management of TB involves high costs in the context of time, finance, human resource etc. The current practice is to treat active TB, vaccinate uninfected populations and improve detection methods (Dye and Williams, 2010; Liew et al., 2015). However, the plateauing trend of TB due to the steady source of TB from populations with LTBI suggests that better approaches need to be established. In the **A**sia **L**atent Tub**ER**culosis (ALTER) meeting held in 2018 in Singapore, 13 TB experts from ASEAN countries, Bangladesh, India, Hong Kong and Taiwan convened, and proposed strategies to improve LTBI management in Asia. The decision made was for the implementation of

customised LTBI management for respective countries. Malaysia has started to screen and treat LTBI among people living with HIV (PLHIV), children <5 years old and other HIV-negative at-risk groups (Huaman et al., 2019; Paton et al., 2019; Goroh et al., 2020b; Rosli, 2022). However, implementation has been limited (Goroh et al., 2020b; Manoharan et al., 2023). Planning, execution and balancing of limited resources effectively and efficiently seem to be key for a successful TB program. The perfect solution however, still remains elusive. Far greater impact can be achieved by combining interventions that attack different aetiological pathways-blocking fast and slow routes (Dye and Williams, 2010; Liew et al., 2015). Emergence of MDRTB and XDRTB is a major public health problem that threatens TB care and control in many countries and reveals weaknesses in TB management and diagnostic services. Case finding and prompt, appropriate treatment to prevent further dissemination of the disease is key to controlling TB. This has been hampered, as with many other infectious diseases, by the continuing emergence of drug resistance (Van Deutekom et al., 2005; Liew et al., 2015).

The role of TB laboratory especially in diagnosing TB/MDRTB cases and supporting TB surveillance activities were much improved by the implementation of quality-assured laboratory tests such AFB microscopy, AFB culture (solid and automated liquid-based culture), MTB identification, DST and nucleic acid amplification test (NAAT). All these diagnostic tests have helped to bacteriologically confirm cases especially DRTB cases (Rosli, 2022). This is especially for TB case notifications, reporting of MDRTB cases and aiding TB infection control activities.

Social characteristics of human populations (Gagneux, 2012), host genetics (Comas and Gagneux, 2009), co-morbidities and environmental conditions, and the quality of TB control programs are crucial determinants of TB (Dye et al., 2009). Local specificity of clones may be explained by recent importation and fast dissemination due to specific pathogenic properties or outbreak conditions, or, somewhat alternatively, due to long-term historical presence in an area (Mokrousov, 2012).

The knowledge gained may be useful for the formulation of effective control programmes to limit the spread of MTB, especially resistant strains. Large-scale genotyping is also needed to get better insights into the biological and geographical diversity of the pathogen and its association with host populations, understanding the origin of, and retracing the evolutionary history and the demography of the pathogen, and some of its major clones in association with multidrug resistance (Gagneux et al., 2006).

2.4.1 WHO Stop TB Strategy and End TB strategy

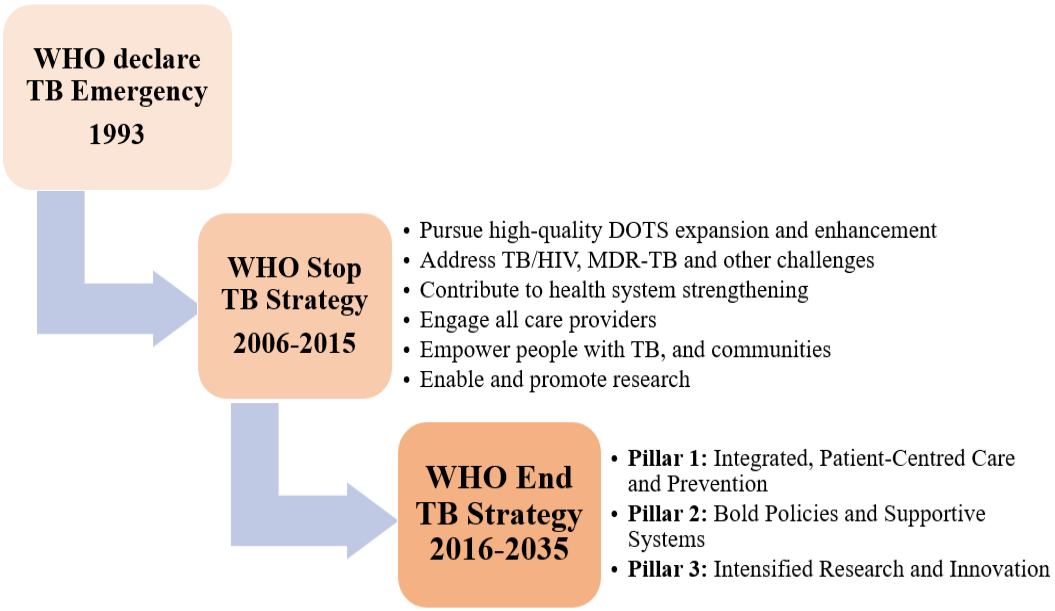


Figure 2.6: WHO Stop TB Strategy and End TB Strategy

The WHO declared TB a global emergency in 1993, and in 2006 the Stop TB Partnership proposed a global plan in the context of an overall vision of a a world free of TB, aiming to save 14 million lives between 2006 and 2015. The expansion and enhancement of the effective daily observed treatment short-course (DOTS) was highlighted in this strategy. The measures that entail include the following:

i. Political commitment with increased and sustained financing

ii. Case detection through quality-assured bacteriology

iii. Standardized treatment with supervision and patient support

iv. An effective drug supply and management system

v. Monitoring and evaluation system, and impact measurement

In 2016, the WHO End TB Strategy was launched, replacing the WHO's Stop TB Strategy, to provide a renewed and comprehensive framework for global TB control efforts. The WHO End TB strategy spans over a 20-year period (2016–2035). It integrates previous global TB control initiatives, particularly the Stop TB Strategy that was launched in 2006. The WHO End TB Strategy addresses persistent challenges and emerging issues in TB prevention, diagnosis, treatment, and care. It also articulated the elimination of catastrophic costs due to TB as the earliest priority, proposing that by 2020, 0% of families should be facing catastrophic costs (WHO, 2014; Uplekar et al., 2015). 'Catastrophic' is defined as total costs of TB care exceeding 20% of annual household income. With 2015 as the baseline, the strategy includes the targets of a 90% reduction in TB deaths and an 80% reduction in the TB incidence by 2030 (WHO, 2014; Uplekar et al., 2015). It also aims to end the global TB epidemic by 2035 through three main pillars:

Pillar 1: Integrated, Patient-Centered Care and Prevention

Pillar 2: Bold Policies and Supportive Systems

Pillar 3: Intensified Research and Innovation

By addressing the root causes of TB and leveraging innovative approaches, the strategy aims to achieve sustainable impact in TB prevention and elimination, and ultimately improving health and well-being worldwide.

2.4.2 TB prevention and management program in Sabah

TB was considered as 'The Biggest Single Killer' in this state with death rate caused by this disease being the highest in government hospitals until the seventies (1970's). Many TB patients were going to the hospital for treatment, but there were no measures and organized efforts in place to control and prevent the transmission of TB at the time. In 1960, a study on the TB disease in Sabah was conducted with the help of the Colombo Plan Advisory Team Australia. Based on this study outcome, the Sabah State TB Control Program was officially launched in 1961 with the aim of reducing the TB burden of disease and deaths caused by MTB in the community (Boxer, 1969; Rosli, 2022).

The National TB Control Program in peninsular Malaysia has been integrated into primary health care services in 1995 and it has been placed under the Disease Control Division, Ministry of Health Malaysia (Dony et al., 2004; Liew et al., 2015; Rosli, 2022). The State TB Control Program is led by the State Health Director with the assistance of the State Deputy Health Director (Public Health) and Respiratory Physicians as the technical advisor. The State TB team consists of the Chief State Assistant Director of Health (TB/Leprosy), who is given full responsibility for all matters of planning, implementation, organization, coordination and technical activities; Disease Control Unit Office (TB/Leprosy), TB treatment manager, BCG manager, laboratory manager and surveillance manager. Besides that, there is a TB team that is set at the district level overseeing the TB management therein.

The National TB Control Strategy include

i) Early detection by screening high-risk groups as well as symptomatic cases in hospitals and health clinics;
ii) Mandatory screening of all foreign workers;
iii) Implementation of the concept of DOTS (Directly observed treatment short course) in all treatment centres;
iv) Detect contact and treatment refusal and/or failure.

Treatment for new cases of positive sputum smears were initiated by medical officers in primary health care (Liew et al., 2015; Koh et al., 2019; Goroh et al., 2020a).

Sabah, in line with the Ministry of Health Malaysia Clinical Practice Guideline has adopted the DOTS control strategy. DST information is important in order to initiate proper TB treatment. Although in Sabah pan-susceptible isolates (~80%) are more prevalent compared to their DRTB counterpart, drug resistance in any form is still cause for alarm.

The TB Control Program has conducted much effort to control this disease employing various strategies. Based on the WHO Stop TB Strategy, The TB Control Unit in the Sabah State Health Department focused on 5 strategies to eliminate TB disease in Sabah (Goroh et al., 2020a; Rosli, 2022).

The strategies are:

i) Strengthening Health System Components, review and upgrading of MyTB database, TB Information System (TBIS) which is systematic

and complete documentation system for all things related to TB management and control;

ii) Improving the TB case detection system especially in the TB laboratory diagnostics and to provide high quality TB treatment management;

iii) Empowering communities in the prevention and control of TB disease;

iv) Reducing the transmission of TB infection to the community; and

v) Promote TB disease research

Contact investigation of patients and cases of treatment failure is under the jurisdiction of the District Health Office which also operates other health programs. TB treatment is initiated at Treatment Centre 1 (PR1) by medical officers at hospitals and health clinics. Patients are referred to the nearest health care facility, Treatment Centre 2 (PR2) that manages DOTS. Follow-up treatment is done at PR2 which consists of a Health Clinic without a medical officer and a Community Clinic (Rural Clinic). There are currently 42 TB PR1, and 200 PR2 in Sabah (Rosli, 2022).

There are differences in the TB disease profiles and prevalence in Sabah when compared to the TB disease landscape in peninsular Malaysia, thus warranting a special guideline for the state of Sabah. It is used concurrently with the existing guidelines published by the Ministry of Health, Malaysia. The use of these guidelines is expected to strengthen the delivery of health services and TB treatment, based on the resources available in the state, with the hope to

achieve the goal of the National TB Control Program to make Malaysia TB free by the year 2035 (MOH, 2012; MOH, 2016; Rosli, 2022)

2.4.3 TB case contact investigation

Passive case detection (PCD) of TB cases has, until recently, been the principal approach to case-finding. In PCD, patients, usually with symptoms e.g. cough go to the clinic or hospital and gets diagnosed with TB. However, the case-detection gap, especially in certain vulnerable populations, along with the persistence of delays in diagnosis and the accompanying continued transmission in the community, highlight the need for a more active approach to detect TB early. Hence, the need to consider systematic screening for active TB in selected risk groups (WHO, 2015). Screening algorithms need to be tailored to specific characteristics of specific populations (WHO, 2015). In Sabah, the number of new TB cases notified and TB notification rates in 2013, 2014 and 2015 are 4515, 127/100,000 population; 4747, 130/100,000 population; and 4497, 120/00,000 population respectively, with Kota Kinabalu, Sandakan, Tawau, Lahad Datu and Keningau recording the top five highest numbers of TB cases in 2013 to 2015 (Goroh et al., 2020a).

Active case detection (ACD) through contact investigation is a principal activity of TB prevention and management programmes. The prevailing approach is the concentric circle or “stone-in-pond” method (Munang, 2019). The underlying principle is that those in closest physical proximity (i.e., shared

air space over the most time) to an infectious case are most likely to be infected, and if they are not then the yield or the likelihood of getting positive cases such as active TB or LTBI from contact investigation those less close is minimal (Veen, 1992). This approach is pragmatic particularly when dealing with outbreaks in congregate settings.

The ultimate target of any TB prevention and management team is to end the spread of TB among the population. Various programs have been instituted in Malaysia to achieve this objective (Dony et al., 2014; Liew et al., 2015; Goroh et al., 2020a; Goroh et al., 2020b). Figure 2.7 shows the TB prevention and management strategies in Sabah (Avoi, 2015). In Malaysia, BCG vaccine is given at birth to confer protection especially against TB meningitis (Dony et al., 2014; Liew et al., 2015). In a population where TB prevalence is relatively high, the general population is at an increased risk of being infected by MTB. In general, only 10% of the people who are infected by MTB would eventually progress into active TB.

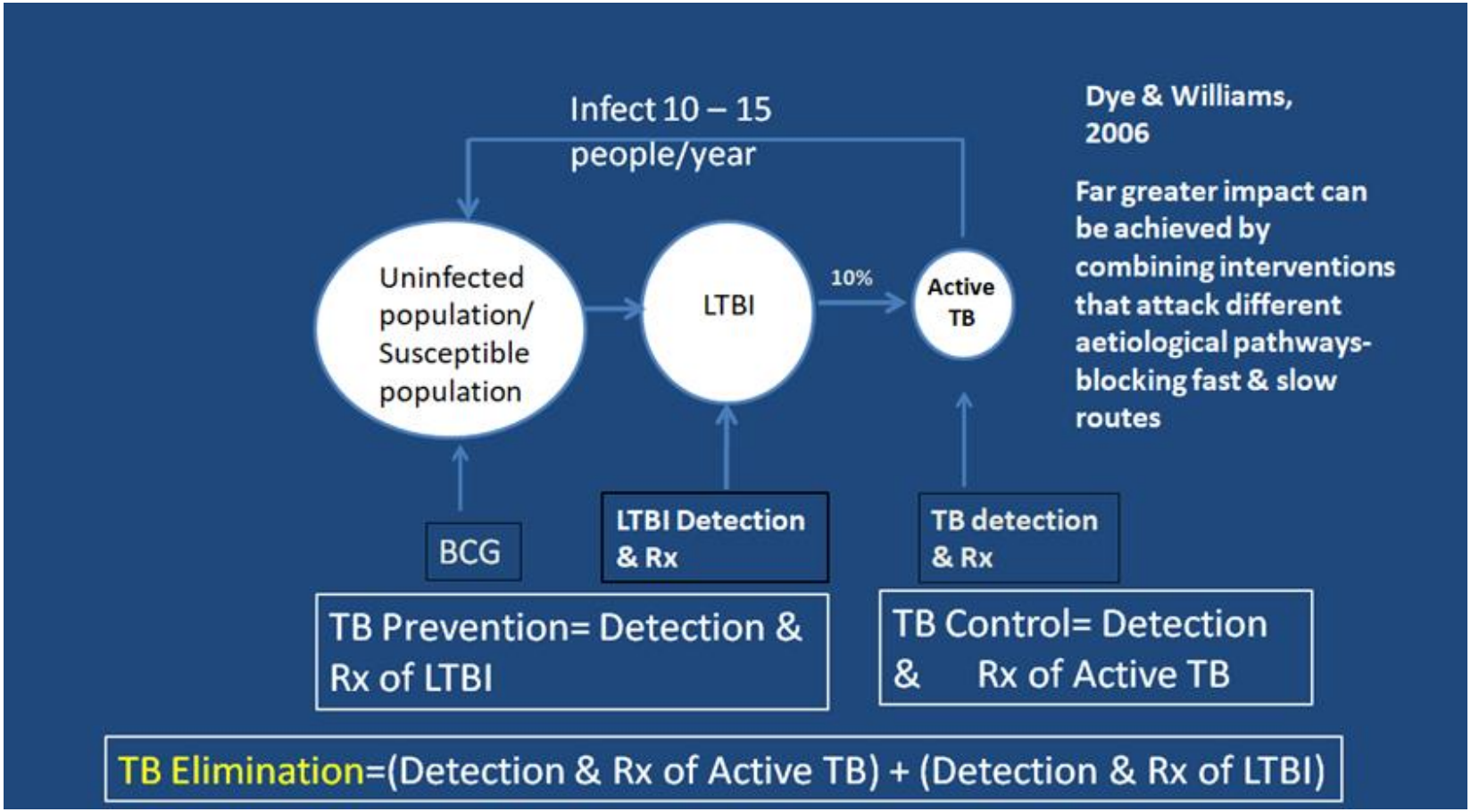


Figure 2.7: TB Prevention and management algorithm in Sabah (Avoi, 2015)

In the course of a year, statistically speaking, a smear positive pulmonary TB patient who is not on anti-TB treatment is able to spread the disease to about 10 to 15 people. Therefore, early detection and initiation of anti-TB treatment is important to ensure a high success rate in treatment, as well as to curb TB transmission in the population (Dony et al., 2014; Avoi, 2015; Liew et al., 2015).

However, if the TB prevention and management activities are only targeted on active TB cases, there will always be a reservoir of possible TB cases in the form of latent TB infection (LTBI) among the population. Dye and Williams (2010) proposed a combined intervention strategy that attacks different aetiological pathways – blocking fast (active TB) and slow (LTBI) routes. Greater impact can be achieved if this approach were implemented (Dye and Williams, 2010; Paton et al., 2019). The WHO End TB strategy aims to reduce TB deaths by 95% and to lower the incidence of new TB cases by 90% between 2015 and 2035 (Uplekar et al., 2015; Koh et al., 2019). To achieve this milestone, especially in high TB-burden regions, it is also crucial to prevent the development of active TB from inadequate treatment of LTBI. During conventional contact investigations, public health workers interview TB patients to obtain the names of other people who may have been exposed to them, and venues in which much exposure may have taken place. One of the weaknesses in conventional contact investigation is TB patients are often unable or reluctant to name all contacts and all places of potential transmission (Goroh, et al., 2020a; Goroh, et al., 2020b). A careful review of public health records, genotyping results, contact investigation logs, estimated infectious periods, and

re-interview of patients in a genotype cluster can uncover additional potential transmission opportunities.

When a TB case is diagnosed and registered at the district level, the case's epidemiological information including HIV status, TB infection risk factors and MDRTB status will be recorded in the District TB Registry, TBIS101A (Buku Daftar TB Daerah). This is part of the MyTB database which is the MOH TB surveillance database. Standard treatment regime for anti-TB treatment which comprise an intensive phase and a maintenance phase lasting for two months and four months respectively is initiated at PR1 (MOH, 2002). Performance indicators such as sputum conversion rate (SCR), short-course chemotherapy (SCC), directly observed treatment (DOT) monitoring, and treatment outcome information can be obtained from this registry. Involvement of non-government organisations are captured in this registry as well (MOH, 2002). The target set by National TB Program is to identify and screen at least 10 contacts for every TB patient. Health inspectors undertake contact investigation for each newly diagnosed TB patient, and all identified contacts (usually among household or work place contacts) are notified by health inspectors to participate in TB screening at the clinic, which requires attendance on several occasions (MOH, 2002, Goroh et al., 2020b). Currently, monitoring of TB contacts is with the use of the TB contact examination monthly registry, TBIS101C (*Daftar bulanan pemeriksaan kontak TB*), which is part of the national TB information system (MOH, 2002; Liew et al., 2015; Rosli, 2022). This registry is maintained by the PR1 handling TB case contact examination. All contacts who came for examination will be recorded. But it is also equally

important to record the names of contacts who did not attend the examination. This is important so that in future, when the contact's name reappears there would be a higher index of suspicion for possible TB and further testing and investigation should be conducted (Munang et al., 2016; Munang et al., 2019; Paul, 2022; Paul et al., 2022). The benefits gained from TBIS101C registry include the information of first-time examination, monitoring of contact absentees, and the workload of contact examination and contact investigation activities at the PR1. Another challenge faced in contact investigation is the low participation in initial and follow-up appointments (Goroh et al., 2020b). Health inspectors provide an effective role in TB contact investigation through direct personal communication to encourage the completion of the TB screening process but these activities need high resource allocations. Further integration with clinical processes, and with workplace and school-based investigations are required as well (Goroh et al., 2020b).

A lot of work has been done to support data collecting at all levels thanks to ICT advancements. The TBIS handbook is now available as an e-book for convenient access (MOH, 2022). Online forms for TBIS and electronic notification are also available for convenient usage. Therefore, for health facilities, the TB prevention and management team and the public health team in general to have access to e-reporting systems, the existing information and communication technology (ICT) infrastructure such as internet connectivity, must keep up with technological advancements. Accurate and readily available information pertaining to TB clustering trends etc. are important to the TB prevention and management programs in order to take prompt action and

implement effective measures to prevent, control, and end the spread of tuberculosis.

TB in big cities is a particular challenge and requires setting-specific research to understand its epidemiology. The availability of a large electronic register of cases and contacts (MyTB) in Sabah, can be exploited to gain insight to inform priorities and resource planning not only for the local TB prevention and management program, but are likely transferable to other settings. It is currently underutilised, but has the potential for in-depth study and exploration to drive further public health activities. The utility of MyTB database as a prediction tool been demonstrated in a study by Awaluddin et al. (2020) when they were able to identify the determinants of treatment success in paediatric TB in Malaysia via a 5-year data analysis obtained from the MyTB database. They discovered that by strengthening contact investigation activities and promoting early identification targeting the youngest children and non-Malaysian children, the treatment success rate among children can be increased. Implementation of new approaches to contact investigation will require increased resources but their feasibility for routine use needs further study.

2.5 Sabah, land below the wind

Figure 2.8: Map showing Sabah (East Malaysia), Philippines (South Philippines) and Indonesia (Kalimantan & Sulawesi) in South East Asia

Sabah is located in East Malaysia, and is positioned in the northern part of Borneo Island. Sabah shares a border with the state of Sarawak to the southwest and Indonesia's Kalimantan region to the south. It is also located across the Sulu Sea from the Philippines (Figure 2.8).

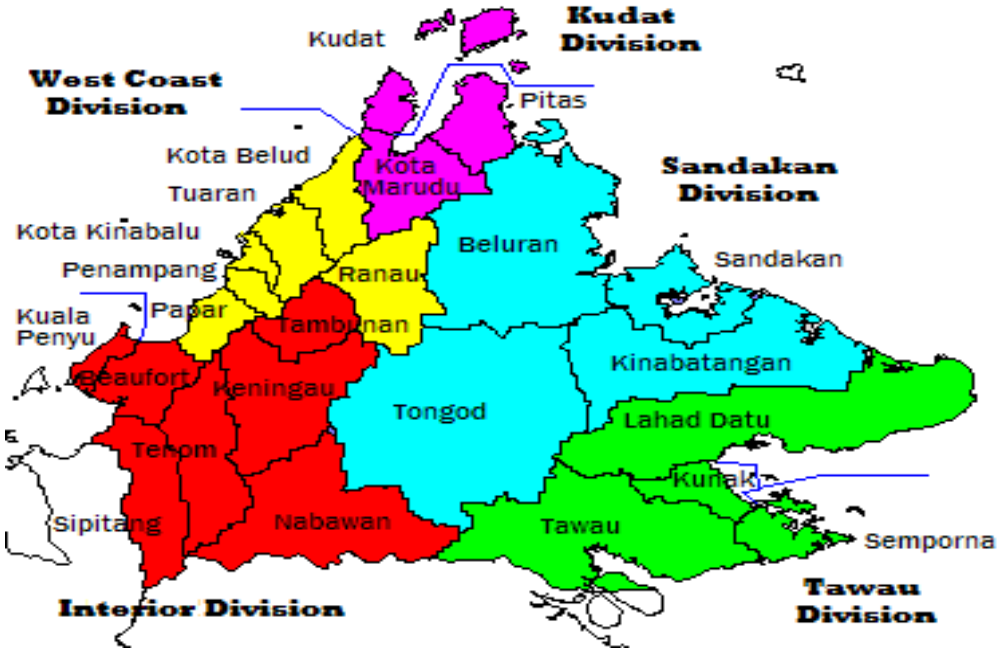


Figure 2.9: Map of Sabah showing the 5 divisions and 25 districts

The state of Sabah spans a land area of 73,631 km square with a population of 3,602,300 (Department of Statistics, 2020). As shown in Figure 2.9, Sabah is divided into 5 division and 25 districts as follows:

1. West Coast Division: Papar, Penampang, Kota Kinabalu, Tuaran, Kota Belud and Ranau
2. Kudat Division: Kota Marudu, Kudat, & Pitas
3. Sandakan Division: Beluran, Tongod, Kinabatangan & Sandakan
4. Tawau Division: Lahad Datu, Kunak, Semporna & Tawau
5. Interior Division: Beaufort, Tenom, Sipitang, Nabawan, Keningau & Tambunan

Thirty-two indigenous groups with over 80 locally-spoken dialects with a wide variation in traditions and cultures make up the people of Sabah. The main ethnic group breakdown is: Kadazan-Dusun: 23.8%; Bajau: 15.6%; Malay: 9.5%; Chinese: 9.1%; Murut: 2.5%; Other Bumiputera: 17.0%; Other non-native Malaysians: 0.6%; and non-Malaysian citizen: 20.5%.

2.5.1 Migrant population in Malaysia

Malaysia has long been a major destination country for migrants (Kaur, 2008). Today, in Southeast Asia, Malaysia is one of the countries that receives the most migrants, with official estimates of 1.4 to 2 million documented migrants and an extra 1.2 to 3.5 million undocumented migrants living there (as reported by the World Bank) between 2018-2020. Migrant population in Malaysia are mostly from Indonesia, Philippines, Bangladesh, Myanmar, Nepal, and other Asian countries in smaller numbers including India, Cambodia, and Lao PDR.

Various industries, including manufacturing, plantations, agriculture, services, and domestic work, employ migrant workers. Migrant workers currently comprise 20% of the workforce in Malaysia, with the majority of low-skilled and semi-skilled employment being held by them (https://www.imi.gov.my/index.php/en/main-services/foreign-worker/). There are strict policies and conditions in place dictating what type of jobs migrant workers are allowed to engage in. The policies for employing foreign workers can be found in https://www.imi.gov.my/index.php/en/main-services/foreign-worker/. The proportion of migrants in peninsular Malaysia and Sabah is not the same due to differences in Federal versus State policies and regulations regarding to this matter. In Sabah, the migrant workers are brought in mostly from the Philippines and Indonesia (Hassan, 2018). The neighbouring countries, Philippines and Indonesia are among the top 30 nations with high TB prevalence. With the high flow of migration in and out of Sabah from these

countries higher TB transmission among different populations is most likely to take place.

Since MTB is said to be closely evolved with human hosts, the MTB lineage profile in the population may reflect this connection. Immigration makes countries with lower TB prevalence vulnerable to strains of MTB that are drug-resistant especially MDRTB or XDRTB, from countries with high disease burdens and sub-par TB control programs (Reed et al., 2009; Gagneux, 2012; Gomgnimbou et al., 2012; Gomgnimbou et al., 2013; Bainomugisa et al., 2021). Strong phylogeographic structure of human-adapted MTB lineages have led to the belief that certain MTB variants might adapt locally to specific human populations (Gagneux, 2012).

2.5.2 TB and the Sabah population

Goroh et al. (2020a) described the TB epidemiology in Sabah in order to determine high risk groups as well as hotspots for TB transmission. There were about 33,000 reported cases of TB in Sabah (128 cases per 100,000 population). They found that TB cases and notification rates varied significantly throughout all 25 districts in Sabah. The Kota Kinabalu district, which is the state capital, reported the highest number of TB cases annually which accounted for about 15% to 18% of the reported cases. The district of Semporna in the east coast, and the district of Pitas at the northern coast, both reported the highest notification rates with higher than 200 cases per 100,000 population in 2018.

Meanwhile, the central eastern district of Kinabatangan (56 cases per 100,000) and interior district of Tongod (53 cases per 100,000), both characterised by large areas of jungle and remote rural communities reported the lowest rates in the study. TB rates increased with age and were highest in older males. Moderate or advanced disease on chest X-ray and sputum smear positivity was high (58 and 81% of cases respectively), corroborating the findings of Rundi et al. (2011) suggesting frequent late diagnosis.

TB notifications in Sabah State is about 20% of Malaysia's total TB notifications, despite representing only 10% of Malaysia's population. Goroh et al. (2020a) recommended that universal health coverage and expansion of GeneXpert® MTB/RIF coverage be made available to reduce barriers to care and early diagnosis and treatment for TB (Goroh et al., 2020a; Bainomugisa et al., 2021). According to Goroh et al. (2020a), the median age of TB cases in Sabah during the same period was 38 years. Case notification rates increased steadily with age from a low of 23 cases per 100,000 among children aged 5–14 years, to a high of 402 cases per 100,000 among adults aged 65 years and older. Only 1.6% of cases were among children aged less than 5 years, and 4.6% of cases were children less than 15 years. Further, the proportion of total cases that were children decreased from 2012 to 2018: by 2018, only 1.2% of cases were < 5 years and 3.6% were < 15 years. This trend is also reflected globally whereby the TB cases among children are low (Cowger et al., 2019; WHO, 2023) while TB cases among the elderly are high (Li et al., 2021; WHO, 2023). The small proportion of TB in younger children could be due to the missing data attributed to the non-specific clinical presentations and lack of sputum samples,

which may lead to underreporting and undertreatment (WHO, 2018; Awaluddin et al., 2020). However, TB among children is one of the important indicators that is being closely monitored by the TB prevention and management unit. LTBI or TB preventive treatment is recommended and offered to children less than 5 years old who are household contacts of people with bacteriologically confirmed PTB (Paton et al., 2019; Goroh et al., 2020b; MOH, 2021; Rosli, 2022).

60% of TB cases were male. In those aged ≥ 55 years the male to female notification ratio is 2.1:1 There was little difference in the number of notifications between males and females among children less than 15 years (Goroh et al.,2020a). William et al. (2015) described the highly advanced TB disease in out-patient settings with low HIV co-infection, echoing the finding of Rundi et al. (2010) in terms of treatment seeking behaviour whereby pulmonary TB with high grade of AFB smear were seen among most TB patients. Bainomugisa et al. (2021) gave insight to the MTB genomic diversity in Sabah, focusing on Kota Kinabalu district, the state capital. They discovered that the majority of MTB strains in Kota Kinabalu were from lineage 1, with a dominance of sub-lineage 1.2.1 (Indo–Oceanic) as well as a possibility of a new clade (1.2.1.X) via comparison with the 480 lineage 1 genomes from Northern Thailand (Palittapongarnpim et al., 2018).

Rashidi Ali et al. (2015) studied the prevalence of MTB drug resistance in Sabah. This group found that mycobacterial culture yielded MTB in 91.9% and non-tuberculous mycobacterium (NTM) in 1.7%. INH-R were found to be

at 4.5%, and current community rates of MDRTB in Sabah are low. Rashidi Ali et al. (2015) also proposed an algorithm for treatment to address INH-R pulmonary TB management issues in Sabah.

Local epidemiological features, population experiences, risk factors among healthcare workers and individuals with HIV co-infection, MTB drug resistance and improvements in detection of extra-pulmonary TB have previously been described (Dony et al., 2004; Jelip et al., 2004; Rashidi Ali et al., 2015; William et al., 2015). Jani et al. (2020) identified and characterized a few MTB isolates (L2 and L4) from Sabah using WGS, while Bainomugisa et al. (2021) provided some insight on how to improve public health measures in the context of TB prevention and management. Both studies have shed some light on the TB molecular epidemiology in Sabah. However, much work still remains to be done.

The TB mortality rate in Sabah had increased significantly from 9.0 per 100,000 population in 2014 to 11.4 per 100,000 population in 2018. The majority of TB deaths occurred in the first two months of treatment (Avoi and Liaw, 2021). TB-related deaths were primarily due to advanced disease or disseminated TB, whereas non-TB-related deaths were primarily due to existing comorbidities.

In 2021, a total of 4,547 cases of TB were reported compared to 4,469 cases in 2020. About 70.1% of the cases were of the smear positive pulmonary TB, and highly infectious. Approximately 5% of the TB cases were among

children under the age of 15 years. An estimated one-third of the cases were detected among non-citizens who were mostly Filipinos, and Indonesians. Treatment interruption rate was reported at 2.97% and mostly observed from districts with high population density such as Kota Kinabalu, Sandakan and Tawau. 53% of these interrupted cases were among citizens.

TB prevention and management in the state is quite unique and challenging as there are stateless population to cater to in addition to the population at large, high socio-economic gradients and geographical structure of the state. These challenges are different from those faced by other states in peninsular Malaysia, even by our neighbouring state, Sarawak (Rosli, 2022). Although Malaysia is not in the Top 30 high TB burden countries in the WHO list, death rate due to TB is the highest in Malaysia as compared to those for other infectious diseases (Avoi and Liaw, 2021).

Sabah has recorded among the highest if not the highest number of TB cases in Malaysia throughout the years. In 2021 Sabah again placed first with a total number of 4,547 cases of TB reported. Selangor trailed by in 2nd place. High risk groups infected with TB are close contacts to TB patients, the immunocompromised such as HIV and diabetes, malnourished, substance abusers, cigarette smokers, people living in overcrowded conditions such as prisons, shelters, homelessness and immigrants. The district with the highest TB cases was Kota Kinabalu with 793 cases followed by Sandakan 485 cases, Tawau 445 cases, Lahad Datu 353 cases, Semporna 334 cases and Keningau 277 cases. 437 mortalities reported in 2021 compared to 321 cases in 2020. 101

cases of TB with HIV and 16 MDR TB cases were reported as well. 24 cases of TB were reported among healthcare workers in 2021. The healthcare workers (HCW) affected are mostly frontliners who have direct with patients (Rosli, 2022).

CHAPTER 3

MATERIALS AND METHODS

3.1 Conceptual framework

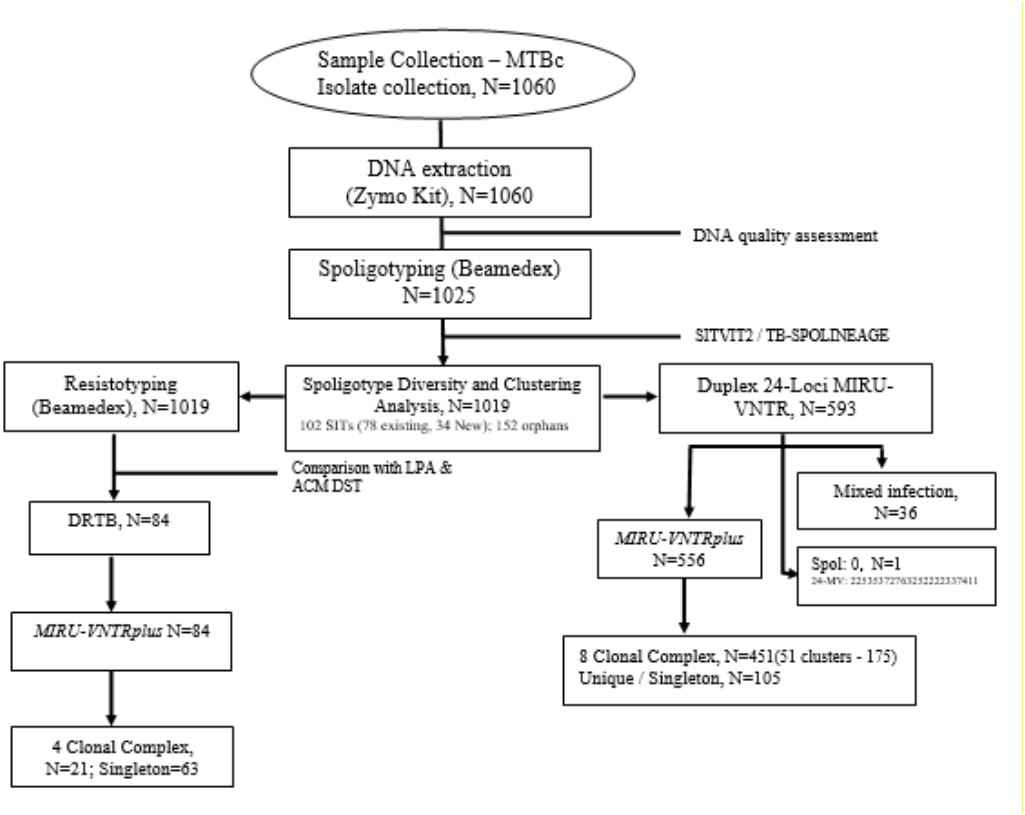


Figure 3.1: Conceptual framework of this study

3.2 Registration of study

Registration of this study in the National Medical Research Register (NMRR) and ethics clearance application (NMRR-16-2780-32850 [IIR]) were conducted at the beginning of the study. Training and execution of study methods were done according to the planned schedule.

3.3 Data mining of TB laboratory diagnostics in Sabah 2012-2018

Basic demography and epidemiological data, as well as TB laboratory results were obtained from SIMKA, MyTB and MKAKK databases. SIMKA (Sistem Informasi Makmal Kesihatan Awam) is a national and regional TB laboratory test results database archived and curated by the National Public Health Laboratory which is also known as Makmal Kesihatan Awam Kebangsaan (MKAK). MyTB is the Malaysian MOH TB surveillance database (national and state level) which is archived and curated by the TB/Leprosy Sector, Ministry of Health Malaysia. MKAKK TB laboratory database is the TB laboratory test results database (pre-2016) archived and curated by MKAKK. Permission to access the database for research purposes have been obtained from the MOH and Sabah state health department. Cross searches were performed for these three TB databases and the number of unique patients were tabulated. The number of tests performed were assumed to reflect the number of patients.

3.4 Sample collection

3.4.1 Epidemiological data

Epidemiological data of matched MTB isolate to MyTB-registered patient for all samples were obtained from MyTB database, SIMKA database and/or individual TB patient case notes as necessary (Paul, 2019). Data

Cleaning was performed to screen out duplication and clerical error. Population data was obtained from the Malaysian Department of Statistics (DOSM, 2020). Population socio-demographics of the isolate-patient matches were studied based on variables obtained from the MyTB database. Patient occupations were categorized according to socio-economic classifications (Connelly et al., 2016). Further analyses were done to correlate spoligopatterns with gender, age group, ethnicity, occupation and co-morbidities. Latitude and longitude coordinates of the cases were obtained with permission from the Sabah TB/Leprosy Control Unit. Cluster maps were built to see the distribution of TB cases in this study. MTB distribution was correlated with epidemiological data. using qGIS version 2.18 Las Palmas, open source, available at: https://www-qgis-org.translate.goog/en/site/forusers/download.html?_x_tr_sl=fr&_x_tr_tl=en&_x_tr_hl=fr

3.4.2 Isolate selection

MTB isolates cultured from patient clinical specimens from the whole state of Sabah were obtained from the MTB culture stock collection in MKAKK. 1060 unique isolates of MyTB-registered patients from all districts in Sabah were obtained from the archived stock culture (2015-2016) and prospective collections (May 2017-May 2018) in MKAKK. The inclusion criteria of the sample population are for the MTB isolates to have been cultured from MyTB-registered TB patients, and furnished with DST results. The sample population represent MTB Isolates and MyTB-registered patients from all

districts in Sabah, except for Kuala Penyu district that did not have any positive culture at the time (Table 3.1).

Table 3.1: Number of MTB isolates sampled by districts in Sabah (random selection of samples, with a MyTB registry match)

NO	DISTRICT	% of TB Patients in Sabah	Number of Isolates (Planned)	Number of Isolates (Actual)
1	KOTA KINABALU	16.7	167	227
2	SANDAKAN	11.5	115	113
3	TAWAU	10.1	101	87
4	LAHAD DATU	6.9	69	58
5	SEMPORNA	6.3	63	74
6	KENINGAU *	6.0	60	7
7	PAPAR	4.5	45	59
8	TUARAN	4.5	45	45
9	PENAMPANG	4.0	40	38
10	KOTA BELUD	3.0	30	36
11	RANAU	2.9	29	26
12	KOTA MARUDU	2.7	27	34
13	KUDAT	2.6	26	32
14	PUTATAN	2.4	24	28
15	KINABATANGAN	2.2	22	20
16	BELURAN	2.2	22	11
17	KUNAK	1.9	19	22
18	BEAUFORT	1.9	19	25
19	TENOM	1.6	16	19
20	PITAS	1.4	14	21
21	NABAWAN	1.2	12	7
22	SIPITANG	1.1	11	11
23	TAMBUNAN	1.0	10	12
24	TONGOD	0.9	9	7
25	KUALA PENYU	0.4	4	0
	TOTAL	100.0	1000	1019

*Facility is a district AFB culture centre and their presumptive pulmonary MTB isolates were transported directly to National Public Health Laboratory (MKAK) for further investigation.

Table 3.2: MTB isolates according to year of isolation and patient country of origin (N=1019)

Year	INDONESIA	MALAYSIA	PHILIPPINES	Total	
2015	2	43	12	57	archived
2016	9	136	38	183	
2017	44	396	180	620	prospective
2018	33	99	27	159	
Total	**88**	**674**	**257**	**1019**	

The MTB isolates in this study were randomly selected from the MTB 2015-2016 archive, and the 2017-2018 routine AFB cultures in MKAKK. The breakdown of isolates according to nationalities are Indonesian (8.6%), Malaysians (66.1%) and Filipinos (25.2%). The numbers are not equally distributed due to the limited availability of the archived isolates.

3.4.3 AFB microscopy

For prospectively collected specimens, 2.0cm x 3.0cm smears were prepared according to WHO guidelines (Lumb et al., 2013; Angra et al., 2017) and stained using the Auramine O stain (NAT, Nazca Scientific SB). Smears were examined at 400x magnification using an LED-based fluorescence microscope (Zeiss Primostar iLED). The grading of smears was done according to WHO / IUATLD standard. For the archived stock, AFB microscopy, culture, drug susceptibility test (DST) results were retrieved from the MKAKK TB and the Public Health Laboratory Information System (SIMKA) database.

3.4.4 AFB culture

Samples were cultured using Modified Ogawa Method, and incubated at 37°C until growth was observed, or for up to 8 weeks. Cultures were examined weekly for growth and were graded according to the WHO/IUATLD standard (Stinson et al., 2014). Culture results were recorded in the laboratory registry (TBIS102B) (MOH, 2002; Liew et al., 2015).

3.4.5 Isolate identification and drug susceptibility testing

MTB isolates (N=1019) were identified via the detection of MPB64 antigen using a rapid lateral flow immune-chromatographic assay (Capilia^{TM} TB-Neo, Tauns Laboratories Incorporated, Fujisawa, Japan). In this study, resistotyping using microbead-based assay (TB-RINT, Beamedex France) was performed on all the samples, and drug susceptibility test (DST) results were compared with phenotypic DST results (Absolute concentration method, ACM). Line Probe Assay (LPA MDRPlus) was also performed for some of the isolates by MKAKK as part of TB case investigation. The LPA MDRplus data is included for comparison.

3.4.6 Line probe assay for the detection of MTB / MDRTB (LPA)

MTB isolates that fulfilled the criteria for testing i.e., DRTB or MDRTB contacts, TB relapse case and cases with retroviral disease were subjected to Line Probe Assay, LPA MDRplus (GenoType MTBDRplus, Hain-Lifescience). The assay was performed according to manufacturer's instruction.

3.4.7 DNA extraction

DNA Extraction of the selected isolates was performed in the MKAKK High Containment TB Laboratory according to strict guidelines (Sola et al.,

2016). DNA was extracted from Ogawa culture slants using ZR Fungal/Bacterial DNA Miniprep (Quick DNA extraction Kit, Zymo Research) following the manufacturer's extraction protocol. DNA quality assessment was done using NanoPhotometer P 300 (Implen) in UTAR. Quality and quantity of the DNA samples were assessed using gel electrophoresis at 0.8% agarose.

3.5 Genotyping of MTB Isolates using a two-tiered classical genotyping

3.5.1 TB-SPRINT (TB SPOL and TB-RINT)

TB-SPRINT is a combination of molecular genetic assays to genotype (spoligotype) MTB (TB-SPOL) and simultaneously identify its genetic resistance to rifampin and isoniazid (TB-RINT). Genotyping was performed on 1060 using the spoligotyping technique (Kamerbeek et al., 1997; Cowan et al, 2004; Gomgnimbou, 2012; Gomgnimbou et al., 2013), whereas RIF and INH resistance were assessed by identification on previously identified single nucleotide polymorphisms (SNPs). Rifampicin resistance interrogation was done for position 516, 526, 531 within *rpoB* gene; and the lesser frequent mutations in the rifampicin resistance determining region, RRDR (81 bp). Meanwhile high-level and low-level resistance to INH was studied by targeting respectively the *katG* gene (encoding for the catalase oxydase; by detection of the 315-codon mutation) and the promoter region of the *inhA* gene (coding for the NADH enoyl ACP reductase; by detecting the -8 and -15 mutations).

Spoligotyping (TB-SPOL) and Resistotyping (TB-RINT) were performed separately on MagPix® as there were overlaps of bead ID for TB SPOL and TB-RINT kits. The available capacity for testing is 50 analytes. The numerical output transferred to MS Excel macro enabled rapid and objective interpretation (Sola et al., 2016). Optimisation and familiarisation of this method were done in the course of 2 weeks. Spoligotyping of the samples at the rate of 48 samples per run took approximately 20 days (~10 hour-work day). This is less than the time needed for membrane-based spoligotyping. The MagPix® is temperature sensitive and a temperature change of ± 5°C would halt the analysis. Good air conditioning was necessary to maintain the laboratory environment temperature, ideally at 20-25°C. Data acquisition may be resumed upon immediate rectification of the temperature.

3.5.1.1 Spoligotyping TB-SPOL (43-Plex) DNA amplification

PCR cocktail of 12.5µL x GoTaq, Promega Mastermix, 2.5µL 17-20µM PCR Dra & Drb Primers mix (TB-SPOL, Beamedex Fr. kit), 2 µL 10ng DNA template and 8 µL nuclease free water was prepared for a total volume of 25 µL per reaction. Amplification of DNA was performed in the Veriti Thermal Cycler (Applied Biosystems ABI) using the following profile: 1 cycle of 96°C for 3 minutes; 25 cycles of 96°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds; 1 cycle of 72°C for 5 minutes. A non-template control (NTC) was included in every run. The PCR products were resolved with gel electrophoresis at 2 % agarose (Sola et al., 2016).

3.5.1.2 RIF-INH typing TB-RINT (16-Plex) DNA amplification

PCR cocktail of 12.5 µL x GoTaq, Promega Mastermix, 1.25 µL Mix RIF/INH Forward primer, 1.25 µL Mix RIF/INH Reverse primer (TB-RINT, Beamedex Fr. kit), 2 µL 10ng DNA template and 8 µL nuclease free water was prepared for a total volume of 25 µL per reaction. Amplification was performed in the Veriti Thermal Cycler (Applied Biosystems [ABI]) using the following profile: 1 cycle of 96°C for 3 minutes; 25 cycles of 96°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds; 1 cycle of 72°C for 5 minutes. A no-template control (NTC) was included in every run. The PCR products were resolved with gel electrophoresis at 2% agarose (Sola et al., 2016).

3.5.1.3 Hybridization for TB-SPOL & TB-RINT

Reagents 1.5xTMAC, 1xTMAC and TE for the hybridization step of TB-SPOL and TB-RINT were prepared according to Appendix C. The reagents were prepared fresh to ensure optimal hybridization. Hybridization steps were done according to steps 1 through 15 (Sola et al., 2016).

1. The coupled-microspheres were resuspended by vortexing for 40 seconds each to ensure proper mixing.
2. The microspheres working mix, MWM (50 beads/µL/target) were prepared in a total volume of 33 µL of 1.5X TMAC (hybridization solution) per sample to analyse, using the coupled bead stocks provided

(TB-SPOL 43-plex, TB-RINT 16-plex) (Magplex: TB-SPOL 0.86 + 32.14; TB-RINT 0.32 + 32.68)

3. The MWM were vortexed for 40 seconds each.
4. 33µL MWM was dispensed in each sample well and control.
5. 17µL of TE, pH=8 was dispensed in the negative control well
6. In the well which contains samples, the biotinylated PCR-amplified DNA and TE were added, pH=8 in a total volume of 17µL (2 µL DNA & 15 µL TE) (Note: 2-5 µL of a robust PCR is generally sufficient for a good hybridization-detection)
7. The samples were mixed gently by pipetting up and down
8. Sealing Tape was placed onto the reaction plate to prevent evaporation and incubated at 98°C from 5 to 10 minutes to denature the amplified biotinylated DNA (PCR product).
9. Hybridization was done at 52°C for 20 minutes in the Veriti Thermal Cycler (Applied Biosystems [ABI]).
10. The plate was centrifuged for 7 minutes at maximum speed (4000rpm on a VWR Megastar 1.6R, with rotor Thermoscientific 75003624 and buckets 50115385 or equivalent) discard as much as possible of the supernatant (25-40µL) by pipetting carefully.
11. The beads were resuspended by adding TMAC in equal amount as the discarded TE in step 10 (25-40 µL).
12. Reporter mix was prepared fresh by adding 4µL of Streptavidin-R-phycoerythrin (1mg/mL stock) to 996 µL of TMAC 1X Hybridization solution.

13. 25µL fresh reporter mix was added to each well and gently mixed by pipetting up and down.
14. The plate was incubated at 52°C for 5 minutes within the MagPix® device.
15. 75µL samples were analysed at 52°C with the MagPix® device following the Software Instructions. Data acquisition was performed as per instructions in the MagPix® XPONENT manual supplied by BMSD (Sola et al., 2016).

3.5.1.4 Quality control

Three controls were systematically performed for internal spoligotyping quality controls.

1. non-template control (NTC): Molecular grade water that has undergone all the PCR and detection process, provide low RFI values to be negative.
2. Positive Control 1, PC1 (*M. bovis*) with spoligopattern: absence of spacers 3,9,16, 39- 43, and presence of all other spacers.
3. Positive Control 2, PC2 (*M. tuberculosis* H37Rv with spoligopattern: absence of spacers 20-21 and 33-36, and presence of all other spacers.

The positive controls, PC1- *M.bovis* BCG P3 DNA and PC2- MTB H37Rv DNA were provided in the TB-SPRINT, Beamedex kit.

3.5.1.5 Computer-assisted interpretation of results (www.beamedex.com)

The raw results obtained on the Magpix®) analysis were provided under Relative Fluorescence Intensity (RFI) Units and were internally compared to the previously determined cut-offs to decide on the presence/absence of the CRISPR (Direct Repeat) locus spacers and on the wild-type/mutated status of first-lane drug resistance (Rif and Inh) associated genes. This is semi-automatically done by an Excel macro (TB-SPRINT-matrix) provided by Beamedex, Fr. which include 5 sheets namely plate map description, raw data, values distribution, TB-SPRINT profiles, and TB-SPRINT final-display.

CSV results file produced by the Exponent® software after conversion to XLS format, were copied from the software. The data were then transferred into the Raw data worksheet and the interpretation was then automatically performed by Macros as can be seen on the SPRINT profiles worksheet with "n" and "o" format and "black square" (positive) or "white square" (negative) display interpretation for CRISPR results or wild-type/mutated status for the drug-resistance mutations. The square-format interpreted results were copied from the SPRINT profiles worksheet and into the SPRINT final display worksheet and drug resistance status will be displayed. This worksheet provided the final display of all profiles and resistance profiles. A final investigator-driven correction / edit of any individual "doubtful" results were introduced in the final display worksheet. Detailed instruction for data acquisition can be obtained from **www.beamedex.com.**

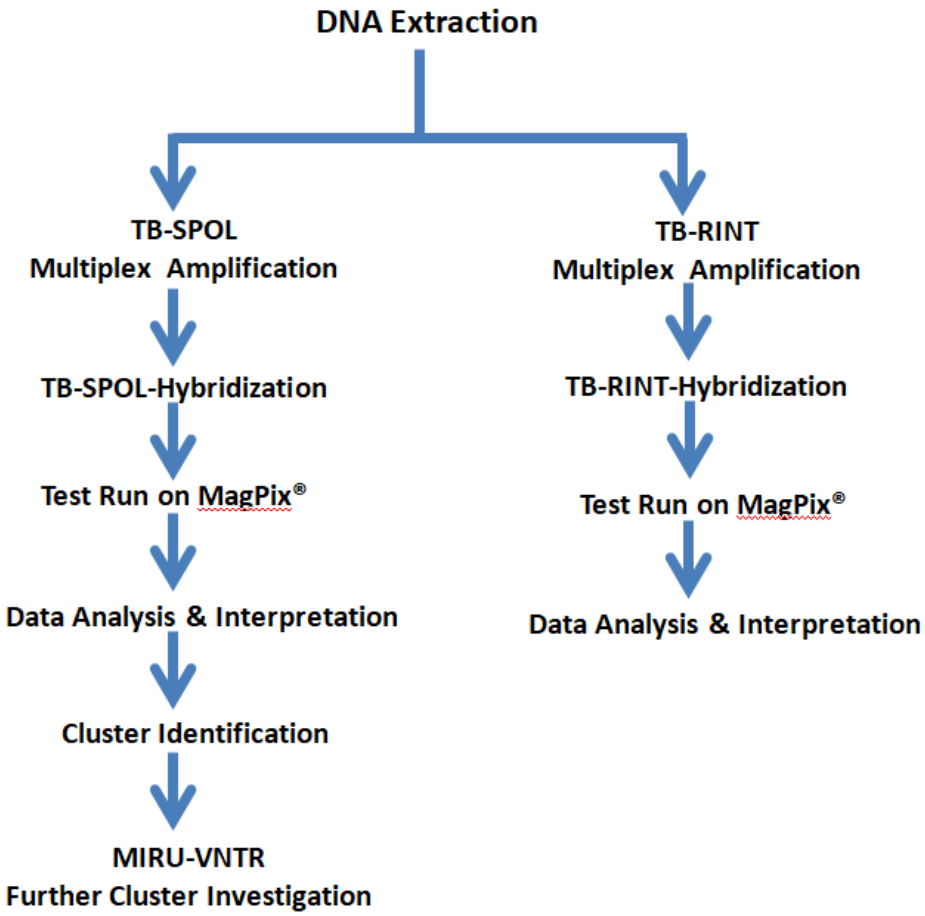


Figure 3.2: Two-tiered genotyping of MTB using spoligotyping and 24-loci MIRU-VNTR, and resistotyping

Figure 3.2 shows the workflow of this study using the two-tiered genotyping method. TB-SPOL (43-plex) and TB-RINT (16-plex) were performed separately on the MagPix® as this platform could only support a maximum of 50-plex at a time.

3.5.2 24-loci MIRU-VNTR typing

24-loci MIRU-VNTR analysis, using a duplex loci PCR amplification method modified by Yasmin et al. (2016) was performed on 593 selected spoligotyped isolates to further discriminate the isolates identified in the clusters of interest (Yasmin et al., 2016; Supply, 2005). These isolates represent all the major spoligotype clusters and orphan spoligotypes with known and unknown lineage signatures in this study. The optimised duplex 24-loci MIRU-VNTR refers to the duplex PCR whereby two, or a pair of VNTR loci that harbours distinct amplicon sizes regardless of the repetition number of each locus are amplified together in a single tube (Table 3.3). This reduces the cost of reagents and consumables into half, as well as the time needed especially to perform PCR and agarose gel electrophoresis (Yasmin et al., 2016).

All primers used in this study were synthesized by Integrated DNA Technologies (Table 3.4). The PCR analyses were done according to the optimised conditions summarized in Table 3.3, which is a slight modification of PCR described previously (Le Flèche et al., 2002; Yasmin et al., 2016). 2µL of concentrated or diluted DNA (5-70 ng) were added as the template. The PCR conditions would have the final concentrations as follows: Tris-HCl (pH ¼ 8.75) 20 mM; KCl 10 mM; (NH4)2SO4 10 mM; MgSO4 2 mM; MgCl2 1.5 mM; Triton 0.1%; dNTP 0.3 mM; Betain 250mM; DMSO 5%; 0.3U Taq DNA polymerase or using GoTaq (Promega); 0.5 or 0.75 mM of each primer (Le Flèche et al., 2002; Yasmin et al., 2016). Amplification program is selected based on the duplex tested as shown in Table 3.3.

4 to 8 µL of PCR product (Table 3.3) were loaded on a 2% agarose gel no further than 4 wells of a 100bp and 50bp DNA ladder (Quick Load Purple DNA Ladder, New England Biolabs) covering the range 100-1000 bp. This was followed by agarose gel electrophoresis run at 70V for 70 minutes.

Allele calling was done using the band size chart in Table 3.5 (Yasmin et al., 2016) and a MIRU-VNTR code was generated for every DNA extract. The phylogenetic link between the isolates from this investigation and the reference isolates in the *MIRU-VNTRplus* was then assigned using the codes in clustering analysis. (http://www.*MIRU-VNTRplus*.org/MIRU/index.faces) (Allix-Béguec et al., 2008; Allix-Béguec et al., 2014).

Table 3.3: Optimised duplex MIRU-VNTR PCR cocktail

Primer X1		D1.1 ETR A	D2.1 ETR B	D3.1 ETR C	D4.1 ETR D-1 / MIRU 04	D5.1 MIRU 02	D6.1 Qub 11b	D7.1 Mtub 30	D8.1 MIRU 24	D9.1 Mtub 29	D10.1 Mtub 34	D11.1 Mtub 04	D12.1 Mtub 21
Primer X2	Volume/rxn	D1.2 MIRU 39	D2.2 Qub 26	D3.2 MIRU 20	D4.2 MIRU 10	D5.2 MIRU 27	D6.2 MIRU 16	D7.2 Miru 23	D8.2 ETR E / MIRU 31	D9.2 MIRU 26	D10.2 MIRU 40	D11.2 Qub 4156	D12.2 Mtub 39
Reagent / Duplex		D1 (x10)	D2 (x10)	D3 (x10)	D4 (x10)	D5 (x10)	D6 (x10)	D7 (x10)	D8 (x10)	D9 (x10)	D10 (x10)	D11 (x10)	D12 (x10)
dNTP (2mM)	2	20	20	20	20	20	20	20	20	20	20	20	20
Buffer Q *10X	1.5	15	15	15	15	15	15	15	15	15	15	15	15
Betain 5M	3	30	30	30	30	30	30	30	30	30	30	30	30
DMSO	0.75	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5	7.5
Taq Pol (0.3/uL)	0.7	7	7	5	7	7	8	8	7	7	5	7	7
Colour Buffer	0.7	7	7	5	7	7	8	8	7	7	5	7	7
NFW	variable 1.85	18.5	18.5	27.5	23.5	23.5	16.5	21.5	23.5	18.5	22.5	18.5	18.5
X.1-F	0.5	7.5	5	5	5	5	7.5	5	5	5	5	5	5
X.1-R	0.5	7.5	5	5	5	5	7.5	5	5	5	5	5	5
X.2-F	0.75	5	7.5	5	5	5	5	5	5	7.5	7.5	7.5	7.5
X.2-R	0.75	5	7.5	5	5	5	5	5	5	7.5	7.5	7.5	7.5
TOTAL Master Mix	11.15	130	130	130	130	130	130	130	130	130	130	130	130
Program		Prog. 1	Prog. 3/1	Prog. 1	Prog. 1	Prog. 1	Prog. 2	Prog. 1	Prog. 2	Prog. 1	Prog. 1	Prog. 2	Prog. 2
DNA Load Volume (uL) /rxn		8	7	3.5	4	4.5	5	4.2	5	4	4.2	5	6.2

Note: The highlighted cell indicates that there are two different programs to be used whereby Prog 3 is used for D2.1 – ETRB, and Prog 1 for D2.2 – Qub26

Table 3.4: Primers for duplex 24-loci MIRU-VNTR panel

Duplex	Loci / Minisatellite Name	Forward Primer	Reverse Primer
1.1	ETR A	ATTTCGATCGGGATGTTGAT	TCGGTCCCATCACCTTCTTA
1.2	MIRU 39	CGCATCGACAAACTGGAGCCAAAC	CGGAAACGTCTACGCCCCACACAT
2.1	ETR B	GCGAACACCAGGACAGCATCATG	GGCATGCCGGTGATCGAGTGG
2.2	Qub 26	GGCCAGGTCCTTCCCGAT	AACGCTCAGCTGTCGGAT
3.1	ETR C	GACTTCAATGCGTTGTTGGA	GTCTTGACCTCCACGAGTGC
3.2	MIRU 20	TCGGAGAGATGCCCTTCGAGTTAG	GGAGACCGCGACCAGGTACTTGTA
4.1	ETR D-1 / MIRU 04	GCGCGAGAGCCCGAACTGC	GCGCAGCAGAAACGTCAGC
4.2	MIRU 10	GTTCTTGACCAACTGCAGTCGTCC	GCCACCTTGGTGATCAGCTACCT
5.1	MIRU 02	TGGACTTGCAGCAATGGACCAACT	TACTCGGACGCCGGCTCAAAAT
5.2	MIRU 27	TCGAAAGCCTCTGCGTGCCAGTAA	GCGATGTGAGCGTGCCACTCAA
6.1	Qub 11b	CGTAAGGGGGATGCGGGAAATAGG	CGAAGTGAATGGTGGTGGCAT
6.2	MIRU 16	TCGGTGATCGGGTCCAGTCCAAGTA	CCCGTCGTGCAGCCCTGGTAC
7.1	Mtub 30	AGTCACCTTTCCTACCACTCGTAAC	ATTAGTAGGGCACTAGCACCTCAAG
7.2	Miru 23	CAGCGAAACGAACTGTGCTATCAC	CGTGTCCGAGCAGAAAAGGGTAT
8.1	MIRU 24	CGACCAAGATGTGCAGGAATACAT	GGGCGAGTTGAGCTCACAGAA
8.2	ETR E / MIRU 31	CTGATTGGCTTCATACGGCTTTA	GTGCCGACGTGGTCTTGAT
9.1	Mtub 29	AACCCATGTCAGCCAGGTTA	ATGATGGCACACCGAAGAAC
9.2	MIRU 26	CCCGCCTTCGAAACGTCGCT	TGGACATAGGCGACCAGGCGAATA
10.1	Mtub 34	GCAGATAACCCGCAGGAATA	GGAGAGGATACGTGGATTTGAG
10.2	MIRU 40	GGGTTGCTGGATGACAACGTGT	GGGTGATCTCGGCGAAATCAGATA
11.1	Mtub 04	GTCCAGGTTGCAAGAGATGG	GGCATCCTCAACAACGGTAG
11.2	Qub 4156	TGACCACGGATTGCTCTAGT	GCCGGCGTCCATGTT
12.1	Mtub 21	AGATCCCAGTTGTCGTCGTC	CAACATCGCCTGGTTCTGTA
12.2	Mtub 39	AATCACGGTAACTTGGGTTGTTT	GATGCATGTTCGACCCGTAG

Source: Yasmin et al. (2016)

Table 3.5: Correspondence table between amplicon sizes and number of repeats for each VNTR, presented for each duplex

VNTR Markers	Position on Chr. in kb	Repeat Size	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ETR-A	2165	75bp		(247)	**322**	**397**	**472**	(547)	(622)	(697*)	(772)	(**847***)	(922)	[997]				
MIRU39	4348	53bp		(593)	**646**	**699***	(752)	[805]	[858*]									
ETR-B	2461	57bp	[121]	**178**	**235**	**292**	(349)	(406)	(463)	[520]	[577]	[634]						
Qub26	4052	111bp		**264**	**375**	**486**	**597**	**708**	**819**	**930**	**1041**	1152	1263	1374	1485			
ETR-C	0577	58bp		[172]	**230**	**288**	**346**	**404**	(462)	(520*)	(578)							
MIRU20	2059	77bp		**514***	**591**	[668]												
MIRU04 (ETRD-1)	0580	77bp		(253)	**330***	**407**	(484)	**561**	(638*)	[715]	[792*]	[869]	[946]					
MIRU10	0960	53bp		[535]	(588)	**643***	**696**	**749**	(802*)	(855)	[908]	[961]	[1013]	[1065]				
MIRU02	0154	53bp		(455)	**508**	(561)	[614]											
MIRU27 (QUB5)	3006	53bp		(551)	(604)	**657**	**709**	[762]										
Qub11b	2163	69bp		[136]	**205**	**274**	**343**	**412**	**481**	(550)	(619*)	(688*)	[757]	[826*]	[895]			
MIRU16	1644	53bp		(618*)	**671***	**724**	**777**	(829*)	(882)	[935]	[988]	[1041]	[1094]	[1147]				
Mtub30	2401	58bp		**261**	**319**	(377)	**435**	(493)	[551]	[609*]								
MIRU23	2531	53bp		[607*]	(660)	(713)	**766**	**819**	**872**	(925)	(978)	[1031]	[1084]	[1137]				
MIRU24	2687	53bp		**447**	**500**	(553)	[606]	[659]	[712]									
MIRU31 (ETR-E)	3192	53bp		(545)	**598**	**651**	**704**	**757**	(810)	(863)	(916)							
Mtub29	2347	57bp		[179]	(236)	**293**	**350**	(407)	[464*]									
MIRU26	2996	51bp	[460*]	(511)	**562**	**613**	**664**	**715**	(766)	**817**	(868)	(919)	[970]	[1021]	[1072]	[1123]		
Mtub34	3171	54bp		(171)	**225**	**279**	(333)	(387)	[441]	[495]								
MIRU40	0802	54bp		**407**	**461**	**515**	**569**	**623**	**677**	(731)	(785)	[839]	[893]					
Mtub04*	0424	51bp	(177)	**218**	**269**	**320**	**371**	**422**	(473)	(524)	(575)	[626*]	[677*]	[728*]				
Qub4156*	4156	59bp	[563]	**622**	**681**	**740***	**799**	[858]	[917]	[976]	[1035]							
Mtub21	1955	57bp		**149**	**206**	**263**	**320**	**377**	(434)	(491)	(548)	(605)	(662)	[719]	[776]	[833]	[890]	
Mtub39*	3690	58bp		**283**	**341**	**399**	**457**	**515**	(573)	(631)	(689)	(747)	(805)	[863]	[921]	[979]	[1037]	[1095]

Amplicon sizes observed less than 6 times (brackets); sizes found in less than 3 times (square brackets); H37Rv Paris strain (underlined); Ambiguous amplicon sizes (proximity between 2 amplicon sizes observed in the same duplex) are highlighted by an asterisk.
Source: Yasmin et al. (2016).

3.5.3 Lineage designation and diversity analysis

A standardized lineage and sub-lineage designation is necessary to allow inter-laboratory comparison of genotyping data, not only on the national level but also globally. The following platforms and databases can be used for this purpose.

3.5.3.1 Spoligotyping analysis using web-based databases SITVIT2, SpolSimilarity Search and SpolLineages tool

There are many web-based database that can be freely accessed online for the purpose of assigning spoligotype and sub-lineages to MTB isolates collection. In this study, three platforms were utilised as listed below.

1. SITVIT2

The SITVIT2 is a simple and user-friendly web-based platform to determine Spoligotype / spoligopattern and to search for the distribution of Spoligotypes worldwide (as is available in the database). Spoligotype patterns in the octal format were keyed in individually, or uploaded as a list in a microscoft excel template provided by the platform onto the SITVIT2 proprietary database of the Institut Pasteur de la Guadeloupe (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/) (Couvin et al., 2017).

Examples on how to generate spoligotype pattern in the binary and octal format is shown in Figure 3.3. The results (octal format) were copy pasted onto any Microsoft office documents or its equivalent, or downloaded as a microsoft excel file. Spoligotype International Type (SIT) were assigned to the isolates if the pattern was shared by 2 or more isolates in the database. Patterns that failed to locate a match in the database, were given the "orphan" designation. According to SITVIT2 guidelines, the different genotypic lineages were either designated as *M. africanum, M. bovis, M. canettii, M. microti, M. pinnipedii*, or MTB lineages *sensu stricto*. Genome-based L1-L7 major lineage and L8-L9 minor labels were added to spoligotype families e.g. SIT19/EAI2-Manila (L1) and SIT1/Beijing (L2) (Couvin, et al., 2019). The definition of macrogeographical regions and sub-regions was done according to the United Nations Scheme (http://unstats.un.org/unsd/methods/m49/m49regin.htm).

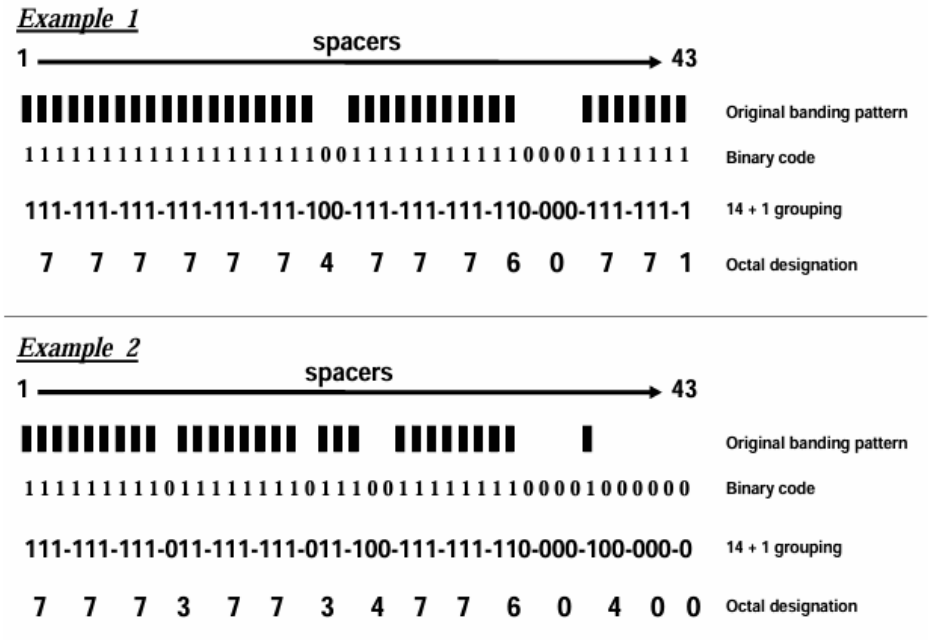


Figure 3.3: Examples of spoligotyping results

Spoligotyping results (membrane or microbead-based assay) are shown as banding patterns based on the presence (1 or n) or absence (0 or o) of spacers in the standard 43-spacer format. The 43-digit binary code is converted to a 15-digit octal (i.e., base 8, having the digits 0-7) designation by a two-step process. First, the 43-digit binary code is divided into 14 sets of three digits (spacers 1 through 42) plus one additional digit (spacer 43). Second, each 3-digit binary set is converted to its octal equivalent, with the final additional digit remaining as 1 or 0. The translation of binary numbers to octal numbers is done as follows: 000 = 0; 001 = 1; 010 = 2; 011 = 3; 100 = 4; 101 = 5; 110 = 6; 111 = 7. Each octal designation is unique, representing one specific banding pattern. The binary code of the spoligotyping pattern can be re-created based on the octal designation (Dale et al., 1999; Navin et al., 2004). The octal designation is the form of the result that is reported by the genotyping laboratories to TB programs.

2. SpolSimilarity Search

Spoligotype octal was keyed in individually, or as a list in a MS Excel template provided by the platform. With information on the presence or absence of 43 spacers, the potential origin of an isolate with a similar spoligopattern, and its global dispersion, a comprehensive overview of similar spoligotype patterns was obtained. Similar spoligotype patterns were compared to examine the dissemination and evolutionary patterns of MTB to distinguish between broad, specialized, and/or limited patterns as well as to identify patterns with significant deleted blocks in this study. Country distribution patterns were also discovered for each spoligotype queried. The results of every query were downloaded or copy pasted onto Microsoft word documents or its equivalent. The SpolSimilaritySearch web-tool is available at: http://www.pasteurguadeloupe.fr:8081/SpolSimilaritySearch)

3. SpolLineages Tool

The SpolLineages Tool was used to predict and determine novel MTB lineages / sub-lineages from spoligotyping and/or MIRU-VNTR profiles which have not yet been reported in the SITVIT2 or other MTB databases using software tools developed by Institut Pasteur de la Guadeloupe and Universite des Antilles. The prediction system is based on:

1) data transformation and decision tree classifiers; and

2) searches for a set of straightforward rules utilizing binary masks through an evolutionary approach specifically built for MTB spoligotype family identification (Couvin et al., 2020).

The SpolLineages Tool software is available for access at: https://github.com/dcouvin/SpolLineages

3.5.3.2 MIRU-VNTR analysis using *MIRU-VNTRplus* and Bionumerics

MIRU-VNTRplus is a highly specialized bioinformatics web tool that can be used to analyse MIRU-VNTR data and public reference databases. To determine the MTB species, lineages, and genotypes, the strains in this study were compared to the 186 *MIRU-VNTRplus* reference strains. Along with the 24 MIRU-VNTR loci copy numbers, spoligopatterns, regions of difference (RD), susceptibility data, single nucleotide polymorphisms (SNPs), and IS6110 RFLP fingerprint images, each reference strain has its own information on species, lineage, and epidemiology (Allix-Béguec et al., 2008; Weniger et al, 2010). The server also maintains the universal expanding nomenclature (MLVA MTB15-9) for naming various MIRU genotypes. Comparisons were made using data from MIRU, spoligo, RD, susceptibility-typing, SNP, or a mix of these data types. Using UPGMA or neighbour-joining clustering methods, the dendrograms were generated based on the relevant distance matrix. Using *MIRU-VNTRplus* and Bionumerics 7.6 (Applied-Maths), the Minimum

Spanning trees (MST) were constructed based on cluster analysis queries for the dataset used in this study. *MIRU-VNTRplus* is available for access at: https://www.miru-vntrplus.org/MIRU/index.faces

For lineage assignment and cluster analysis, the genotyping data was uploaded to the *MIRU-VNTRplus* platform at www.miru-vntrplus.org (Allix-Béguec et al., 2008; Weniger et al., 2010). The "similarity search" module, which compares the combined spoligotype and MIRU genotypes against a database of 186 reference strains that represent key global MTB complex lineages, was used to determine the strain lineage for each isolate (Allix-Béguec et al., 2008; Weniger et al., 2010). Lineage matching was performed using MIRU and spoligotype sequences and category distance metrics. The global collection of TB lineages (SITVIT2, SpolSimilaritySearch, and SpolLineages Tool) was then used to compare the characterized lineages. "Orphan" strains are those that have no matching spoligopattern in SITVIT2. (Demay et al., 2012; Rodwell et al., 2012).

3.6 Genotyping-complemented case investigation studies

A retrospective genotype-complemented TB contact investigation study was conducted in order to see if there was any predictive power or capability to determine epidemiological links to complement conventional contact investigation. TB cases with genotyped isolates using the two-tiered scheme were studied. Cluster analyses of cases using spoligotyping and MIRU-VNTR

data were done. When two or more MTB isolates from various patients evaluated during the study period had the same spoligotype and 24-loci MIRU-VNTR genotypes, the TB cases were considered to be clustered and were referred to as such. Singletons are strains that do not have genotypes that match any of the other isolates in the study (Demay et al., 2012; Rodwell et al., 2012).

The genetic relatedness of clustered strains was determined by analyzing the frequency of locus variations using the *MIRU-VNTRplus* platform's "minimum spanning tree" (MST) module. The MST of MTB isolates (n=556) was built using the *MIRU-VNTRplus* platform. The evolutionary connections between the MTB spoligotypes were depicted by MST. The degree of changes necessary to switch from one allele to another is used to connect each genotype on the phylogenetic tree. The distance between the patterns is indicated by the length of the branches. Thicker lines imply a single change, whereas thinner lines indicate 2 or 3 changes, and the thickness of the lines indicates how many spacer changes there are between two patterns. The MTB strains with related MIRU genotypes were grouped and structured into clonal complexes (CC) by the MST analysis (Allix-Beguec et al., 2008; Weniger et al., 2010). In this investigation, CC were defined as collections of strains with a maximum of two locus variations (DLV2) from either the central strain or any other strain in the complex (Allix-Béguec et al., 2008). CC and exact 24-loci MIRU-VNTR clustering were then analysed in terms of distribution of cases by sub-lineages, nationality, location, and occupation. Transmission and reactivation rates were calculated for conventional contact investigation and genotype-complemented contact investigation.

Predictive values of notification interval between first case and second TB case in term of resulting cluster size were studied. Molecular clonality of the strains was considered as direct proof of epidemiological link, identify reservoirs and risk factors for targeted control in the transmission study. Information obtained from conventional contact investigation was compared with the information derived from phylogenetic and geospatial analyses. Genotype and geospatial clusters were compared with conventional contact investigation clusters in order to see if there were any overlaps in information or generation of new information. A comparison between the operational and theoretical method of genotype-complemented contact investigations was performed. The feasibility, value and limitation of integrating genotyping in the existing TB surveillance was assessed.

3.7 Statistical analysis

Statistical Package for the Social Sciences (SPSS) version 22.0 was used to conduct the statistical analysis. Using Chi-square analysis and Fisher's exact test, the relationship between MTB clades and patient demographic and treatment resistance criteria was evaluated. The 95% confidence interval (95% CI) was determined and P values of 0.05 were deemed statistically significant. The age of the patients was determined using the mean and standard deviation (SD).

CHAPTER 4

RESULTS

4.1 General performance of AFB culture and DST in Sabah

The Kota Kinabalu Public Health Laboratory, MKAKK reported approximately 11000 AFB cultures were performed annually for an estimated 5000 patients with 75% culture coverage of registered TB patients throughout 2014 to 2016 (Paul, 2019). Goroh et al. (2020a) reported similar findings for the period of 2012 to 2018 in their work on Sabah TB epidemiology.

AFB culture positivity is generally about 50-60%, as culture were requested for different purposes, and at different phases of TB investigations e.g., diagnostics for presumptive TB cases, baseline culture for newly diagnosed TB patients, high DRTB / MDRTB risk cases and MDRTB case monitoring. Approximately 5% of the growth positive AFB cultures exhibited poor growth due to various factors which might include low bacterial load, exposure to high temperatures and other harsh environments during transportation to AFB culture facilities due to logistic issues (Paul et al., 2022). Another reason for poor AFB growth could be associated with drug resistance-related MTB fitness issues (Borrell and Trauner, 2017; Gagneux, 2009).

AFB cultures were identified as MTB in about 95% of the cases. For the same time period, DST is done for just under 90% of the total MTB isolates after eliminating isolates not fulfilling the standard requirement for DST whereby a minimum growth of MTB colonies with a 1+ grading is necessary for a successful DST (Stinson et al., 2014). About 2% of the positive AFB cultures were identified as non-tuberculous mycobacteria (NTM). The Sabah AFB culture performance for the period of 2014 to 2016 is shown in Table 4.1.

Table 4.1: AFB culture performance in Sabah 2014-2016 (by patient)

Parameter / Year	2014	2015	2016
patient with at least 1 specimen sent for AFB culture	3427 (72.5%)	3366 (75.4%)	3883 (78.4%)
patient with NO specimen sent for AFB culture	1301 (27.5%)	1098 (24.6%)	1070 (21.6%)
Total number of patient	4728 (100%)	4464 (100%)	4953 (100%)
patients registered ≥2 times at different times in the same year (with AFB culture) - change of diagnosis / defaulters	19 (0.40%)	17 (0.38%)	20 (0.40%)
Total patient with AFB growth (% AFB growth/culture)	1887 (55.1%)	1998 (59.4%)	2154 (55.5%)
Patient with MTB growth (%MTB growth/AFB growth)	1782 (94.4%)	1819 (91.0%)	2027 (94.1%)
DST done (%DST done/MTB identified)	1587 (89.1%)	1690 (88.5%)	1722 (85.0%)
Patient with NTM growth (% NTM/AFB growth)	26 (1.4%)	49 (2.5%)	42 (1.9%)
Patient with mixed MTB and NTM growth (% [MTB+NTM]/AFB growth)	3 (0.16%)	4 (0.20%)	8 (0.37%)
Poor growth of AFB (unable to proceed testing) - issues related to fitness cost?	79 (4.2%)	130 (6.5%)	85 (3.9%)

Table 4.2: AFB smear grading in this study

Origin/ Grading	Negative	% Negative (Total)	Scanty	% Scanty (Total)	High Grade	% High Grade (Total)	Grand Total
INDONESIA	18	21.7 (7.8)	13	15.7 (6.8)	52	62.7 (10.1)	83
MALAYSIA	169	27.0 (73.2)	144	23.0 (75.0)	313	50 (60.7)	626
PHILIPPINES	44	19.1 (19.0)	35	15.2 (18.2)	151	65.7 (29.3)	230
Grand Total	231	24.60	192	20.4	516	55.0	939

WHO/IUATLD AFB grading for LED-based fluorescence microscopy, 400x magnification (Lumb et al., 2013; Angra et al., 2017)
Negative: 0; Scanty: 1-2 AFB in 1 length (need confirmation) / 3-24 AFB in 1 length; 1+:1-6 AFB in 1 length; 2+: 7-60 AFB in 1 length; 3+: More than 60 AFB in 1 length

A high-grade smear is reported when the microscopy reading is ≥1+ (Stinson et al., 2014). Percentage of high-grade smears among non-Malaysian is higher compared to Malaysian. AFB Smear: Pulmonary smear positive 925 (91.2%), smear negative 54 (5.3%); Extra-pulmonary smear positive 0 (0%), smear negative 18 (1.8%); Pulmonary and extra-pulmonary smear positive 7 (0.7%), smear negative 10 (0.99%) (Table 4.2). High grade smears usually indicate that patients were diagnosed with severe disease and may have endured lengthy diagnostic delays (Goroh et al., 2020a; Rundi, 2010; Rundi et al., 2011).

Low rates of HIV co-infection, relatively high smoking rates (>30%), and a 7.4% incidence of diabetes were present in the study sample population. These findings offer crucial understandings to direct regional TB control initiatives (Dony et al., 2004; William et al., 2015; Liew et al., 2015). In this study, MTB isolates were grown from samples obtained from 1019 patients in Sabah as follows: Malaysia 674 (66.2%); Philippines 257 (25.2%); Indonesia 88 (8.6%). The male:female ratio is 1.6:1. Age distribution of the sample population is as follows: <1 year, 1 (0.1%); 1-4 years, 1 (0.1%); 5-14 years, 13 (1.3%); 15-24 years, 206 (20.1%); 25-34 years, 235 (23.0%); 35-44 years, 195 (19.1%); 45-54 years, 165 (16.2%); 55-64 years, 127 (12.5%); >65 years, 77 (7.5%). Refer Table 4.3 for data summary.

Table 4.3: Sample population details (N=1019)

Characteristics			p-Value
Sex			
Male (%)	625 (61.3)		
Female (%)	394 (38.7)		
Age distribution			
<1 year	1 (0.1)		
1-4 years	1 (0.1)		
5-14 years	13 (1.3)		
15-24 years	206 (20.2)		
25-34 years	235 (23.1)		
35-44 years	195 (19.1)		
45-54 years	165 (16.2)		
55-64 years	127 (12.5)		
>65 years	77 (7.6)		
Mean age	39.2 ±16.4		
Male	40.8 ± 16.2		P<0.0001
Female	36.5 ± 16.3		
AFB Smear (%)	Positive smear	Negative smear	
Pulmonary	925 (91.2)	54 (5.3)	
Extrapulmonary	0 (0)	18 (1.8)	
Pulmonary & Extrapulmonary	7 (0.7)	10 (0.99)	
Nationality (%)			
Malaysia	674 (66.2)		
Philippines	257 (25.2)		
Indonesia	88 (8.6)		
Type of Case (%)			
New Case	976 (95.8)		
Relapse / Retreatment	43 (4.2)		

In this study, quite a number of patients have advanced TB disease at diagnosis. Percentage of high-grade smears at diagnosis for locals is 50%, and it is more than 60% for foreigners (Table 4.2 and Table 4.3). Health promotion activities in terms TB awareness and the need to go to healthcare facilities or health clinics to get checked for TB. There is no case of PLHIV in this study sample. TB HIV co-infection in Sabah is reported to be low by (William et al., 2015). This study reported percentages of two co-morbidities, diabetes and smoking at 7.5%, and 33.3% respectively.

4.1.1 Sample population by occupation in Sabah

Occupation was categorized according to Connelly et al. (2016). Figure 4.1 shows that the MTB isolates were mostly isolated from population that are Unemployed 18%, home maker 13%, blue collar labourer 11%, farming and fisheries 10% and entrepreneur and self-employed 6%. The driver category which included taxi, bus and heavy vehicle drivers who were highly mobile contributed 3% to the total population in this study. Information on the nature of occupation is quite important as it gives an idea of the patients' possible exposure to the population in the context of TB prevention and management. Further information on the occupation stratification by nationality can be seen in Appendix G.

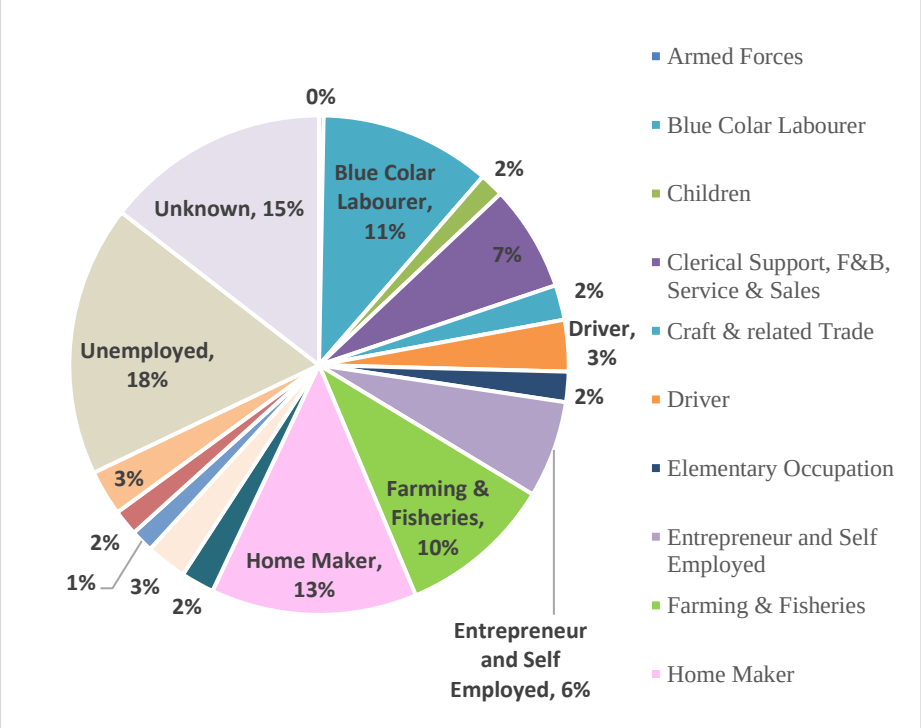


Figure 4.1: Stratification of population by occupation

4.2 Genetic diversity, population structure and clustering of MTB in Sabah

4.2.1 Spoligotype of MTB isolates in Sabah

This study detected 102 SITs (Spoligotype international types) in total with 68 existing SITs and 34 newly created SITs (SIT4181 to SIT4216) as well as 153 orphan spoligotypes. The most frequently found pattern is SIT19/EAI2-Manila and most orphans (22.2%) in this study that fulfil the binary spoligo rules using the SpolLineage Tool (Couvin et al., 2020) are related to this genotype (Table 4.4). Orphan spoligotypes represent patterns that have been found only once and that have not been reported in the past (Navarro et al., 2011; Rodwell et al., 2012; Conceição et al., 2017). They may indicate recent and/or sporadic TB infections in the studied area. A spoligopattern profile is considered a SIT when this pattern had been found at least twice in the same or different study (Demay et al., 2012; Rodwell et al., 2012).

When two or more strains corresponding to the novel pattern are discovered in a study utilizing the SpolLineages Tool and the SITVIT2 web-based database, new SITs are established; both proprietary databases of the Institut Pasteur de la Guadeloupe (Demay et al., 2012; Couvin et al., 2019a; Couvin et al., 2020). A singleton is a general term that refers to an MTB isolate that is not clustered in the study, but it may have a match or clustered in the database whereby a SIT has already been given. A singleton could also refer to a single orphan MTB spoligotype in the study (Demay et al., 2012; Rodwell et al., 2012).

The lineage distributions from 1019 spoligotyped isolates are: EAI (n=811), Unknown (n=119), T (n=25), Beijing (n=25), LAM (n=17), Manu (n=9), AFRI (n=4), H (n=4), X (n=3), Turkey (n=1), and Atypical (n=1) (Table 4.4). SIT19/EAI2-Manila is seen in every district. The three most populated cities, Kota Kinabalu, Tawau, and Sandakan, displayed the greatest distribution of spoligotypes and orphan types as well as a significant incidence of the novel SITs: SIT 4181/EAI2-Manila, SIT4186/Unknown and SIT4204/Unknown. The sub-lineages and SITs were assigned using SpolLineages Tool (Couvin et al., 2020). Distribution of MTB spoligotypes is higher in cities because of a larger population size and gathering of different communities. However, there are also other factors that can cause this diversity.

Spoligotyping was able to group the samples into 55 clusters (2-427 isolates/cluster). As discovered by Molina-Moya et al. (2017) utilization of TB-SPRINT spoligotyping was able to confirm the MIRU-VNTR analysis in revealing the existence of two different MTB populations in a sample. As a result, TB-SPRINT may be helpful for identifying mixed infections, a crucial component of managing TB patients. According to the Hunter Gaston discriminatory index (HGDI), spoligotyping in this study has a relatively high discriminating power of 0.8145 which means that this genotyping method is able to identify genetic differences between isolates, allowing for good clustering analysis and contact investigation (Hunter and Gaston, 1988).

Table 4.4: Sub-lineages and SIT of MTB in this study

Sub-lineage	SIT	SIT	Total
Atypical	1	2669 (1)	1
AFRI	2	4215 (3), Orphan (1)	4
Beijing	2	1 (24), 2610 (1)	25
EAI1-SOM	5 + 5	48 (6), 349 (2), 734 (1), 744 (1), 3355 (1), Orphan (5)	16
EAI2	2	4182 (2), 4214 (1)	3
EAI2 Manila	33 + 34	19 (428), 287 (74), 483 (13), 756 (5), 758 (2), 894 (7), 895 (2), 897 (41), 1169 (7), 1171 (1), 1490 (1), 1501 (1), 1511 (1), 1775 (1), 1781 (6), 1992 (1), 2704 (1), 4181 (14), 4183 (7), 4184 (4), 4187 (5), 4189 (1), 4191 (2), 4193 (3), 4194 (1), 4196 (5), 4197 (3), 4200 (3), 4201 (1), 4202 (2), 4210 (2), 4213 (2), 4216 (1), Orphan (34)	682
EAI3-IND	3	11 (1), 1097 (1), 4212 (1)	3
EAI5	15 + 23	28 (2), 152 (1), 236 (1), 413 (4), 892 (1), 1365 (46), 1408 (1), 1489 (2), 3953 (1), 3993 (1), 4188 (2), 4190 (2), 4199 (2), 4203 (3), 4211 (4), Orphan (23)	96
EAI6-BGD1	3 + 8	591 (1), 882 (1), 3483 (1), Orphan (8)	11
H3	3 + 1	50 (1), 512 (1), 655 (1), Orphan (1)	4
LAM	2	29 (1), 249 (1)	2
LAM1	1	Orphan (1)	1
LAM2	1	17 (5)	5
LAM6	1	Orphan (1)	1
LAM8	1	290 (2)	2
LAM9	3 + 2	42 (2), 866 (1), 891 (1), Orphan (2)	6
Manu_ancestor	1	523 (1)	1
Manu1	6	Orphan (6)	6
Manu3	1 + 1	1378 (1), Orphan (1)	2
T	1	102 (1)	1
T1	6 + 6	51 (1), 53 (5), 612 (1), 1079 (3), 1494 (2), 2786 (1), Orphan (6)	19
T3	1 + 1	37 (1), Orphan (1)	2
T4-CEU1	1	39 (1)	1
T-H37Rv	1	4206 (2)	2
Turkey	1	Orphan (1)	1
Unknown	13 + 62	32 (1), 623 (1), 821 (1), 1196 (2), 1509 (1), 4185 (1), 4186 (16), 4192 (2), 4195 (3), 4198 (2), 4204 (22), 4208 (3), 4209 (2), Orphan (62)	119
X1	1	336 (2)	2
X3	1	92 (1)	1
TOTAL			1019

Note: Number following the sign '+" denotes the number of orphan-MTB spoligotypes. 119 isolates from Unknown lineage (5 established SITs and 8 new SITs with clustering).

4.2.1.1 Predominant MTB SITs in Sabah

There are nine predominant MTB SITs in Sabah (Table 4.5). The word predominant refers to MTB strains that are consistently found in high frequencies within the particular population or area. The predominant SITs are: SIT19/EAI2 Manila (41.9%); SIT287/EAI2 Manila (7.3%); SIT1365/EAI5 (4.5%); SIT897/EAI2 Manila (4.0%); SIT1/Beijing (2.4%); SIT4204(new)/Unknown (2.2%); SIT4186(new)/Unknown (1.6%); SIT4181(new)/EAI2 Manila (1.4%); and SIT483/EAI2 Manila (1.3%). EAI2-Manila STI19 is seen in all districts in Sabah. The three most populous cities, Kota Kinabalu, Tawau, and Sandakan, have the widest range of spoligotype and orphan MTB patterns.

4.2.1.2 Newly created MTB SITs in Sabah

The 34 newly created SITs discovered in this study, as shown in Table 4.6, comprised the lineages as follows: AFRI (1 SIT); EAI2 (2 SITs); EAI2 Manila (16 SITs); EAI3-IND (1 SIT); EAI5 (5 SITs); T-H37Rv (1 SIT); Unknown (8 SITs). There are 27 clustered SITs, and 7 unique SITs of which 3/7 were found in US, most probably associated with Filipinos. The new SITs are grouped into 4 clonal complexes with some singletons (Table 4.7).

Table 4.5: Predominant MTB SITs in this study

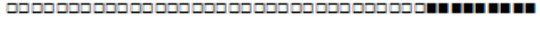
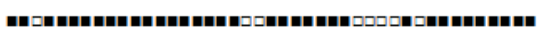
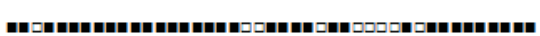
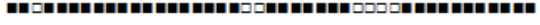
SIT	Spoligotype Description	Octal code	Nb in study	% in study	% in study vs. SITVIT database	Lineage	Distribution in countries	Distribution in Regions
1		000000000003771	24	2.36	0.23	Beijing	CN=18.95, US=18.72, JP=10.76, ZA=7.74, RU=6.79, VN=3.64, IN=3.58, MY=2.99, PE=2.93	Eastern Asia=31.9, North America=18.75, Southeastern Asia=10.2, Austral Africa=7.74, Northern Asia=6.79, Southern Asia=5.93, Northern Europe=3.41, South America=3.39, Western Asia=2.78, Western Europe=2.4
19		677777477413771	428	41.9	31.96	EAI2-Manila	MY=35.85, US=31.06, PH=8.23, MX=5.84, IT=2.92, TW=2.69, SA=2.17	Southeastern Asia=46.18, North America=31.06, Central America=6.21, Eastern Asia=4.57, Southern Europe=2.92, Western Asia=2.92, Western Europe=2.1, Northern Europe=2.02
287		677777477413751	74	7.26	79.57	EAI2-Manila	MY=82.8, US=7.53, PH=2.15, MX=2.15	Southeastern Asia=86.02, North America=7.53, Central America=2.15, Western Europe=2.15
483		677777477413701	13	1.28	39.39	EAI2-Manila	MY=45.45, US=33.33, SA=3.03, IT=3.03, TN=3.03, MM=3.03, PH=3.03, GF=3.03, FR=3.03	Southeastern Asia=51.52, North America=33.33, Southern Europe=3.03, Western Asia=3.03, North Africa=3.03, Western Europe=3.03, South America=3.03
897		676003477413771	41	4.02	82.0	EAI2-Manila	MY=84.0, US=12.0, AU=2.0, SG=2.0	Southeastern Asia=86.0, North America=12.0, Australasia=2.0
1365		000000007413771	46	4.51	88.46	EAI5	MY=92.31, PH=3.85, TZ=1.92, SG=1.92	Southeastern Asia=98.08, East Africa=1.92
4181		677777475413771	14	1.37	93.33	EAI2-Manila	MY=93.33, TW=6.67	Southeastern Asia=93.33, Eastern Asia=6.67
4186		677777477417771	16	1.57	94.12	Unknown	MY=94.12, SA=5.88	Southeastern Asia=94.12, Western Asia=5.88
4204		677777577617771	22	2.16	100.0	Unknown	MY=100	Southeastern Asia=100.00

Table 4.6: Newly created MTB SITs in Sabah (4181-4216)

SIT	Spoligotype Description	Octal code	Nb in study	% in study	% in study vs. SITVIT database	Lineage	Country Distribution	Unique or Clustered	Study Country of Origin Distribution
4181	■■□■■■■■■■■■■■■■■■□□■■■■□■■□□□□■□■■■■■■■■■	677777475413771	14	1.37	93.33	EAI2-Manila	MY=14, TW=1	Clustered	INA (1), MAS (8), PHP (5)
4182	■■□■■■■■■■■■□□□□□□□□□□■■■■■□□□□■□■■■■■■■■■	677740017413771	2	0.2	66.67	EAI2	MY=2, US=1	Clustered	MAS (2)
4183	■■□■■■■■■■■■■■□■■■□□■■■■■■■□□□□■□■■■■■■■■■	677773477413771	7	0.69	87.5	EAI2-Manila	MY=7, US=1	Clustered	MAS (7)
4184	■■□■■■■■■■■□□□□□□■■□□■■■■■■■□□□□■□■■■■■■■■■	677601477413771	4	0.39	100.0	EAI2-Manila	MY=4	Clustered	MAS (3), PHP (1)
4185	■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■□■■■■■■■■■	777777777775771	1	0.1	50.0	Unknown	CN=1, MY=1	Unique	MAS (1)
4186	■■□■■■■■■■■■■■■■■■■□□■■■■■■■□□□□■■■■■■■■■■■	677777477417771	16	1.57	94.12	Unknown	MY=16, SA=1	Clustered	INA (1), MAS (12), PHP (3)
4187	■■□■■■■■■■■■■■■■■■□□■■■■□■■□□□□■□■■■■■■■■■□	677777475413770	5	0.49	100.0	EAI2-Manila	MY=5	Clustered	INA (2), MAS (2), PHP (1)
4188	□□□□□□□□□□□□□□■□□□□□□□□□■■■■□□□□■□■■■■■■■■■	000010007413771	2	0.2	100.0	EAI5	MY=2	Clustered	MAS (2), PHP (2)
4189	■■□■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■□□□□□□□■	677777477413001	1	0.1	50.0	EAI2-Manila	MY=1, US=1	Unique	MAS (1)
4190	■□□■■■■■□□□□□□□□■■■□□■■■■■■■□□□□■□■■■■■■■■■	476003477413771	2	0.2	66.67	EAI5	MY=3	Clustered	MAS (1), PHP (1)
4191	■■□■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■□■■□■□	677777477413550	2	0.2	100.0	EAI2-Manila	MY=2	Clustered	MAS (2)
4192	■■□■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■■■■■■	677777577777771	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (2)
4193	■■□■■■■■■■■■■■■■■■□□■■■■■□■□□□□■□■■■■■■■■■	677777476413771	3	0.29	100.0	EAI2-Manila	MY=3	Clustered	MAS (2), PHP (1)
4194	■■□■■□□■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■■■■■■	663777477413771	1	0.1	50.0	EAI2-Manila	MY=1, US=1	Unique	PHP (1)
4195	■■□■■■■■■■■■□■■■■■■■□□□□■■■■■■■■□□□■■■■■■■■■	677737607761771	3	0.29	100.0	Unknown	MY=3	Clustered	MAS (2), PHP (1)
4196	■■□■■■■■■■■■□■■■■■■□□■■■■■■■□□□□■□■■■■■■□■■	677737477413751	5	0.49	100.0	EAI2-Manila	MY=5	Clustered	INA (2), MAS (2), PHP (1)
4197	■■□■■■■■■■■■■■■■■■■□□■■□■■■■□□□□■□■■■■■■■■■	677777467413771	3	0.29	100.0	EAI2-Manila	MY=3	Clustered	MAS (1), PHP (2)

Table 4.6 (continued): Newly created MTB SITs in Sabah (4181-4216)

SIT	Spoligotype Description	Octal code	Nb in study	% in study	% in study vs. SITVIT database	Lineage	Country Distribution	Unique or Clustered	Study Country of Origin Distribution
4198	■■□■■■■■■■■■■■■■■■□□■■■■■□□□□□□□□□■■■■■□■■	677777476001751	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (2)
4199	□□□■□□□□□□□□□□□□□□□□□□□□□■■■■□□□□■□■■■■■■■■	040000007413771	2	0.2	100.0	EAI5	MY=2	Clustered	MAS (2)
4200	■■□■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□□■■■■■■■■	677777477411771	3	0.29	100.0	EAI2-Manila	MY=3	Clustered	MAS (2), PHP (1)
4201	■■□■□■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■■□■■	657777477413751	1	0.1	50.0	EAI2-Manila	MY=2	Unique	INA (1)
4202	■■□■■■■■□□□□□□□□■■■□□■□■■■■■□□□□■□■■■■■■■■	676003457413771	2	0.2	100.0	EAI2-Manila	MY=2	Clustered	INA (1), MAS (1)
4203	■■□■■■■■□□□□□□□□■□□□□■■■■■■■□□□□■□■■■■■■■■	676002077413771	3	0.29	100.0	EAI5	MY=3	Clustered	MAS (2), PHP (1)
4204	■■□■■■■■■■■■■■■■■□■■■■■■■■□□□■■■■■■■■■■■■	677777577617771	22	2.16	100.0	Unknown	MY=22	Clustered	INA (2), MAS (13), PHP (7)
4206	■■■■■■■■■■■■■■■■■■□□■■■■■□■■■■■□□□□■■■■■■■	777777476760771	3	0.29	100.0	T-H37Rv	MY=3	Clustered	MAS (2)
4208	■■□■■■■■■■■■■■■■■■□■■■■■■■■□□□■■■■■■■■□■■	677777577617751	3	0.29	100.0	Unknown	MY=3	Clustered	MAS (3)
4209	□□□□□□□□□□□□□□□□□□□□□□□□□■■■■□□□□■■■■■■■■■	000000007417771	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (1), PHP (1)
4210	■■□■■■■■■■■■■■■■■■□□■■■■■□■□□□□■□■■■■■■□■■	677777476413751	2	0.2	100.0	EAI2-Manila	MY=2	Clustered	MAS (2)
4211	□□□□□□□□□□□□□□□□□□□□□□□□□■□■■□□□□■□■■■■■■■■	000000005413771	4	0.39	100.0	EAI5	MY=4	Clustered	MAS (2), PHP (2)
4212	■□□■■■■■■■■■■■■■■■■■■■■■□■■□□□□■□■■□□□■■■■	477777775413071	1	0.1	50.0	EAI3-IND	GB=1, MY=1	Unique	MAS (1)
4213	■■□■■■■■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■□■□■■	677777477413651	2	0.2	100.0	EAI2-Manila	MY=2	Clustered	MAS (2)
4214	■■□■■■■■■■■■■■■■■□□□□□■■■■■■□□□□■□■■■■■■■■■	677776037413771	1	0.1	50.0	EAI2	MY=1, US=1	Unique	MAS (1)
4215	■■□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□■■■■	600000000000071	3	0.29	100.0	AFRI	MY=3	Clustered	MAS (3)
4216	■■□■■□□□■■■■■■■■■■■□□■■■■■■■□□□□■□■■■■■■■■■	661777477413771	1	0.1	50.0	EAI2-Manila	IT=1, MY=1	Unique	MAS (1)

Note: Yellow fields denote SITs only reported in Sabah, Malaysia to date. MAS-Malaysia, INA-Indonesia, PHP-Philippines

Table 4.7: New MTB SIT and clonal complex by 24-loci MIRU-VNTR in Sabah, East Malaysia

Origin / Clonal Complex	4181	4182	4183	4184	4185	4186	4187	4188	4189	4190	4191	4192	4193	4194	4195	4196	4197	4198	4199	4200	4201	4202	4203	4204	4206	4208	4209	4210	4211	4212	4213	4214	4215	4216	Total
CC1	1						1														1															3
CC3					1																															1
Spol only						1	1									2						1		2												7
INA	**1**				**1**	**1**	**2**									**2**					**1**	**1**		**2**												**11**
CC1	3	1	2			4			1	1	1	1	2			1	1	1	1	1		1	1	6			1	1	1		1	1	2	1	37
CC2						1																													1
CC4															1																				1
S			1			1					1								1				1		2	1									8
Spol only	5	1	4	3		6	2					1			1	1		1		1				7		2		1	1	1	1		1		40
MAS	**8**	**2**	**7**	**3**		**12**	**2**		**1**	**1**	**2**	**2**	**2**		**2**	**2**	**1**	**2**	**2**	**2**		**1**	**2**	**13**	**2**	**3**	**1**	**2**	**2**	**1**	**2**	**1**	**3**	**1**	**87**
CC1	3			1		1	1	1					1	1		1	2			1			1	2					2							18
CC4															1																				1
S																								2											2
Spol only	2					2		1		1														3			1								10
PHP	**5**			**1**		**3**	**1**	**2**		**1**			**1**	**1**	**1**	**1**	**2**			**1**			**1**	**7**			**1**		**2**						**31**
Total	**14**	**2**	**7**	**4**	**1**	**16**	**5**	**2**	**1**	**2**	**2**	**2**	**3**	**1**	**3**	**5**	**3**	**2**	**2**	**3**	**1**	**2**	**3**	**22**	**2**	**3**	**2**	**2**	**4**	**1**	**2**	**1**	**3**	**1**	**129**

Table 4.8: Possible members of new MTB lineages in Sabah, East Malaysia

SIT	Spoligotype Description	Octal code	Nb in study	% in study	% in study vs. SITVIT database	Lineage	Country Distribution	Unique or Clustered	Study Country of Origin Distribution
4192	■■□■■■■■■■■■■■■■■■□■■■■■■■■■■■■■■■■■■■■■■■■	677777577777771	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (2)
4195	■■□■■■■■■■■■□■■■■■■■□□□□■■■■■■■■■□□□□■■■■■■■■■	677737607761771	3	0.29	100.0	Unknown	MY=3	Clustered	MAS (2), PHP (1)
4198	■■□■■■■■■■■■■■■■■■■□□■■■■■■□□□□□□□□□■■■■■■□■■	677777476001751	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (2)
4204	■■□■■■■■■■■■■■■■■■■□■■■■■■■■■□□□■■■■■■■■■■■■	677777577617771	22	2.16	100.0	Unknown	MY=22	Clustered	INA (2), MAS (13), PHP (7)
4208	■■□■■■■■■■■■■■■■■■■□■■■■■■■■■□□□■■■■■■■■■□■■	677777577617751	3	0.29	100.0	Unknown	MY=3	Clustered	MAS (3)
4209	□□□□□□□□□□□□□□□□□□□□□□□□■■■■□□□□■■■■■■■■■■■■	000000007417771	2	0.2	100.0	Unknown	MY=2	Clustered	MAS (1), PHP (1)

4.2.1.3 Comparison of new SITs from EAI2-Manila and Unknown lineage

Figure 4.2 shows the geographical distribution of members of the Unknown lineage. The different colours represent the eight different CC. Of the eight SITs (4185, 4186, 4192, 4195, 4198, 4204, 4208, and 4209) from the Unknown lineage (Table 4.4) two SITs have been reported in China (4185) and Saudi Arabia (4186) respectively, while six clustered SITs (Table 4.8) are reported here for the first time.

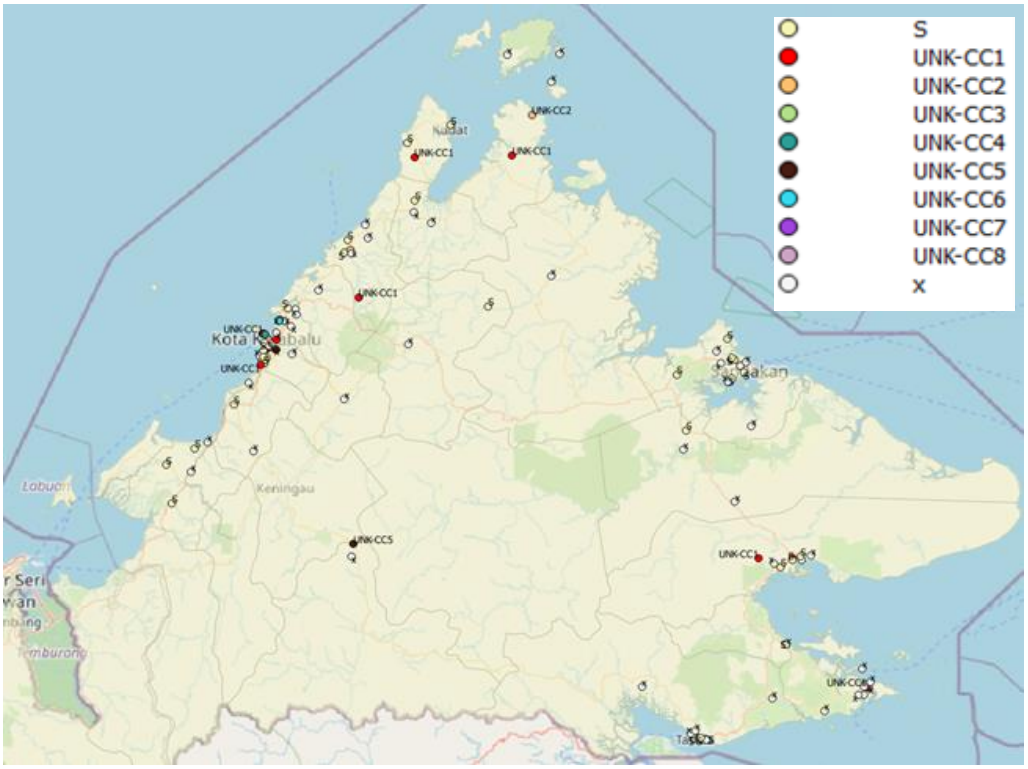


Figure 4.2 MTB isolates from Unknown Lineage were grouped into 8 clonal complexes (CC) involving 119 isolates

Also noteworthy is the association of two SITs (4186 and 4204) with all three countries in this study; two SITs (4188 and 4194) Malaysians and Filipinos and the remaining four SITS (4185, 4192, 4198, and 4208) are associated with Malaysians only at this point. These isolates which are found to be clustered by spoligotyping method do not adhere to any known rules using the SpolLineages Tool, TBMiner and *MIRU-VNTRplus* hence they are grouped together in the Unknown lineage. The clustering of these isolates points towards the possibility of a new lineage. WGS analysis needs to conducted on these isolates, to verify this claim of a new lineage. Tables 4.6 and Table 4.8 exhibited the new MTB SITs discovered in this study. Focusing on the EAI2-Manila and Unknown lineages, the MST of the 25 newly created MTB SITs in Figure 4.3 showed that the new spoligotypes were grouped into three CC and 12 singletons. Since deletion of spacers is irreversible, the newer spoligotypes are determined by looking at the missing spacers. CC2 and CC3 are grouped into the Unknown lineage, meanwhile CC1 showed local adaptations of SIT19/EAI2-Manila which rendered six different spoligotypes with one additional missing spacer at different positions. Two of the new EAI2-Manila spoligotypes, SIT4193 and SIT4181, lost another spacer resulting in the spoligotypes SIT4210 and SIT 4187 respectively. Table 4.9 shows that four MIRU-VNTR loci (ETR-B, Mtub2, QUB11b and ETR-A) are highly discriminatory especially for EAI genotypes. These four loci can be used for preliminary investigations especially during outbreaks.

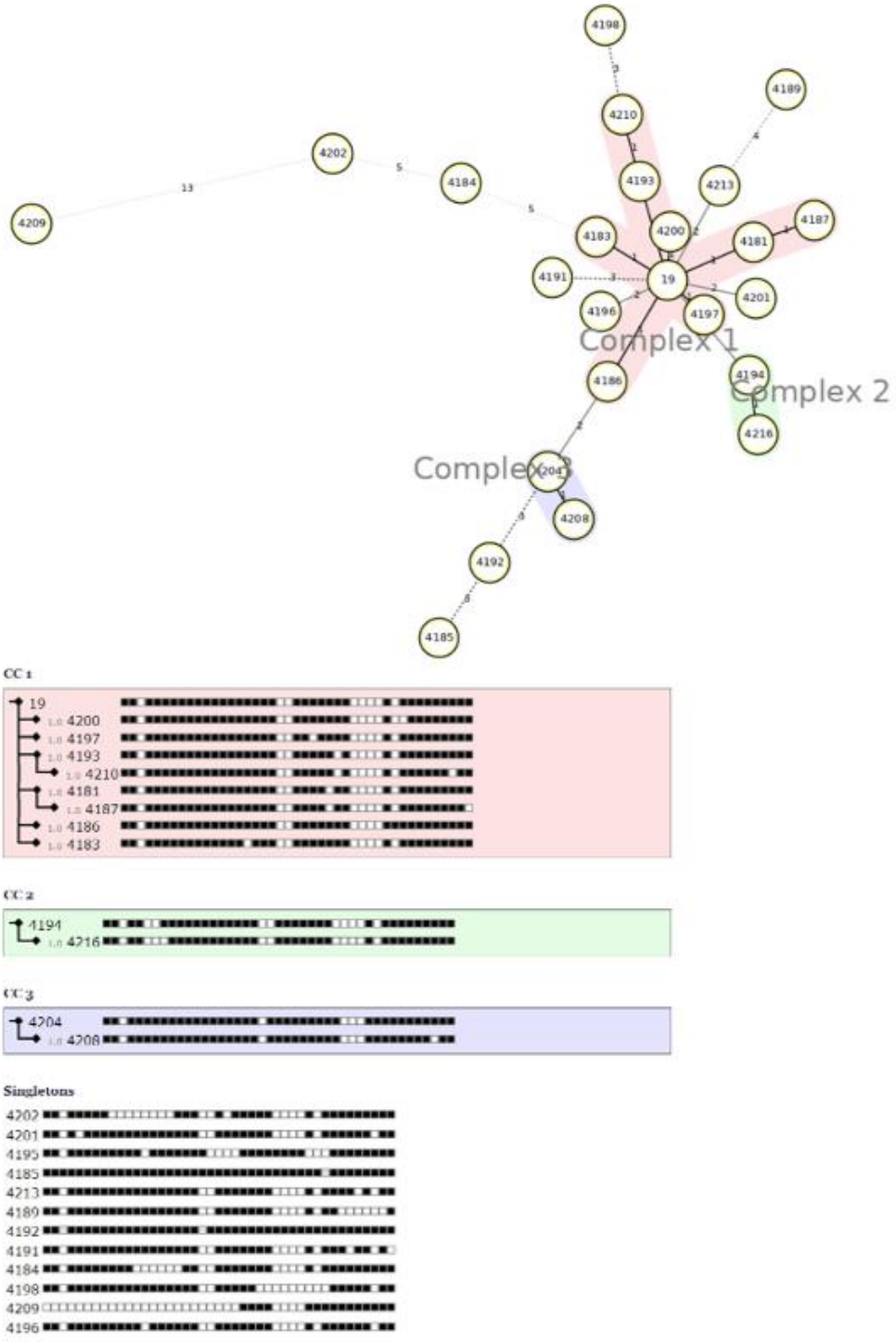


Figure 4.3: MST of the newly created MTB SITs from the EAI2-Manila and Unknown lineage

4.2.1.4 MTB spoligotype among three nationalities in Sabah

It is observed that there are shared as well as exclusive SITs among the three nationalities studied in the Sabah population (Figure 4.4 and Figure 4.5). Malaysians (n=674), Indonesians (n=88) and Filipinos (n=257) all share 10 SITs which include 5 new SITs (sub-lineages EAI2-Manila, EAI5, Beijing, T1, EAI1-SOM and EAI6-BGD1).

The Indonesian population in this study has 12 exclusive SITs (including 1 new SIT) and share 5 SITs (including 1 new SIT) with Malaysians. The sub-lineages unique to the Indonesian sample population in this study are: SIT29/LAM, H (SIT512/H3 and SIT655/H3) and T (SIT102/T, SIT2786/T1 and SIT37/T3). It is interesting to note that the Haarlem sub-lineages (first described in Netherlands) are only found among the Indonesians in this study. The Filipino population in this study on the other hand has 5 exclusive SITs (4 EAI: 3EAI2 Manila, 1 EAI5; and 1 Unknown) and share 18 SITs [(14 EAI of which 10 EAI2-Manila, 1 LAM, 1T, 2 UNK); 9 NEW-2 Unknown] with Malaysians. Malaysian population has 52 exclusive SITs (16 NEW SITs whereby 4 are from Unknown lineage. Sub-lineages unique to Malaysians: AFRI, EAI2, EAI3-IND, S, T4-CEU1, T-H37Rv, Turkey, X1 and X3). It is interesting to note that the sub-lineage grouping among the Filipinos is almost a perfect sub-set to that of the Malaysian sub-lineage types, with 18 shared SITs and only 5 SITs unique to Filipinos (as compared to 12 SITs unique to Indonesians). Although all three nationalities live and work in close proximities to each other (Hassan, 2018; DOSM, 2020; Peters, 2022), in the context of TB

infection, it does seem that the Filipinos share more similarities in terms of MTB strains with the Malaysian population in Sabah. The age range of 25 to 34 years has the highest rates of TB cases (23%) and the greatest number of MTB strains (19 sub-lineages and 52 SITs including 13 new SITs).

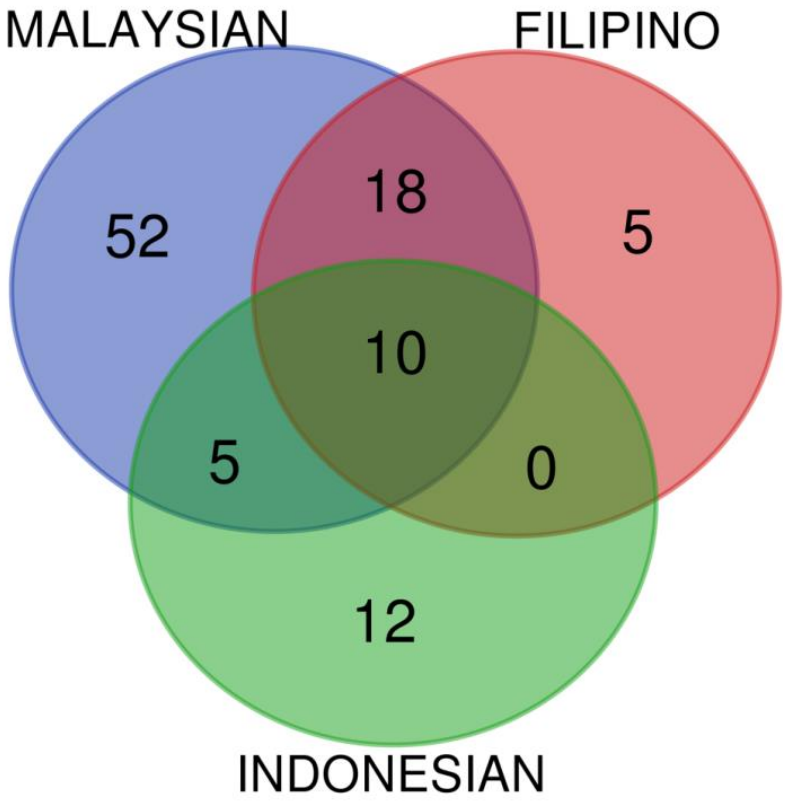


Figure 4.4: VENN Diagram of MTB SIT distribution among Malaysian, Filipino and Indonesian in Sabah

Spoligotyping data comparison (Figure 4.6) was done between this study and two previous studies conducted in Malaysia (Ngeow et al., 2006; Ismail et al., 2014). Spoligotyping study by Ngeow et al. (2006) on 689 isolates from the whole of Malaysia, albeit with low representation of isolates from Borneo, reported that the predominant lineage is the Beijing family, and ST1/Beijing is the predominant strain (32.95%) identified in all geographic

areas in peninsular Malaysia, followed by ST19/EAI2-Manila (6.24%) which is mostly found in natives in East Malaysia. T lineage ST51/T1 (3.05%) is mostly found in the southern states of peninsular Malaysia namely Johor and Melaka. Ngeow et al. (2006) also reported the presence of ST745/EAI1-SOM (2.75%) which is exclusive to Malays in the east-coast of Malaysia. This specific finding was corroborated by (Ismail et al., 2014) who later proposed to rename this strain to SIT745/EAI1-MYS (Couvin et al., 2019a). The genotyping study on 220 MTB isolates (Ismail et al., 2014) from Kelantan and Kuala Lumpur showed that the Beijing strain SIT1/Beijing is the most frequently recognized strain (25.5%), with the ancestral EAI lineage being the more predominant lineage. SIT89/EAI2-Nonthaburi, mostly found in neighbouring Thailand was also detected in the Kelantan sample population. Ismail et al. (2014) study was limited to these two areas and does not reflect the actual distribution and dispersal of MTB lineage in other parts of Malaysia. The 22 isolates in this study from the T lineage are mostly distributed in the east coast of Sabah [Interior (1), Kudat (1), Sandakan (5), Tawau (12) and West Coast (3)], comprising Indonesians and Malaysians mostly from the communities or localities with high Bugis population (Indonesian 8, Malaysian 11 and Filipino 3).

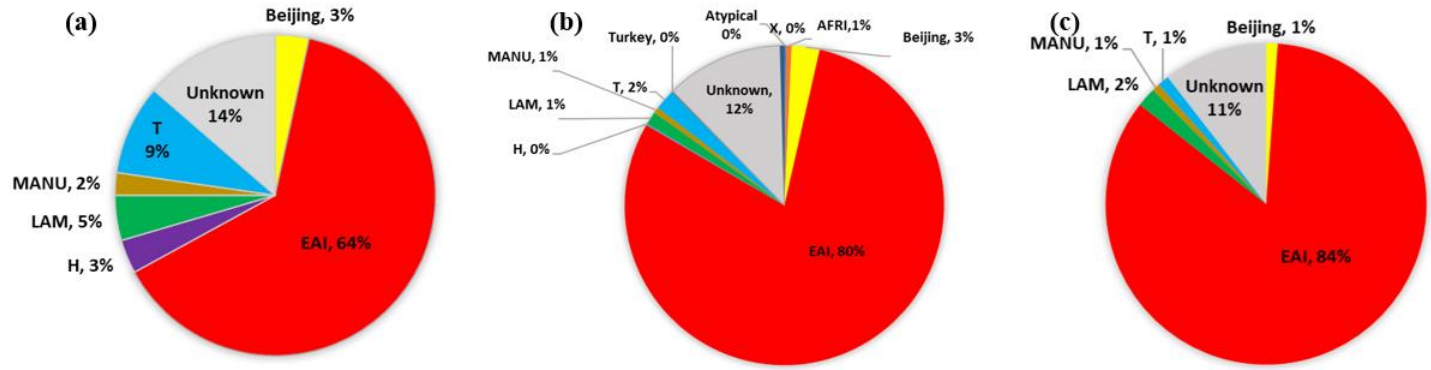


Figure 4.5: Comparison of MTB lineages in Sabah by country of origin – (a) Indonesia, (b) Malaysia, and (c) Philippines

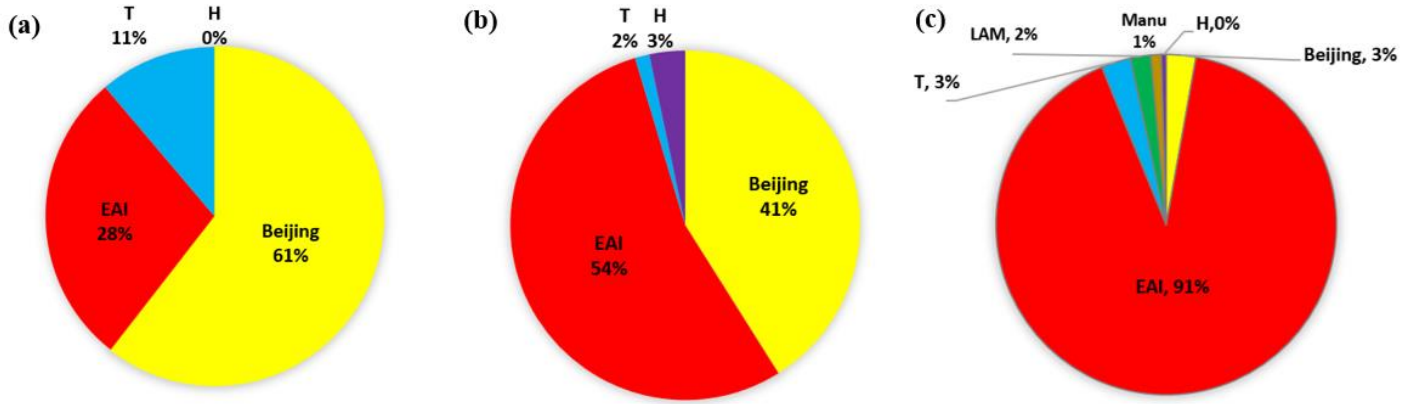


Figure 4.6: Sub-lineages by spoligotyping (a) Ngeow et al., 2006, MTB isolates from Malaysia (with relatively low representation from Borneo (N=689); (b) Ismail et al., 2014, MTB Isolates from Kuala Lumpur & Kelantan (N=220); (c) This study (N=1019), MTB Isolates from Sabah

4.2.2 24-loci MIRU-VNTR typing panel for Sabah MTB population

The MIRU-VNTR typing gave a high HGDI value of 0. 997 ($h \geq 0.6$ highly discriminant, $0.3 \leq h \leq 0.6$ moderately discriminant and $h < 0.3$ poorly discriminant). The loci 2461_ETR-B, 1955_ Mtub21, 2163b_QUB11b, 2165_ETR-A are highly discriminant for Sabah MTB isolates especially for EAI2-Manila. These loci could be used as preliminary screening tool for outbreak investigations. However, they are not very discriminatory for the Beijing lineage (Table 4.9). By employing 24-loci MIRU-VNTR and/or spoligotype, depending on availability, and using human expert interpretation of spoligotype signatures, lineage assignation was carried out.

Table 4.9: Comparison of MIRU VNTR HGDI among different lineages of MTB isolates in Sabah, East Malaysia

MIRU-VNTR loci	24	15	12	*ALL(N=556)*			BEIJING (n=15)			EAI (n=451)			UNKNOWN (n=55)			Others (n=35)		
				*h**	*Conclusion*	Total Samples	*h**	*Conclusion*	Total Samples	*h**	*Conclusion*	Total Samples	*h**	*Conclusion*	Total Samples	*h**	*Conclusion*	Total Samples
154_MIRU02	x		x	0.01	PD	556	0	**PD***	15	0.04	PD	451	0.75	HD	55	0.07	PD	35
424_Mtub04	x	x		0.29	PD	555	0.45	MD	15	0.10	PD	451	**0.88**	**HD**	55	0.72	HD	35
577_ETRC	x	x		0.10	PD	556	0	**PD***	15	0.07	PD	451	0.78	HD	55	0.23	PD	35
580_MIRU04_ETRD	x	x	x	0.35	MD	555	0	**PD***	15	0.18	PD	450	**0.9**	**HD**	55	0.62	HD	35
802_MIRU40	x	x	x	0.28	PD	553	0.13	PD	15	0.11	PD	448	**0.89**	**HD**	55	0.72	HD	35
960_MIRU10	x	x	x	0.34	MD	556	0.34	MD	15	0.23	PD	451	**0.87**	**HD**	55	0.62	HD	35
1644_MIRU16	x	x	x	0.13	PD	556	0	**PD***	15	0.06	PD	451	**0.81**	**HD**	55	0.46	MD	35
1955_Mtub21	x	x		0.70	HD	552	0.68	HD	15	0.63	HD	449	**0.95**	**HD**	55	0.84	HD	35
2059_MIRU20	x		x	0.18	PD	555	0	**PD***	15	0.19	PD	450	0.79	HD	55	0.18	PD	35
2163b_QUB11b	x	x		0.67	HD	543	0.76	HD	15	0.61	HD	441	**0.94**	**HD**	55	0.83	HD	35
2165_ETRA	x	x		0.65	HD	544	0.51	MD	11	0.65	HD	444	**0.89**	**HD**	55	0.72	HD	35
2347_Mtub29	x			0.28	PD	556	0	**PD***	15	0.08	PD	451	**0.86**	**HD**	55	0.53	MD	35
2401_Mtub30	x	x		0.10	PD	556	0.13	PD	15	0	**PD***	451	0.77	HD	55	0.31	MD	35
2461_ETRB	x			0.72	HD	554	0	**PD***	15	0.66	HD	451	0.93	HD	55	0.71	HD	35
2531_MIRU23	x		x	0.32	MD	555	0.42	MD	15	0.22	PD	450	**0.87**	**HD**	55	0.53	MD	35
2687_MIRU24	x		x	0.21	PD	556	0	**PD***	15	0.02	PD	451	**0.85**	**HD**	55	0.50	MD	35
2996_MIRU26	x	x	x	0.34	MD	556	0.26	PD	15	0.13	PD	451	**0.92**	**HD**	55	0.79	HD	35
3007_MIRU27_QUB5	x		x	0.24	PD	556	0.34	MD	15	0.23	PD	451	0.77	HD	55	0.24	PD	35
3171_Mtub34	x			0.04	PD	552	0	**PD***	15	0.04	PD	448	0.76	HD	55	0.17	PD	35
3192_MIRU31_ETRE	x	x	x	0.41	MD	556	0.34	MD	15	0.27	PD	451	**0.89**	**HD**	55	0.63	HD	35
3690_Mtub39	x	x		0.30	MD	556	0.36	MD	15	0.13	PD	451	**0.9**	**HD**	55	0.74	HD	35
4052_QUB26	x	x		0.49	MD	555	0.65	HD	15	0.38	MD	450	**0.92**	**HD**	55	0.81	HD	35
4156_QUB4156	x	x		0.20	PD	555	0.51	MD	15	0.01	PD	451	**0.84**	**HD**	55	0.55	MD	35
4348_MIRU39	x		x	0.51	MD	556	0.45	MD	15	0.47	MD	451	**0.88**	**HD**	55	0.55	MD	35

* "highly discriminant" HD ($h > 0.6$), "moderately discriminant" MD ($0.3 \leq h \leq 0.6$) or "poorly discriminant" PD ($h < 0.3$). PD* h=0
Note: the figure for total samples varies as they exclude 'No band' samples for the respective loci. Others include LAM, T, H, X, AFRI

4.3 MTB clustering and clonal complex clustering

Clustering and clonal complex clustering occurs when at least two TB cases have matching or closely matched MTB genotypes to suggest that a recent or active transmission from a common source, or because of reactivation of a latent infection acquired from the same source in the past has taken place (Allix-Béguec et al., 2008; de Vries et al., 2009; Weniger et al., 2010; García De Viedma and Pérez-Lago, 2018). The size of the cluster also indicates the fitness and the high transmissibility rate of the MTB strain involved (Zhang and Yew, 2015; Borrell and Trauner, 2017).

Table 4.10 shows the clustering analysis using both spoligotyping and 24-loci MIRU-VNTR. When spoligotyping alone (HGDI=0.814) is used, the clustering rate is 0.748. When 24-loci MIRU-VNTR with a higher discriminatory power (HGDI=0.997) is used on its own, the clustering rate is 0.223. And when both spoligotyping and 24-loci MIRU-VNTR are combined, the HGDI becomes 0.998 and the clustering rate is now 0.162. By using the two-tiered genotyping approach we are able to get a better estimate of the TB transmission.

4.3.1 24-loci MIRU VNTR clonal complex cluster

Of the 556 isolates (non-mixed infection) analysed by 24-loci MIRU-VNTR, eight clonal complexes made up of genetically related strains with no more than two MIRU locus differences (DLV2 - twofold locus variance) between them were formed from the 451 isolates. There were two isolates in clonal complexes 6 and 7 (CC6 and CC7) and 422 isolates in clonal complex 1 (CC1) respectively (Table 4.11). CC1 is the biggest and the most diverse clonal complex and spanned throughout the whole state of Sabah (Figure 4.7). By utilizing *MIRU-VNTRplus* phylogenetical analysis it was found that there are 16 major branches in the CC1 minimum spanning tree (MST) with branch 12 having the most members. There are a few noteworthy clusters especially the AFRI, Beijing, and the MIT24-397 EAI2 Manila major cluster (n=17).

Table 4.10: Clustering analysis of mycobacterial interspersed repetitive units - variable number tandem repeat (MIRU-VNTR) and spoligotyping

Typing Methods	n	No. of different patterns	No. of clusters	No. of clusters isolates	No. of unique isolates	Size of clusters	HGDI	clustering rate, CR [(nc-c)/n]	% in cluster
Spoligotyping43	1020	256	55	818	202	2-427	0.814	0.748	80.2
MIRU-VNTR 12 loci	556	169	50	437	119	2-126	0.928	0.696	78.6
MIRU-VNTR 15 loci	556	337	56	275	281	2-44	0.988	0.394	49.5
MIRU-VNTR 24-loci	556	432	51	175	381	2-20	0.997	0.223	31.5
Spoligotyping43 & 12 Loci MIRU-VNTR	556	261	63	358	198	2-93	0.965	0.531	64.4
Spoligotyping43 & 15 Loci MIRU-VNTR	556	418	48	186	370	2-20	0.996	0.248	33.5
Spoligotyping43 & 24-loci MIRU-VNTR	556	466	41	131	425	2-12	0.998	0.162	23.6

The clustering rate (CR=[nc-c]/n), where CR is the clustering rate, nc is the total number of clustered isolates, c is the total number of clusters, and n is the total number of isolates, is used to estimate recent transmission. Based on combined 24-loci MV and spoligotyping, the proportion of active infection to reactivation is 0.162: 0.838 (clustering rate); 23.6:76.4 (% in cluster) respectively

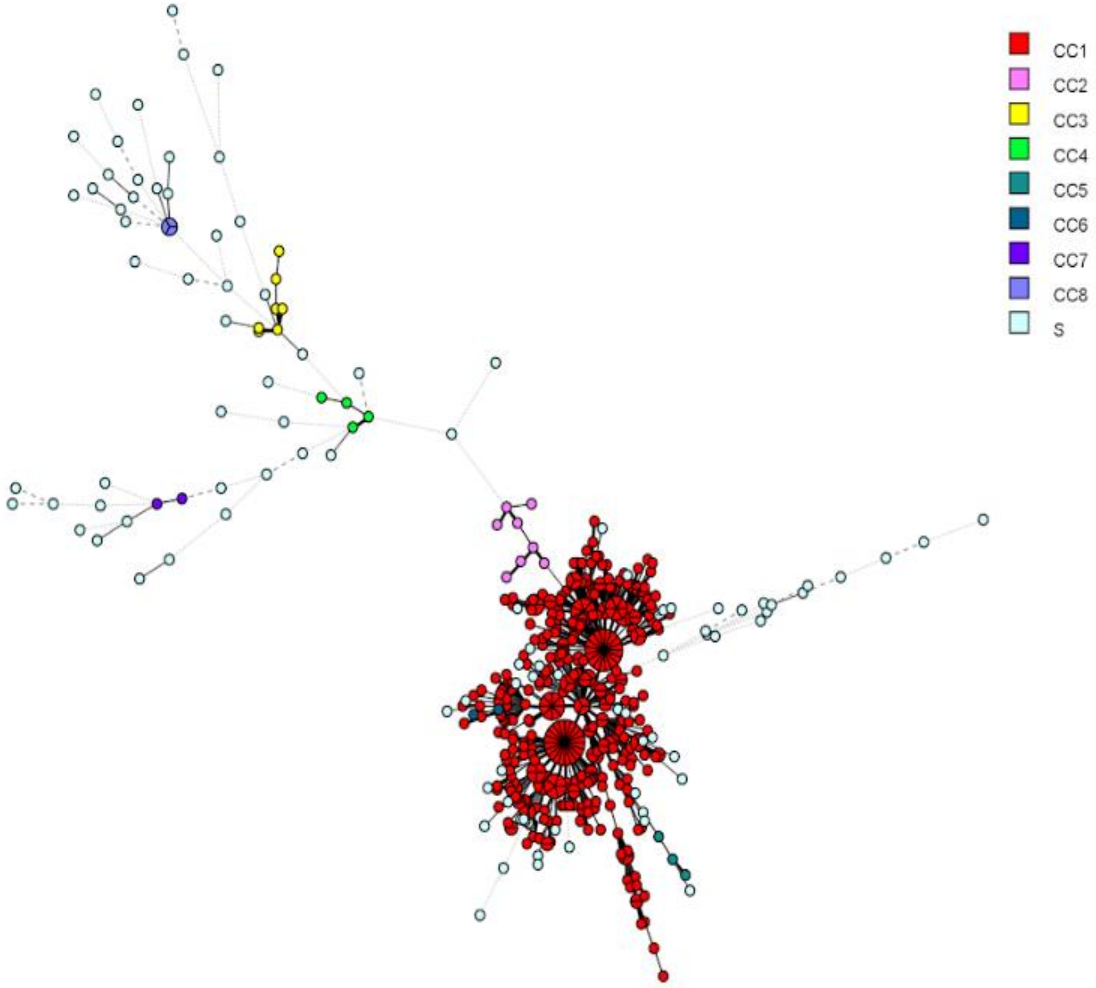


Figure 4.7: 8 Clonal complexes (CC) of MTB in Sabah (Bionumerics 7.6)

CC1 (red) is the biggest clonal complex

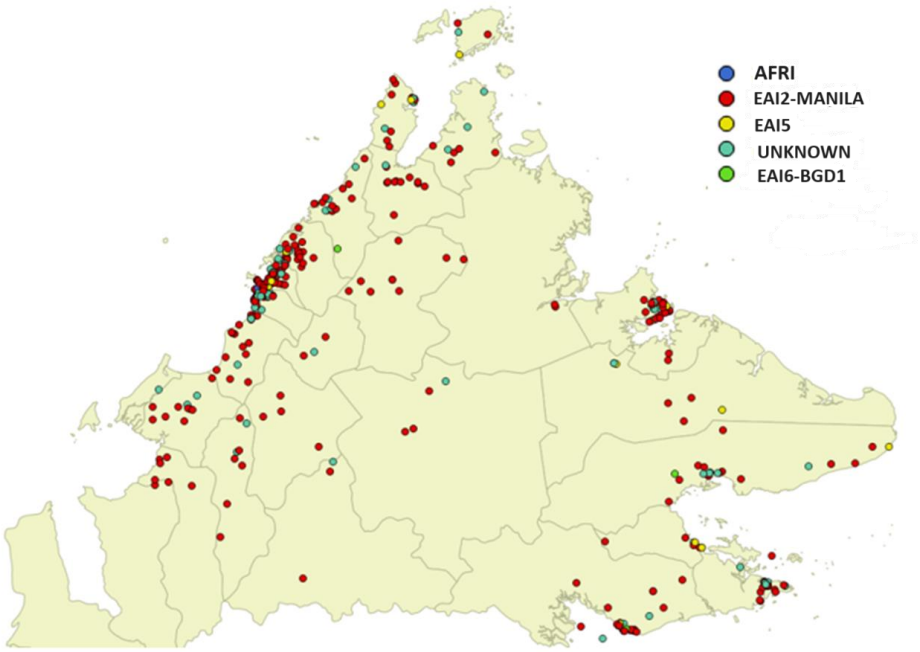


Figure 4.8: Clonal complex 1 (CC1) distribution by sub-lineage

The MTB isolates were divided into eight clonal complexes (CC) in this study using 24-loci MIRU-VNTR (n=556) (Table 4.11). Members of clonal complexes may differ in 1-2 loci. The CC sizes were n=2-422/CC. CC1 is the biggest clonal complex with 422 isolates comprising AFRI (2), EAI2-Manila (359) EAI5 (32), EAI6-BGD1 (2) and Unknown (25). They are distributed all over the state (Figure 4.8). All the CC were profiled to see the connection and relatedness. The smallest CC is CC7 with two isolates both from the Unknown sub-lineage. CC2 containing seven members were confined to Sandakan district involving two Filipinos and five Malaysians. Noteworthy clusters are the AFRI, Beijing, and the MIT24-397 EAI2-Manila major cluster. The loci 2461_ETR-B, 1955_ Mtub21, 2163b_QUB11b, 2165_ETR-A are highly discriminant for Sabah MTB isolates especially for EAI2-Manila, but not very discriminatory for the Beijing lineage. These loci could be used for preliminary screening of outbreaks (Table 4.9).

Table 4.11: Clonal complex clustering pattern of MTB by country of origin (n=556)

Clonal Complex / Country of Origin	INA (%)	MAS (%)	PHP (%)	Total (%)
CC1	24 (58.5)	284 (75.9)	114 (80.9)	422 (75.9)
CC2	0 (0)	6 (1.6)	2 (1.4)	8 (1.4)
CC3	1 (2.4)	6 (1.6)	0 (0)	7 (1.3)
CC4	0 (0)	2 (0.5)	2 (1.4)	4 (0.7)
CC5	1 (2.4)	2 (0.5)	0 (0)	3 (0.5)
CC6	0 (0)	0 (0)	2 (1.4)	2 (0.4)
CC7	1 (2.4)	1 (0.3)	0 (0)	2 (0.4)
CC8	0 (0)	1 (0.3)	2 (1.4)	3 (0.5)
S	14 (34.1)	72 (19.3)	19 (13.5)	103 (18.5)
Total	41	374	141	554

4.3.2 24-loci MIRU VNTR exact clusters MIT24-391 to MIT24-441

175 isolates were grouped into 51 clusters with the exact 24-loci MIRU-VNTR profile (n=2-20/cluster), while 381 isolates were assigned as singletons. The geographical distribution of the 51 MIT24 clusters is shown in Figure 4.9.

In the 51 MIT24 Clusters, it was anticipated that isolates from the same 24-loci MIRU-VNTR cluster would have the same spoligotype patterns. However, this was not so, especially for the bigger clusters (Table 4.15). A more stringent definition of cluster in this context would mean that the spoligopattern and 24-loci MIRU-VNTR should be exactly the same. Thus, instead of 51 exact 24-loci MIRU-VNTR clusters, there are 40 GenTypes clusters in this study with cluster size n=2-12/cluster. The biggest cluster is SIT19/EAI2-Manila – MIT24 397 whereby n=12 (Appendix I).

175 isolates grouped into 51 clusters of 24-loci MIRU-VNTR exact profile (n=2-17/cluster), 381 isolates were assigned as singletons of which some are known to be orphans. Most orphans were derived from EAI lineage, majority from EAI2-Manila sub-lineage. Of the 51 exact 24-loci MIRU-VNTR profile clusters (MIT24 391-441) found in this study, 50 were grouped into CC1 while only MIT24 434 cluster was grouped into CC8 (Beijing sub-lineage). The largest exact 24-loci MIRU-VNTR clusters are MIT24 396 (n=20) and MIT24 397 (n=17).

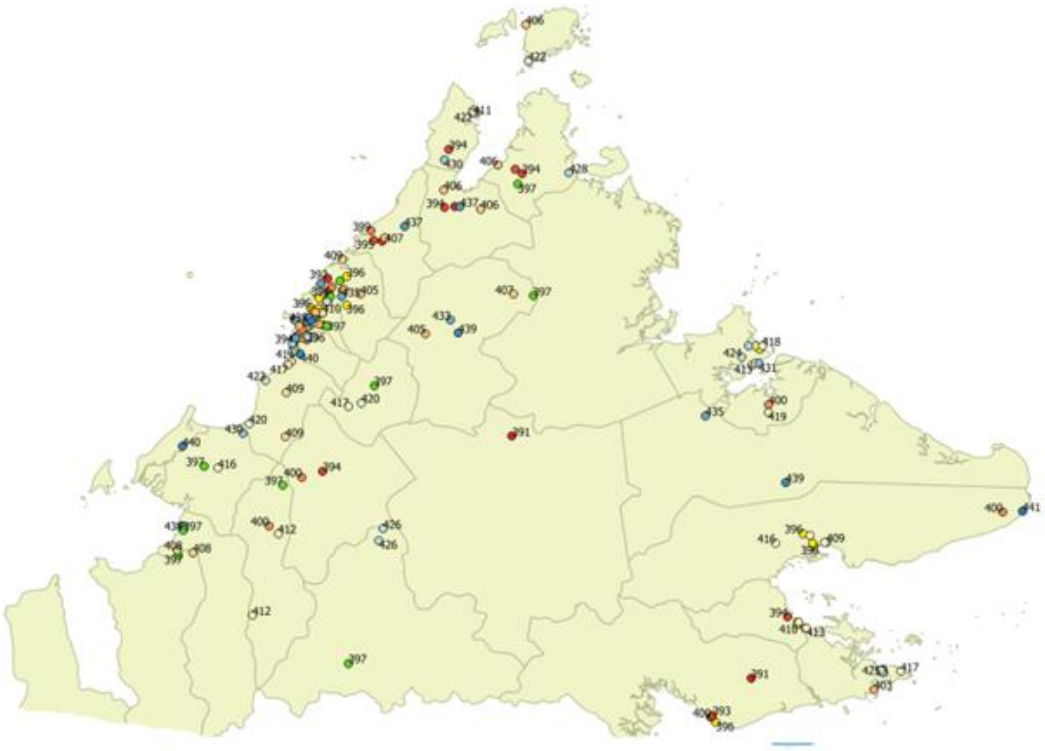


Figure 4.9: Distribution of 51 MTB MIT24 exact 24-loci MIRU-VNTR clusters

The exact MIT24 Exact 24-loci MIRU-VNTR clusters are distributed all throughout the state especially in the densely populated areas in the west coast division. 50 of 51 clusters are from CC1. The only non-CC1 MIT24 cluster is MIT24-434, Beijing genotype from CC8 with 3 member isolates located in Kota Kinabalu (n=2) and Semporna (n=1). The two biggest exact MIT24 clusters are MIT24-396 with 20 member isolates (yellow), and MIT24-397 with 17 member isolates (yellow green).

4.4 Genotype-complemented TB contact investigation with conventional TB contact investigation

4.4.1 Transmission and reactivation rate

In conventional contact investigation for this dataset, patient isolates were matched with TBIS101C database from 2015-2018. Four clusters (n=2/cluster) established by conventional contact investigation were seen in this data set.

When conventional contact investigation was utilised, the active infection was calculated as 6.1%; meanwhile reactivation was calculated as 93.9%. Active infection and reactivation for Genotyping complemented contact investigation was calculated as 23.6% and 76.4% respectively. Molecular clustering by spoligotyping was able to confirm all 4 clusters with matching isolates (n=2/cluster) which were previously detected by conventional contact investigation method in this study population.

The clustering rate for 24-loci MIRU-VNTR is 0.223 (recent transmission) indicating that most TB cases in Sabah are caused by reactivation (1-0.223) as opposed to recent transmission. Clustering rate when calculated using spoligotyping alone is 0.748. When calculated using the combined methods (MIRU-VNTR and spoligotyping) the clustering rate is 0.162.

4.4.1.1 TB notification interval between Case #1 and Case #2 as predictor of cluster size

Notification interval (notification time gap) of ≤2 months between the first two cases in a cluster (Izumi et al. 2019) or ≤3 months between the first two cases in a cluster, Case #1 and Case #2 (Kam et al., 2022) is said to be a significant predictor of large genotype clusters (n≥6) within 3 years (Izumi et al., 2019; Kam et al., 2022). In this study, a notification interval of 1 month or less between Case#1 and Case#2 in the same cluster in a given year gave rise to big clonal complex clusters (n=6-225/CC) as shown in Table 4.12. For CC1: in the years 2015 to 2018, notification intervals between Case#1 and Case#2 were 4 days (2015), 20 days (2016), 1 day (2017) and 1 day (2018) respectively. All of the notification intervals were less than 1 month. A notification interval of >1 month resulted in small clonal complexes (n=2-5/CC). Therefore, this means that if a TB cluster in a district in Sabah has a notification interval of ≤1 month, it is highly likely that more TB cases will be discovered as this figure is an indicator of a big cluster. More resources will need to be allocated in terms of contact investigation to these areas. This further highlights the importance of genotyping in targeted interventions in the context of TB prevention and management.

Table 4.12: TB notification time interval between Case #1 and Case #2 in clonal complex clusters

Clonal Complex	Based on Actual Time (DLV2)			
	2015	**2016**	**2017**	**2018**
CC1	<1 month (4 days), n=30	<1 month (20 days), n=88	<1 month (1 day), n=225	<1 month (1 day), n=79
CC2	Nil	Nil	<1 month (7 days), n=6	<2 months (43 days), n=2
CC3	n=1	n=1	>3 months (152 days), n=5	Nil
CC4	>3 months (140 days), n=2	Nil	n=1	n=1
CC5	Nil	n=1	>3 months (160 days), n=1	>3 months (236 days), n=1
CC6	Nil	Nil	n=1	>3 months (356 days), n=1
CC7	Nil	Nil	<3 months (82 days), n=2	Nil
CC8	n=1	Nil	>1 month (35 days), n=2	Nil

4.4.2 Proposed genotype-complemented contact investigation

Based on the clustering analyses performed using the two-tiered genotyping method in this study, genotyping-complemented contact investigation is proposed to be conducted in two ways:

A) Direct cluster approach (DCA)

B) Prioritised cluster approach (PCA)

In the direct method, TB contact investigation to be conducted could be selected directly without further transformation from the case population listed in A1-A4 (Table 4.13). In the prioritised cluster approach, the clustering data is transformed using the steps laid out in B1 and B2 (Table 4.13).

Table 4.13: Direct cluster approach (DCA) and prioritised cluster approach (PCA) - simple heuristic approach for genotype complemented TB contact investigation in this study

Approach	Strategy
	A. Direct Cluster Approach
A1 - Largest clonal complex, CC1, n=422 A2 - Largest branch 12, n=274 A3 - MIT24 51 clusters, n=175 A4 - Large MIT24 clusters, e.g., MIT24 397, n=17	Choose feasible method(s). Focus resources for contact investigation on all cases accordingly
	B. Prioritised Cluster Approach
B1 - Super Clonal Complex	1. Determine factor(s) for analysis (Occupation) 2. Choose groups with highest likelihood of transmission (size, mobility) Unemployed (111) Blue Collar (66) Driver (21) Farmer & Fishermen (45) Home Maker (88) 3. Perform cluster analysis for every group 4. Combine largest clusters of all groups and reanalyse as 1 single group 5. Select largest clonal complex for further investigation
B2 – Expanded Clonal Complex Cluster Investigation	1. Determine factor(s) for analysis (e.g. Nationality) 2. Perform cluster analysis 3. Expand analysis to other cases with common / overlapping features

4.4.3 Direct cluster approach for TB case-contact investigation

The direct cluster approach (DCA) comprises four strategies, A1 to A4 (Table 4.13), with A1 (largest clonal complex) focusing contact investigation efforts on the largest number of TB index patients, which incurs the highest operational cost in this approach (n=422). This is followed by A2, largest branch (n=274), A3, all 51 exact MIT24 cluster (n=175), and A4, one of the largest MIT24 cluster MIT397 (n=17). Selection of the strategies is done based on the available resources in the TB prevention and management unit.

4.4.3.1 Direct cluster approach A1 (DCA-A1) largest clonal complex, CC1 n=422

Contact investigation activities could be done on all CC1 members (n=422) as it is the largest CC in this study and it is distributed all over the state (Table 4.14). Likelihood of exposure to contacts yet to be identified is high.

Table 4.14: Clonal complexes derived from 24-loci MIRU-VNTR DLV2

CC	Name	Number	Distribution
CC1	Largest Clonal Complex	422	All district
CC2	Sandakan Clonal Complex	8	Restricted to Sandakan only
CC3	MAS-INA Clonal Complex	7	1 INA, 6 MAS (2015-2018) EAI5, Manu3, T1-2, UNK-2
CC4	MAS-PHP Clonal Complex	4	KK and KM; LAM2, T3 & UNK
CC5	INA-MAS I Clonal Complex	3	BFT, LDU & PITAS, EAI2-MANILA and EAI5
CC6	FILIPINOS Clonal Complex	2	KK, LDU
CC7	INA-MAS II Clonal Complex	2	SDK and TWU
CC8	Beijing Clonal Complex	3	KK, SPN

4.4.3.2 Direct cluster approach A2 (DCA-A2) Largest branch 12 of CC1, n=274

Figure 4.10: Figure showing branch 12 (n=274) which is a part of CC1 (a) dendogram of the 593 MTB isolates, (b) branch 12 of CC1 (n=274), (c) detailed members of branch 12 of CC1.

The largest branch in CC1, which is branch 12 is a suitable sample population for contact investigation (Figure 4.10). This branch alone carries 9 DRTB cases which include one MDRTB cases as well as eight INH-R cases, and it is spread to the major districts in Sabah. An interesting observation is that seven of the eight INH-R cases carry the *katG* gene mutation. This is a strong indicator of possible direct transmission (de Vries et al., 2009; Leavitt et al., 2022). Resources could be focused on investigating and studying in depth the behavioural pattern of the members of branch 12 (DCA-2).

4.4.3.3 Direct cluster approach A3 (DCA-A3) 51 MIT24 24-loci MIRU-VNTR clusters, n=175

There are 51 clusters with the exact 24-loci MIRU-VNTR profile (Table 4.15). In the whole sample size of 175 for this specific population, no epidemiological link was established based on conventional contact investigation information in the TB contact registry (TBIS101C) from 2015 to 2018 except for 1 cluster MySBH0605 and MySBH0608 (n=2). However, the genetic relatedness and the distribution of the MTB highly indicate that the strains are highly transmissible (de Vries et al., 2009; García De Viedma and Pérez-Lago, 2018; Leavitt et al., 2022). Based on the nature of occupation and age group MySBH0608 could also be linked to MySBH0609 and MySBH0535 (Figure 4.11). All 3 cases are from the MIT24 397 cluster. The large MIT24 Cluster, MIT24-397, n=17 (Table 4.16) therefore is naturally one of the best sample populations on which to conduct in-depth contact investigation.

One interesting thing to highlight about the 51 MIT24 cluster is the baffling observation of the same exact 24-loci MIRU-VNTR clusters exhibiting different spoligotype patterns, although the patterns tend to be quite similar. Attributing this to homoplasy could be a reasonable explanation. In their study, Azé et al. (2015) observed that isolates with various spoligotype characteristics were frequently obtained throughout a range of years.

Table 4.15: 51 Clusters 24-loci MIRU-VNTR MIT24 of MTB in Sabah, n=175

No	SIT (Number of Isolates)	Total	24MIRU Profile	MIT24 Cluster
1	19 (2); Orphan (1)	3	2145243A2843266222342712	391
2	19 (2); 894 (1); 1169 (2); Orphan (1)	6	2145243A2923276222342713	392
3	19 (1); 1365 (1)	2	2145243A2943266222342612	393
4	19 (7)	7	2145243A2973276223342713	394
5	Orphan (1); 4192 (1)	2	2145243A2823276222342713	395
6	19 (11); 897 (6); 4200 (1); 4202 (1); Orphan (1)	20	2145243A2943266223342712	396
7	19 (12); 1490 (1); 1501 (1); 4197 (1); Orphan (2)	17	2145243A2923276223342713	397
8	4181 (1); 19 (2)	3	2145233A2843266223342713	398
9	Orphan (1); 19 (2)	3	2145243A2923276223342613	399
10	287 (7); 4201 (1)	8	2145243A1943266223342713	400
11	483 (3)	3	2145243A2943266223332713	401
12	19 (3)	3	2145243A2843266223332813	402
13	897 (1); 19 (3)	4	21452439293243266223342712	403
14	19 (3)	3	2145243A2973276213342713	404
15	19 (3)	3	2145243A2923276213342713	405
16	19 (4)	4	2A45243A2973246223342713	406
17	19 (2)	2	2145243A2923276223322713	407
18	19 (2)	2	21452439292326622334 2713	408
19	19 (5)	5	2145243A2823276223342713	409
20	19 (3)	3	2145243A2773266223342713	410
21	19 (2)	2	21452438292327622334 2713	411
22	19 (2)	2	2145243A2973266223342713	412
23	Orphan (1); 1365 (4)	5	2145233A2943266223342712	413
24	19 (3)	3	2145243A2963266223342712	414
25	19 (1); 1169 (1)	2	21452439292327622334 2713	415
26	19 (2); 4186 (1)	3	2145243A2943266223342713	416
27	19 (3)	3	21442236274325622334 2713	417
28	4197 (1); 19 (1)	2	21452439284326622334 2713	418
29	756 (2); 897 (1)	3	2145243A2443266223342712	419
30	1171 (1); 4200 (1)	2	2145243A2843266223342712	420
31	4196 (1); 287 (2)	3	21452435164326622334 2713	421
32	1365 (2)	2	2144243A2943266223342712	422
33	19 (1); Orphan (1)	2	21442238294325622334 2713	423
34	19 (2)	2	2145243A2873266223332813	424
35	Orphan (2)	2	2145233A2844266243342712	425
36	4204 (1); 19 (1)	2	21452434292327622334 8713	426
37	4204 (1); 19 (1)	2	21452434292327622334 7713	427
38	19 (2)	2	2145243A2673276223342513	428
39	Orphan (1); 19 (1)	2	2145243A2743246223342713	429
40	287 (2)	2	21452439282327622334 2713	430
41	19 (2)	2	2145243A2A73266223342713	431
42	19 (2)	2	2145253A2963266223342712	432
43	19 (2)	2	2145243A2743266223342713	433
44	1 (3)	3	244233352544425173353823	434
45	19 (2)	2	2145243A2923286213342713	435
46	4211 (1); Orphan (1)	2	21452439264326522334 2712	436
47	19 (1); 287 (1); 483 (1)	3	2145243A2A43266223342713	437
48	4181 (1); 19 (1)	2	21452439297327622334 2713	438

No	SIT (Number of Isolates)	Total	24MIRU Profile	MIT24 Cluster
49	287 (1); 19 (1)	2	2145243A2943266223342613	439
50	4215 (2)	2	2145243A2942266223342713	440
51	1365 (2)	2	21452535284326622334 2712	441

In their analysis, they found that most spoligotype patterns with more deletions were generally posterior, or more evolved relative to patterns with less deletions within the 24- loci MIRU-VNTR clusters. This proves that convergence in 24-loci MIRU-VNTR data for related isolates is not as uncommon as previously thought and that unusual mutations in spoligotype pattern without changes in the 24-loci MIRU-VNTR set may occur (Azé et al., 2015).

4.4.3.4 Direct cluster approach A4 (DCA-A4) large MIT24 Clusters, MIT24 397, n=17

In DCA-A4, genotype-complemented contact investigation can be conducted on one of the largest MIT24 cluster which is MIT24-397, n=17 (Table 4.16). According to the TB contact registry (TBIS101C) the index case CE05 has a known contact CE10. Other than this known index-contact relationship, no other epidemiological link has been established among the members of the MIT24-397 cluster. However, based on the nature of work and age group, two other members of the MIT24-397 BZ02 and CE01 have been selected as possible contacts in this contact investigation exercise. Figure 4.11 shows the possible links between these four TB cases. A focused investigation on this sub-cluster is likely to reveal new links, contacts and TB cases (active and latent).

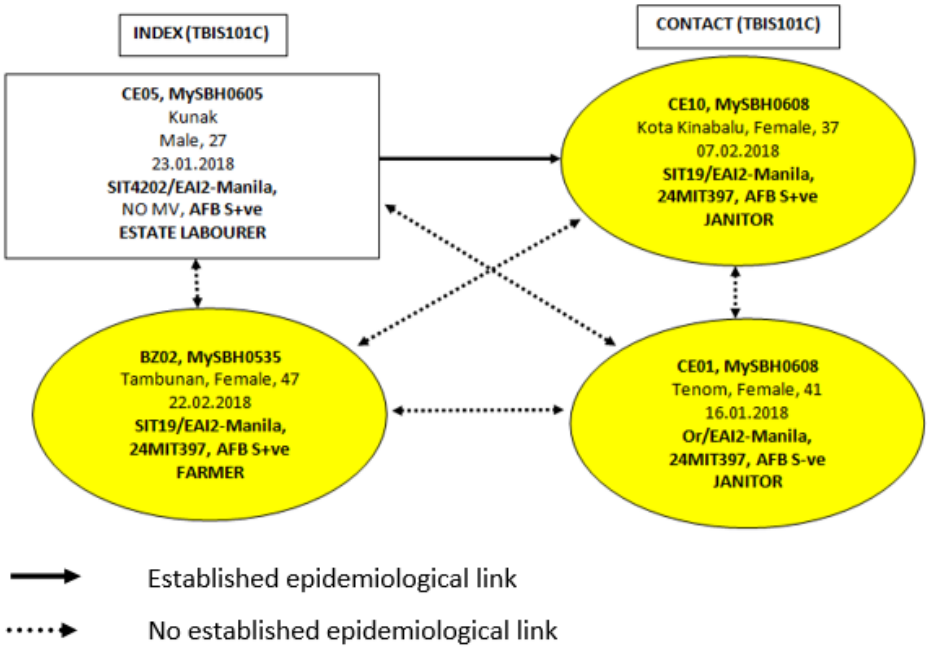


Figure 4.11: Combination of conventional and genotype-complemented TB case contact investigation (MIT24-397)

Table 4.16: MIT24-397 (24-loci MIRU-VNTR profile 2145243A2923276223342713)

No	Patient ID	Date of Diagnosis	Microscopy	Sublineage	SIT	Status	Country of Birth	District	Nature of Occupation	Spoligotyping	DST
1	MYSBH_0366	30/03/2015	Scanty	EAI2-Manila	4197	Index	PHP	KOTA KINABALU	Home Maker	[redacted]	1
2	MYSBH_0264	10/10/2016	Scanty	EAI2-Manila	Orphan	Index	PHP	KUNAK	Unemployed	[redacted]	1
3	MYSBH_0248	02/11/2016	NA	EAI2-Manila	19	Index	MAS	SIPITANG	Unknown	[redacted]	1
4	MYSBH_0315	22/12/2016	NA	EAI2-Manila	1490	Index	MAS	PITAS	Student	[redacted]	1
5	MYSBH_0395	07/01/2017	Scanty	EAI2-Manila	19	Index	MAS	RANAU	Self Employed	[redacted]	1
6	MYSBH_0330	09/02/2017	3+	EAI2-Manila	19	Index	MAS	KOTA KINABALU	Factory Labourer	[redacted]	1
7	MYSBH_0324	03/03/2017	Negative	EAI2-Manila	1501	Index	MAS	SIPITANG	Healthcare	[redacted]	1
8	MYSBH_0193	14/03/2017	1+	EAI2-Manila	19	Index	MAS	TUARAN	Home Maker	[redacted]	1
9	MYSBH_0872	19/05/2017	Negative	EAI2-Manila	19	Index	MAS	TUARAN	Driver	[redacted]	1
10	MYSBH_0072	19/05/2017	3+	EAI2-Manila	19	Index	MAS	TUARAN	Driver	[redacted]	1
11	MYSBH_0602	10/01/2018	Scanty	EAI2-Manila	19	Index	MAS	BEAUFORT	Unemployed	[redacted]	1
12	**MYSBH_0608**	**16/01/2018**	**Negative**	EAI2-Manila	Orphan	Contact	MAS	TENOM	Janitor	[redacted]	1
13	MYSBH_0558	26/01/2018	2+	EAI2-Manila	19	Index	MAS	NABAWAN	Unemployed	[redacted]	1
14	**MYSBH_0609**	**07/02/2018**	**1+**	EAI2-Manila	19	Contact	MAS	KOTA KINABALU	Janitor	[redacted]	3
15	**MYSBH_0535**	**22/02/2018**	**1+**	EAI2-Manila	19	Contact	MAS	TAMBUNAN	Farmer	[redacted]	1
16	MYSBH_0538	11/03/2018	3+	EAI2-Manila	19	Index	MAS	TUARAN	Unemployed	[redacted]	1
17	MYSBH_0539	16/04/2018	Scanty	EAI2-Manila	19	Index	MAS	KOTA KINABALU	Unemployed	[redacted]	1

No patient was found to be contact via conventional contact tracing within this cluster

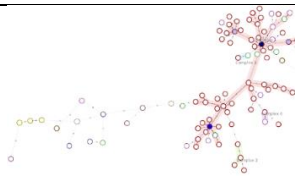
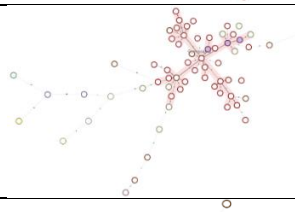
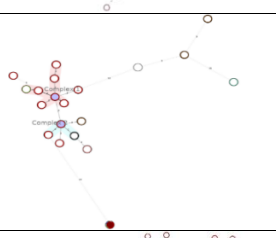
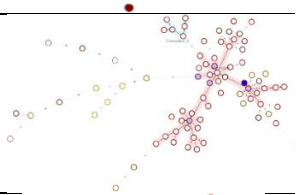
4.4.4 Prioritized cluster approach (PCA) for TB case-contact investigation

In this approach, the bigger or more important CC clusters as determined by the TB prevention and management team based on relevant factors are reanalyzed by a second-round (nested) clustering exercise. Contact investigation activities will then be applied on the resulting CC clusters.

4.4.4.1 Prioritised cluster approach B1 (PCA-B1)- Super clonal complex (CC) cluster investigation

When the CCs are analyzed altogether, individually and then superimposed, the resulting group of cases are expected to yield the highest possibility of transmission. In this example, occupation was selected as the factor to be investigated. All cases from the occupation categories selected based on population size / number and mobility (highest likelihood of transmission): Unknown (n=111), Blue collar worker (n=66), driver (n=21), Farmer and fishermen (n=45) and home maker (n=88). Unknown cases were excluded in the analysis. Individual occupation categories were first analysed using *MIRU-VNTRplus* (Table 4.17). All resulting CC from each category were pooled into one dataset and reanalysed using *MIRU-VNTRplus*. The biggest clonal complex from each occupation category were then combined and reanalysed as a single group using *MIRU-VNTRplus* (Figure 4.12). The resulting largest CC was selected for contact investigation.

Table 4.17: Prioritised cluster approach PCA-B1, step 2 - 24-loci MIRU-VNTR clonal complexes of TB cases for prioritization for contact investigation

Patient Occupation	No. of Isolates	No. of CC	% in CC	% Singleton	% in MIT24	Minimum Spanning Tree
Unemployed	105	3	64.8	35.2	30.5	
Blue-Collar	60	1	65	35	16	
Driver	19	2	36.8	42.1	36.8	
Home Maker	84	2	61.9	38.1	33.3	
Farmer & Fisherman	27	1	74.1	25.9	18.5	

Clonal complexing and clustering indicate that recent transmission of TB has taken place. The bigger the size of the cluster implies that many cases are involved. Geo-mapping of the location of the cases would show how extensive the transmission is. Since most places in Sabah are accessible by road, big clusters or clonal complexes are not at all unexpected. TB cases are usually managed within districts, unless there's an official request for TB patients to be transferred inter-district. There is only one clonal complex for the blue-collar worker. This implies that it is highly likely that they run in the same specific

circuit e.g., construction projects, estates, and social spaces. Therefore, further understanding of the dynamics of TB transmission, particularly in an urban setting, can be aided by activity space-based spatial analysis for potential TB transmission venues such as workplaces, railway stations, malls, and market places (Izumi et al., 2015).

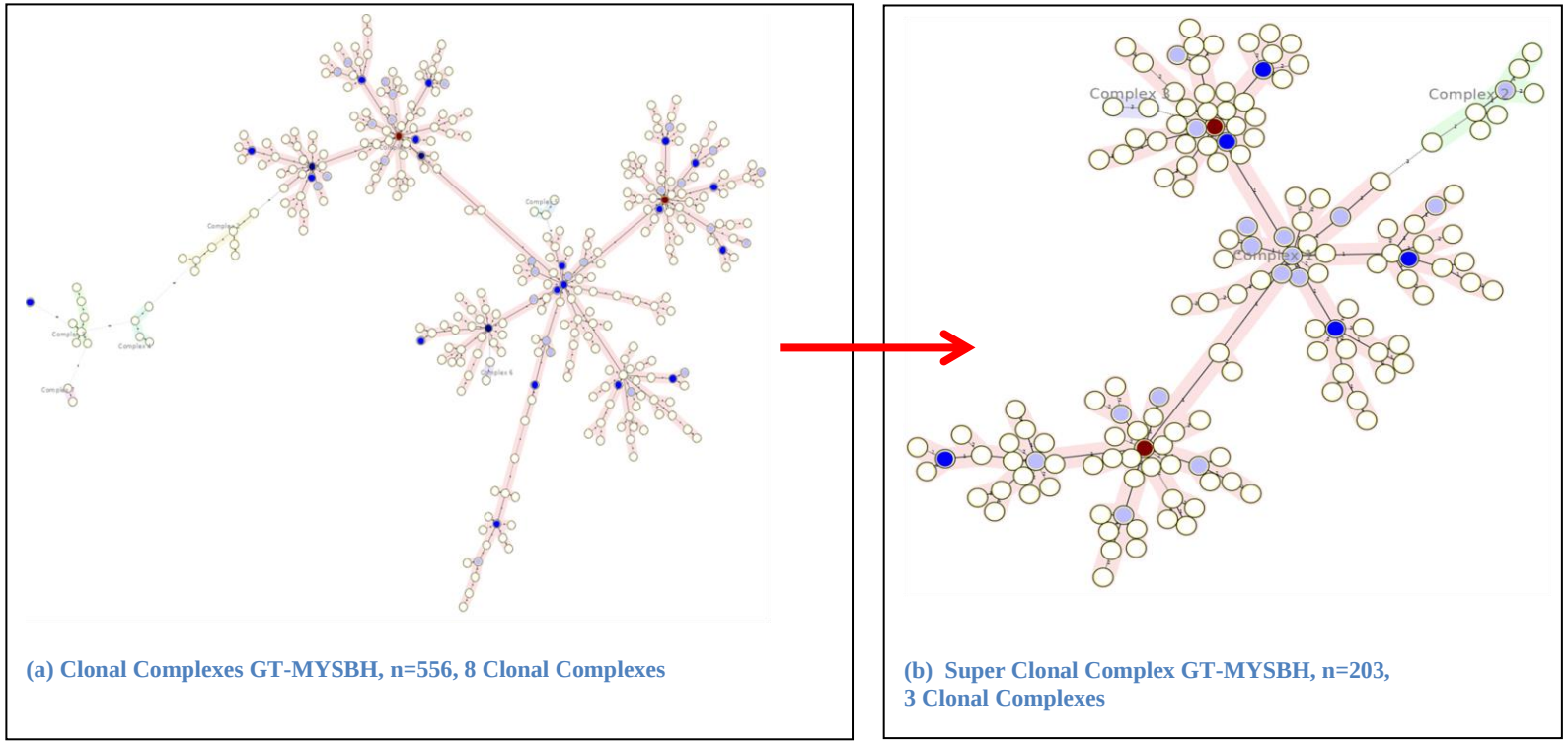


Figure 4.12: Prioritised Cluster Approach 1(PCA-B1) for genotype complemented contact investigation, (a) DLV2 24-loci MIRU-VNTR Clonal Complexes, (b) Super Clonal Complex of isolates with highest likelihood of transmission

4.4.4.2 Expanded clonal complex cluster investigation (PCA-B2)

In this strategy, nationality was chosen as the factor for analysis, and specifically the Filipino (PHP) nationality. 24-loci MIRU-VNTR was performed and MST was built for the three nationalities in this study (Table 4.13-B2 and Figure 4.13).

Figure 4.13 shows that all 3 nationalities in the Sabah study population showed both star and chain-like network in the MST. The difference is in the proportion of the star and chain-like pattern in the MST network. The Filipino network showed greater star-like pattern compared to chain-like network, implying greater recent transmission in the state (Figure 4.13-a). The Indonesian network showed more chain-like pattern as opposed to the star-like pattern, indicating reactivation (Figure 4.13-b) while Malaysians in this study had an equal ratio of star and chain-like pattern, meaning both active transmission and reactivation are occurring concurrently (Figure 4.13-c). It is presumed that a star-like network topology and low diversity are likely signs of the genotype's recent spread in the area, indicating the existence of many secondary cases caused directly by a single index case. In contrast, a higher level of diversity and a network's topology that resembles a chain or, more specifically, a cloud may indicate a longer evolutionary history (Yin et al., 2016, García De Viedma and Pérez-Lago, 2018).

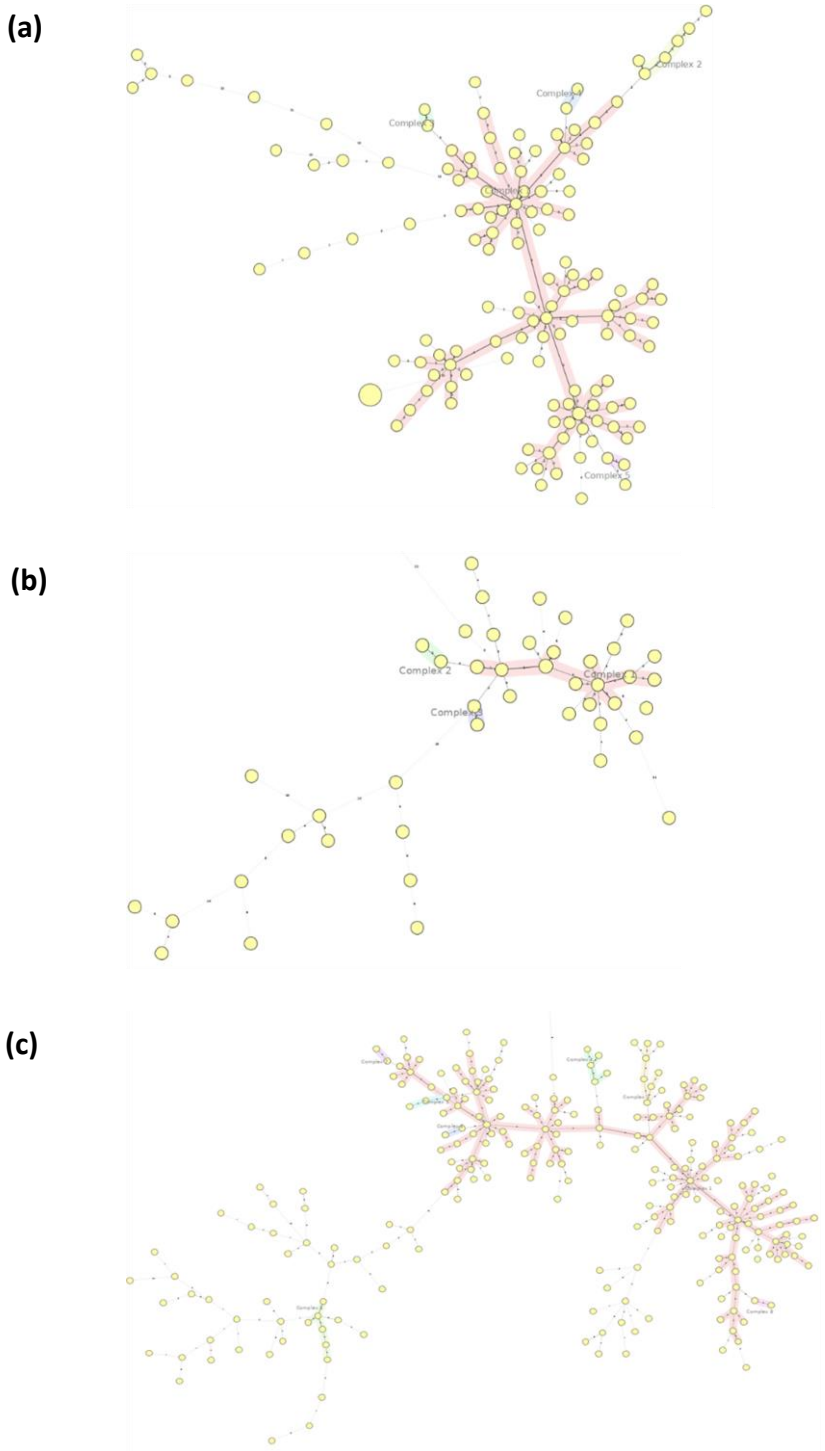


Figure 4.13: Minimum Spanning Tree of MTBC in Sabah, East Malaysia among(a) Filipinos, CC=5; (b) Indonesian, CC=3; (c) Malaysian, CC=8. Clusters and clonal complexes indicate recent transmission while singletons indicate reactivation

24-loci MIRU-VNTR were performed on the PHP isolates, and two CC were obtained CC1-PHP and CC2-PHP. CC2-PHP (n=5) was chosen for further investigation and simplicity of analysis because all its members are Filipinos (Figure 4.14). No epidemiological link was established among the five members based on conventional contact investigation (TBIS101C register). CC1-PHP is a large CC and is a subset of the main CC1 (n=422) which is the largest clonal complex in this study. When CC2-PHP (Figure 4.14) was analysed individually, MySBH0400 is the index case for this clonal complex. Thus, the possible direction of transmission is proposed in Figure 4.14. However, upon scrutiny, two of the CC2-PHP members were also part of the MIT24 clusters MySBH0400 – MIT24-417 and MySBH0577 – MIT24-423.

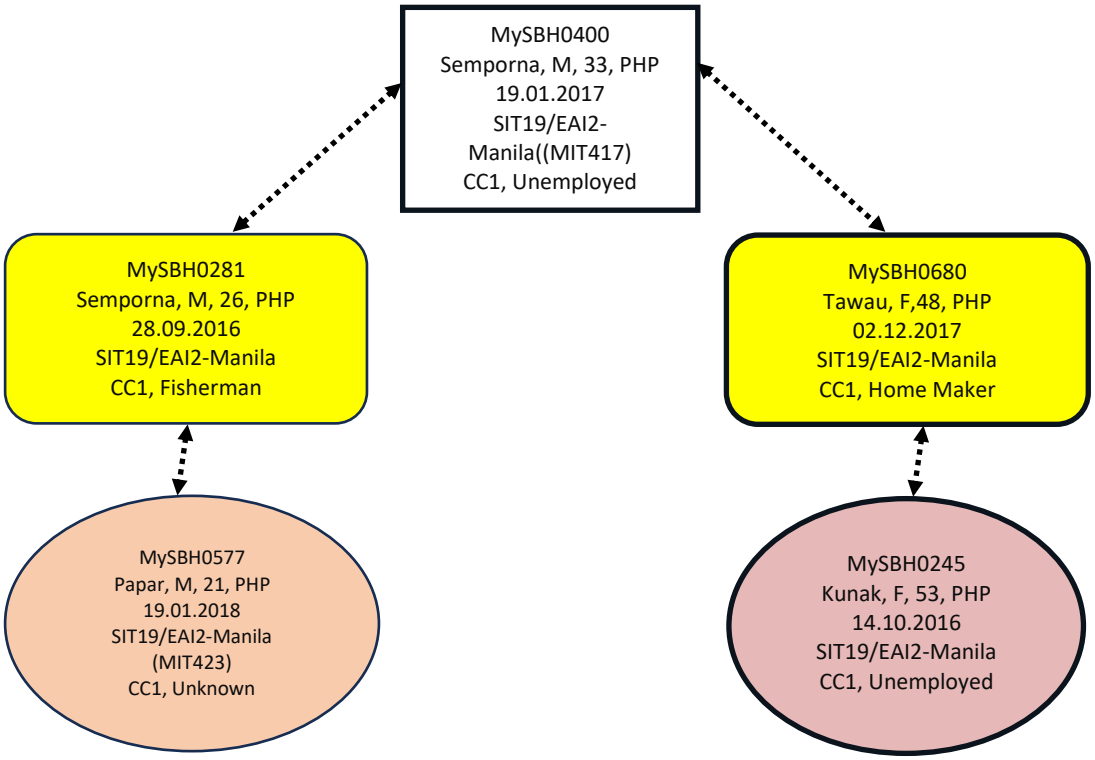


Figure 4.14: Possible transmission of TB CC2-PHP

Table 4.18: Filipino (PHP) – Malaysian (MAS) MTB combined clonal complex

Sample ID	Country	Notification	District	Sex	Age	SIT / SUB-LINEAGE	MIT24	OCCUPATION
MYSBH_0754	MAS	16/07/2016	TAMBUNAN	M	21	SIT19/EAI2-Manila	417	Craft & related Trade
MYSBH_0281	PHP	28/09/2016	SEMPORNA	M	26	SIT19/EAI2-Manila		Fisheries
MYSBH_0245	PHP	14/10/2016	KUNAK	F	53	SIT19/EAI2-Manila		Unemployed
MYSBH_0358	MAS	19/10/2016	PAPAR	F	37	SIT19/EAI2-Manila	417	Home Maker
MYSBH_0400	PHP	19/01/2017	SEMPORNA	M	33	SIT19/EAI2-Manila	417	Unemployed
MYSBH_0680	PHP	02/12/2017	TAWAU	F	48	SIT19/EAI2-Manila		Home Maker
MYSBH_0577	PHP	19/01/2018	PAPAR	M	21	SIT19/EAI2-Manila	423	Unknown
MYSBH_0615	MAS	19/01/2018	KOTA KINABALU	F	20	Orph/EAI2-Manila	423	Service & Sales

MIT417 and MIT423 differs by 1 locus, 1955_Mtub21: 6 and 8 respectively. MySBH0281 (possible index) differs with MIT417 by 1 locus, 2163b_QUB11b: 9 and 7 respectively; and differs with MIT423 by 1 locus, 1955_ Mtub21: 6 and 8 respectively

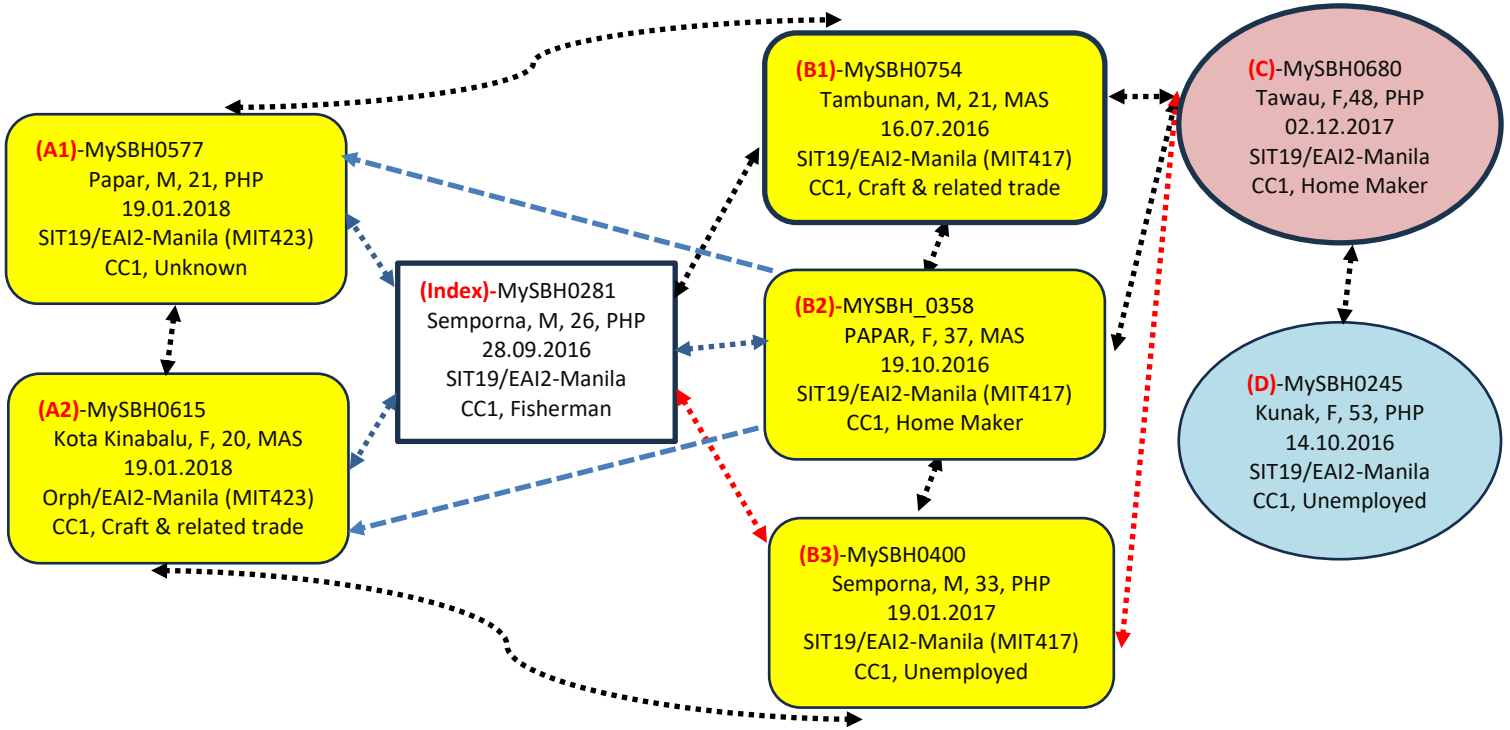


Figure 4.15: Possible transmission of PHP–MAS MTB combined clonal complex (CC PHP-MAS)

The information on the relevant MIT24 clusters allowed the expansion of analysis to other members of the respective MIT24 clusters. The other MIT24 cluster members were combined with the initial CC2-PHP and the expanded group which now comprised PHP-MAS (Table 4.18) were reanalysed using *MIRU-VNTRplus*. The new MST built showed that the index case has now shifted from MySBH0400 to MySBH0281 (Figure 4.15).

Figure 4.15 shows the many possible routes of transmission as indicated by the different coloured arrows for the new combined CC PHP-MAS (4.14-a). The index case is MySBH0281; MIT24-423 cluster is labelled as A: A1-MySBH0577 and A2-MySBH0615; MIT24-417 cluster is labelled as B: B1-MySBH0754, B2-MySBH0358 and B3-MySBH0400; C: MySBH0680 and D: MySBH0245. However, referring to Figure 4.15, two examples of the most likely scenarios depicting the routes of transmission based on the date of diagnosis, location and ethnicity (with the assumption of higher intra-ethnicity interactions) are suggested as follows:

Scenario 1: Index → A1 → A2

Scenario 2: Index → B3 → C

In scenario 1, the index case, a Filipino fisherman from the Semporna district (Tawau division) diagnosed with TB in September 2016 with a SIT19/EAI2-Manila spoligotype may have come into contact with A1 who is a Filipino man (employment status unknown) in Papar with a 2+ AFB microscopy grading. The distance between Semporna and Papar is about 520km. It is possible that the transmission event might have happened in Semporna (possible

place of origin of index and A1) or it could also happen at a well-known fish market in Papar district that sells fish and seafood from all over the state (including the east coast). A1 in turn might have infected A2, a female Malaysian in Kota Kinabalu (~40km or half an hour drive away from Papar) with a Negative AFB microscopy result. Both A1 and A2 were diagnosed with TB on the same day about two years after the index was diagnosed. This two-year gap before diagnosis might mean that the original MTB strain might have mutated slightly in A1 resulting a change at MIRU-VNTR locus 1955_Mtub21 from 6 repeats (index) to 8 repeats (A1). An AFB smear grading of 2+ implies that A1 was diagnosed with a more advanced TB compared to A2. Both A1 and A2 have the same 24-loci MIRU-VNTR profile (MIT24-423), but A2 has a SIT19/EAI2-Manila spoligotype that has a spacer-24 deletion making it an Orphan/EAI2-Manila spoligotype. This means that the MTB has changed slightly during incubation in A2. Although the relationship between the index, A1 and A2 were not captured in the TBIS records, this possibility cannot be ruled out and it warrants a more in-depth investigation. This also shows that the knowledge about the shared social space is important to make informed decision to guide the investigation and interventions necessary (Izumi et al., 2015).

In scenario 2, the index might have infected B3, a Filipino man from the same district who is unemployed and was diagnosed with TB four months after the index' TB diagnosis. B3 has a MIT24-417 MIRU-VNTR profile which has a single locus difference at 2163b_QUB11b, from 9 (index) repeats to 7 repeats (B3). B3 then infects C, a home maker in Tawau (which is in the same division

as Semporna) at about 10 months after being diagnosed with TB. The difference in their MIRU-VNTR profile is only at 2165_ETR-A locus, from 4 repeats (B3) to 7 repeats (C). However, there is still a possibility that the transmission could have occurred in the opposite direction whereby C was infected by D. This is the best we can do with the available epidemiological and genotyping data. Should universal genotyping be conducted, a more comprehensive and accurate picture of the transmission routes can be generated.

If we were to conduct enhanced epidemiological investigation, EEI (Rodwell et al., 2012; Munang, 2019) we would be able to determine transmission and direction of transmission at a higher rate. In this instance, a work-play diary which records social network, places frequented etc. in order to establish pattern of behaviour would greatly enhance the contact investigation strategy whereby possible TB cases could be detected much earlier as demonstrated by Munang et al. (2016). This could be seen as a breach of privacy and personal freedom. However, this gray area needs to be addressed tactfully in order to control spread of disease (Richards et al., 2012; Simonsen et al., 2016; Bardosh et al., 2019; Aiello et al., 2020).

4.5 Relationship between MTB sub-lineages and drug-resistance

The routine testing for MTB drug susceptibility in Sabah was done using the Absolute Concentration Method, ACM (Barry et al., 1998; Rigouts et al., 2013; Stinson et al., 2014). ACM is a widely used, simple and robust standardised phenotypic DST method performed using drug-containing solid-media (DCSM) at a fixed or absolute concentration (Barry et al., 1998; Rigouts et al., 2013; Stinson et al., 2014). A standardised inoculum of MTB isolates is cultured onto the DCSM. However, it is time consuming, usually taking up to 4 weeks to be completed (Barry et al., 1998; Rigouts et al., 2013; Stinson et al., 2014). From the total sample of 1019 MTB isolate, 899 (86.4%) were pan-susceptible, 6 (0.6%) cases were MDRTB, 78 (7.7%) reported mono and poly drug-resistance, while 55 (5.4%) had failed drug susceptibility testing (FD).

Out of the 556 MTB genotyped using 24-loci MIRU-VNTR, 43 were found to exhibit drug resistance (CC1=20, INH-R=9 [*katG*=7], MDRTB=2 [diff in Mtub04 and ETRB, five months difference], S-R=9; S=20, INH-R=6 [2 *katG*], RRTB=1. CC1 Branch 12 using 24-loci MIRU-VNTR was found to be the biggest branch in the largest Clonal Complex. 9 DRTB cases (including one MDRTB cases, seven INH-R *katG* mutation, and one INH-R *inhA* promoter mutation) were detected in this branch alone.

Table 4.19: Distribution of MTB drug-resistance by country of origin in Sabah

Country	Total Cases of TB cases (%)	Number of DRTB Cases (% Ov, % WG)	MDRTB (% Ov, % WG)	I-R (% Ov, % WG)	R-R (% Ov, % WG)	S-R (% Ov, % WG)
Indonesia	88 (8.6)	9 (0.9, 10.2)	0 (0,0)	2 (0.2, 22.2)	1 (0.1, 11.1)	6 (0.6, 66.7)
Malaysia	674 (66.1)	47 (4.6, 7.0)	5 (0.5, 10.6)	18 (1.8, 38.3)	1 (0.1, 2.1)	23 (2.3, 48.9)
Philippines	257 (25.2)	28 (2.7, 10.9)	1 (0.1, 3.6)	14 (1.4, 50)	2 (0.2,7.1)	11 (1.1, 39.3)
Total	**1019 (100)**	**84 (8.2)**	**6 (0.6, 7.1)**	**34 (3.3, 40.5)**	**4 (0.4, 4.8)**	**40 (3.9, 47.6)**

Note: OV - overall / total; WG – within group / proportionate

Table 4.19 summarizes the distribution of MTB drug resistance by country of origin in Sabah, East Malaysia. Overall, there were a total of DRTB 84 isolates in this study: 40 Streptomycin mono-resistant cases (ACM only); 31 INH mono-resistant cases, three STR (ACM only) and INH polyresistant cases; three RRTB cases, and six MDRTB cases. Although Malaysians record the higher absolute number of DRTB cases, Filipinos (10.9%) and Indonesians (10.2%) have higher percentages of DRTB within their respective populations.

The drug resistance profile for this sample population (Phenotypic Method, ACM to Molecular method) are as follows: Failed DST - 55 (5.39%): 1 (0.10%); Pan-susceptible 897 (88%): 982 (96.37%), MDRTB - 4 (0.39%):6 (0.59%) – both methods detected three common MDRTB and another 1 and 2 different MDRTB isolates respectively; DRTB (mono and poly drug resistance) - 63 (6.18%):31 (3.04%), ACM also include 41 cases of Streptomycin resistance. Molecular detection only targets region of resistance for the two main components of the first-line anti-TB medication namely isoniazid and rifampicin.

4.5.1 Mono drug-resistant MTB

4.5.1.1 Isoniazid (INH) resistant MTB

Isoniazid (INH) resistance was found in 34 isolates in this study population: 31 INH mono-resistance; three polydrug-resistant (INH and STR-R) isolates - 1/3 has *inhA*-prom_mut-15_T (Table 4.20). Another four cases

were detected by LPA MDRplus with three *inhA* mutation and one *katG* mutation which were missed by ACM and TB-RINT.

High level INH resistance (*katG* gene mutation, 14 cases) is just slightly higher compared to low level INH resistance (*inhA* gene mutation, 13 cases). Among the 34 INH resistant (INH-R) cases, two of them are from the Beijing sub-lineage (one *katG* gene mutation and one with *inhA* prom gene mutation, both from Tawau Division). 24 of the INH-R cases involve isolates from the EAI sub-lineage, meanwhile T1 and LAM9 has one each, and Unknown sub-lineage has six cases of INH resistance-from all over the state especially from the west coast (Table 4.20).

Overall, ACM detected 18 INH-R cases (1.77%), whereby six cases (four EAI2 Manila, one Unknown, one T) were missed by molecular methods (0.59%); Molecular methods (TB-RINT and LPA MDRplus) detected 27 INH-R cases (2.65%), of which 16 cases were missed by ACM (1.57%). The cases that were missed by ACM comprises seven cases with *katG* gene mutation (4 EAI2 Manila, two Unknown, one Beijing) and 9 cases with *inhA* gene mutation (one Beijing, one EAI2, 1one EAI2 Manila, five EAI5, one Unknown).

Although the sample number is small, it is interesting to note that *katG* gene mutation were associated to slightly more failed DST (FD) cases as compared to *inhA* gene mutation (4 FD to 2 FD). The rest of the missed INH-R cases were reported as susceptible to all antibiotics in the anti-TB regime tested.

It is also interesting to note that the pattern of INH resistance seems to have a slight difference in its geographical distribution in this study. Both INH resistance conferring gene mutations can be found in the west coast especially Kota Kinabalu district where it is the busiest economic hub of the state. Low level INH resistance (*inhA* gene mutation) seems to be more prevalent in Sandakan area (Sandakan, Beluran and Lahad Datu) close to Philippines waters where Filipino population is higher; meanwhile *katG* gene mutation 315 is more predominant in Tawau area (Tawau, Semporna and Kunak districts). The relatively higher prevalence of *inhA* mutation in Tawau division for Semporna (well known for beautiful islands) could perhaps be due to importation of cases from domestic and international tourists. One case of INH-R with *katG* mutation is found on the Malaysian part of Pulau Sebatik (Pulau Sebatik is shared by Malaysia and Indonesia). This particular TB case is a Malaysian female Sabah native, possibly of Bugis descent. Judging by the close proximity to Indonesia, it is tempting to say that *katG* gene mutation in Sabah might have been brought from Indonesia. However, most patients with *katG* gene mutation are Filipinos (9 of 14, seven EAI2 Manila, one EAI5 and one Unknown) and Malaysians (four of 14; one Beijing, one EAI, one LAM and one Unknown). Only one patient from Indonesia has MTB with the *katG* mutation (Unknown lineage), and this patient is residing in the west coast (Tuaran). Higher transmission spread among Filipinos warrants a closer look at how transmission occurs among this population.

Table 4.20: MTB isoniazid resistance in Sabah, East Malaysia

No	Isolate (SIT/sub-lineage)	ACM-DST (1-0-0)	TB-RINT (0-1-0)	LPA-MDR Plus (0-0-1)	Final DST (this study)
1	MYSBH_0016 (SIT287/EAI2-Manila)	SI-R	inhA-prom_mut-15_T I-R	Not Done	SI-R (1-1-0) inhA-prom_mut-15_T
2	MYSBH_0041 (Orphan/EAI2-Manila)	All-S	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
3	MYSBH_0066 (SIT19/EAI2-Manila)	FD	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
4	MYSBH_0067 (SIT4182 / EAI2)	FD	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T
5	MYSBH_0820 (SIT1 / Beijing)	FD	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
6	MYSBH_0073 (SIT4181 / EAI2-Manila)	All-S	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
7	MYSBH_0136 (SIT19/EAI2-Manila)	I-R	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
8	MYSBH_0137 (SIT19/EAI2-Manila)	I-R	Not Detected	Not Done	I-R (1-0-0)
9	MYSBH_0147 (SIT19/EAI2-Manila)	I-R	Not Detected	Not Done	I-R (1-0-0)
10	MYSBH_0204 (SIT19/EAI2-Manila)	SI-R	Not Detected	Not Done	SI-R (1-0-0)
11	MYSBH_0233 (Orphan / Unknown)	FD	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
12	MYSBH_0251 (SIT894 / EAI2-Manila)	I-R	Not Detected	Not Done	I-R (1-0-0)
13	MYSBH_0400 (SIT19 / EAI2-Manila)	I-R	katG_315_mutC I-R	katG mutation	I-R (1-1-1) katG_315_mutC

Note: S-R, streptomycin resistant; I-R, isoniazid resistant; SI-R, streptomycin and isoniazid resistant; FD – failed DST

Table 4.20 (continued): MTB isoniazid resistance in Sabah, East Malaysia

No	Isolate (SIT/sub-lineage)	ACM-DST (1-0-0)	TB-RINT (0-1-0)	LPA-MDR Plus (0-0-1)	Final DST (this study)
14	MYSBH_0414 (SIT4186/Unknown)	All-S	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T
15	MYSBH_0425 (SIT1365/EAI5)	All-S	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T
16	MYSBH_0431 / (SIT1079/T1)	I-R	inhA-prom_mut-15_T I-R	Not Done	I-R (1-1-0) inhA-prom_mut-15_T
17	MYSBH_0451 / (SIT4199/EAI5)	All-S	Not Detected	inhA mutation	I-R (0-0-1) inhA
18	MYSBH_0691 / (SIT897/EAI2-Manila)	SI-R	katG_315_mutC I-R	Not Done	I-R (1-1-0) katG_315_mutC
19	MYSBH_0453 / (Orphan/EAI5)	All-S	Not Detected	inhA mutation	I-R (0-0-1) inhA
20	MYSBH_0469 / (Orphan/Unknown)	I-R	Not Detected	Not Done	I-R (1-0-0)
21	MYSBH_0568 / (Orphan/Unknown)	SI-R	Not Detected	Not Done	SI-R (1-0-0)
22	MYSBH_0583 / (SIT1365/ EAI5)	All-S	Not Detected	inhA mutation	I-R (0-0-1) inhA
23	MYSBH_0620 / (SIT19/EAI2-Manila)	I-R	inhA-prom_mut-15_T I-R	Not Done	I-R (1-1-0) inhA-prom_mut-15_T
24	MYSBH_0653 / (SIT4204/Unknown)	SI-R	katG_315_mutC I-R	Not Done	I-R (1-1-0) katG_315_mutC
25	MYSBH_0790 / (SIT897/EAI2-Manila)	FD	katG_315_mutC I-R	Not Done	I-R (0-1-0) katG_315_mutC
26	MYSBH_0771 / (SIT19/EAI2-Manila)	All-S	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T
27	MYSBH_0785 / (SIT1/Beijing)	All-S	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T

Note: S-R, streptomycin resistant; I-R, isoniazid resistant; SI-R, streptomycin and isoniazid resistant; FD – failed DST

Table 4.20 (continued): MTB isoniazid resistance in Sabah, East Malaysia

No	Isolate (SIT/sub-lineage)	ACM-DST (1-0-0)	TB-RINT (0-1-0)	LPA-MDR Plus (0-0-1)	Final DST (this study)
28	MYSBH_0108 / (SIT1365/EAI5)	I-R	katG_315_mutC I-R	Not Done	I-R (1-1-0) katG_315_mutC
29	MYSBH_0452 / (Orphan /Unknown)	All-S	Not Detected	katG mutation	I-R (0-0-1) katG
30	MYSBH_0847 / (Orphan/EAI5)	FD	inhA-prom_mut-15_T I-R	Not Done	I-R (0-1-0) inhA-prom_mut-15_T
31	MYSBH_0934 / (SIT483/EAI2-Manila)	I-R	katG_315_mutC I-R	Not Done	I-R (1-1-0) katG_315_mutC
32	MYSBH_0943 / (SIT1365/EAI5)	I-R	inhA-prom_mut-15_T I-R	Not Done	I-R (1-1-0) inhA-prom_mut-15_T
33	MYSBH_0993 / (SIT891/LAM9)	I-R	katG_315_mutC I-R	Not Done	I-R (1-1-0) katG_315_mutC
34	MYSBH_1014 / (SIT19/EAI2-Manila)	I-R	Not Detected	Not Done	I-R (1-0-0)

Note: S-R, streptomycin resistant; I-R, isoniazid resistant; SI-R, streptomycin and isoniazid resistant; FD – failed DST

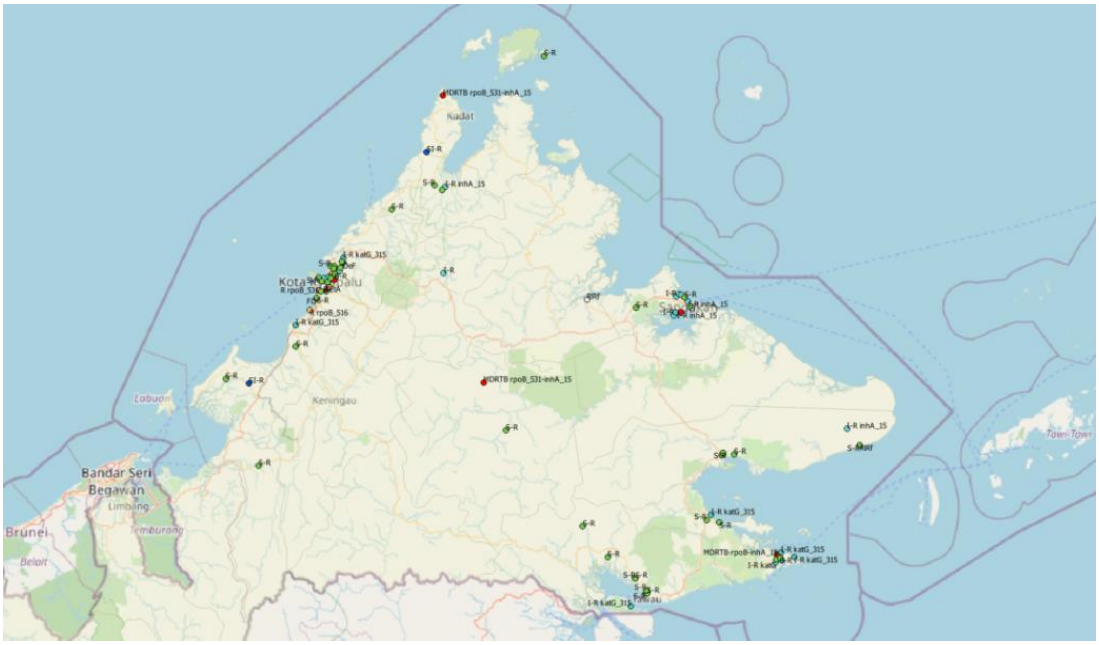


Figure 4.16: Drug resistant TB (DRTB) distribution in Sabah

***Kudat division**: 1 inhA mutation (Orphan-EAI5), 1 I-R (ACM); **Lahad Datu area**: 1 inhA mutation (T1); **Sandakan division**: 4 inhA mutation (2 EAI2 Manila, 1 EAI5, 1 EAI2newSIT); 2 katG mutation (1 LAM9, 1 unknown newSIT); 2 INH-R (EAI2-Manila);**Tawau division** – 3 inhA mutation (2 EAI5-1 new SIT, 1 Beijing – all in Semporna – beautiful Tourist Spots- likely imported); 7 katG mutation (1 Beijing, 4 EAI2-Manila plus 1 new SIT, 1 EAI5 & 1 orphan); **West Coast** – 4 inhA mutation (2 EAI5-1orphan, 1 EAI2 Manila, 1 Unknown new SIT); 5 katG mutation (4 EAI2-Manila with 1 orphan, 1 Unknown-orphan); 4 INH-R (3 EAI2-Manila, 1 Unknown-orphan)*

Figure 4.16 shows the distribution of the DRTB (n=84) cases in this study. Most of them are in areas with relatively high population densities. MDRTB cases (red) are found in the west coast, Kudat division, interior division and the Sandakan division.

4.5.1.2 Rifampicin resistant MTB (RRTB)

Four cases of Rifampicin mono-resistance (RRTB) were found in the west coast of Sabah (Table 4.21). As in the cases of INH mono-resistance, the phenotypic and genotypic methods showed some discrepancies in the RRTB results. The four RRTB cases include two EAI, one Beijing and one unknown sub-lineages as follows: MYSBH_0277, *rpoB*_516_mut_GTC R-R; SIT1365/EAI5, Papar (2016); MYSBH_0677, R-R (ACM), SIT1/Beijing, Kota Kinabalu (2017); MYSBH_0736, *rpoB*_531_mut_TTG R-R, Orphan/Unknown, Kota Kinabalu (2017); MYSBH_0779, *rpoB*_526_mut_TAC R-R, SIT897/EAI2-Manila, Penampang (2015). One RRTB case (MYSBH_0677, SIT1/Beijing) was detected by ACM but not detected by molecular method, while another RRTB case MYSBH_0277 *rpoB*_516_mut_GTC was reported as Streptomycin resistant using ACM DST.

4.5.2 Multi drug-resistant MTB (MDRTB)

Six MDRTB cases were detected in this study, 0.59% (Table 4.22). Five of the MDRTB cases were detected by direct molecular method (TB-RINT and LPA MDRplus) all with the same mutations (*rpoB*_531_mut_TTG; *inhA*-prom_mut-15_T), meanwhile one case, MYSBH 0566, SIT7/LAM2 was detected by ACM only (MDRTB) and partially supported by LPA MDRplus (*katG* gene mutation only). The MDRTB cases comprised Malaysians (3 EAI2 Manila and 1 LAM2) and 2 Filipinos (1 EAI5 and 1 EAI3 IND).

Table 4.21: MTB rifampicin resistance in Sabah, East Malaysia

Isolate (SIT/sub-lineage)	ACM-DST (1-0-0)	TB-RINT (0-1-0)	LPA MDRplus (0-0-1)	Final DST (this study)
MYSBH_0277 (SIT1365/ EAI5)	S-R	rpoB_516_mut_GTC	Not Done	R-R (0-1-0) rpoB_516_mut_GTC
MYSBH_0677 (SIT1 / Beijing)	R-R	Not Detected	Not Detected	R-R (1-0-0)
MYSBH_0736 (Orphan/ Unknown)	R-R	rpoB_531_mut_TTG	Not Detected	R-R (1-1-0) rpoB_531_mut_TTG
MYSBH_0779 (SIT897/ EAI2-Manila)	R-R	rpoB_526_mut_TAC	Not Done	R-R (1-1-0) rpoB_526_mut_TAC

Note: S-R, streptomycin resistant; R-R, rifampicin resistant. LPA MDRplus requires higher concentration of MTB DNA

Table 4.22: Multi drug-resistant MTB (MDRTB) in Sabah, East Malaysia

Isolate	ACM-DST (1-0-0)	TB-RINT (0-1-0)	LPA MDRplus (0-0-1)	Final DST (this study)
MYSBH_0290 (SIT19/EAI2-Manila)	MDRTB	*rpoB*_531_mut_TTG; *inhA*-prom_mut-15_T MDRTB	Not Detected	MDRTB (1-1-0) *rpoB*531_mut_TTG; *inhA*_prom_mut-15_T
MYSBH_0482 (SIT19/EAI2-Manila)	MDRTB	*rpoB*531_mut_TTG; *inhA*_prom_mut-15_T	MTB Detected (*rpoB* mutation; *inhA* gene mutation)	MDRTB (1-1-1) *rpoB*531_mut_TTG; *inhA*_prom_mut-15_T
MYSBH_0566 (SIT17/LAM2)	MDRTB	No gene mutation detected	MTB Detected (*rpoB* mutation not detected; *katG* gene mutation)	MDRTB (1-0-0.5) *katG*
MYSBH_0628 (SIT3953/EAI5)	SI-R	*inhA*-prom_mut-15_T I-R	MTB Detected (*rpoB* mutation; *inhA* gene mutation)	MDRTB (0.5-0.5-1) *rpoB*; *inhA*_prom_mut-15_T
MYSBH_0751 (SIT19/EAI2 Manila)	All-S	*rpoB*_531_mut_TTG; *inhA*-prom_mut-15_T MDRTB	Not Detected	MDRTB (1-1-0) *rpoB*531_mut_TTG; *inhA*_prom_mut-15_T
MYSBH_0862 (SIT4212/EAI3 IND)	MDRTB	*rpoB*_531_mut_TTG; *inhA*-prom_mut-15_T MDRTB	Not Detected	MDRTB (0.5-1-0) *rpoB*_531_mut_TTG; *inhA*-prom_mut-15_T

Note: SI-R, streptomycin and isoniazid resistant, R-Rifampicin. TB-RINT was able to detect mutations rendering MDRTB but not detected in ACM-DST – No further information on patient.

4.5.3 Drug-resistant MTB (DRTB) clonal complex

DRTB cases were grouped into 4 clonal complexes by the Minimum Spanning Tree (MST) derived from 24-loci MIRU-VNTR, and spoligotyping data using the *MIRU-VNTRplus* platform.

There are two MDRTB cases from the same DR-clonal complex (CC2): MYSBH_0751 was detected four months earlier than MYSBH_290. They were different only at two loci namely 424_Mtub04 and 2461_ETR-B. Both cases have the same MDRTB gene mutations (*rpoB*531_mut_TTG; *inhA*_prom_mut-15_T). However, there is no established epidemiological link to connect these two cases (Figure 4.17).

Clonal complex clustering by *MIRUVNTRplus* was able to reveal a possible link between two MDRTB cases. The MDRTB cases

MYSBH_0751 (21452437297326623342713)

MYSBH_0290 (26452437297324623342713)

are both SIT19/EAI2 Manila, both carried the same gene mutations conferring MDRTB (*rpoB*531_mut_TTG; *inhA*_prom_mut-15_T) (refer Table 4.22 and figure 4.17). According to conventional case investigation these two cases have no epidemiological link. Both cases were reported in 2016. Although the distance between the district of Kudat and Penampang is more than 180 km, but these districts are accessible by land and it is only a 3.5-hour drive away. This

is another example of how genotyping could complement conventional contact investigation. Although it must be mentioned that clonal complexing of DRTB cases is quite low.

Another interesting issue to look into pertains to isolates MYSBH_0628 (two alleles at loci 4348_MIRU39 – 2 and 1) and MYSBH_0862, EAI3-IND (2 alleles at loci 3690_Mtub39 – 4 and 2). Both were grouped as singletons in the DRTB Clonal complex, and both were officially reported as SI-R (polydrug-resistant MTB / non-MDRTB). TB-RINT detected both *rpoB* and *inhA* gene mutations (MDRTB). Line probe assay (LPA) for the detection of MDRTB were not performed for these cases.

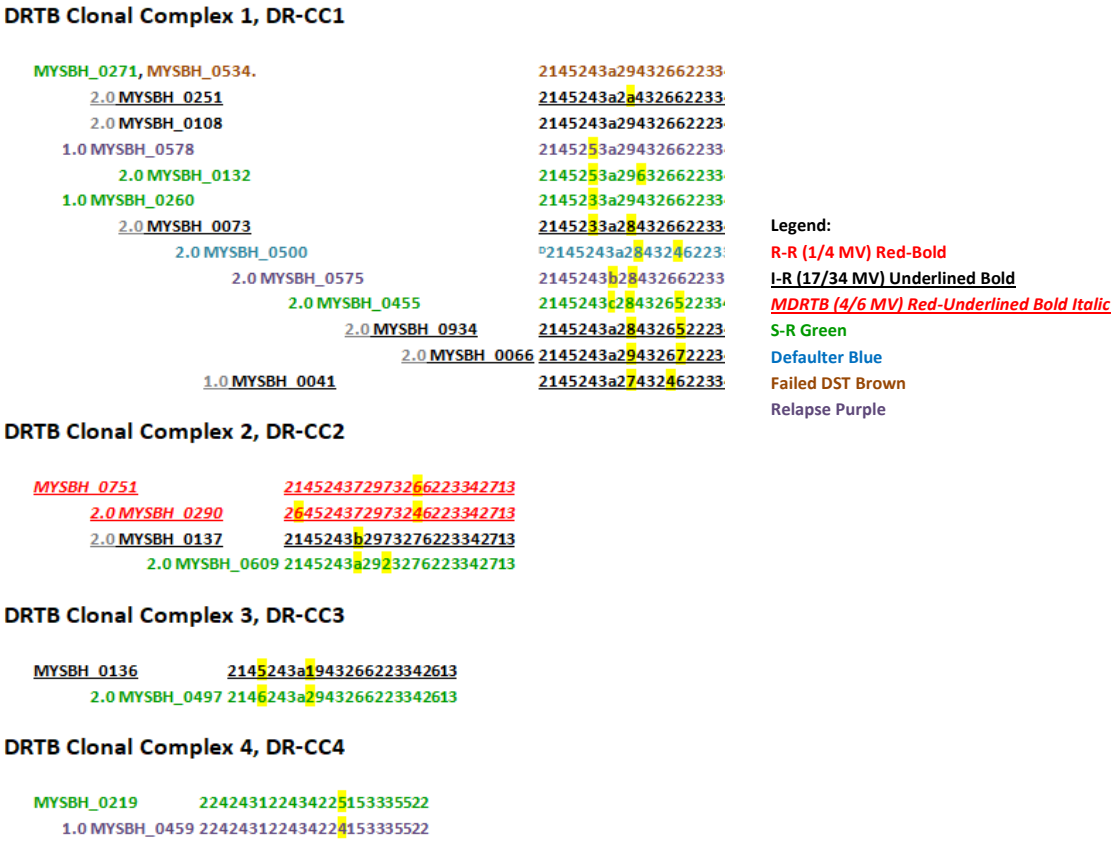


Figure 4.17: Minimum spanning tree (MST) of DRTB clonal complexes of MTB isolates in Sabah

4.6 Relationship between sub-lineages and TB relapse cases

Based on MyTB and TBIS data, TB relapse or retreatment cases in Sabah exhibited a downward trend, 3.85% in 2012 to 0.86% in 2017 (Table 4.23). In this study, there are 961 new cases (94.3%), 43 relapse cases (4.2%; 2016 – 9 cases; 2017 – 25 cases; 2018 – 9 cases) and 15 (1.5%) retreatment cases. and 15 (1.5%)**.** Of the 43 relapse cases, 83.7% are Malaysians, 11.6% Filipinos and 4.7% Indonesians. Most of the relapse cases comprised EAI2-Manila (69.8%) sub-lineages, especially SIT19/EAI2-Manila (39.5%) (Figure 4.18).

Relapse cases among Malaysians also include new EAI2-Manila SIT 4181 and 4183, new EAI5 SIT 4199 and new Unknown SIT 4204. In this study, there is an interesting case of a possible relapse whereby a patient M/22/PHP/Unemployed had 2 MTB cultures isolated 1 year apart (February 2017 vs January 2018). The patient completed treatment within the standard time of 6 months. The isolates were both SIT19/EAI2 Manila with the 24-loci MIRU-VNTR differing only at the ETR-A loci (7 and 4 respectively).

MYSBH_0163 2145243B2873266223342713

MYSBH_0575 2145243B2843266223342713

Table 4.23: Relapse / retreatment cases in Sabah, 2012-2017

Year	Relapse / Reinfection Cases	Total Number of Cases	% Relapse / Reinfection Cases
2012	169	4390	3.85
2013	131	4526	2.89
2014	114	4747	2.40
2015	122	4464	2.73
2016	98	4953	1.98
2017	44	5105	0.86

On the whole, this study was not able to detect whether any case of reinfection has taken place. This is due to the initial intention of the study to only include individual / non-duplicate subjects at any time during the sample collection. Only one 'duplicate' case, as mentioned earlier, was included by chance. The subject in question had used different identification documents during registration at the PR1, and this case is a relapse case with a difference in the MIRU-VNTR locus ETR-A. Perhaps this could be the focus of future studies.

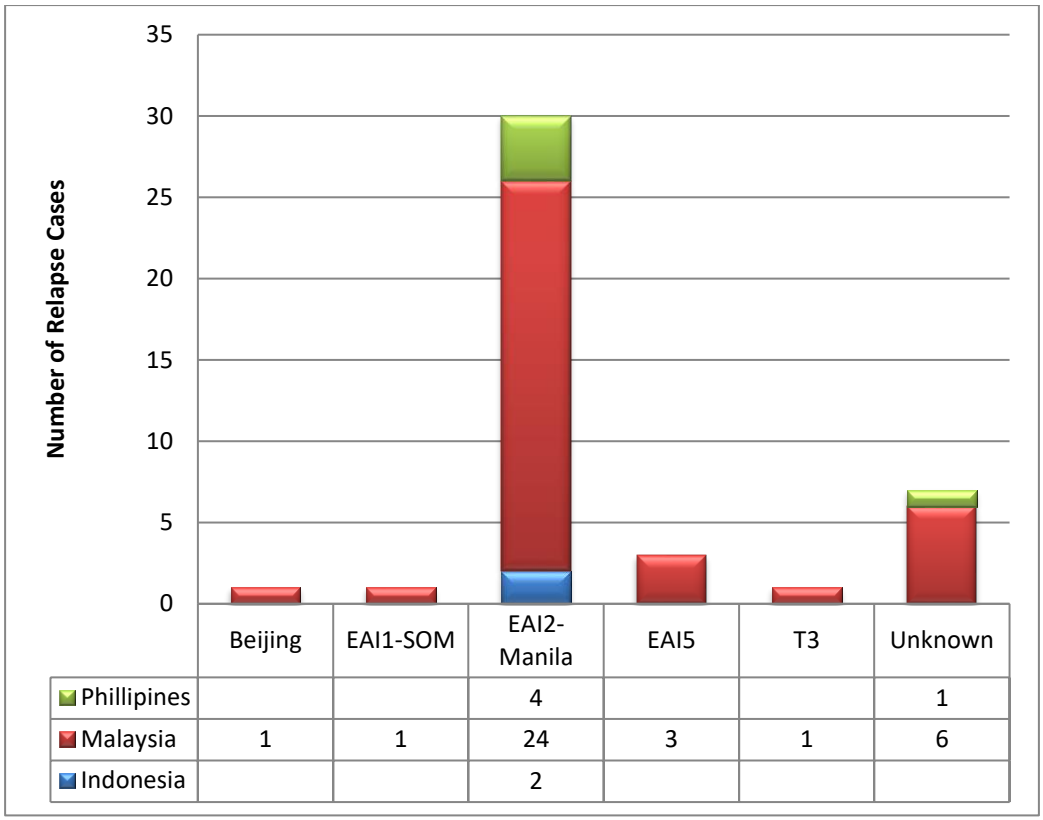


Figure 4.18: TB relapse cases by nationality and MTB sub-lineage in Sabah, East Malaysia

4.7 Relationship between sub-lineages and mixed infection of MTB

In the 24-loci MIRU-VNTR typing of 593 isolates, 36 isolates were found to have multiple bands / alleles rendering them as mixed infections. However, the presence of the multiple MIRU-VNTR bands indicating mixed infection did not affect the spoligotyping results. The spoligotyping signals still show clear, direct interpretable reading without chaotic readings. Most mixed infection MTB cases in this study come from the EAI and Unknown lineage.

Mixed infection cases were not analysed for clustering together with the other pure spoligotypes / MIRU-VNTR because of the presence of multiple alleles in the isolates. An extra screening step was required whereby the different possible spoligotype / MIRU-VNTR versions of MTB isolates were determined and matched with the clusters and MIRU-VNTR profiles from the initial 556 sample clustering analysis.

All 36 cases of mixed infection occurred in newly diagnosed TB cases. The mixed infection cases comprised Malaysians (n=21, 2.1% [overall], 3.1% [within group]), Filipinos (n=10, 1.0% [overall], 3.9% [within group]) and Indonesians (n=6, 0.6% [overall] 6.8% [within group]). There are 11 sub-lineages involving 20 SITs involved in the mixed infections, with EAI2 Manila garnering the highest number of mixed infections at 37.8%, followed by EAI5 (8, 21.6%) and Unknown (7, 18.9%). Beijing case accounts for 2.7% of the mixed infection cases. Orphan SITs and SIT19/EAI2 Manila recorded the highest numbers of mixed infections with eight (21.6%) and six (16.2%) cases

respectively. MIRU-VNTR locus Mtub21is found to have the most occurrence of multiple alleles (Table 4.24). The blue-collar workers and the unemployed also recorded higher numbers of mixed infections (Figure 4.19).

Table 4.24: MIRU-VNTR Loci with Multiple Allele

LOCI	Number
1955_ Mtub21	10
2163b_QUB11b	4
4348_MIRU39	4
2165_ETRA	3
3171_Mtub34	2
3690_Mtub39	2
424_Mtub04	1
577_ETRC	1
802_MIRU40	1
960_MIRU10	1
2461_ETRB	1
4052_QUB26	1
577_ETRC; 2531_MIRU23	1
580_MIRU04_ETRD; 802_MIRU40; 3192_MIRU31_ETRE	1
960_MIRU10; 4348_MIRU39	1
1644_ MIRU16; 2461_ETRB; 4348_MIRU39	1
1955_ Mtub21; 2163b_QUB11b	1
Grand Total	**36**

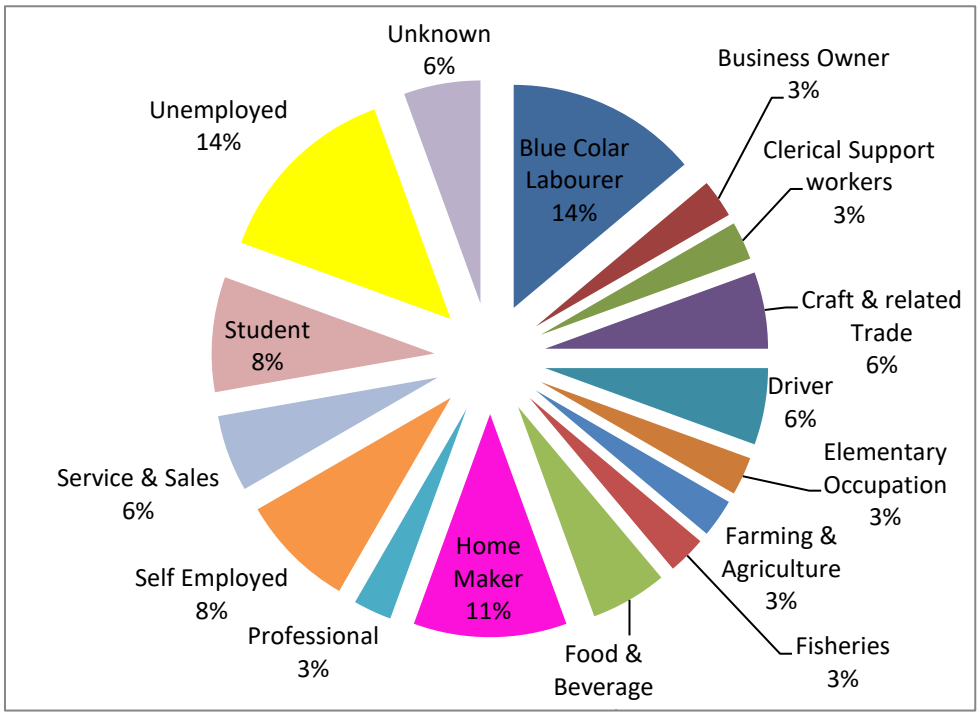


Figure 4.19: Distribution of mixed TB infection cases by occupation

Mixed infection cases were reanalysed to see if these cases might be clonally complexed or clustered with other isolates in this study. All possible versions of 24-loci MIRU-VNTR profile were listed for all 36 cases and they were

matched against all existing 24-loci MIRU-VNTR profiles in this study. Mixed infection cases possible versions that had the exact same profile or had a single locus variant (SLV1) as compared with other isolate MIT24 are presented in Table 4.25. One possible new MIT24 cluster was created, one possible new clonal complex, six possible MIT24 new cluster members, and six possible new CC1 members were discovered (Table 4.25). Three cases with six to eight possible version of 24-loci MIRU-VNTR profile MySBH0632 (INA, six version), MySBH0953 (MAS, eight version) and MySBH0309 (PHP, eight version) were observed (Table 4.26). A TB hot spot usually refers to a physical area or a sub-population with significantly higher TB cases compared to its surrounding areas or populations. It is also usually densely populated with poor living conditions (Theron et al., 2015; Robsky et al., 2020; Brooks et al., 2022).

Table 4.25: Possible transmission pattern for TB cases with mixed infections (1-14)

No	Cluster / CC	ID / Cluster	MIRU24
1	Possible New Cluster MIT24-442	MYSBH0578	2145253**A**2**9**43266223342712
		MYSBH_0592 – 2 (960_MIRU10 multiple allele)	2145253A2943266223342712
2	Possible New Clonal Complex 9	MYSBH_0889	214534392923275223242713
		MYSBH_0221 – 1, (1955_ Mtub21 multiple allele), DLV2	214534392923276223342713
3	MIT24-437, and possible new cluster member	MIT24 437	2145243A2A43266223342713
		MYSBH_0251 – 1, (4348_MIRU39 multiple allele)	2145243A2A43266223342713
4	MIT24-422, and possible new CC1 member	MIT24 422 (42)	2144243A2943266223342712
		MYSBH_0673 – 2, (1955_ Mtub21 multiple allele), SLV1 MIT24 422	2144243A2943266243342712
5	MIT24-413, and possible new cluster member	MIT24 413	2145233A2943266223342712
		MYSBH_0199 – 1 (802_MIRU40 multiple allele)	2145233A2943266223342712
6	Possible new CC1 member	MYSBH_0311	2145243428232762233**4**2713
		MYSBH_0102-2, (2163b_QUB11b multiple allele) SLV1	2145243428232762233**5**2713
7	Possible new CC1 members	MYSBH_0756, (2461_ETRB & 4052_QUB26 difference) DLV2 770 - 1	2145243429432**4**6223347**7**13
		MYSBH_0726 (4052_QUB26 difference) SLV1 770 -1	2145243429432662233478**8**13
		MYSBH_0770 – 1 (577_ETRC multiple allele)	2145243429432**6**6223347**4**13
8	Possible new CC1 member	MYSBH_0020 (3007_MIRU27_QUB5 & 3690_Mtub39 difference) DLV2 623-1	2145243819432662**2**34**2**713
		MYSBH_0562 (3690_Mtub39 difference) SLV1 623-1	2145243819432662233**2**713
		MYSBH_0623 – 1 (2165_ETR-A multiple allele) SLV1 117, DLV2 116	2145243819432662233**1**713
9	MIT24-411, and new cluster member	MIT24 411	2145243829232762233427 13
		MYSBH_0223 - 2 (1955_ Mtub21 difference, multiple allele)	2145243829232762233427 13

Table 4.25: Possible transmission pattern for TB cases with mixed infections (1-14)

No	Cluster / CC	ID / Cluster	MIRU24
10	Possible new CC1 member	MYSBH_0162	2145243A2643266222**3**3**2**712
		MYSBH_0332 - 2(3007_MIRU27_QUB5 & 3192_MIRU31_ETRE difference; 2163b_QUB11b multiple allele) DLV2 198	2145243A2643266223**3**3**4**2712
11	MIT24-433, and possible new cluster member	MIT24 433	2145243A2743266223342713
		MYSBH_0393 – 1 (4348_MIRU39 multiple allele)	2145243A2743266223342713
12	MIT24-397, and possible new cluster member	MIT24 397	2145243A2923276223342713
		MYSBH_0223 – 1 (1955_ Mtub21 multiple allele)	2145243A2923276223342713
13	MIT24 393, and possible new CC1 member	MIT24 393	2145243A2943266222342**6**12
		MYSBH_0118 - 1 (4052_QUB26 multiple allele difference)	2145243A2943266222342**4**12
14	MIT24-396, and possible new cluster members	MIT24 396	2145243A2943266223342712
		MYSBH_0332 - 1 (2163b_QUB11b difference)	2145243A2**9**43266223342712
		MYSBH_0592 - 1 (960_MIRU10 difference)	21452**4**3A2943266223342712

Note: Bold and underscored characters are the locus version that may be clustered

Table 4.26: Mixed infection MTB by nationality in Sabah, East Malaysia

Nationality (Isolates 24MV)	No. of Mixed Infection (%)	MTB Isolates	Remark
INA (41)	6 (14.6%)	MySBH0603, MySBH0623, **MySBH0632**, MySBH0673, MySBH0774, MySBH0789	**MySBH0632 -** 6 possible combinations of MTB 24-loci MIRU-VNTR profile (possible exposure to ≥2 strain of MTB)
MAS (374)	20 (5.3%)	MySBH0102, MySBH0135, MySBH0199, MySBH0221, MySBH0223, MySBH0354, MySBH0393, MySBH0592, MySBH0607, MySBH0626, MySBH0627, MySBH0630, MySBH0679, MySBH0689, MySBH0690, MySBH0717, MySBH0775, MySBH0862, **MySBH0953**, MySBH0987	**MySBH0953** – 8 possible combinations of MTB 24-loci MIRU-VNTR profile (possible exposure to ≥2 strain of MTB)
PHP (141)	10 (7.1%)	MySBH0122, MySBH0332, MySBH0251, **MySBH0309**, MySBH0488, MySBH0628, MySBH0770, MySBH0713, MySBH0118, MySBH0714	**MySBH0309** – 8 possible combinations of MTB 24-loci MIRU-VNTR profile (possible exposure to ≥2 strain of MTB)

4.8 Orphan MTB in Sabah

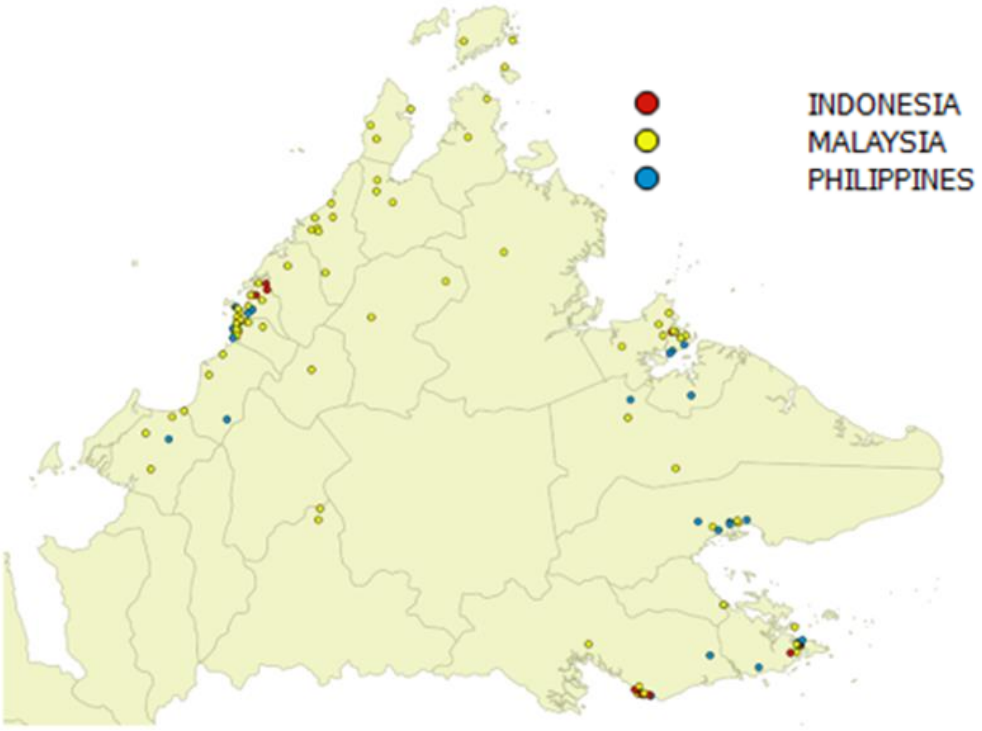


Figure 4.20: Distribution of orphan MTB Spoligotypes in Sabah, East Malaysia

The distribution of MTB orphan spoligotypes are seen to be higher in the more densely populated areas in the whole state except for the interior division. The MTB orphans (n=153) were mostly derived from the lineages: EAI (45, 29.41%) especially EAI2 Manila (34, 22.22%); Unknown (63, 41.18%), T (7, 4.58%) and MANU (6, 3.92%). Overall, there were 14 sub-lineages detected, and all 3 nationalities shared seven of the MTB sub-lineages.

EAI1-SOM, EAI2-Manila, EAI5, EAI6-BGD1, LAM9, Manu1 and T1 were the shared sub-lineages among orphan MTB of the three nationalities (Figure 4.20). One orphan MTB from the sub-lineage LAM1 was isolated from

a Filipino patient, meanwhile one other orphan MTB from sub-lineage LAM6 was isolated from an Indonesian patient. There are four H3 cases in this study, and one had an orphan spoligotype.

Of the 153 Orphans, 74 isolates were randomly selected to be genotyped using 24-loci MIRU-VNTR. This grouped the MTB orphans into 5 clonal complexes and singletons (Figure 4.21). However, further study needs to be done on their relatedness as there are no known epidemiological links based on TBIS101C database. Table 4.27 shows that Malaysian (92, 13.6%) and Filipino (36, 14.0%) strains render similar orphan percentages while Indonesian orphan MTB percentage is about double the number (25, 28.4%). Figure 4.22 shows the orphan isolates among the three nationalities proportionate to their MTB lineages.

MTB orphan strain DNA sequencing and additional sub-typing of spoligotype clusters using WGS would have provided more phylogenetic inferences. However, WGS was not able to be conducted due to degradation of the DNA extracts attributed to shortfall in sample storage as well as breakdown in cold chain during transportation from site of DNA extraction MKAKK (Sabah, East Malaysia) to UTAR Pathogen Laboratory (Selangor, Central-West Malaysia).

Table 4.27: Orphan MTB isolates by patient nationality in Sabah, East Malaysia

Country	No. of Isolates	% Isolates	No. of Orphans	% Orphans (within group)
Malaysia	674	66.1	92	13.6
Philippines	257	25.2	36	14.0
Indonesia	88	8.6	25	28.4
Total	1019	100	153	15.0

Note: Orphan isolates from unknown and known lineages without SIT

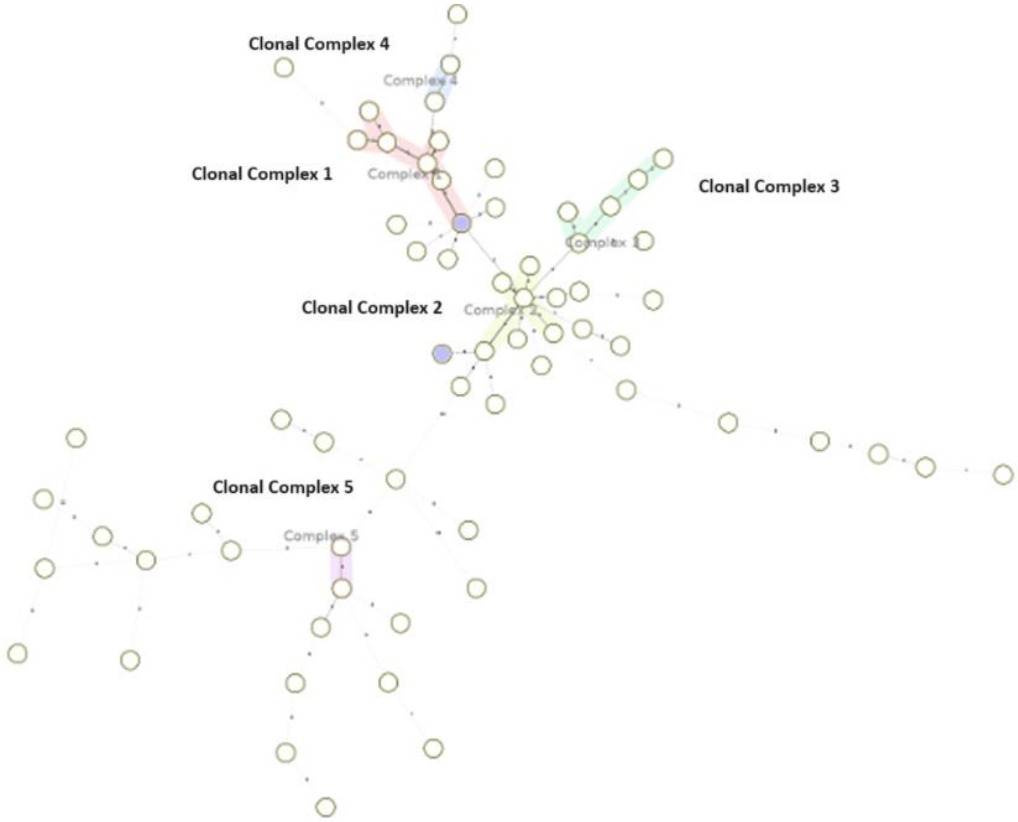


Figure 4.21: Minimum spanning tree of orphan MTB in Sabah, East Malaysia.

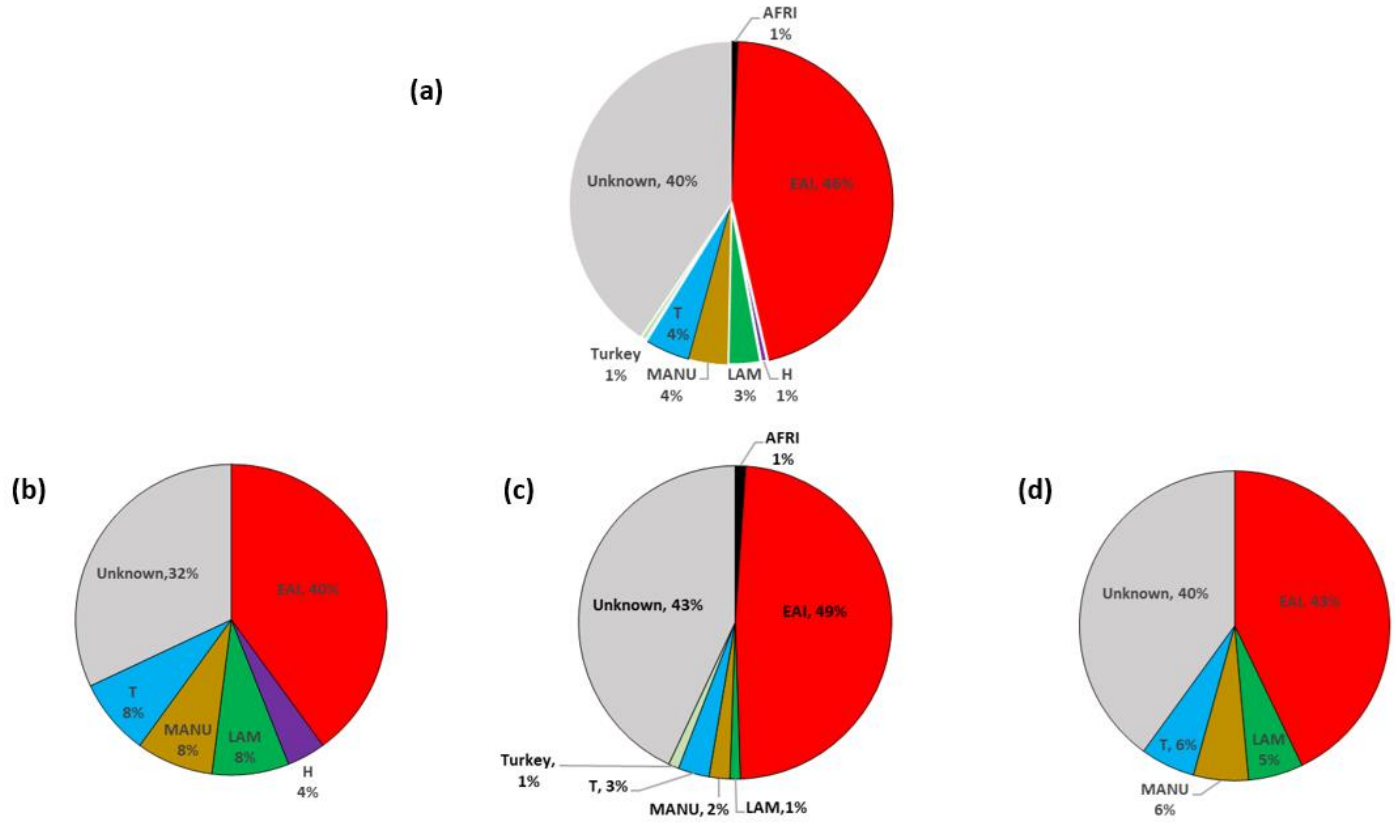


Figure 4.22: Distribution of orphan MTBC spoligotype by nationalities in Sabah, East Malaysia a) Total orphan MTB Spoligotype; b) orphan MTBC spoligotype of Indonesians; c) Orphan MTBC spoligotype of Malaysians, East Malaysia d) Orphan MTBC spoligotype of Filipinos

CHAPTER 5

DISCUSSIONS

5.1 Diversity and distribution of MTB genotypes in Malaysia

MTB strains in Malaysia have been characterized by various genotyping techniques such spoligotyping, MIRU-VNTR and WGS (Ngeow et al., 2006; Ismail et al., 2014; Fakhruzzaman et al., 2019; Jani et al., 2020; Bainomugisa et al., 2021). These studies have revealed the presence of multiple MTB lineages and genotypes such as Beijing, EAI, CAS, H, LAM and T which are circulating in the country. The distribution of MTB genotypes in Malaysia varies regionally and temporally (Ngeow et al., 2006; Ismail et al., 2014). Population demographics, migration patterns, socioeconomic status, and healthcare infrastructure are factors that can influence the prevalence and spread of the different MTB strains. A deep understanding of the genetic diversity and transmission dynamics of MTB strains is therefore crucial for the development of effective TB prevention and management strategies.

5.1.1 Thirty-four newly created MTB SITs

The 34 newly created SITs discovered in this study comprised the lineages as follows: AFRI (1 SIT); EAI2 (2 SITs); EAI2 Manila (16 SITs);

EAI3-IND (one SIT); EAI5 (five SITs); T-H37Rv (one SIT); Unknown (eight SITs). There are 27 clustered SITs, and seven unique SITs of which 3/7 were found in the US, most probably associated with Filipinos who may have relatives in the US (Table 4.6). The new SITs are grouped into five clonal complexes with some singletons. Most of the newly created SITs are from the sub-lineage EAI2-Manila, which is the predominant MTB family in Sabah. Based on SpolSimilarity Search and SITVIT2 database, it is possible that Filipino contacts or domestic workers-relatives in Saudi Arabia may have brought one new spoligotype (SIT 4186) to Sabah, or the other way round. It is also interesting to see other 'lesser known' EAI2-Manila (new SITs) are also found in US.

The identification of newly created spoligotypes: SIT 4182, 4183 (clustered in this study); 4189, 4194 and 4214 (singleton spoligotypes) derived from strain EAI2 and EAI2-Manila which have only been reported as orphans in the US as isolated cases, highly suggests that these strains could have been brought in from the Philippines, or perhaps eventually evolved locally in Sabah especially among the Filipino communities comprising talented artisans, workers in the tourism and entertainment sector as well as domestic workers who may have relatives or contacts in the US. Based on these observations, it could also be suggested that the sub-lineage profile could be used to a certain extent, to determine the degree of close interaction or assimilation in the population when there is no distinct association between the sub-lineage and specific host population group (Reed et al., 2009; Brites and Gagneux, 2017).

5.1.2 Newly created SITs from unknown lineage

Clustering of eight new SITs from the unknown lineage by spoligotyping suggests that they could have arisen from different origins (Figure 4.3). The similarities in the spoligotyping pattern of the new SITs in this study from the Unknown Lineage of MTB, point to the possibility that these SITs may be a part of an emerging lineage possibly unique to this region (Table 4.8). SpolLineages Tool was unable to resolve the queries for these SITs. WGS analysis would need to be conducted to confirm this claim for new lineage.

5.1.3 Differences in lineage profiles in peninsular Malaysia and Sabah

Dale et al. (1999) and Ngeow et al. (2006) mentioned that the lineage profiles in peninsular Malaysia and Borneo are markedly different in terms of the MTB genotype prevalence. There are more EAIs in Sabah, and the EAI strains are quite different from the ones in peninsular Malaysia especially EAI2 Manila SITs, including a few other strains not previously reported (Ngeow et al., 2006, Ismail et al., 2014; Nur et al., 2019; Zamri et al., 2022) (Figure 4.5 and Figure 4.6).

A genotyping study by Ngeow et al. (2006) reported that the primary MTB lineage in Malaysia is the Beijing family, and ST1/Beijing is the predominant strain (32.95%) identified in all geographic areas in peninsular Malaysia, followed by ST19/EAI2-Manila (6.24%) which is mostly found in

natives in East Malaysia. Ismail et al. (2014) study showed that the Beijing lineage is second in dominance to the ancestral EAI strain. However, Ismail et al. (2014) study is limited to Kelantan and Kuala Lumpur, and does not reflect the actual distribution and dispersal of MTB lineage in other parts of Malaysia. The Beijing sub-lineage only makes up 2.36% of the strains in this study (Figure 4.6). Parwati et al. (2008) reported that in Indonesia, the Beijing genotype family was the most common lineage (31.2%). However, when geographical distribution with significant geographic variation was considered, they found that the numbers are vastly different: 33.0% Beijing genotype in Java compared to 14.3% in Timor. This is consistent with the observation in Ho Chi Minh City, Vietnam that had a 50% frequency of Beijing genotype strains, but the Mekong Delta only had a 30% frequency (Parwati et al., 2008). Extremely crowded populations perhaps may favour the propagation of strains with the Beijing genotype. Hence the reason why there are more DRTB cases among Beijing genotypes in Central Malaysia as compared to Sabah, East Malaysia (Ngeow et al., 2006; Fakhruzzaman et al., 2019; Zamri et al., 2022) (Table 4.20, Table 4.21 and Table 4.22). This geographical separation and difference in population density may also be some of the reasons to explain the variation in MTB Beijing genotype spread in Sabah, in addition to the controlled entry of migrant workers in Sabah. The physical remoteness in Manila, Philippines may be a favourable environment for the clonal expansion of MTB where the majority of MTB in Manila belonged to the lineage 1 Manila clade (EAI2-Manila), while the minority belonged to lineages 4 (European-American) and 2 (East Asian) (Phelan et al., 2019).

In this study, EAI especially EAI2-Manila is overly represented. This is perhaps an indication of their success or fitness in the study environment. Thus, these strains may have a selective advantage over other strains in Sabah. It is possible that the geographically confined local adaption of the EAI2-Manila strains to Sabah population is indicative of its ecological specialisation, which is the ability to evolve and adapt to the human host genetics in their endemic regions (Stucki et al., 2016; Borrell and Trauner, 2017; Gagneux, 2018).

In their four-year study on 208 MTB isolates from Kota Kinabalu district, the capital of Sabah, Bainomugisa et al. (2021) found that the MTB strains were diverse, with several lineages 1 strains dominating the landscape. WGS analysis showed that 93.8% of the strains were L1, with the rest from L2 (3.8%) or L4 (2.4%). The strains of lineage 1 were varied, with sub-lineage 1.2.1 dominating (192, 98 %). Lineage 1.2.1.3 isolates were geographically most widely distributed. Some linked TB cases spanned the whole four-year study duration. This corroborates the finding in this study whereby CC1, the biggest clonal complex includes linked cases from 2015 to 2018. Bainomugisa et al. (2021) also found that transmission clusters were not commonly identified, and this is likely attributable to incomplete sampling. Another reason why universal genotyping should be implemented. Similar to this study's findings, Bainomugisa et al. (2021) showed that clustering occurred throughout the community, not just confined to households or within the districts. MTB lineages or sub-lineages were not associated with patient ethnicity.

5.1.4 There is no significant association between MTB sub-lineages and ethnicity in Sabah

To understand MTB migration and evolution, there is a need to understand the people that make up the community, population, state and country. Gagneux et al. (2006) mentioned that there's a significant MTB lineage-people affinity thus specific lineage-people infection clustering. Reed et al. (2009) tried to counter this statement by saying that perhaps this is due to the non-assimilation of the population groups (immigrants and locals in the Canadian / European setting at the time). Sabah, Poni, or British North Borneo as it is used to be known has had a fascinating history when it comes to geopolitics and nation building. The on again and off again claim of the heirs of Sulu sultanate over Sabah as being the rightful owners of the state first filed in 1961 revealed that there has been a strong long-standing tie between the Philippines and Sabah.

The possibility of assimilation and intertwining of the Filipino and local Sabahan natives is undeniable as evidenced in the peculiarities of some aspects of the local culture and language. The Indonesian communities in Sabah mostly comprised the Bugis and Toraja communities from Sulawesi. Most work as traders (especially in markets), farmers, fishermen, estate workers etc. Other Indonesian community with high population numbers in Sabah are the Timorese from Nusa Tenggara Timur who mostly work in the agriculture sector (Kurus, 1998; Hassan, 2018; Peters, 2022).

T1 is seen at higher frequencies among Indonesian especially from Makassar, Sulawesi. In Timor, the EAI and LAM were the predominant genotype families (Parwati et al., 2008). It is interesting to note that in earlier works by Ngeow et al. (2006) all ST51/T1 strains in the study hailed from Johor and Melaka states which are known to have among the highest Bugis populations in Malaysia, and possibly still having close ties with relatives in Makassar, Sulawesi. Parwati et al. (2008) and Sasmono et al. (2012) reported considerable number of isolates from T sub-lineage in Makassar, Sulawesi. Most T lineage cases in Sabah are also identified among the Bugis and the Bugis-adjacent populations.

The Chinese population make up approximately 9% of the total population in Sabah (DOSM, 2020). In the early 1900s the British brought in Chinese traders mostly Hakka, and some labourers of Hokkien and Teochew descent (brought in via Labuan) in order to drive the economy in Sabah (Wong, 2005). Until now the Chinese population are much involved in trade and business in Sabah. However, in Tawau especially, it is the Bugis traders who compete head-to-head with the Chinese in terms of business and trading. Many inter-marriages between Bugis and Chinese are seen as well especially in the east coast of Sabah. Ngeow et al. (2006) reported that the Beijing genotypes have been found mostly among the Chinese population in peninsular Malaysia (Ngeow et al., 2006). In this study of the 28 subjects from the Chinese population in Sabah, only 2 (7.14%) have the Beijing genotype compared with the 15 (53.57%) other samples with the predominant EAI2-Manila genotype.

The population mix in the West Coast of Sabah is higher due to greater work and economic opportunities. In Sabah, Indonesians (16.1%) and Filipinos (12.6%) outnumber other migrant populations (Hassan, 2018). Considering that there are more Indonesians as compared to Filipinos in Sabah, it is interesting to note that the TB case numbers among Indonesians in this study is only 1/3 of the numbers made by the Filipinos (INA:PHP: MAS - 88:257:674). High rate of inter marriages between Sabahan-Filipino especially in Sandakan; Sabahan-Indonesian especially in Tawau indicate that the populations studied are highly assimilated. There are many naturalized Indonesian and Filipino in Sabah (especially 2nd and 3rd generation of Indonesian and Filipino migrants) with some having reached prominence in society, especially in the arena of Education, Business, Politics and Government. The spread of MTB sub-lineages is quite similar among the three nationalities in this study population as there are some shared sub-lineages.

However, upon closer inspection, it is observed that clonal complex clustering among Malaysian and Filipino are higher compared to Indonesians. There is a higher reactivation rate among Indonesians, and there are more exclusive SITs among Indonesians not shared by Malaysian and Filipinos which might be recently imported genotypes brought into Sabah from Indonesia. Shared sub-lineages between Filipino-Malaysian is greater compared to Indonesian-Malaysian. Sabahan-Filipino relationship has existed way back since the Sulu sultanate rule in the 1800s, and the political migration commencing in the 1970s. It is logical to assume that there would be sufficient time for at least limited genetic assimilation to take place. The close genetic-

relatedness between local Sabahans and Filipinos could also be the reason why there is similarity of isolate sub-lineage infection with 18 shared SITs, and only 6 unique SITs among Filipinos (Figure 4.4).

At the outset, the observation is that MTB of different lineages affect different populations, but as subtly nuanced by Reed et al. (2009), the MTB lineage's affinity towards specific populations could possibly dissipate as the different peoples become more assimilated. In this study, there seem to be no significant association between the sub-lineages and ethnicity. However, this could be due to high assimilation of the population of different ethnicities at varying degrees. A deeper look into the differences related to the sub-cultural, economic, education, health factors of the Indonesian, Filipino and Malaysian populations may well prove to be valuable in relation to formulating effective TB control strategies. This actually does imply possible different behavioural factors.

One interesting observation to note is that the MANU, an ancient sub-lineage (with some predominance in Indian subcontinent) is also seen in this study (including orphan MANU spoligotypes). The biggest MTB sub-lineage in Sabah, the EAI2 Manila is well spread throughout all the age groups. The Unknown lineage however is seen especially among ages 15-24, 25-34 and 35-44. These age groups are usually associated with high mobility and work-productive. Philippines and Indonesia are included in the top 30 TB high burden countries. With the high flow of migration in and out of Sabah from these countries higher TB transmission among different populations is most likely to

take place. It is interesting to note that observations in metropolitan setting where diverse human and pathogen genotypes mix showed that the sympatric host-pathogen associations remained stable. Brites and Gagneux (2017) mentioned that sympatry patterns in MTB transmission could be driven by the preferential social mixing among people of same ethnicity. But with assimilations and closer inter-population interactions between migrants and citizens such as that seen in intermarriages, investigation of MTB evolutionary development from this angle could prove to be worthwhile.

5.2 There is no correlation between MTB sub-lineages and DRTB in Sabah

This study is one of the first large scale studies at the state level to look into the actual cases of drug-resistant TB (DRTB) focusing on INH and RIF, and to describe the prevalence of DRTB in terms of associated mutations. DST information and accurate information on the types of mutations present are important to initiate proper TB treatment (MOH, 2012; MOH, 2016). The routine testing for MTB drug susceptibility in Sabah is done using the ACM (Stinson et al., 2014). In general, the prevalence of drug resistance in Sabah is still relatively low (Table 4.20, Table 4.21 and Table 4.22). Logically, the more prevalent lineage would have higher numbers of DRTB cases due to higher likelihood of DR acquisition through transmission (primary resistance) or selection pressure (acquired resistance), and this is the observation in this study. There seem to be no clear relationship between DRTB and sub-lineage. However, DRTB, failed DST and relapse cases need to be closely monitored for

possible links as they have similar underlying implications and potential impact. It is also important to improve the training of laboratory scientists and technologists in terms of possible DRTB by identifying early signs of DRTB to ensure sufficient and early detection of cases of this nature.

The EAI2-Manila isolates which makes up the majority of the isolates in this study did not show significant association with patient's age, sex, TB severity, bacillary load and phenotypic drug resistance profile. This is corroborated by earlier studies on EAI2-Manila genotypes (Phelan et al., 2019; Montoya et al., 2022). The mutations found in the drug resistance genes in this study resembled several of the mutations described in earlier studies that sequenced Philippine MTB strains (Phelan et al., 2019; Montoya et al., 2022).

Although the number of Beijing isolates is small in this study, n=25; it is interesting to note that isolates with any resistance accounted for 16% of the total Beijing strains (DRTB proportion to Beijing) in this study, while EAI2-Manila-the most numerous strains had a 4.7% proportion of isolates with any drug resistance. In this context, perhaps Beijing genotypes do appear to have higher occurrence of resistance compared to EAI2-Manila (n=682). Central Malaysia show that L2 seem to be more predominant via DRTB study (Zamri et al., 2022). Some argued that it could be due to geographical separation, urban setting, and higher population density (Mokrousov et al., 2004; Parwati et al., 2008; Lisdawati et al., 2015; Yin et al., 2016).

5.2.1 MTB isoniazid (INH) resistance in Sabah

Given that not all resistance loci are currently targeted by the TB-SPRINT technique for INH, some resistant isolates were undetected (Gomgnimbou et al., 2012; Gomgnimbou et al., 2013) (Table 4.20). This shows that there are possibly other gene mutation that confers resistance in regions other than the ones that have been studied. Prevalence of high level and low-level resistance in INH is different from one setting to another. It is highly dependent on the MTB transmission. If inter-border travel takes place at a high rate, it can be expected that the MTB genotype and resistotype in a state or country would reflect that of its surrounding neighbours (Borrell and Trauner, 2017; Cirillo et al., 2017). A study by Narmandakh et al. (2020) in Mongolia found that in a sample of n=409 INH-R, *inhA* mutation (-C15T mutation) is more prevalent than *katG* mutation (S315T) in isolates identified by the GenoType MTBDRplus assay. Gagneux (2009) reported that the *inhA* promoter mutation, are observed more significantly among the ancestral lineage, meanwhile the genome-based Euro-American lineage is associated with the *katG* S315T mutation, maintaining *katG* activity. This shows that the differences in the genetics between the MTB lineages could influence their propensity to have different isoniazid resistance conferring mutations (Gagneux, 2009). In this study, it is found that the ratio between *katG* gene mutation (high level INH resistance) and *inhA*-prom gene mutation (low-level INH resistance) is about 1:1.

5.2.2 MTB rifampicin (RIF) resistance prevalence is low in Sabah

The *rpoB* gene mutations found in Sabah are *rpoB* 516 GTC, *rpoB* 531 TTG, *rpoB* 526 TAC as detected using TB-RINT. All four mono-resistant RRTB cases in this study were found in the west coast of Sabah (Table 4.21). As in the cases of INH mono-resistance, the phenotypic and genotypic methods also showed some discrepancies in the RRTB results. In Phelan et al. (2019) WGS study on 178 isolates in Manila Philippines, it is found that 80.3% of the MTB strains are EAI2-Manila with L4 Euro-American n=33 and East Asian (Beijing) n=2, it is natural to assume that the types of gene mutations conferring resistance would be the same as well. However, the *rpoB* mutation conferring rifampicin resistance they listed were: S450L, H445Y, 1297_1298insTTC, 1295_1303del, D435F, S441L. According to Cirillo et al. (2017) D516Y is one of the 'disputed' mutations alongside L511P, L533P and H526N. These mutations may test differently on solid and liquid DST media, and are usually connected with poor clinical results and need to be seen as conferring resistance. The most prevalent mutation seen in MDRTB, which is linked to resistance to all rifamycins, is *rpoB* S531L. This is exhibited well in this study as four of six MDRTB cases (3 Malaysians and 1 Filipino) reported *rpoB*_531_mut_TTG; *inhA*_prom_mut-15_T gene mutation (Table 4.22). A number of isolates detected as carrying resistance-causing mutations but susceptible isolates possibly attributed to low-level resistance; and no detected mutations but resistant isolates due to possible mutations not previously reported were observed for both isoniazid and rifampicin resistance associated gene mutation. These observations were also corroborated by Zamri et al. (2022).

5.2.3 Multi drug-resistant TB (MDRTB) prevalence is low in Sabah

Goroh et al. (2020a) reported that Sabah has a low prevalence of MDRTB (0.3% of TB patients). Drug-sensitive TB treatment success was 83% (interquartile range: 81-85%), while that for MDRTB was 36% (interquartile range: 25-45%). In this study, five of the six MDRTB (83%) cases in this study carried the same gene mutation profile: *rpoB* 531_mut_TTG; and *inhA*_prom_mut-15_T, suggesting primary MDRTB (Table 4.22). This raises a concern about the ongoing transmission of MDRTB. Although the genotypes do not seem to be related, or at least distantly related, the question to ask perhaps is could it be because the genotyping information of the links are missing? Thus, making a hard case for the implementation of universal genotyping.

Previous studies by Borrell and Gagneux (2011) and Cirillo et al. (2017) also reported that *katG* gene mutation is associated with MDRTB. INH resistance in the Beijing sub-lineage is usually associated with *katG* gene mutation (Gagneux, 2009; Skiba et al., 2015; Zamri et al., 2022; Mokrousov et al., 2023). Of the six MDRTB cases detected in this study, only one MDRTB case carried a *katG* gene mutation, and this MDRTB case is the only one involving a LAM sub-lineage (SIT17/LAM2). The other five MDRTB cases are of EAI lineage (3 SIT19/EAI2 Manila, 1 SIT3953/EAI5 and 1 SIT4212/EAI3 IND) and all carried the *inhA* gene mutation (*inhA*-prom_mut-15_T). It is interesting to note that the only other isolate exhibiting isoniazid resistance from the LAM sub-lineage also carried a *katG* gene mutation (SIT891/LAM9) (Table 4.20).

In the recent WHO survey in 2015, 4.6% of 10.4 million TB cases are MDRTB. The overall rate seems to be low but there are significant variations between and within countries. MDRTB and hotspots are defined as regions with more than 5% prevalence of MDRTB. Localized high incidence rates have been recorded, primarily in China, India, and the Russian Federation (36% of global MDRTB). Varying prevalence of MDRTB across the globe is due to factors such as: variably effective control programs, environmental factors. presence of co-morbidities and patient related (Cirillo et al., 2017; Walker et al., 2022).

When it comes to MDRTB, the impact of the implementation of GeneXpert® MTB/RIF in Sabah must also be highlighted. It is truly a game changer for TB patient management in terms of the speed with which MTB and RRTB (proxy indicator for MDRTB) cases are detected. The wait for DST result has been reduced from several months to a matter of a couple of days if not hours. At present, its implementation is still relatively costly. GeneXpert® MTB/RIF can only be utilised for cases that meet strict criteria e.g. PLHIV, relapse case, MDRTB contact and new DRTB on anti-TB treatment (Rosli, 2022). Indeed, ensuring universal health coverage and expansion of GeneXpert® coverage is recommended to reduce barriers to care and early diagnosis and treatment for TB (Goroh et al., 2020a).

5.2.4 Universal genotyping of all patient isolate is important to systematically gauge, and to establish the initial level of MTB drug resistance in a population

An understanding of MTB drug resistance is important for the clinical team to give proper management and for the public health team to control transmission. A baseline DST and genotyping for all newly diagnosed TB cases are important information for patient management and TB prevention activities. Since treatment of TB takes at least six months, it is important to know whether or not the MTB isolate from patients carries any drug resistance. Although Sabah currently has low prevalence of DRTB (Rashidi Ali et al., 2015; Goroh et al., 2020a; Bainomugisa et al., 2021), it is still important to monitor the trends to ensure optimal TB prevention and management.

Consequent MTB isolate post-baseline indicates new episode of infection which could be relapse / reactivation or reinfection. In some cases, patients may have had an initial MTB infection that isolated drug-susceptible MTB but are detected as DRTB in the 'relapse' episode (García De Viedma and Pérez-Lago, 2018; Leavitt et al., 2022). A standard surveillance for DRTB needs to be conducted to learn more about the prevalence and geographic distribution of anti-TB drug resistance (Guerrin-Tran et al., 2006; Shao et al., 2011). Data comparison with other available MTB database, and identification of potential risk factors for MDRTB need to be conducted to provide policy makers with recommendations for better organized TB control programs (Guerrin-Tran et al., 2006). In Malaysia, we have SIMKA (Public Health Laboratory Information

System) and MyTB database that stores valuable information on TB. We need to analyse these available data to enable the formulation of actionable strategies in the context of TB prevention and management.

5.2.5 Phenotypic (ACM) detected DRTB need to be counter-tested using molecular DST, and vice versa

Upon close scrutiny, detection of DRTB showed some differences when phenotypic and genotypic test modalities were used (Table 4.20, Table 4.21 and Table 4.22). It is also possible however that the INH-R conferring mutations were not expressed phenotypically in the supposedly 'missed' INH-R cases. Percentage of agreement between TB-RINT and ACM is 35.29%. There is a possibility that some gene mutations will not be detected by existing commercial kits as most were developed for MTB isolates archived by western or European laboratories. To date, China, Taiwan and South Korea are spearheading research in this region. There is also a need for regionally developed DST for ASEAN and other shared regions with similar diversity of lineages and MTB strains.

There were 18 missed cases of DRTB in this study: 16 I-R (of which six are failed DST: four katG gene mutation and two *inhA* gene mutation), one MDRTB (*rpoB*531_mut_TTG; *inhA*_prom_mut-15_T) and one R-R (*rpoB*_516_mut_GTC). One of the important tell-tale signs of possible DRTB, especially *rpoB* and *inhA* / *katG* gene mutation related resistance (backbone of

anti-TB treatment), often time is the poor growth of isolate. Multiple samples may be sent but repeated cultures show poor or no growth (Gagneux et al., 2006; Borrell and Gagneux, 2011; Paul, 2019). There is conflicting, yet complementing DST results when phenotypic and genotypic DST methods were compared, reiterating the fact that no single test is perfect.

Thus, where possible and applicable, especially in cases involving DRTB isolates, it is recommended to perform both ACM and Molecular DST tests for a more conclusive result. Interpretation correlated with clinical presentation is of utmost importance. Silent mutations, or rather compensating mutations could also be an interesting angle to study. It is therefore important to conduct studies on local MTB isolates to elucidate the actual type or perhaps unique gene-mutation conferring resistance in the region to aid TB prevention and management measures.

Cases with failed DST should be tested using molecular method to rule out possibility of DRTB. Phenotypic DST is important in determining the active infection at the moment, guiding basic clinical management of the patient. Information on the type of isoniazid resistance (high level resistance-*katG* gene mutation, or low-level gene mutation-*inhA* gene mutation) would also help determine the correct dosing of Isoniazid for the TB patient involved. More insight could be unveiled with the use of WGS and NGS especially pertaining to the mutations related to MTB drug resistance (Walker et al. 2015; Zignol et al., 2018).

More DRTB cases were found in districts with big cities and higher densities, bigger economic opportunities such as Kota Kinabalu, Sandakan and Semporna (high tourism activities). In addition to being crucial for early treatment and the related prevention of TB transmission, quick identification of TB disease and assessment of DRTB profile is also very important for the management of settings with limited financial resources. Both genotypic and phenotypic DST are needed for better treatment and management in the long run. Genotyping would add value in determining the general trend or potential for resistance which is vital for long term management.

5.2.6 Challenges in Culture-based MTB DST

On average, conventional DST usually takes about 35-40 days to be completed (after culture and identification tests are done). Molecular DST could be done within one to two working days. Even if tests are done in batches for cost effectiveness reasons, results could be obtained much faster as compared to conventional DST. Difficult cases which account for just fewer than 10% of the cases would take up to five months and sometimes even more than a year for substantial test results to be released (Paul, 2019). Thus, the utilisation of molecular assays to reduce missed DRTB cases as well as improvement in the ability of laboratory scientists to recognize possible DRTB cases must go hand in hand.

Another challenge faced by DST laboratories is the seemingly 'changing' results for MTB DST with every test performed, be it the same sample or different samples sent. This is one of the examples how genotyping could actually be used to guide decision for TB patient management which entails the complexities of changing treatment regime, cost implication and the necessity of close monitoring. Multiple samples are usually sent for 'Presumptive DRTB' and 'Confirmed MDRTB' (monthly samples) patients in order to monitor patient progress. These samples are good for longitudinal studies of MDRTB/XDRTB in future. Genotyping and WGS are indeed the way forward. This warrants the allocation of funds for the universal genotyping of all the MTB isolates in Sabah.

Despite the fact that there are still many problems, a paradigm-shift in TB prevention and management from a culture-based (culturing of samples with long laboratory turn-around time) to a specimen-based (direct testing from specimen e.g. molecular testing with shorter laboratory turn-around time) strategy appears to be imminent. For the time being, genotypic testing cannot be fully implemented to replace phenotypic DST due to cost constraints, but it can be an additional test (Gomgnimbou et al., 2012; Gomgnimbou et al., 2013). The use of both phenotypic and genotypic DST would give more reliable results for the treatment and management of MTB patients, especially MDRTB and XDRTB cases. Indeed, both molecular DST and universal genotyping should be the way forward.

5.3 Relapse / reinfection cases of MTB in Sabah

Although most studies reported Beijing lineages are associated with higher relapse cases and longer clearance time (Parwati et al., 2010; Huyen et al., 2013), this study does not seem to corroborate those findings as there is only one Beijing relapse case (2.3%) compared to the 79% relapse cases involving EAI. Perhaps this is due to the relatively small number of Beijing genotype in this study (n=25, 2.5%). Based on MyTB and TBIS data (2012-2018), TB relapse or retreatment cases in Sabah is seen to be on a decline. This perhaps reflects the improvement in TB case management in the state in terms of monitoring the TB cases, especially the non-compliant cases. The strategy employed was to quickly re-treat the patients, and implement good follow up with HCW appointed DOTS supervisor. Family DOTS supervisors were appointed for selected cases, needing to be vetted by the Family Medicine Specialist (FMS), District TB Program Manager (DTPM) or Respiratory Specialist / Consultant.

At present, it is difficult to determine if a relapse case is indeed a true relapse case as it could also be a re-infection with a different strain of MTB. Thus, the need for a baseline genotype database (spoligotype and MIRU-VNTR) for all TB patients is imperative to determine if there are any changes in the MTB strain should there be new infections in the future. This information will be able to guide physicians in terms of management and making clinical decisions. This information would also give epidemiologists a sense of direction in terms of effective and efficient surveillance and contact investigation.

5.4 Mixed infections of MTB in Sabah

Mixed Infection of MTB is defined as presence of ≥2 alleles in the MIRU-VNTR loci (de Vries et al., 2009). In the 24-loci MIRU-VNTR typing about 6% of the isolates were found to have multiple bands / alleles rendering them as mixed infections. However, the presence of the multiple MIRU-VNTR bands indicating mixed infection did not affect the spoligotyping results. It is possible that the mixed infection was caused by multiple strains of the same genotype; or different strains of MTB with the dominant spoligotypes masking the overall result by exhibiting positive signals for spacers which are supposedly negative for other strains. e.g. in the case of a mixed growth of Beijing genotype SIT1 and EAI5: Beijing genotype shows an absence of signal for the first 34 spacers and positive signal for the last 9 spacers, while SIT1365/EAI5 shows positive signals for spacers 25-28, 33 and the last 9 spacers. In this case, SIT1365/EAI5 is the more 'dominant' genotype by having more positive signals.

Most mixed infection MTB cases in this study come from the EAI and Unknown lineage. 'Mixed infections' in new TB cases suggest infection with more than one MTB strain acquired in a TB Hot-Spot The mixed infection isolates showed greater proportion of sub-lineage diversity as compared to overall MTB isolates in this study. All mixed infection cases in this study occurred in newly diagnosed TB cases with EAI2 Manila sub-lineage recording most of the mixed infections at 37.8%, followed by EAI5 (21.6%) and Unknown (7, 18.9%). Beijing case accounts for 2.7% of the mixed infection cases. The

reason why mixed infections only happen in newly diagnosed TB cases is not understood as yet and is subject to further research. In terms of SITs, Orphans and SIT19/EAI2 Manila recorded the highest numbers of mixed infections with 8 (21.6%) and 6 (16.2%) cases respectively. It could not be ascertained however if the cases are truly mixed infections due to laboratory cross contamination, endogenous or exogenous infection. A mixed infection may also indicate that a patient is exposed to high 'concentration' of TB cases, or that a patient is living in a TB Hot-spot (Navarro et al., 2011; García De Viedma and Pérez-Lago, 2018).

Mixed infections usually see a delay in the result release due to difficulties in determining the accurate drug susceptibility results. Apparently the 'mixed infection' gave different results with different test batches, and types of tests performed. Much could be studied from this angle of TB mixed infections. Findings of this study pertaining to mixed infection are corroborated by studies previously conducted (Huyen et al., 2012; Muwonge et al., 2013; Shin et al., 2018), whereby mixed infections were mostly found in newly diagnosed cases. There was also no association between mixed infection and the sex and age of patients. In the reanalysis of the mixed infection cases, three cases with 6 to 8 possible version of 24-loci MIRU-VNTR profile have been observed (Table 4.25). Similar cases like these three cases may perhaps be used as a proxy for mixed infection in the population. The greater the mix (more multiple alleles at multiple loci), it could be assumed that the case could have been exposed to many TB cases rendering it a possible super-index TB case. It

could also be assumed that these three cases may have come from hot-spot areas (Table 4.25.

In the case of mixed infections, especially in DRTB cases, without genotyping-resistotyping we wouldn't be able to confirm the presence of a mixed infection. Treatment is usually given based on DST results. Sometimes it can be difficult to decide on proper management of TB patients due to seemingly changing DST results most probably caused by mixed infection. DRTB diagnosis delays may be made worse by the failure to effectively identify resistance patterns in mixed infection cases. This could have an impact on the specific patient as well as the spread of drug-resistant strains. Van Rie et al. (2005) suggested three mechanisms by which during treatment, mixed infections may result in modifications to drug-susceptibility patterns:

1. The first-line antibiotics lowered the drug-susceptible strain population during the initial treatment period while allowing the drug-resistant strain population to increase. This is the mechanism of selection through antibiotic pressure. The patient is then changed from a drug-susceptible TB case to a DRTB case as a result.

2. Mechanism of selection when antibiotic pressure is not present: The underlying drug-susceptible strain population reemerged as the dominant population when the antibiotic pressure was relieved due to poor adherence or default. Unknown and perhaps deceptive, the "overgrowth" of the drug-susceptible population has been documented. This may be due to a difference

in "fitness" between the populations of these patients who are drug-susceptible and drug-resistant (Zhang and Yew, 2015; Borrell and Trauner, 2017).

3. Selection mechanism brought on by a decreased antibiotic pressure: The underlying drug-susceptible strain population was seen to reappear after the introduction of second-line therapy changed the antibiotic pressure. It is common knowledge that second-line medicines are less effective at killing bacteria than first-line antibiotics. Thus, Van Rie et al. (2005) supports the hypothesis that medications targeting both drug-susceptible and drug-resistant sub-populations may be necessary for the treatment of individuals with MDRTB.

The relevance of mixed infections for the patient and for TB control is not completely clear. Mixed infections could potentially accelerate the emergence of MDR TB isolates. Moreover, the high proportions of mixed infections in an area where TB is endemic may have implications for prophylactic approaches, such as new vaccines, for the control of TB (Stavrum et al., 2009).

5.5 Orphan MTB showed higher proportional sub-lineage variation as compared to total MTB genotypes in Sabah

An Orphan MTB spoligotype is defined as a single isolate that doesn't correspond to any reported SIT in the database repository (Demay et al., 2012; Couvin et al., 2020a). Prevalence of many MTB orphans may be a sign of sporadic or recent TB infection, and could well be an indication of long-time presence of precursor MTB strains in the studied area. They may have micro-evolved through a series of mutations which eventually resulted in orphan genotypes (Demay et al., 2012; Couvin, David, et al., 2019a; Conceição et al., 2021). The difference among the lineages of the isolates with regard to geographical distribution and demographic variables were analysed.

Indonesian MTB orphans are found to be twice the proportion (within group) as compared to Malaysian and Filipino MTB orphans (Table 4.27) suggesting the possibilities that:

a) most Indonesian TB infections in Sabah are reactivations of previous infections contracted in Indonesia, or
b) most Indonesian TB infections in Sabah are caused by MTB strains brought in from Indonesia in recent years.

Whichever the reason, both these possibilities are corroborated by the theory of overseas-seeding of MTB postulated by Goroh et al. (2020a). In the same light, several earlier investigations using universal genotyping were

complementary whereby they discovered that most immigrant (recently immigrated) TB cases were not clustered. This finding indicated that these cases were the result of reactivations of exposures in the country of origin. Thus, any clustering due to recent transmission among these immigrants would be low (de Vries et al., 2009). However, this also reiterates the fact that clustering or recent transmission among immigrants could mean that the said immigrant population are interacting closely with the locals. In other words, "transmission permeability" has taken place (García De Viedma and Pérez-Lago, 2018) (Figure 4.22). Therefore, universal genotyping has a huge potential for deciphering the complexity and peculiarities of TB transmission dynamics in various populations (García De Viedma and Pérez-Lago, 2018).

5.6 MTB cluster analysis and genotype-complemented TB contact investigation

When paired with epidemiological information, the results of TB genotyping can be used to identify people with TB disease who have recently been involved in the same chain of transmission. It can tell whether a person has TB disease as a result of a recent infection or a former infection (García De Viedma and Pérez-Lago, 2018; Izumi et al., 2019; Kam et al., 2022). The acquired knowledge may be helpful in developing efficient TB prevention and management programs to stop the spread of MTB, particularly resistant strains. Although conventional contact investigation seems laborious and sometimes seemingly inefficient, it can be cost effective if it is equipped with additional

information such as genotypic links (genetic relatedness of MTB strains). In this context, classical genotyping is helpful to complement the conventional contact investigation (Table 4.9, Table 4.10, Table 4.11, Table 4.13). Genotype-complemented TB case contact investigation is necessary because, some of the challenges encountered in conventional contact tracing include memory bias among TB cases and case contacts, as well as omission of information for fear of safety. Conventional contact investigation is usually limited to close household and workplace contact due to inadequate resources in terms of man-power, financial resource, and time. Thus, genotyping has several advantages (provided universal genotyping is implemented) as follows:

1. Able to establish additional epidemiological links based on MTB genetic-relatedness.
2. Able to determine mixed infection especially in DRTB cases.
3. Able to differentiate relapse or reinfection cases.

All these information can be used to guide targeted TB prevention and management interventions.

5.6.1 Conventional contact investigation needs to be revamped with the use of genotyping

Currently, monitoring of TB contacts is with the use of the TBIS101C registry, the national TB information system (Liew et al., 2015; Rosli, 2022). In

order to find additional probable active TB cases and secondary cases that may also have been exposed to the index case, the traditional method of detecting TB transmission events was based on contact investigation (Munang et al., 2016). This method required interviewing every TB patient. The research is based on a concentric-rings strategy, which prioritizes closer interactions in contact investigation tasks before moving on to contacts with lower exposure frequency and length. This approach is helpful when exposure events take place at work or at home, but it has significant drawbacks when tracking exposure events connected to leisure activities or arising from more random or infrequent connections. (García De Viedma and Pérez-Lago, 2018).

This conventional method of contact investigation is quite laborious and tricky as it deals with a lot of recall bias on the part of the TB contacts. Patients, especially non-citizens are sometimes reluctant to give information on people in their sphere of influence and contact especially fellow immigrants as this has an implication on their livelihood and socio-economic status. In cases like these, genotyping will be able to complement contact investigation, and possibly advocate predictive patient management based on genotypes.

This study's classical TB genotyping approach discovered a surprisingly high amount of strain relatedness, which is suggestive of continued transmission (Figure 4.7, Figure 4.8, Figure 4.9). To comprehend the variety of strains circulating in Sabah, both genotypic and epidemiological data were analysed. among Malaysian, Filipinos and Indonesians). The traditional "people, time, place" method works well for diseases with a brief incubation period, but it is

less effective for chronic illnesses like TB, which may develop from a latent infection to an active disease over a period of months to many years.

5.6.2 Clustering rate need to be monitored to assess recent transmission, and reactivation of MTB in Sabah

TB transmission and reactivation rates for conventional contact investigation are 6.1% and 93.9% respectively (Table 4.10). The low transmission and high reactivation rates show or give the impression that TB control activities are effective. TB transmission and reactivation rates using genotyping complemented contact investigation are 23.6% and 76.4% respectively (Table 4.10). Based on the genotyping-complemented figure, TB transmission is higher than previously thought when using conventional contact investigation. The Sabah TB prevention and management program is performing well, as evidenced by the low overall transmission rate. However, there is still an urgent need to further bring down the numbers of MTB cases as suggested by (Dye and Williams, 2010) which is to address LTBI cases in addition to battling the active TB cases. Therefore, the monitoring of clustering and transmission rates can be helpful in containing epidemics and guiding public health activities, such as evaluating the efficacy of TB control measures or allocating funds for contact investigation (de Vries et al. 2009). The determination of the rate of transmission may also have broader social repercussions (de Vries et al., 2009).

5.6.3 Notification interval should be considered as an indicator in the TB surveillance system

In this study a notification interval of ≤1 month between Case#1 and Case#2 in the same cluster resulted in big clonal complex clusters (n=6-225/CC) (Table 4.12). This is slightly different compared to the findings of Izumi et al. (2019) and Kam et al. (2022). The relatively short notification interval implies that there is a higher rate of TB transmission taking place in Sabah especially in the areas where the clustered cases are reported. In this situation, universal genotyping would have been better able to capture the direction and estimated rate of transmission. In order to better understand TB transmission dynamics, the notification interval parameter should also be considered to be included in the TB surveillance system in Sabah. This would give an early alert warning to the TB Control team, and help guide the formulation of case and contact investigation strategies.

5.6.4 Implementation of DCA and PCA has the potential of reducing 60% to 80% of the cost allocation for TB case contact investigation

Implementation of DCA and PCA has the potential of reducing approximately 60% and 80% of the cost respectively while yielding the maximum outcome / output in the contact investigation and TB prevention activities (Figure 4.12. Figure 4.15, Table 4.13). When dealing with high workloads in the field, one is usually pressed for time as patient management

and contact investigations are done according to a tight timeline. Thus, this ranked direct approach (without further analyses on the primary clustering result) is proposed to guide TB prevention and management personnel on which cases should be given priority for contact investigation that would give possible best outcomes based on the resources that is available at hand. The prioritised cluster approach is an expanded search that is expected to get the best possible genotype-complemented contact investigation strategy that is still cost effective. Resources will be prioritised proportionately, focusing on the overlaps or common cases that are considered as important transmission nodes or possible group of index cases.

If we were to look at the A1, Direct Cluster Approach the direct selection of CC1, the largest clonal complex in this study, would reduce the number of contact investigations to be conducted from 1019 to 593 to 422, which is a total reduction of 58.6% (Table 4.13). If we look at the Super Clonal Complex, Prioritised Cluster Approach, we would be 'stacking' the biggest clonal complexes of selected groups with the highest likelihood of exposure to build a super clonal complex that is theoretically the group of cases that would have been the kick-starters of the biggest TB clonal complex in this study (Figure 4.12). This approach would reduce the contact investigation activities from 1019 to 593 to 203, which is a total workload reduction of 80%. This means that an estimated 80% reduction of cost could be expected should this approach be implemented. In a setting with significant budget-constraints, this has massive cost saving implications. According to Pareto analysis, 80% of a project's benefits can be obtained by doing 20% of the effort, or, in the opposite case,

80% of issues can be linked to 20% of the causes. Although the figure 20-80 is just an estimate in pareto analysis, this ratio is seen when the Prioritised Cluster Approach is applied.

5.6.5 Proposed accessory surveillance modalities for early detection of TB

In order to improve TB prevention and management activities, stakeholders will need to be creative and resourceful in terms of finding affordable alternatives to achieve the desired outcomes. These additional measures will hopefully be able to support the existing (laboratory diagnostics and TBIS system) and proposed (genotyping complemented TB contact investigation) TB prevention and management activities.

One of the ways is to optimize utilization of TB patient and TB contact data in the TBIS registries. Repetitive appearance of contact names (≥2 times) in the TB contact registry (TBIS101C) should be closely monitored, and considered for further investigation, as this reflects high possibility of TB exposure which would eventually lead to TB infection.

Diagnostic modalities such as concentrated sputum smear microscopy to detect paucibacillary cases, and/or interferon gamma release assay IGRA to identify LTBI should be considered in these situations (Auguste et al., 2017; Paul et al., 2022). In a concentrated specimen smear microscopy (CSSM) study

(Paul, 2019; Paul et al., 2022), 6.5% of close contact of index cases were found to be AFB smear positive using CSSM, but negative AFB smear using conventional direct sputum smear microscopy (DSSM).

Enhanced epidemiological investigation (EEI) whereby contact investigation is expanded to patient or subject's greater social network. This approach was proposed by Munang et al. (2016; 2019) whereby TB contact investigation could be conducted by prioritising contacts and communities by using novel contact-based criteria in addition to more conventional demographic and clinical ones. In their investigation, the researchers discovered that recurring cases originating from the pool of known patients treated for active or LTBI had a 1.5% likelihood at five years and a 2.7% probability at 10 years. According to Munang (2019), repeated contacts had a twofold increased risk of developing TB when they were investigated compared to non-repeated contacts. (active cases: 3.9% versus 1.6%; latent disease: 10.7% versus 3.7%).

Logically, it is imperative and practicable to incorporate the close monitoring of repeated cases and contacts as one of the key indicators in the TB transmission surveillance as it does not incur significant additional costs, but is able to provide a parameter for the measurement of TB prevention and management performance in Sabah. Maintaining good and consistent recording of patient data is of utmost importance. The well-established Malaysian TB Information System (TBIS) Registers such as TBIS101A (active TB cases) and TBIS101C (case contacts) offers a rare chance to look into the local epidemiology and assist with service planning (MOH, 2002; Dony et al., 2004;

Jelip et al., 2004; Liew et al., 2015). New knowledge and understanding may be produced by the extensive TB contact data related to cases transmission of TB in this environment. Continuous training to improve necessary skills in terms of data management need to be considered as well (Liew et al., 2015).

5.6.6 Cost considerations for MTB genotyping

Dealings with allocations, reagent and asset procurement for laboratory diagnostics is quite the rigmarole. Value of test diagnostics in terms of TB management and control is usually calculated as how much money is spent for a TB case to be detected. A major portion of TB control program allocation has already been channelled to TB laboratory diagnostics. Thus, the idea of increasing allocation even more for genotyping does not seem like a wise idea. In terms of molecular epidemiology, genotyping is able to give information on evolutionary development of MTB strains, especially information on the circulating MTB strains in regions of interest (Brites and Gagneux, 2017; Merker et al., 2017). Improved epidemiological predictions are made possible by studying the evolution of MTB because it advances our knowledge of basic biology, particularly with regard to antibiotic resistance. This in turn would facilitate the prediction of treatment outcome, DST profile, and subsequently aid TB control measures and complement contact investigation activities (Merker et al., 2017; Walker et al., 2018; Merker et al., 2020; Walker et al., 2022).

Genotyping is costly (Table 5.1), therefore, at present, it is only utilised when there are reports of outbreaks. The National TB Reference Laboratory in Sungai Buloh, Malaysia (MKAK) currently conducts genotyping for 10% of positive cultures (1+ growth) from all the states / regions with public health laboratories. States in East and West Malaysia have different MTB genotype profile as shown in Figure 4.6. A paradigm-shift among stakeholders from looking at these costs as expenditures to investments towards controlling and eventually eliminating the disease in future is necessary (Zwerling et al., 2014; Laurence et al., 2015; Vassall, 2016).

Controlling TB is one of the 19 development investments that should be prioritized on a worldwide scale, according to a team of eminent economists' global research for the Copenhagen Consensus on the United Nations' Sustainable Development Goals. Vassall (2016) looked into the rationale for investment and came to the conclusion that, globally, every dollar invested in TB control would result in societal benefits of about $43. When we consider a region with a very high incidence, it is evident that prompt diagnosis and encouraging medication compliance will not only save lives but also result in significant financial gains (Laurence et al., 2015; Vassall, 2016). The genomics approach will need to come hand in hand with the implementation of other downstream TB prevention and management activities such as expanded contact investigation or enhanced epidemiologic investigation (Munang, 2016; Munang, 2019). The utilization of genotyping or the genomic investigations are scalable and ultimately have been shown to be cost saving in some laboratory settings (Witney et al., 2016; Shea et al., 2017). The information afforded by

the genomics approach especially on MTB transmission dynamics, and drug resistance profile will reap its economic value through the (i) prevention of onwards transmission; (ii) reduced time to individually targeted therapy; (iii) reduced hospitalisation and/or death, or (iv) improved efficiency in public health activities.

Table 5.1: Cost for microbead-based spoligotyping (N=1060) and 24-loci MIRU-VNTR (N=593)

Item	Spoligotyping (RM)	24-loci MIRU-VNTR (RM)	DNA Extraction (RM)	Apparatus (One Off)
Consumable	6,896.00	3,448.00	3,045.00	5,014.00
Reagent & Chemical	136,040.00	17,767.00	23,820.00	
Total	142,936.00	21,215.00	26,865.00	5,014.00
Cost/test	134.85	35.78	25.34	

Note: Cost is calculated based on rates in 2017 RM1=USD0.2328; 2018: RM1= USD0.2479

At the time of study, cost for genotyping which include the reagents and consumables are shown in Table 5.1. Based on prices for commercial services as of August 2014, conventional spoligotyping costs US$26, regular MIRU-VNTR typing costs US$49, and WGS costs US$330 (Stucki, 2014). The two-tiered classical genotyping method cost is about 25% of the cost for WGS. Complicated DRTB cases will undoubtedly need WGS as changes can be seen in longitudinal monitoring of cases (de Vries et al., 2009; Navarro et al., 2013; Pérez-Lago et al., 2015).

5.7 Feasibility of application of two-tiered classical genotyping in Sabah

5.7.1 Advantages of spoligotyping in TB molecular epidemiology in Sabah

Spoligotyping is a simple, fast and relatively low-cost method among the classical genotyping methods. It is able to give preliminary clustering information by grouping isolates of cases into sub-lineages or clades. However, due to this typing method's high levels of homoplasy and low-resolution power, the precision for sub-lineage identification is significantly reduced (Barnes and Cave, 2003). Napier et al. (2023) studied the congruence between spoligotype-based and SNP-based genotyping. Their study found that spoligotyping is suited for low resolution surveillance and WGS/SNP-based typing is good for high resolution studies. WGS technology has led to characterization using SNP-based sub-lineage nomenclature. The majority of spoligotypes (L1-L7 including animal lineage) showed perfect correlation with the major MTB lineages. However, at the sub-lineage level, only 65% spoligotypes at the second or higher levels of the hierarchy were fully correlated with sub-lineages. (Napier et al., 2023). Although spoligotyping does not possess high precision, it is able to function as a simple, fast and reasonably good first-tier screening tool for MTB especially in a large-scale operation such as outbreaks or surveillance exercises.

5.7.2 Importance of 24-loci MIRU-VNTR in TB case contact investigation in Sabah

Since Sabah has a significantly high proportion of EAI2-Manila spoligotype, it is challenging to determine if the isolates are epidemiologically-linked based on sub-lineage information alone (Table 4.4). Therefore, a 24-loci MIRU-VNTR with its high discriminatory power and better able to differentiate and discriminate between closely related isolates carrying the exact same 24-loci MIRU-VNTR profile ID, or isolates differing only in one to two alleles should be considered for implementation (Figure 4.7, Table 4.9, Table 4.10, Table 4.11, Table 4.13). High levels of strain similarity are a sign of continued transmission. To comprehend the diversity of MTB strains circulating among individuals in Sabah, Malaysia, genomic and epidemiological data were studied.

The duplex 24-loci MIRU-VNTR method modified by Yasmin et al. (2016) employing optimised agarose gel electrophoresis method has much reduced cost and made it affordable especially for high TB burden low-income countries. Various studies on MIRU-VNTR showed that the discriminatory power of the loci panel varies in different countries / regions, possibly attributed to the different proportion or predominance of MTB lineages and genotypes in the area of interest (Wada et al., 2007; Allix-Béguec et al., 2014; Skiba et al., 2015; Yin et al., 2016; Shi et al., 2018; Ghavidel et al., 2019; Mokrousov et al., 2023). Henan, the most populous province in China, with predominantly Beijing genotype (85%) reported that the combination of the 10 most discriminatory loci: QUB11b, Mtub21, MIRU26, QUB26, Mtub04, MIRU10,

ETR-F, ETR-E, MIRU39 and ETR-A rendered a HGDI close to that of the 26-loci MIRU-VNTR (Shi et al., 2018), and is therefore suitable for use in regions highly prevalent with Beijing genotypes. EAI genotypes, especially EAI2-Manila are found in high frequencies in this study (Table 4.4). This study also found that the 4 loci: QUB11b, Mtub21, ETR-A and ETR-B exhibited the highest discriminatory power when tested on MTB isolates in Sabah, especially with EAI (Table 4.9). This shows that this method can be customised according to the genotyping needs of the area of interest, and the allocation of the operational budget as well.

At present, there are varied perception and conclusion as related to the usefulness and functionality of MIRU-VNTR. Wyllie et al. (2018) and Munang (2019) who did remarkable work in WGS studies and TB contact investigation respectively, argued against the routine use of MIRU-VNTR in TB prevention and management as it has the potential to overestimate isolate relatedness. But it also has to be taken into consideration that their study setting had TB prevalence that is much lower than this study setting. Indeed, WGS is the way forward when it comes to molecular epidemiology and especially in dealing with complicated cases such as DRTB, MDRTB and XDRTB.

5.7.3 The practicability of classical genotyping in TB prevention and management in Sabah

MIRU-VNTR is not the perfect option, however, it also needs to be highlighted that many global databases on TB are still largely using classical genotyping data such as spoligotyping, and 24-loci MIRU-VNTR. In the Malaysian context, universities and Institute for Medical Research (IMR) have already embarked on the genomics, however this is still mostly done for research purposes. When it comes to field-work, especially for the purpose of TB epidemiology and contact investigation for TB cases, this two-tiered genotyping is a practical solution to get the work done. WGS at the mass scale is not yet feasible at the moment. At the very least, 24-loci MIRU-VNTR gives an idea or direction of how contact investigation could and should be done. It is also able to screen the isolates and render the best candidates for further testing for WGS etc. in a cost-effective manner, enabling relevant and significant inference of all things MTB in future. Ultimately, cost will still be the deciding factor (Table 5.1). The US CDC has recently announced that the new isolates will only be genotyped using WGS instead of RFLP / spoligotyping and MIRU-VNTR (CDC, 2022). Singapore has also done the same switch earlier (Kam et al., 2022). Perhaps a compromise could be met. Until the costs to conduct WGS is no longer an issue, classical genotyping could be employed in the operational setting, while WGS studies could be conducted in state-of-the-art referral or research laboratories in IMR and CDC Malaysia (government setting), and institutions of higher learning (private setting) with the condition that the information will be shared and made available in an agreed manner.

At this juncture, in the context of programmatic TB contact investigation, and cost restriction concerns, WGS need not be fully employed as yet as classical genotyping could get the preliminary work done especially in hot-spots and highly populated areas. Implementation of classical genotyping will be able to render valuable information and give insight to the evolutionary development and transmission dynamics of the MTB strains in this region. This move will also minimize cost significantly and may prove to be cost effective in the long run. When more complex in-depth questions pertaining to highly unique MTB features in the public health or clinical setting arises, WGS will be conducted efficiently on the pre-selected strains for the best outcome. However, it cannot be denied that full WGS implementation for genotyping and in-depth phylogenetic and molecular epidemiology studies is indeed the way forward.

5.8 Limitations of research

Whole genome sequencing (WGS) was not done due to degradation of DNA samples. Thus, many in-depth discussion and inferences could not be done due to the lack of genomic information. Universal genotyping was also not done thus, many possible connections between cases were not able to be identified. Sample size for measurement of resistance and other detailed parameters are small. Distribution of samples conveniently collected from the small archived isolates 2015-2016, and the prospective isolates 2017-2018 were not equal due to technical limitations.

CHAPTER 6

CONCLUSIONS

6.1 Overall study conclusions

This study discovered some new information and reiterated findings and facts previously reported as well.

SIT19/EAI2 Manila SIT 19 is the predominant spoligotype in Sabah. Spoligotype profiles are quite similar among the 3 nationalities (Malaysian, Indonesian and Filipino). Filipino spoligotypes are almost a perfect subset of Malaysian spoligotypes with only a difference of 5 SITs in EAI2 Manila (3 SITs), EAI5 (1 SIT) and 1 SIT from the unknown lineage. Indonesians have more exclusive SITs, namely EAI1-SOM (2 SITs), EAI2 Manila (3 SITs), EAI5 (1 SIT), H3 (2 SITs), LAM (1 SIT), and T (3 SITs).

Spoligotype spread in Sabah is different compared to peninsular Malaysia in terms of sub-lineage proportion whereby there are more EAI strains, especially EAI2 Manila variety of SITs versus Beijing strain. The MTB spoligotypes proportion (%) for Sabah: peninsular Malaysia are as follows:

EAI : 92%:30%

Beijing: 3%:59%

T : 2%:11%

LAM : 2%:<1%

34 new spoligotypes have been discovered in this study. Many of them are derived from EAI2-Manila (16 of 34 SITs). Of the 8 SITs from the Unknown lineage, 2 SITs have been reported in China and Saudi Arabia respectively, while 6 SITs are reported for the first time in this study. 2 SITs (4186 and 4204) are associated with all 3 nationalities; 2 SITs (4188 and 4194) Malaysians and Filipinos and the remaining 4 SITS (4185, 4192, 4198, 4208) are associated with Malaysians only. The similarities in the spoligotyping pattern of the new SITs in this study point to the possibility that these SITs may be a part of an emerging lineage possibly unique to this region (not yet ascertainable using available methods e.g., SpolLineages Tool, SITVIT2, SpolSimilarity Search and SpotClust databases).

Microbead-based spoligotyping is a suitable genotyping method for preliminary screening and clustering of TB cases in outbreak investigation and case contact investigation. Genotyping demonstrated epidemiological links not previously reported which allowed the a) corroboration of or support clustering of cases established by conventional TB case contact investigation (generally assumed to be from the same source of TB infection); and b) determination of inter-district associations of TB strains (state-wide or nationwide).

Better definition of the proportion of recent transmission versus reactivation TB cases in Sabah were reported. When conventional method of case contact investigation is used, the rate of active infection and reactivation respectively are 6.1% and 93.9%. The figures for genotyping complemented contact investigation are 23.6% and 76.4% respectively.

Highly discriminatory 24-loci MIRU-VNTR typing method is able to cluster MTB isolated from patients. ETR-A, ETR-B, Mtub21 and QUB11b have the highest discriminatory power especially among the EAI lineage and EAI2-Manila sub-lineages in Sabah MTB population.

Drug Resistance (8.2%) and MDRTB (0.6%) levels are still relatively low in Sabah. INH-R is at 3.3%, and is nine times more than RRTB. High level INH resistance (*katG* gene mutation) was found to occur at the same ratio as low-level INH-R (*inhA*-prom gene mutation). Although there is no clear relationship between DRTB and sub-lineage, cases of DRTB, failed DST and relapses need to be closely monitored for possible links as they have similar underlying implications and potential impact.

MTB orphans were mostly derived from the lineages: EAI (45.8%) especially EAI2 Manila (22.2%) and Unknown (41.2%). 24-loci MIRU-VNTR clustering analysis grouped 74 randomly selected orphan isolates into five clonal complexes as well as singletons. Malaysian (13.8%) and Filipino (13.6%) strains render similar proportionate percentages of orphan while Indonesian MTB orphan percentages recorded double the number

(28.4%) supporting the likelihood of most Indonesian MTB as possible reactivation of old cases seeded in Indonesia, or new strains possibly recently brought in from Indonesia.

All 36 cases of ‘mixed infections’ occur in newly diagnosed TB cases, revealing the possibility that TB patients may be infected with, or have a disease caused by more than 1 MTB strain and exposed to highly prevalent TB cases in possible TB Hot-spot environment. Mixed infections in DRTB could potentially cause delays in result release due to the difficulties in determining the accurate drug susceptibility results rendered by conflicting DST outcomes with different test batch, and types of tests performed.

Another challenge faced by DST laboratories is the seemingly ‘changing’ results for MTB DST with every test performed, be it the same sample or different samples sent. This is one of the examples how genotyping could actually be used to guide decision for TB patient management which entails the complexities of changing treatment regime, cost implication and the necessity of close monitoring.

6.2 Recommendations to the Sabah State Health Department

Based on this study, there are four activities recommended to the State Health Department for the future of TB laboratory services in the greater context of TB Control.

Universal genotyping on all TB cases (both pan-susceptible and drug-resistant MTB) should be performed to get a baseline information of MTB genotype. Although it does not give conclusive evidence, this will give a good idea as to which direction TB control program need to move towards. Genotyping will help to determine as to whether or not re-infection or reactivation of cases took place, thus guiding treatment and management strategies. Analysis of genotype-complemented contact investigation would enable earlier detection of outbreaks and rapid control, easy identification of incorrect TB diagnoses based on false-positive culture results. Discovery of previously unknown epidemiological links in new as well as unusual and / or inter-district transmission settings will be detected more readily.

To update the Nationwide MTB Laboratory Algorithm: Genotyping of baseline (new) positive MTB cultures, and MTB database compilation for future reference. This information will be used to monitor trends as a predictive surveillance and outbreak

management tool, as well as an alternative laboratory surveillance method on drug resistance.

To devise and improve training of laboratory scientists and technologists to recognize possible DRTB by identifying early signs of DRTB to ensure sufficient and early detection of cases of this nature. Refresher courses for clinicians and the TB management team on DRTB and MDRTB also needs to be reviewed so as to enable early recognition of possible DRTB cases among patients. By doing this, DRTB and MDRTB diagnosis and appropriate treatment initiation could be done much earlier. This will also reduce even further transmission of primary drug resistance.

To establish Smart Partnerships between Universities and Ministry of Health in a symbiotic manner whereby information discovered by universities through research and exploration of MTB and other aetiologic agents relevant to the state can be implemented as appropriate in the field by government stakeholders such as hospital and public health laboratories. Workable memorandum of understanding needs to be in place to achieve this goal.

6.3 Future studies

Addressing the global TB burden requires enhancement in biomedical technology, in conjunction with socio-economic and cultural re-orientation for change. Multi-sectoral coordination and action involving the health and other governmental sectors are needed (Koh et al., 2019). Being a socially-related disease, TB research also requires a bio-social approach, whereby socio-economic, cultural and environmental factors need to be scrutinized. Additional research incorporating the implementation of science, health economics, and social epidemiology is needed to identify and design feasible, scalable social protection interventions that are context-specific. Effective policies need to be established to give national TB programmes the support and mandate necessary to address the socioeconomic consequences of illness for their patients (Koh et al., 2019). In light of this vision, these six future studies are proposed to be conducted following this study to complement TB prevention and management, and to facilitate the reduction of the number of TB cases in the state.

Trend monitoring of the MTB strains and determining rate of change via longitudinal TB case monitoring (from latency, active to relapse or re-infection) in Sabah needs to be conducted to deepen our understanding of MTB drug resistance and transmission.

Future studies on MTB orphan strains and subtypes of spoligotype and MIRU-VNTR clusters in Sabah using WGS to provide more phylogenetic

inferences especially on the rate of evolution to determine epidemiological link and trend of clustering.

Improvement on Case Contact investigation utilising custom-defined, enhanced questionnaire which include extended daily social activities to capture possible TB cases earlier within 'High Risk' populations.

Application of the proposed Direct and Prioritised Culture Approach for TB case contact investigation described in this study to determine its effectiveness in terms of case detection and cost effectiveness.

A coordinated nationwide survey using WGS to provide more detail on the distribution of different genotype families in Malaysia, and a study of possible associations between host and mycobacterial genetics to establish if differences in MTB population structures are caused by evolutionary adaptation of particular mycobacterial lineages to certain human populations on a national level.

A Knowledge-Attitude / Perception-Practice (KAP) study on the similarities and differences in race / ethnicity-related subcultures and practices pertaining to healthcare in general, and MTB specifically.

LIST OF REFERENCES

Aiello, A.E., Renson, A. and Zivich, P., 2020. Social media-and internet-based disease surveillance for public health. *Annual review of public health*, 41, p.101.

Aldridge, R.W., Zenner, D., White, P.J., Williamson, E.J., Muzyamba, M.C., Dhavan, P., Mosca, D., Thomas, H.L., Lalor, M.K., Abubakar, I. and Hayward, A.C., 2016. Tuberculosis in migrants moving from high-incidence to low-incidence countries: a population-based cohort study of 519 955 migrants screened before entry to England, Wales, and Northern Ireland. *The Lancet*, *388*(10059), pp.2510-2518.

Allix-Béguec, C., Harmsen, D., Weniger, T., Supply, P. and Niemann, S., 2008. Evaluation and strategy for use of MIRU-VNTR plus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *Journal of clinical microbiology*, *46*(8), pp.2692-2699.

Allix-Béguec, C., Wahl, C., Hanekom, M., Nikolayevskyy, V., Drobniewski, F., Maeda, S., Campos-Herrero, I., Mokrousov, I., Niemann, S., Kontsevaya, I. and Rastogi, N., 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), pp.164-172.

Alonso, H., Samper, S., Martín, C. and Otal, I., 2013. Mapping IS 6110 in high-copy number Mycobacterium tuberculosis strains shows specific insertion points in the Beijing genotype. *BMC genomics*, 14, pp.1-11.

Angra, P., Ridderhof, J., Tahseen, S. and Van Deun, A., 2016. Read the new microscopy handbook: even the Ziehl-Neelsen technique has changed. The international journal of tuberculosis and lung disease: the official journal of the *International Union against Tuberculosis and Lung Disease*, 20(4), 567-567.

Arriaza, B. T., Salo, W., Aufderheide, A. C. and Holcomb, T. A., 1995. Pre-Columbian tuberculosis in Northern Chile: Molecular and skeletal evidence. *American journal of physical anthropology*, 98(1), 37-45.

Auguste, P., Tsertsvadze, A., Pink, J., Court, R., McCarthy, N., Sutcliffe, P. and Clarke, A., 2017. Comparing interferon-gamma release assays with tuberculin skin test for identifying latent tuberculosis infection that progresses to active tuberculosis: systematic review and meta-analysis. *BMC infectious diseases*, *17*, pp.1-13.

Avoi, R., 2015. TB Prevention and management algorithm in Sabah. TB/Leprosy Control Unit, Sabah State Health Department *(*Unpublished data*).*

Avoi, R. and Liaw, Y.C., 2021. Tuberculosis death epidemiology and its associated risk factors in sabah, malaysia. *International Journal of Environmental Research and Public Health*, 18(18).

Awaluddin, S.M., Ismail, N., Zakaria, Y., Yasin, S.M., Razali, A., Mutalip, M.H.A., Lodz, N.A., Musa, K.I., Kusnin, F. and Aris, T., 2020. Characteristics of paediatric patients with tuberculosis and associated determinants of treatment success in Malaysia using the MyTB version 2.1 database over five years. *BMC public health*, 20, pp.1-9.

Azé, J., Sola, C., Zhang, J., Lafosse-Marin, F., Yasmin, M., Siddiqui, R., Kremer, K., van Soolingen, D. and Refrégier, G., 2015. Genomics and machine learning for taxonomy consensus: the *Mycobacterium tuberculosis* complex paradigm. *PloS one*, *10*(7), p.e0130912.

Bainomugisa, A., Meumann, E.M., Rajahram, G.S., Ong, R.T.H., Coin, L., Paul, D.C., William, T., Coulter, C. and Ralph, A.P., 2021. Genomic epidemiology of tuberculosis in eastern Malaysia: insights for strengthening public health responses. *Microbial genomics*, *7*(5), p.000573.

Bandera, A., Gori, A., Catozzi, L., Esposti, A. D., Marchetti, G., Molteni, C., Ferrario, G., Codecasa, L., Penati, V., Matteelli, A. and Franzetti, F., 2001. Molecular epidemiology study of exogenous reinfection in an area with a low incidence of tuberculosis. *Journal of clinical microbiology*, 39(6), 2213-2218.

Balaban, N.Q., 2011. Persistence: mechanisms for triggering and enhancing phenotypic variability. *Current opinion in genetics & development*, *21*(6), pp.768-775.

Barcellos, R.B., De Almeida, I.N., Da Silva, E.C., Gomes, H.M., de Assis Figueredo, L.J., Halon, M.L., Dalla Costa, E.R., Dos Santos, Í.R., Vater, M.C., Alves, S. and da Silva Carvalho, W., 2019. Multicenter evaluation of TB-SPRINT 59-Plex Beamedex®: accuracy and cost analysis. *BMC Infectious Diseases, 19*, pp.1-8.

Bardosh, K., Vries, D.D., Stellmach, D., Thorlie, A., Cremers, L. and Kinsman, J., 2019. Towards people-centred epidemic preparedness & response: from knowledge to action.

Barnes, P.F. and Cave, M.D., 2003. Molecular epidemiology of tuberculosis. *New England Journal of Medicine*, *349*(12), pp.1149-1156.

Borrell, S. and Gagneux, S., 2011. Strain diversity, epistasis and the evolution of drug resistance in *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection*, *17*(6), pp.815-820.

Borrell, S. and Trauner, A., 2017. Strain diversity and the evolution of antibiotic resistance. In: Gagneux, S., (ed). *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*. New York: Springer LLC, pp.263-279.

Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., Zoppo, M., Orgeur, M., Khanna, V., Hardy, D. and Mangenot, S., 2020. TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis lineages. *Nature communications*, 11(1), p.684.

Boxer, A.H., 1969. *Experts in Asia: An Inquiry into Australian Technical Assistance*, Canberra.

Brandis, G., Wrande, M., Liljas, L. and Hughes, D., 2012. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Molecular Microbiology*, 85(1), pp.142–151.

Brites, D. and Gagneux, S., 2015. Co-evolution of *Mycobacterium tuberculosis* and *Homo sapiens*. *Immunological Reviews*, 264(1), pp.6–24.

Brites, D. and Gagneux, S., 2017. The nature and evolution of genomic diversity in the *Mycobacterium tuberculosis* complex. In: Gagneux, S., (ed). *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*. New York: Springer LLC, pp.1-26.

Brossier, F., Veziris, N., Truffot-Pernot, C., Jarlier, V. and Sougakoff, W., 2006. Performance of the genotype MTBDR line probe assay for detection of resistance to rifampin and isoniazid in strains of Mycobacterium tuberculosis with low-and high-level resistance. *Journal of clinical microbiology*, *44*(10), pp.3659-3664.

Brooks, M.B., Millones, A.K., Puma, D., Contreras, C., Jimenez, J., Tzelios, C., Jenkins, H.E., Yuen, C.M., Keshavjee, S., Lecca, L. and Becerra, M.C., 2022. Mapping local hot spots with routine tuberculosis data: A pragmatic approach to identify spatial variability. *Plos one*, *17*(3), p.e0265826.

Broset, E., Martín, C. and Gonzalo-Asensio, J., 2015. Evolutionary landscape of the *Mycobacterium tuberculosis* complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *MBio*, *6*(5), pp.10-1128.

Brudey, K., Driscoll, J.R., Rigouts, L., Prodinger, W.M., Gori, A., Al-Hajoj, S.A., Allix, C., Aristimuño, L., Arora, J., Baumanis, V. and Binder, L., 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC microbiology*, *6*, pp.1-17.

CDC, 2006. TB Program Evaluation Handbook TB Program Evaluation Handbook: Introduction to Program Evaluation. *CDC*, p.34.

CDC, 2022. Transition to whole-genome multilocus sequence typing (wgMLST) for TB cluster detection. *Division of Tuberculosis Elimination*. CDC Atlanta.

Chakaya, J., Petersen, E., Nantanda, R., Mungai, B.N., Migliori, G.B., Amanullah, F., Lungu, P., Ntoumi, F., Kumarasamy, N., Maeurer, M. and Zumla, A., 2022. The WHO Global Tuberculosis 2021 Report–not so good news and turning the tide back to End TB. *International Journal of Infectious Diseases*, *124*, pp. S26-S29.

Chen, Y.Y., Chang, J.R., Huang, W.F., Hsu, C.H., Cheng, H.Y., Sun, J.R., Kuo, S.C., Su, I.J., Lin, M.S., Chen, W. and Dou, H.Y., 2017. Genetic diversity of the *Mycobacterium tuberculosis* East African–Indian family in three tropical Asian countries. *Journal of Microbiology, Immunology and Infection*, *50*(6), pp.886-892.

Chihota, V.N., Niehaus, A., Streicher, E.M., Wang, X., Sampson, S.L., Mason, P., Källenius, G., Mfinanga, S.G., Pillay, M., Klopper, M. and Kasongo, W., 2018. Geospatial distribution of *Mycobacterium tuberculosis* genotypes in Africa. *PLoS One*, *13*(8), p.e0200632.

Cirillo, D.M., Miotto, P. and Tortoli, E., 2017. Evolution of phenotypic and molecular drug susceptibility testing. In: Gagneux, S., (ed). *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*. New York: Springer LLC, pp. 221–246.

Click, E. S., Moonan, P. K., Winston, C. A., Cowan, L. S., and Oeltmann, J. E., 2012. Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. *Clinical Infectious Diseases*, 54(2), 211-219.

Cohn, D. L., Bustreo, F., & Raviglione, M. C., 1997. Drug-resistant tuberculosis: review of the worldwide situation and the WHO/IUATLD global surveillance project. *Clinical infectious diseases*, 24(Supplement_1), S121-S130.

Coitinho, C., Greif, G., Robello, C., Laserra, P., Willery, E., and Supply, P., 2014. Rapidly progressing tuberculosis outbreak in a very low risk group. *European Respiratory Journal*, 43(3), 903-906.

Comas, I. and Gagneux, S., 2009. The past and future of tuberculosis research. *PLoS Pathogens*, 5(10), pp.1–7.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G. and Yeboah-Manu, D., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature genetics*, *45*(10), pp.1176-1182.

Conceição, E.C., Rastogi, N., Couvin, D., Lopes, M.L., Furlaneto, I.P., Gomes, H.M., Vasconcellos, S.E.G., Suffys, P.N., Schneider, M.P.C., de Sousa, M.S. and Sola, C., 2017. Genetic diversity of *Mycobacterium tuberculosis* from Pará, Brazil, reveals a higher frequency of ancestral strains than previously reported in South America. *Infection, Genetics and Evolution*, *56*, pp.62-72.

Conceição, E.C., Salvato, R.S., Gomes, K.M., Guimarães, A.E.D.S., da Conceição, M.L., Souza e Guimarães, R.J.D.P., Sharma, A., Furlaneto, I.P., Barcellos, R.B., Bollela, V.R. and Anselmo, L.M.P., 2021. Molecular epidemiology of *Mycobacterium tuberculosis* in Brazil before the whole genome sequencing era: a literature review. *Memórias do Instituto Oswaldo Cruz*, *116*, p.e200517.

Connelly, R., Gayle, V. and Lambert, P.S., 2016. A Review of occupation-based social classifications for social survey research. *Methodological Innovations*, 9, p.205979911663800. Available at: http://journals.sagepub.com/doi/10.1177/2059799116638003.

Coscolla, M. and Gagneux, S., 2010. Does M. tuberculosis genomic diversity explain disease diversity? *Drug Discovery Today: Disease Mechanisms*, 7(1), pp.e43–e59. Available at: http://dx.doi.org/10.1016/j.ddmec.2010.09.004.

Coscolla, M. and Gagneux, S., 2014. Consequences of genomic diversity in *Mycobacterium tuberculosis*. *Seminars in Immunology*, 26(6), pp.431–444. Available at: http://dx.doi.org/10.1016/j.smim.2014.09.012.

Couvin, D., Zozio, T. and Rastogi, N., 2017. SpolSimilaritySearch – A web tool to compare and search similarities between spoligotypes of *Mycobacterium tuberculosis* complex. *Tuberculosis*, 105, pp.49–52.

Couvin, D., David, A., Zozio, T. and Rastogi, N., 2019a. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infection, Genetics and Evolution*, 72, pp.31–43.

Couvin, D., Reynaud, Y. and Rastogi, N., 2019b. Two tales: Worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of *Mycobacterium tuberculosis* underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. *PLoS ONE*, 14(7).

Couvin, D., Segretier, W., Stattner, E. and Rastogi, N., 2020. Novel methods included in SpolLineages tool for fast and precise prediction of *Mycobacterium tuberculosis* complex spoligotype families. *Database*, 2020, p.baaa108.

Cowan, L.S., Diem, L., Brake, M.C. and Crawford, J.T., 2004. Transfer of a *Mycobacterium tuberculosis* Genotyping Method, Spoligotyping, from a Reverse Line-Blot Hybridization, Membrane-Based Assay to the Luminex Multianalyte Profiling System. *Journal of Clinical Microbiology*, 42(1), pp.474–477.

Cowger, T.L., Wortham, J.M. and Burton, D.C., 2019. Epidemiology of tuberculosis among children and adolescents in the USA, 2007–17: an analysis of national surveillance data. The Lancet Public Health, 4(10), pp.e506-e516.

Cox, H. S., Niemann, S., Ismailov, G., Doshetov, D., Orozco, J. D., Blok, L., Rüsch Gerdes, S. and Kebede, Y., 2007. Risk of acquired drug resistance during short-course directly observed treatment of tuberculosis in an area with high levels of drug resistance. *Clinical Infectious Diseases*, 44(11), 1421-1427.

Dale, J. W., R. Mat Nor, S. Ramayah, T. Hock Tang, and Z.F.Z., 1999. Molecular epidemiology of tuberculosis in Malaysia. *J. Clin. Microbiol*, 37(5), pp.1265–1268.

Dale, J. W., Al-Ghusein, H., Al-Hashmi, S., Butcher, P., Dickens, A. L., Drobniewski, F., Forbes, K. J., Gillespie et al., S. H., Lamprecht, D., McHugh, T. D., Pitman, R., Rastogi, N., Smith, A. T., Sola, C. and Yesilkaya, H., 2003. Evolutionary relationships among strains of Mycobacterium tuberculosis with few copies of IS 6110. *Journal of bacteriology*, 185(8), 2555-2562.

Daniel, T.M., Bates, J.H. and Downes, K.A., 1994. History of tuberculosis. *Tuberculosis: pathogenesis, protection, and control*, pp.13-24.

Daniel, T.M., 2006. The history of tuberculosis. *Respiratory Medicine*, 100(11), pp.1862–1870.

Dean, A. S., Cox, H. and Zignol, M., 2017. Epidemiology of drug-resistant tuberculosis. In: Gagneux, S., (ed). *Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control*. New York: Springer, 209-220.who

Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., Mokrousov, I., Sola, C., Zozio, T. and 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), pp.755-766.

Department of Statistics (DOSM), 2020. Current Population Estimates. *Department of Statistics Malaysia*, (December), pp.1–141. Available at: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=aWJZRkJ4UEdKcUZpT2tVT090Snpydz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09.

Devaux, I., Kremer, K., Heersma, H. and van Soolingen, D., 2009. Clusters of multidrug-resistant *Mycobacterium tuberculosis* cases, Europe. *Emerging Infectious Diseases*, 15(7), pp.1052–1060.

De Vos, M., Müller, B., Borrell, S., Black, P. A., Van Helden, P. D., Warren, R. M., Gagneux, S. and Victor, T. C., 2013. Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. *Antimicrobial agents and chemotherapy*, 57(2), 827-832.

de Vries, G., van Hest, R.A., Burdo, C.C., van Soolingen, D. and Richardus, J.H., 2009. A *Mycobacterium tuberculosis* cluster demonstrating the use of genotyping in urban tuberculosis control. *BMC infectious diseases*, *9*, pp.1-8.

Dony, J.F., Ahmad, J. and Tiong, Y.K., 2004. Epidemiology of tuberculosis and leprosy, Sabah, Malaysia. *Tuberculosis*, 84(1–2), pp.8–18.

Douglas, J.T., Qian, L., Montoya, J.C., Musser, J.M., Van Embden, J.D., van Soolingen, D. and Kremer, K., 2003. Characterization of the Manila family of *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology*, *41*(6), pp.2723-2726.

Dye, C. and Williams, B.G., 2010. The population dynamics and control of tuberculosis. *Science*, 328(5980), pp.856–861.

England, S., Espinal, M. and Raviglione, M., 2008. Global Plan to Stop TB 2006–2015. Clinical Tuberculosis. 4th Edition. London: Hodder Arnold, 489-505.

Espinal, M. A., Laszlo, A., Simonsen, L., Boulahbal, F., Kim, S. J., Reniero, A., Hoffner, S., Rieder, H. L., Binkin, N., Dye, C., Williams, R. and Raviglione, M. C., 2001. Global trends in resistance to antituberculosis drugs. *New England Journal of Medicine*, 344(17), 1294-1303.

Fakhruzzaman, M.N.N., Abidin, N.Z., Aziz, Z.A., Lim, W.F., Richard, J.J., Noorliza, M.N., Hani, M.H., Norhayati, R., Zamzurina, A.B., Zuraina, M.Y.F. and Hisyam, M.J., 2019. Diversified lineages and drug-resistance profiles of clinical isolates of *Mycobacterium tuberculosis* complex in Malaysia. *The International Journal of Mycobacteriology*, *8*(4), pp.320-328.

Filliol, I., Sola, C., and Rastogi, N., 2000. Detection of a previously unamplified spacer within the DR locus of *Mycobacterium tuberculosis*: epidemiological implications. *Journal of clinical microbiology*, 38(3), 1231-1234.

Filliol, I., Driscoll, J. R., van Soolingen, D., Kreiswirth, B. N., Kremer, K., Valétudie, G., Dang, D. A., Barlow, R., Banerjee, D., Bifani, P. J., Brudey, K., Cataldi, A., Cooksey, R. C., Cousins, D. V., Dale, J. W., Dellagostin, O. A., Drobniewski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M. C., Haas, W. H., Heersma, H., Kassa-Kelembho, E., Ho, M.L., Makristathis, A., Mammina, C., Martin, G., Moström, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Niobe-Eyangoh, S. N., Pape, J. W., Rasolofo-Razanamparany, V., Ridell, M., Rossetti, M. L., Stauffer, F., Suffys, P. N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J. H., Sola, C. and Rastogi, N., 2003. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *Journal of clinical microbiology*, 41(5), 1963–1970.

Freschi, L., Vargas Jr, R., Husain, A., Kamal, S.M., Skrahina, A., Tahseen, S., Ismail, N., Barbova, A., Niemann, S., Cirillo, D.M. and Dean, A.S., 2021. Population structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. *Nature communications*, *12*(1), p.6099.

Gagneux, S., Long, C.D., Small, P.M., Van, T., Schoolnik, G.K. and Bohannan, B.J., 2006. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science*, *312*(5782), pp.1944-1946.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., De Jong, B.C., Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C. and Hilty, M., 2006. Variable host–pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *103*(8), pp.2869-2873.

Gagneux, S. and Small, P.M., 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infectious Diseases*, 7(5), pp.328–337.

Gagneux, S., 2009. Fitness cost of drug resistance in *Mycobacterium tuberculosis. Clinical Microbiology and Infection*, *15*, pp.66-68.

Gagneux, S., 2012. Host–pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), pp.850-859.
Gagneux, S., 2018. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*, *16*(4), pp.202-213.

Garcia de Viedma, D., Marín, M., Hernangómez, S., Díaz, M., Serrano, M. J. R., Alcalá, L. and Bouza, E., 2002. Tuberculosis recurrences: reinfection plays a role in a population whose clinical/epidemiological characteristics do not favor reinfection. *Archives of internal medicine*, 162(16), 1873-1879.

García De Viedma, D. and Pérez-Lago, L., 2018. The evolution of genotyping strategies to detect, analyze, and control transmission of tuberculosis. *Microbiology spectrum*, *6*(5), pp.10-1128.

Gautam, S., Qureshi, K. A., Jameel Pasha, S. B., Dhanasekaran, S., Aspatwar, A., Parkkila, S., Alanazi, S., Atiya, A. and Venugopal, D., 2023. Medicinal plants as therapeutic alternatives to combat *Mycobacterium tuberculosis*: A comprehensive review. *Antibiotics*, 12(3), 541.

Ghavidel, M., Tadayon, K., Mosavari, N., Nourian, K., BahramiTaghanaki, H.R., Mohammadi, G.R., Rashtibaf, M. and Ghazvini, K., 2019. Introducing the best six loci in mycobacterial interspersed repetitive unit-variable-number tandem repeat (MIRU-VNTR) typing for *Mycobacterium tuberculosis* genotyping. *Reports of Biochemistry & Molecular Biology*, *8*(3), p.335.

Gomgnimbou, M.K., Abadia, E., Zhang, J., Refrégier, G., Panaiotov, S., Bachiyska, E. and Sola, C., 2012. "Spoligoriftyping," a dual-priming-oligonucleotide-based direct-hybridization assay for tuberculosis control with a multianalyte microbead-based hybridization system. *Journal of clinical microbiology*, *50*(10), pp.3172-3179.

Gomgnimbou, M.K., Hernández-Neuta, I., Panaiotov, S., Bachiyska, E., Palomino, J.C., Martin, A., del Portillo, P., Refregier, G. and Sola, C., 2013. Tuberculosis-spoligo-rifampin-isoniazid typing: an all-in-one assay technique for surveillance and control of multidrug-resistant tuberculosis on Luminex devices. *Journal of clinical microbiology*, *51(11)*, pp.3527-3534.

Good, R.C. & Shinnick, T.M., 1998. Mycobacterium. In *Topley & Wilson's microbiology and microbial infections, systemic bacteriology, 9th ed., vol. 2*. London: Edward Arnold, pp. 549–576.

Goroh, M.M.D., Rajahram, G.S., Avoi, R., Van Den Boogaard, C.H., William, T., Ralph, A.P. and Lowbridge, C., 2020a. Epidemiology of tuberculosis in Sabah, Malaysia, 2012–2018. *Infectious diseases of poverty*, *9*, pp.1-11.

Goroh, M.M.D., van den Boogaard, C.H., Ibrahim, M.Y., Tha, N.O., Swe, Robinson, F., Lukman, K.A., Jeffree, M.S., William, T. and Ralph, A.P., 2020b. Factors affecting continued participation in tuberculosis contact investigation in a low-income, high-burden setting. *Tropical Medicine and Infectious Disease*, *5*(3), p.124.

Goroh, M.M.D., van den Boogaard, C.H., Lukman, K.A., Lowbridge, C., Juin, W.K., William, T., Jeffree, M.S. and Ralph, A.P., 2023. Factors affecting implementation of tuberculosis contact investigation and tuberculosis preventive therapy among children in Sabah, East Malaysia: A qualitative study. *Plos one*, *18*(5), p.e0285534.

Groenen, P. M., Bunschoten, A. E., Soolingen, D. V. and Errtbden, J. D. V., 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.

Guerrin-Tran, E., Thiolet, J.M., Rousseau, C., Henry, S., Poirier, C., Che, D., Vinas, J.M., Jarlier, V. and Robert, J., 2006. An evaluation of data quality in a network for surveillance of Mycobacterium tuberculosis resistance to antituberculosis drugs in Ile-de-France region-2001–2002. *European journal of epidemiology*, 21, pp.783-785.

Gutierrez, M. C., Ahmed, N., Willery, E., Narayanan, S., Hasnain, S. E., Chauhan, D. S., Katoch, V. M., Vincent, V., Locht, C. and Supply, P., 2006. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerging infectious diseases*, 12(9), 1367

Guyeux, C., Senelle, G., Le Meur, A., Supply, P., Gaudin, C., Phelan, J. E., Clark, T. G., Rigouts, L., de Jong, B., Sola, C. and Refrégier, G., 2024. Newly Identified *Mycobacterium africanum* Lineage 10, Central Africa. *Emerging Infectious Diseases*, 30(3), 560-563. https://doi.org/10.3201/eid3003.231466

Hassan, K., 2018. The dependency of economy of Sabah on foreign workers: A study using labour force survey, 2017. DOSM://www.dosm.gov.my/v1/uploads/files/7_Publication/Research_Papers/3_%20JP%20SABAH%20-%20The%20dependency%20of%20economy%20of%20Sabah%20on%20workers.pdf

Hatherell, H.A., Colijn, C., Stagg, H.R., Jackson, C., Winter, J.R. and Abubakar, I., 2016. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC medicine*, *14*, pp.1-13.

Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W. and Gagneux, S., 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS biology*, *6*(12), p.e311.
Hussain, T., 2007. Leprosy and tuberculosis: an insight-review. *Critical reviews in microbiology*, *33*(1), pp.15-66.

Hunter, P.R. and 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), pp.2465-2466.

Hunter, P.R., 1990. Reproducibility and indices of discriminatory power of microbial typing methods. *Journal of clinical microbiology*, 28(9), pp.1903-1905.

Huyen, M.N., Kremer, K., Lan, N.T., Cobelens, F.G., Buu, T.N., Dung, N.H., Caws, M., Tiemersma, E.W. and van Soolingen, D., 2012. Mixed tuberculosis infections in rural South Vietnam. *Journal of clinical microbiology*, *50*(5), pp.1586-1592.

Huyen, M.N., Buu, T.N., Tiemersma, E., Lan, N.T., Dung, N.H., Kremer, K., Soolingen, D.V. and Cobelens, F.G., 2013. Tuberculosis relapse in Vietnam is significantly associated with Mycobacterium tuberculosis Beijing genotype infections. *The Journal of infectious diseases*, 207(10), pp.1516-1524.

Ismail, F., Couvin, D., Farakhin, I., Abdul Rahman, Z., Rastogi, N. and Suraiya, S., 2014. Study of Mycobacterium tuberculosis complex genotypic diversity in Malaysia reveals a predominance of ancestral East-African-Indian lineage with a Malaysia-specific signature. *PLoS One*, *9*(12), p.e114832.

Izumi, K., Ohkado, A., Uchimura, K., Murase, Y., Tatsumi, Y., Kayebeta, A., Watanabe, Y. and Ishikawa, N., 2015. Detection of tuberculosis infection hotspots using activity spaces based spatial approach in an urban Tokyo, from 2003 to 2011. *PloS one*, *10*(9), p.e0138831.

Izumi, K., Murase, Y., Uchimura, K., Kaebeta, A., Ishihara, K., Kaguraoka, S., Takii, T. and Ohkado, A., 2019. Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study. *BMJ open*, *9*(5), p.e029295.

Jani, J., Abu Bakar, S.F., Mustapha, Z.A., Ling, C.K., Teo, R. and Ahmed, K., 2020. Identification and Characterization of *Mycobacterium tuberculosis* Beijing Genotype Strain SBH163, Isolated in Sabah, Malaysia. *Microbiology resource announcements*, 9(2), pp.e01322-19.

Jelip, J., Mathew, G.G., Yusin, T., Dony, J.F., Singh, N., Ashaari, M., Lajanin, N., Ratnam, C.S., Ibrahim, M.Y. and Gopinath, D., 2004. Risk factors of tuberculosis among health care workers in Sabah, Malaysia. *Tuberculosis*, *84*(1-2), pp.19-23.

Jiao, W.W., Mokrousov, I., Sun, G.Z., Guo, Y.J., Vyazovaya, A., Narvskaya, O. and Shen, A.D., 2008. Evaluation of new variable-number tandem-repeat systems for typing *Mycobacterium tuberculosis* with Beijing genotype isolates from Beijing, China. *Journal of clinical microbiology*, *46*(3), pp.1045-1049.

Kam, K.M. and Yip, C.W., 2001. Surveillance of *Mycobacterium tuberculosis* drug resistance in Hong Kong, 1986-1999, after the implementation of directly observed treatment. *International Journal of Tuberculosis and Lung Disease*, 5(9), pp.815–823.

Kam, M.L.W., Lim, L.K., Sng, L.H., Ong, R.T.H., Lee, V.J., Wang, Y.T. and Chee, C.B., 2022. Country-wide genotyping of *Mycobacterium tuberculosis* complex in Singapore, 2011–2017. *Tuberculosis*, *134*, p.102204.

Kamerbeek, J., Schouls, L.E.O., Kolk, A., Van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M. and van Embden, J., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of clinical microbiology*, *35*(4), pp.907-914.

Kanduma, E., McHugh, T.D. and Gillespie, S.H., 2003. Molecular methods for Mycobacterium tuberculosis strain typing: a users guide. Journal of applied microbiology, 94(5), pp.781-791.

Kato-Maeda, M., Metcalfe, J.Z. and Flores, L., 2011. Genotyping of Mycobacterium tuberculosis: application in epidemiologic studies. Future microbiology, 6(2), pp.203-216.

Kaur, A., 2008. International migration and governance in Malaysia: Policy and performance. *UNEAC Asia Papers*, *22*, pp.4-18.

Knight, G.M., Zimic, M., Funk, S., Gilman, R.H., Friedland, J.S. and Grandjean, L., 2018. The relative fitness of drug-resistant *Mycobacterium tuberculosis*: a modelling study of household transmission in Peru. *Journal of The Royal Society Interface*, *15*(143), p.20180025.

Koh, H.L., Mohamed Naim, A.K., Noorliza, M.N. and Teh, S.Y., 2019. Tuberculosis Elimination in Malaysia by 2035: Linkages and Implications to SDGs. *International Journal of Social Science and Humanity*, 9(4).

Koleske, B.N., Jacobs, W.R. and Bishai, W.R., 2023. The *Mycobacterium tuberculosis* genome at 25 years: lessons and lingering questions. The Journal of clinical investigation, 133(19).

Kremer, K., Van Soolingen, D., Frothingham, R., Haas, W. H., Hermans, P. W. M., Martin, C., Palittapongarnpim, P., Plikaytis, B., Riley, W., Yakrus, M. A., Musser, M. and Van Embden, J. D. A., 1999. Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. *Journal of clinical microbiology*, 37(8), 2607-2618.

Krishnan, N., Malaga, W., Constant, P., Caws, M., Thi Hoang Chau, T., Salmons, J., Thi Ngoc Lan, N., Bang, N.D., Daffe, M., Young, D.B. and Robertson, B.D., 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. *PloS one*, *6*(9), p.e23870.

Kurus, B., 1998. Migrant labor: the Sabah experience. *Asian and Pacific Migration Journal*, *7*(2-3), 281-295.

Laurence, Y. V., Griffiths, U.K. and Vassall, A., 2015. Costs to Health Services and the Patient of Treating Tuberculosis: A Systematic Literature Review. *PharmacoEconomics*, 33(9), pp.939–955.

Leavitt, S.V., Horsburgh Jr, C.R., Lee, R.S., Tibbs, A.M., White, L.F. and Jenkins, H.E., 2022. What can genetic relatedness tell us about risk factors for tuberculosis transmission? *Epidemiology*, *33*(1), pp.55-64.

Le Flèche, P., Fabre, M., Denoeud, F., Koeck, J.L. and Vergnaud, G., 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC microbiology*, *2*, pp.1-12.

Legrand, E., Filliol, I., Sola, C. and Rastogi, N., 2001. Use of spoligotyping to study the evolution of the direct repeat locus by IS 6110 transpositions in *Mycobacterium tuberculosis*. *Journal of clinical microbiology*, 39(4), 1595-1599.

Levin-Reisman, I., Brauner, A., Ronin, I. and Balaban, N.Q., 2019. Epistasis between antibiotic tolerance, persistence, and resistance mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 116(29), pp.14734–14739.

Li, Q.J., Jiao, W.W., Yin, Q.Q., Li, Y.J., Li, J.Q., Xu, F., Sun, L., Xiao, J., Qi, H., Wang, T. and Mokrousov, I., 2017. Positive epistasis of major low-cost drug resistance mutations *rpoB*531-TTG and *katG*315-ACC depends on the phylogenetic background of *Mycobacterium tuberculosis* strains. *International journal of antimicrobial agents*, *49*(6), pp.757-762.

Li, S.J., Li, Y.F., Song, W.M., Zhang, Q.Y., Liu, S.Q., Xu, T.T., An, Q.Q., Liu, J.Y. and Li, H.C., 2021. Population aging and trends of pulmonary tuberculosis incidence in the elderly. *BMC Infectious Diseases*, *21*, pp.1-10.

Liew, S.M., Khoo, E.M., Ho, B.K., Lee, Y.K., Mimi, O., Fazlina, M.Y., Asmah, R., Lee, W.K., Harmy, M.Y., Chinna, K. and Jiloris, F.D., 2015. Tuberculosis in Malaysia: predictors of treatment outcomes in a national registry. *The International Journal of Tuberculosis and Lung Disease*, *19*(7), pp.764-771.

Lisdawati, V., Puspandari, N., Rif'ati, L., Soekarno, T., K, S., Ratnasari, L., Izzatun, N. and Parwati, I., 2015. Molecular epidemiology study of *Mycobacterium tuberculosis* and its susceptibility to anti-tuberculosis drugs in Indonesia. *BMC infectious diseases*, *15*, pp.1-8.

Lumb, R., Van Deun, A., Bastian, I. and Fitz-Gerald, M., 2013. Laboratory diagnosis of tuberculosis by sputum microscopy: The Handbook. SA Pathology.

Manoharan, A., Koh, W.M., K, M. and Khoo, E.M., 2023. Facilitators and barriers to latent tuberculosis infection treatment among primary healthcare workers in Malaysia: a qualitative study. *BMC health services research*, 23(1), p.914.

Marwal, A. and Gaur, R.K., 2020. Molecular markers: tool for genetic analysis. In *Animal biotechnology* (pp. 353-372). Academic Press.

Mazars, E., Lesjean, S., Banuls, A.L., Gilbert, M., Vincent, V., Gicquel, B., Tibayrenc, M., Locht, C. and Supply, P., 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. *Proceedings of the national academy of sciences*, *98*(4), pp.1901-1906.

McIvor, A., Koornhof, H. and Kana, B.D., 2017. Relapse, re-infection and mixed infections in tuberculosis disease. *Pathogens and disease*, 75(3).

Meehan, C.J., Moris, P., Kohl, T.A., Pečerska, J., Akter, S., Merker, M., Utpatel, C., Beckert, P., Gehre, F., Lempens, P. and Stadler, T., 2018. The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. *EBioMedicine*, *37*, pp.410-416.

Merker, M., Feuerriegel, S., Cox, H., Borrell, S., Gagneux, S., Rüsch-Gerdes, S., Kohl, T.A. and Niemann, S., 2015. Understanding *Mycobacterium tuberculosis* evolution in times of antibiotic treatment–A whole genome approach. *International Journal of Mycobacteriology*, *4*, p.2.

Merker, M., Kohl, T.A., Niemann, S. and Supply, P., 2017. The evolution of strain typing in the *Mycobacterium tuberculosis* complex. In: *Advances in Experimental Medicine and Biology*. Springer New York LLC, pp. 43–78.

Merker, M., Kohl, T.A., Barilar, I., Andres, S., Fowler, P.W., Chryssanthou, E., Ängeby, K., Jureen, P., Moradigaravand, D., Parkhill, J. and Peacock, S.J., 2020. Phylogenetically informative mutations in genes implicated in antibiotic resistance in Mycobacterium tuberculosis complex. *Genome Medicine*, *12*, pp.1-8.

Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., Jackson, C., Rauzier, J., Bifani, P., Warren, R. and Rasolofo, V., 2011. Phylogeny of Mycobacterium tuberculosis Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. *PloS one*, *6*(1), p.e16020.

Ministry of Health Malaysia (MOH), 2002. Manual Sistem Maklumat Tibi Kebangsaan. In National Tuberculosis Information System [TBIS] Manual; National Information System Development Committee Dcd, Ed.; Ministry of Health Malaysia: Putrajaya, Malaysia.

Ministry of Health Malaysia (MOH), 2012. *Clinical Practice Guidelines: Management of Tuberculosis* 3rd Ed. Abdul Rahaman, J.A., (ed.), Bahagian Perkembangan Perubatan, Malaysia Health Technology Assessment Section (MaHTAS), Putrajaya.

Ministry of Health Malaysia (MOH), 2016. *Clinical Practice Guidelines: Management of Drug Resistant Tuberculosis* 1st Ed. Sivaraman Kannan, K.K., (ed.), TB/Leprosy Sector, Disease Control Division, Putrajaya.

Ministry of Health Malaysia (MOH), 2021. *Clinical Practice Guidelines: Management of Tuberculosis* 4th Ed. Goon, A.K., (ed.), Bahagian Perkembangan Perubatan, Malaysia Health Technology Assessment Section (MaHTAS), Putrajaya.

Mokrousov, I., Narvskaya, O., Limeschenko, E., Vyazovaya, A., Otten, T. and Vyshnevskiy, B., 2004. Analysis of the allelic diversity of the mycobacterial interspersed repetitive units in Mycobacterium tuberculosis strains of the Beijing family: practical implications and evolutionary considerations. *Journal of clinical microbiology*, *42*(6), pp.2438-2444.

Mokrousov, I., Ly, H.M., Otten, T., Lan, N.N., Vyshnevskyi, B., Hoffner, S. and Narvskaya, O., 2005. Origin and primary dispersal of the Mycobacterium tuberculosis Beijing genotype: clues from human phylogeography. *Genome Research*, *15*(10), pp.1357-1364.

Mokrousov, I., Narvskaya, O., Vyazovaya, A., Otten, T., Jiao, W. W., Gomes, L. L., Suffys, P.N., Shen, A.D. and Vishnevsky, B., 2012. Russian "successful" clone B0/W148 of *Mycobacterium tuberculosis* Beijing genotype: a multiplex PCR assay for rapid detection and global screening. *Journal of Clinical Microbiology*, 50(11), 3757-3759.

Mokrousov, I., Vyazovaya, A., Iwamoto, T., Skiba, Y., Pole, I., Zhdanova, S., Arikawa, K., Sinkov, V., Umpeleva, T., Valcheva, V., Alvarex-Figueroa, M., Ranka, R., Jansone, I., Ogarkov, O. and Narvskaya, O., 2016. Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: Human traces across pathogen's phylogeography. *Molecular phylogenetics and evolution*, 99, 133-143.

Mokrousov, I., 2021. Ubiquitous and multifaceted: SIT53 spoligotype does not correlate with any particular family of *Mycobacterium tuberculosis. Tuberculosis*, 126, 102024.

Mokrousov, I., Vyazovaya, A., Shitikov, E., Badleeva, M., Belopolskaya, O., Bespiatykh, D., Gerasimova, A., Ioannidis, P., Jiao, W., Khromova, P. and Masharsky, A., 2023. Insight into pathogenomics and phylogeography of hypervirulent and highly-lethal *Mycobacterium tuberculosis* strain cluster. *BMC Infectious Diseases*, *23*(1), p.426.

Molina-Moya, B., Gomgnimbou, M.K., Lafoz, C., Lacoma, A., Prat, C., Refregier, G., Samper, S., Dominguez, J. and Sola, C., 2017. Molecular characterization of *Mycobacterium tuberculosis* strains with TB-SPRINT. *The American journal of tropical medicine and hygiene*, *97*(3), p.806.

Montoya, J.C., Malabad, J.C.M., Ang, C.F., Reyes, L.T., Basilio, R.P., Lim, D.R., Amarillo, M.L.E., Ama, M.C.G., Phelan, J.E., Hibberd, M.L. and Clark, T.G., 2022. Molecular characterization of drug-resistant *Mycobacterium tuberculosis* among Filipino patients derived from the national tuberculosis prevalence survey Philippines 2016. *Tuberculosis*, *135*, p.102211.

Munang, M.L., Browne, C., Evans, J.T., Smith, E.G., Hawkey, P.M., Welch, S.B., Kaur, H. and Dedicoat, M.J., 2016. Programmatic utility of tuberculosis cluster investigation using a social network approach in Birmingham, United Kingdom. *The International Journal of Tuberculosis and Lung Disease*, *20*(10), pp.1300-1305.

Munang, M.L., 2019. *Persistence of tuberculosis inferred from case and contact networks in Birmingham, UK*. PhD thesis, University of Warwick, United Kingdom.

Musser, J. M., Kapur, V., Williams, D. L., Kreiswirth, B. N., Van Soolingen, D. and Van Embden, J. D., 1996. Characterization of the catalase-peroxidase gene (*katG*) and inhA locus in isoniazid-resistant and-susceptible strains of *Mycobacterium tuberculosis* by automated DNA sequencing: restricted array of mutations associated with drug resistance. *The Journal of infectious diseases*, 173(1), 196-202.

Muwonge, A., Kankya, C., Olea-Popelka, F., Biffa, D., Ssengooba, W., Berit, D., Skjerve, E.and Johansen, T. B., 2013. Molecular investigation of multiple strain infections in patients with tuberculosis in Mubende district, Uganda. *Infection, Genetics and Evolution*, *17*, pp.16-22.

Napier, G., Couvin, D., Refrégier, G., Guyeux, C., Meehan, C.J., Sola, C., Campino, S., Phelan, J. and Clark, T.G., 2023. Comparison of in silico predicted *Mycobacterium tuberculosis* spoligotypes and lineages from whole genome sequencing data. *Scientific Reports*, *13*(1), p.11368.

Narmandakh, E., Tumenbayar, O., Borolzoi, T., Erkhembayar, B., Boldoo, T., Dambaa, N., Burneebaatar, B., Nymadawa, N., Mitarai, S., Jav, S. and Chiang, C.Y., 2020. Genetic mutations associated with isoniazid resistance in *Mycobacterium tuberculosis* in Mongolia. *Antimicrobial Agents and Chemotherapy*, *64*(7), pp.10-1128.

Navarro, Y., Herranz, M., Pérez-Lago, L., Martinez Lirola, M., Ruiz-Serrano, M.J., Bouza, E. and Garcia de Viedma, D., 2011. Systematic survey of clonal complexity in tuberculosis at a populational level and detailed characterization of the isolates involved. *Journal of clinical microbiology*, *49*(12), pp.4131-4137.

Navarro, Y., Perez-Lago, L., Sislema, F., Herranz, M., de Juan, L., Bouza, E. and Garcia-de-Viedma, D., 2013. Unmasking subtle differences in the infectivity of microevolved *Mycobacterium tuberculosis* variants coinfecting the same patient. *International Journal of Medical Microbiology*, *303*(8), pp.693-696.

Navin, T., Wallace, C., Etkind, S. and Rosenblum, L., 2004. Guide to the application of genotyping to tuberculosis prevention and control; handbook for TB controllers, epidemiologists, laboratorians, and other program staff.

Ngeow, Y.F., Cheong, S.F., Yap, S.F. and Kuppusamy, I., 2006, Tuberculosis in Malaysia: Population Structure and Genetic Diversity among Malaysian Isolates. In: *Molecular Epidemiology and Population Genetics of Tuberculosis*. YF Ngeow, SF Yap (eds). Academy of Sciences Malaysia.

Nicol, M.P. and Wilkinson, R.J., 2008. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(10), pp.955–965.

Niemann, S., Merker, M., Kohl, T. and Supply, P., 2016. Impact of Genetic Diversity on the Biology of *Mycobacterium tuberculosis* Complex Strains., pp.1–18.

Palittapongarnpim, P., Ajawatanawong, P., Viratyosin, W., Smittipat, N., Disratthakit, A., Mahasirimongkol, S., Yanai, H., Yamada, N., Nedsuwan, S., Imasanguan, W. and Kantipong, P., 2018. Evidence for host-bacterial co-evolution via genome sequence analysis of 480 Thai *Mycobacterium tuberculosis* lineage 1 isolates. *Scientific reports*, 8(1), p.11597.

Palomino, J.C. and Martin, A., 2014. Drug resistance mechanisms in *Mycobacterium tuberculosis*. *Antibiotics*, 3(3), pp.317–340.

Parwati, I., Van Crevel, R., Sudiro, M., Alisjahbana, B., Pakasi, T., Kremer, K., Van Der Zanden, A. and van Soolingen, D., 2008. *Mycobacterium tuberculosis* population structures differ significantly on two Indonesian Islands. *Journal of Clinical Microbiology*, *46*(11), pp.3639-3645.

Parwati, I., van Crevel, R. and van Soolingen, D., 2010. Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. The Lancet infectious diseases, 10(2), pp.103-111.

Paul, D.C., 2019. TB Laboratory Test database 2015-2018, Kota Kinabalu Public Health Laboratory, Sabah State Health Department (Unpublished Data).

Paul, D.C., Ngeow, Y.F., Yap, S.F., Dony, J.F., Avoi, R., Mohammad, R. and Ng, H.F., 2022. Concentrated specimen smear microscopy utilising a polymer membrane sandwich filtration vessel for the detection of acid-fast bacilli in health facilities in Sabah, East Malaysia. *Tuberculosis*, *133*, p.102183.

Paul, D.C., 2022. 'Concentrated sputum smear microscopy project – Supplementary Data', Universiti Tunku Abdul Rahman & Kota Kinabalu Public Health Laboratory, Sabah State Health Department (Unpublished Data).

Pérez-Lago, L., Lirola, M.M., Navarro, Y., Herranz, M., Ruiz-Serrano, M.J., Bouza, E. and García-de-Viedma, D., 2015. Co-infection with drug-susceptible and reactivated latent multidrug-resistant *Mycobacterium tuberculosis*. *Emerging Infectious Diseases*, *21*(11), p.2098.

Peters, D., 2022. Growth and Distribution of Non-Citizen Population. In: Tey, N.P., Lai, S.L., Chan, J.K.L. (eds). *Demographic and Socioeconomic Changes in Sabah*. Kota Kinabalu: Universiti Malaysia Sabah Press. pp 353-372.

Phelan, J.E., Lim, D.R., Mitarai, S., de Sessions, P.F., Tujan, M.A.A., Reyes, L.T., Medado, I.A.P., Palparan, A.G., Naim, A.N.M., Jie, S. and Segubre-Mercado, E., 2019. *Mycobacterium tuberculosis* whole genome sequencing provides insights into the Manila strain and drug-resistance mutations in the Philippines. *Scientific reports*, *9*(1), p.9305.

Poblete, J., Montoya, J., Ang, C., Rivera, P., Kato, S., Keicho, N., Maeda, S., Miyabayashi, A., Lopez, A. L., Aldaba, J., Sylimo, P., Geraldino, X., Salonga, A. and Ama, C., 2021. Genetic Diversity of Mycobacterium tuberculosis Isolates from a Rural Community in the Philippines. *Philippine Journal of Science*, 150(4).

Portevin, D., Gagneux, S., Comas, I. and Young, D., 2011. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathogens*, 7(3).

Procop, G. W., Church, D. L., Hall, G. S., & Janda, W. M., 2020. Mycobacteria. In *Koneman's color atlas and textbook of diagnostic microbiology*, 7th ed. Jones & Bartlett Learning, pp. 1219-1268

Ramaswamy, S. and Musser, J.M., 1998. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tubercle and Lung disease*, *79*(1), pp.3-29.

Rastogi, N. and Sola, C., 2007. Molecular Evolution of the *Mycobacterium tuberculosis* Complex. *In: Tuberculosis 2007: From Basic Science to Patient Care.*, (April 2017), p.Palomino JC, Leão SC, Ritacco V. (eds). p53-91. Available at: http://www.tuberculosistextbook.com/.

Rashid Ali, M.R.S., Parameswaran, U., William, T., Bird, E., Wilkes, C.S., Lee, W.K., Yeo, T.W., Anstey, N.M. and Ralph, A.P., 2015. A prospective study of tuberculosis drug susceptibility in Sabah, Malaysia, and an algorithm for management of isoniazid resistance. *Journal of tropical medicine*, *2015*.

Reed, M.B., Pichler, V.K., McIntosh, F., Mattia, A., Fallow, A., Masala, S., Domenech, P., Zwerling, A., Thibert, L., Menzies, D. and Schwartzman, K., 2009. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *Journal of clinical microbiology*, *47*(4), pp.1119-1128.

Reiling, N., Homolka, S., Walter, K., Brandenburg, J., Niwinski, L., Ernst, M., Herzmann, C., Lange, C., Diel, R., Ehlers, S. and Niemann, S., 2013. Clade-specific virulence patterns of *Mycobacterium tuberculosis* complex strains in human primary macrophages and aerogenically infected mice. *MBio*, *4*(4), pp.10-1128.

Richards, N.M., 2012. The dangers of surveillance. Harv. L. Rev., 126, p.1934.

Robsky, K.O., Kitonsa, P.J., Mukiibi, J., Nakasolya, O., Isooba, D., Nalutaaya, A., Salvatore, P.P., Kendall, E.A., Katamba, A. and Dowdy, D., 2020. Spatial distribution of people diagnosed with tuberculosis through routine and active case finding: a community-based study in Kampala, Uganda. *Infectious diseases of poverty*, *9*, pp.1-10.

Rodwell, T.C., Kapasi, A.J., Barnes, R.F.W. and Moser, K.S., 2012. Factors associated with genotype clustering of *Mycobacterium tuberculosis* isolates in an ethnically diverse region of southern California, United States. *Infection, Genetics and Evolution*, 12(8), pp.1917–1925.

Rosli, M.K.T, 2022. *Sabah's handbook on management of tuberculosis in adults by Respiratory Department, QEH & TB and Leprosy Sector, JKN Sabah* 1st ed. Rosli, M.K., Sivaraman Kannan, K.K., Ramarmuty, H.Y. and Nai Chien, H., (eds.), Sabah State Health Department, Kota Kinabalu.

Rozo-Anaya, J.C. and Ribón, W., 2010. Molecular tools for Mycobacterium tuberculosis genotyping. Revista de Salud Pública, 12(3), pp.510-521.

Rundi, C., 2010. Understanding tuberculosis: Perspectives and experiences of the people of Sabah, East Malaysia. *Journal of Health, Population and Nutrition*, 28(2), pp.114–123.

Rundi, C., Fielding, K., Godfrey-Faussett, P., Rodrigues, L.C. and Mangtani, P., 2011. Delays in seeking treatment for symptomatic tuberculosis in Sabah, East Malaysia: factors for patient delay. *The International journal of tuberculosis and lung disease*, *15*(9), pp.1231-1238.

Sasmono, R.T., Massi, M.N., Setianingsih, T.Y., Wahyuni, S., Halik, H., Yusuf, I., Dick, T., Sola, C., Bifani, P.J. and Phyu, S., 2012. Heterogeneity of *Mycobacterium tuberculosis* strains in Makassar, Indonesia. *The International journal of tuberculosis and lung disease*, *16*(11), pp.1441-1448.

Sebban, M., Mokrousov, I., Rastogi, N., and Sola, C., 2002. A data-mining approach to spacer oligonucleotide typing of *Mycobacterium tuberculosis. Bioinformatics*, 18(2), 235-243.

Shakespeare, W., 2015. Romeo and Juliet (Act 2, Scene 2, line 43). London: Penguin Classics.

Shanmugam, S., Selvakumar, N. and Narayanan, S., 2011. Drug resistance among different genotypes of Mycobacterium tuberculosis isolated from patients from Tiruvallur, South India. *Infection, Genetics and Evolution*, 11(5), pp.980-986.

Shao, Y., Yang, D., Xu, W., Lu, W., Song, H., Dai, Y., Shen, H. and Wang, J., 2011. Epidemiology of anti-tuberculosis drug resistance in a Chinese population: current situation and challenges ahead. *BMC public health*, *11*, pp.1-10.

Shea, J., Halse, T. A., Lapierre, P., Shudt, M., Kohlerschmidt, D., Van Roey, P., Limberger, R., Taylor, J., Escuyer, V. and Musser, K. A., 2017. Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York State. *Journal of clinical microbiology*, 55(6), 1871-1882.

Shen, X., Yang, C., Wu, J., Lin, S., Gao, X., Wu, Z., Tian, J. Y., Gan, M. Y., Luo, T., Wang, L. L., Yu, C. L., Mei, J., Pan, Q. C., DeRiemer, K., Yuan, Z.A. and Gao, Q., 2017. Recurrent tuberculosis in an urban area in China: relapse or exogenous reinfection? *Tuberculosis*, 103, 97-104.

Shi, J., Zheng, D., Zhu, Y., Ma, X., Wang, S., Li, H. and Xing, J., 2018. Role of MIRU-VNTR and spoligotyping in assessing the genetic diversity of *Mycobacterium tuberculosis* in Henan Province, China. *BMC infectious diseases*, *18*, pp.1-12.

Shin, S. S., Modongo, C., Baik, Y., Allender, C., Lemmer, D., Colman, R. E., Engelthaler, D.M, Warren, R.M. and Zetola, N. M., 2018. Mixed *Mycobacterium tuberculosis*–strain infections are associated with poor treatment outcomes among patients with newly diagnosed tuberculosis, independent of pretreatment heteroresistance. *The Journal of infectious diseases*, 218(12), 1974-1982.

Shorten, R.J., 2011. *The molecular epidemiology of Mycobacterium tuberculosis in north London*. PhD thesis, University College London, United Kingdom.

Simonsen, L., Gog, J.R., Olson, D. and Viboud, C., 2016. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. The Journal of infectious diseases, 214(suppl_4), pp.S380-S385.

Skiba, Y., Mokrousov, I., Ismagulova, G., Maltseva, E., Yurkevich, N., Bismilda, V., Chingissova, L., Abildaev, T. and Aitkhozhina, N., 2015. Molecular snapshot of *Mycobacterium tuberculosis* population in Kazakhstan: a country-wide study. *Tuberculosis*, *95*(5), pp.538-546.

Slayden, R. A., and Barry III, C. E., 2000. The genetics and biochemistry of isoniazid resistance in *Mycobacterium tuberculosis*. *Microbes and Infection*, 2(6), 659-669.

Sola, C., Horgen, L., Goh, K. S., and Rastogi, N., 1997. Molecular fingerprinting of *Mycobacterium tuberculosis* on a Caribbean Island with IS6110 and DRr probes. *Journal of clinical microbiology*, 35(4), 843-846.

Sola, C., Filliol, I., Legrand, E., Mokrousov, I., and Rastogi, N., 2001. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS 1081, IS 6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. *Journal of molecular evolution*, 53, 680-689.

Sola, C., Gomgnimbou, M. and Gomes, H.M., 2016. User's Manual TB-SPRINT (59-Plex) spoligotyping TB-SPOL (43-Plex) and RIF-INH typing TB-RINT (16-Plex)., pp.1–10.

Somphavong, S., Berland, J.L., Gauthier, M., Vu, T.T., Nguyen, Q.H., Iem, V., Vongvichit, P., Inthavong, D., Akkhavong, V., Chanthavilay, P. and Soundala, S., 2019. First insights into the genetic characteristics and drug resistance of *Mycobacterium tuberculosis* population collected during the first national tuberculosis prevalence survey of Lao PDR (2010–2011). *BMC Infectious Diseases*, *19*, pp.1-10.

Sougakoff, W., 2011. Molecular epidemiology of multidrug-resistant strains of *Mycobacterium tuberculosis*. *Clinical Microbiology and Infection*, *17*(6), pp.800-805.

Sreevatsan, S., Pan, X.I., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S. and Musser, J.M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences*, *94*(18), pp.9869-9874.

Stavrum, R., Mphahlele, M., Øvreås, K., Muthivhi, T., Fourie, P. B., Weyer, K., and 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.

Stinson, K.W., Eisenach, K., Kayes, S., Matsumoto, M., Siddiqi, S. and Nakashima, S., 2014. Mycobacteriology Laboratory Manual, global laboratory initiative advancing TB diagnosis. *Global Laboratory Initiative of Stop TB Partnership*.

Stucki, D., 2015. *Transmission and evolution of" Mycobacterium tuberculosis" studied by whole genome sequencing and single nucleotide polymorphism-typing*. PhD thesis, University of Basel, Switzerland.

Stucki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaihwa, L., Borrell, S., Luo, T. and Gao, Q., 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature genetics*, *48*(12), pp.1535-1543.

Sun, Y.J., Bellamy, R., Lee, A.S., Ng, S.T., Ravindran, S., Wong, S.Y., Locht, C., Supply, P. and Paton, N.I., 2004. Use of mycobacterial interspersed repetitive unit-variable-number tandem repeat typing to examine genetic diversity of *Mycobacterium tuberculosis* in Singapore. *Journal of clinical microbiology*, *42*(5), pp.1986-1993.

Supply, P., Mazars, E., Lesjean, S., Vincent, V., Gicquel, B. and Locht, C., 2000. Variable human minisatellite-like regions in the Mycobacterium *tuberculosis* genome. *Molecular microbiology*, 36(3), 762-771.

Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D. and 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), pp.3563-3571.

Supply, P., 2005. Multilocus variable number tandem repeat genotyping of *Mycobacterium tuberculosis*. *Technical Guide*, *6*, pp.1-74.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., Savine, E., De Haas, P., Van Deutekom, H., Roring, S. and Bifani, P., 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *Journal of clinical microbiology*, *44*(12), pp.4498-4510.

Theron, G., Jenkins, H.E., Cobelens, F., Abubakar, I., Khan, A.J., Cohen, T. and Dowdy, D.W., 2015. Data for action: collection and use of local data to end tuberculosis. *The Lancet*, *386*(10010), pp.2324-2333.

Thomas, S.K., Iravatham, C.C., Moni, B.H., Kumar, A., Archana, B.V., Majid, M., Priyadarshini, Y., Rani, P.S., Valluri, V., Hasnain, S.E. and Ahmed, N., 2011. Modern and ancestral genotypes of Mycobacterium tuberculosis from Andhra Pradesh, India. *PloS one*, 6(11), p.e27584.

Torrea, G., Levee, G., Grimont, P., Martin, C., Chanteau, S. and Gicquel, B., 1995. Chromosomal DNA fingerprinting analysis using the insertion sequence IS6110 and the repetitive element DR as strain-specific markers for epidemiological study of tuberculosis in French Polynesia. *Journal of Clinical Microbiology*, 33(7), 1899-1904.

Uplekar, M., Weil, D., Lonnroth, K., Jaramillo, E., Lienhardt, C., Dias, H. M.,Falzon, D., Floyd, K., Gargioni, G., Getahun, H., Gilpin, C., Glaziou, P.,GRzemska, M., Mirzayev, F., Nakatani, H. and Raviglione, M., 2015. WHO's new end TB strategy. *The Lancet*, 385(9979), 1799-1801.

Violeta, V., Mokrousov, I., Narvskaya, O., Rastogi, N. and Markova, N., 2009. Genetic Diversity of *Mycobacterium tuberculosis* Population in Bulgaria. Genetic Diversity, 69-104.

Van Deutekom, H., Supply, P., de Haas, P.E., Willery, E., Hoijng, S.P., Locht, C., Coutinho, R.A. and van Soolingen, D., 2005. Molecular typing of *Mycobacterium tuberculosis* by mycobacterial interspersed repetitive unit-variable-number tandem repeat analysis, a more accurate method for identifying epidemiological links between patients with tuberculosis. *Journal of clinical microbiology*, *43*(9), pp.4473-4479.

Van Doorn, G. S. and Dieckmann, U., 2006. The long-term evolution of multilocus traits under frequency-dependent disruptive selection. *Evolution*, 60(11), 2226-2238.

Van Rie, A., Victor, T.C., Richardson, M., Johnson, R., van der Spuy, G.D., Murray, E.J., Beyers, N., van Pittius, N.C.G., van Helden, P.D. and Warren, R.M., 2005. Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns. *American journal of respiratory and critical care medicine*, *172*(5), pp.636-642.

Vassall, A., 2016. *Analyzing the benefit-cost ratio of interventions to improve the outcomes for those with drug-susceptible and multi-drug resistant Tuberculosis (MDR-TB) Bangladesh Priorities*

Veen, J., 1992. Microepidemics of tuberculosis: the stone-in-the-pond principle. *Tubercle and lung disease*, *73*(2), pp.73-76.

Vitol, I., Driscoll, J., Kreiswirth, B., Kurepina, N. and Bennett, K.P., 2006. Identifying Mycobacterium tuberculosis complex strain families using spoligotypes. *Infection, Genetics and Evolution*, 6(6), pp.491-504.

Vynnycky, E., Borgdorff, M. W., van Soolingen, D. and Fine, P. E., 2003. Annual Mycobacterium tuberculosis infection risk and interpretation of clustering statistics. Emerging infectious diseases, 9(2), 176-183.

Wada, T., Maeda, S., Hase, A. and Kobayashi, K., 2007. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan., pp.1052–1057.

Walker, T.M., Kohl, T.A., Omar, S.V., Hedge, J., Elias, C.D.O., Bradley, P., Iqbal, Z., Feuerriegel, S., Niehaus, K.E., Wilson, D.J. and Clifton, D.A., 2015. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *The Lancet infectious diseases*, 15(10), pp.1193-1202.

Walker, T.M., Merker, M., Knoblauch, A.M., Helbling, P., Schoch, O.D., Van Der Werf, M.J., Kranzer, K., Fiebig, L., Kröger, S., Haas, W. and Hoffmann, H., 2018. A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *The Lancet Infectious Diseases*, *18*(4), pp.431-440.

Walker, T.M., Miotto, P., Köser, C.U., Fowler, P.W., Knaggs, J., Iqbal, Z., Hunt, M., Chindelevitch, L., Farhat, M.R., Cirillo, D.M. and Comas, I., 2022. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *The Lancet Microbe*, *3*(4), pp.e265-e273.

Weniger, T., Krawczyk, J., Supply, P., Niemann, S. and Harmsen, D., 2010. MIRU-VNTR plus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. *Nucleic acids research*, *38*(suppl_2), pp.W326-W331.

William, T., Parameswaran, U., Lee, W.K., Yeo, T.W., Anstey, N.M. and Ralph, A.P., 2015. Pulmonary tuberculosis in outpatients in Sabah, Malaysia: advanced disease but low incidence of HIV co-infection. *BMC infectious diseases*, *15*, pp.1-9.

Witney, A. A., Cosgrove, C. A., Arnold, A., Hinds, J., Stoker, N. G., and Butcher, P. D., 2016. Clinical use of whole genome sequencing for *Mycobacterium tuberculosis*. *BMC medicine*, 14, 1-7.

Wong, D. K. T., 2005. Historical Sabah: The Chinese. Kota Kinabalu: Natural History Publications (Borneo) SB.

World Health Organization (WHO), 2015. Systematic screening for active tuberculosis: an operational guide. Geneva: World Health Organization

World Health Organization (WHO), 2018. Roadmap towards ending TB in children and adolescents. Geneva: World Health Organization

World Health Organization (WHO), 2020. *Global tuberculosis report 2020.*, Geneva: World Health Organization

World Health Organization (WHO), 2021. *Meeting report of the WHO expert consultation on the definition of extensively drug-resistant tuberculosis, 27-29 October 2020*. Geneva: World Health Organization

World Health Organization (WHO), 2023. *Global tuberculosis report 2023.*, Geneva: World Health Organization

Wyllie, D.H., Davidson, J.A., Smith, E.G., Rathod, P., Crook, D.W., Peto, T.E., Robinson, E., Walker, T. and Campbell, C., 2018. A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying *Mycobacterium tuberculosis* transmission: a prospective observational cohort study. *EBioMedicine*, *34*, pp.122-130.

Yasmin, M., Le Moullec, S., Siddiqui, R.T., De Beer, J., Sola, C. and Refrégier, G., 2016. Quick and cheap MIRU-VNTR typing of *Mycobacterium tuberculosis* species complex using duplex PCR. *Tuberculosis*, *101*, pp.160-163.

Yin, Q.Q., Liu, H.C., Jiao, W.W., Li, Q.J., Han, R., Tian, J.L., Liu, Z.G., Zhao, X.Q., Li, Y.J., Wan, K.L. and Shen, A.D., 2016. Evolutionary history and ongoing transmission of phylogenetic sublineages of *Mycobacterium tuberculosis* Beijing genotype in China. *Scientific reports*, *6*(1), p.34353.

Zachariah, R., Harries, A. D., Srinath, S., Ram, S., Viney, K., Singogo, E., Lal, P., Mendoza-Ticona, A., Sreenivas, A., Aung, N.W., Sharath, B.N., Kanyerere, H., van Soelen, N., Kirui, N., Ali, E., Hinderaker, S.G., Bissell, K., Enarson, D.A. and Edginton, M. E., 2012. Language in tuberculosis services: can we change to patient-centred terminology and stop the paradigm of blaming the patients? [Perspectives]. *The International journal of tuberculosis and lung disease*, 16(6), 714-717.

Zamri, H. F., Ruzan, I. N., Ramli, S. R., & Ahmad, N., 2022. Predominance of the East-Asian Beijing genotype in a *Mycobacterium tuberculosis* drug-resistant population in Central Malaysia. *Journal of Global Antimicrobial Resistance*, *30*, 302-307.

Zhang, J., Abadia, E., Refregier, G., Tafaj, S., Boschiroli, M.L., Guillard, B., Andremont, A., Ruimy, R. and Sola, C., 2010. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping'with new spacers and a microbead-based hybridization assay. *Journal of medical microbiology*, *59*(3), pp.285-294.

Zhang, Y. and Yew, W.W., 2015. Mechanisms of drug resistance in *Mycobacterium tuberculosis*: Update 2015. *International Journal of Tuberculosis and Lung Disease*, 19(11), pp.1276–1289.

Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H. and Nerlich, A.G., 2004. Molecular identification and characterization of *Mycobacterium tuberculosis* complex in ancient Egyptian mummies. *International Journal of Osteoarchaeology*, *14*(5), pp.404-413.

Zwerling, A., White, R.G., Vassall, A., Cohen, T., Dowdy, D.W. and Houben, R.M., 2014. Modeling of novel diagnostic strategies for active tuberculosis–a systematic review: current practices and recommendations. *PLoS One*, *9*(10), p.e110558.

APPENDICES

Appendix A

List of reagents and consumables for Spoligotyping and 24-loci MIRU-VNTR

Table : List of Reagents and Consumables for Spoligotyping & 24-Loci-MIRU VNTR

No	Description	Category	Unit Price (RM)	Quantity	Total Price	DNA Extraction	TB-SPRINT		MIRU-VNTR	
							PCR TB-SPRINT	Hybridization TB-SPRINT	PCR MIRU-VNTR	Electro-phoresis MIRU-VNTR
1	Micropipette 0.5-10uL	Apparatus	778.00	2	1,556.00		x	x	x	x
2	Micropipette 10-100uL	Apparatus	778.00	2	1,556.00	x	x	x	x	x
3	Multichannel Micropipette 0.5-10uL (8 Channel)	Apparatus	3,300.00	1	3,300.00					x
4	10X TBE Buffer	Chemical	223.00	2	446.00					x
5	0.5M EDTA, pH8.0	Chemical	221.00	1	221.00			x		
6	1M MgSO4	Chemical	105.00	1	105.00				x	
7	1M Tris-HCl, pH8.0	Chemical	305.00	1	305.00			x		
8	20% Sarkosyl	Chemical	203.00	1	203.00			x		
9	5M TMAC	Chemical	344.00	1	344.00			x		
10	Beta-mercaptoethanol (Molecular Grade)	Chemical	104.00	1	104.00	x				
11	Nucleic Acid Staining Solution	Chemical	269.00	8	2,152.00					x
12	Streptavidin Phycoerythrin Conjugate (SAPE)	Chemical	375.00	1	375.00			x		
13	Triton 100X	Chemical	74.00	1	74.00				x	
14	0.2 mL PCR tubes	Cons.	82.00	9	738.00		x	x	x	
15	0.2mL PCR 8-strip tubes	Cons.	2,363.00	1	2,363.00					x
16	10µL non-filtered tips	Cons.	500.00	2	1,000.00	x	x	x	x	x
17	100µL non-filtered tips	Cons.	450.00	2	900.00	x	x	x	x	x
18	1000µL non-filtered tips	Cons.	858.00	2	1,716.00	x	x	x	x	
19	PCR rack	Cons.	158.00	1	158.00		x	x	x	
20	Sealing tape (for 96-well PCR plate)	Cons.	700.00	1	700.00			x		
21	Sterile 10µL filtered tips	Cons.	1,268.00	4	5,072.00		x	x	x	x
22	Sterile 100µL non-filtered tips	Cons.	36.00	11	396.00	x		x		x
23	Sterile 1000uL filtered tips	Cons.	174.00	15	2,610.00	x				
24	Sterile disposable loop, 10µL	Cons.	145.00	3	435.00	x				
25	Sterile Microcentrifuge Tube 1.5mL	Cons.	28.00	9	252.00	x		x		x
26	Thermowell 96-well PCR plate (polycarbonate)	Cons.	950.00	2	1,900.00			x		
27	Agarose Powder	Media	1,377.00	3	4,131.00					x
28	Disposable Isolation Gown	PPE	125.00	1	125.00	x	x	x	x	x
29	Disposable latex gloves, non-powder	PPE	145.00	1	145.00	x	x	x	x	x
30	100bp DNA Ladder	Reagent	965.00	2	1,930.00					x
31	20bp DNA Ladder	Reagent	534.00	2	1,068.00					x
32	50bp DNA Ladder	Reagent	407.00	2	814.00					x
33	Beamedex TB SPRINT Kit (Cat. No. BM-TB SPRINT 50-500) x2	Reagent	128,700.00	1	128,700.00			x		
34	dNTP	Reagent	218.00	3	654.00				x	
35	Fungal/Bacterial DNA MiniPrep Extraction Kit	Reagent	1,078.00	22	23,716.00	x		x		x
36	PCR Master Mix (Taq)	Reagent	1,646.00	2	3,292.00		x			
37	Taq Polymerase (with dye)	Reagent	310.00	13	4,030.00				x	
TOTAL					197,586.00					

Appendix B

List of items needed for TB SPRINT

Source: Sola et al. (2016)

Microspheres coupled with specific probes (59-Plex for TB-SPRINT, in two mixes 43-Plex (TB-SPOL)16-Plex (TB-RINT). These reagents are provided with Beamedex® batch number sheet and quality control data to guarantee their quality

1. TB-SPOL primer mix (Fw, Rv) 20 µM**
2. TB-RINT primer mix (Rw, Rv) 17µM**
3. Streptavidin-phycoerythrin solution (Roche or Interchim, 1 mg/ml)
4. dNTP mix 2mM
5. **TB-SPOL Positive controls:** H37Rv DNA (positive control 1) M. bovis BCG P3 DNA (positive control 2)
6. **TB-RINT Positive controls:** MDR1 (*rpoB* 526TAC, *inhA* -15T) and MDR2 (*rpoB* 516GTC Mutation, *katG* 315ACC)

Others not provided by Beamedex Fr.:

- Thermal cycler (ex: Biorad DNA engine) and *Taq* Polymerase, and Taq buffer, MgCl2 to perform a PCR (*e.g*., Promega GoTaq)
- Luminex MagPix® (magnetic microspheres only) device with computer and Exponent® or BioPlex & Manager® softwares
- Disposable gloves
- Thermowell 96 Plate Model Polycarbonate (e.g.: Costar, 6509, Corning International, NY14831)
- Adjustable pipettes for 10, 20, 200 and 1000 µL
- Disposable sterile pipette tips (with filter preferably)
- DNA extraction reagents
- PCR tubes (DNAse and RNAse free)
- Ultrasonic bath (ex: Branson 200)
- Centrifuge for 96-well Plates (e.g., VWR Megastar 1.6R, with appropriate rotor Thermoscientific ref. 75003624 and
- buckets ref. 50115385)
- Vortex

Appendix C

Correspondence of spacers with the microsphere sets for TB-SPRINT

Spacer	Oligonucleotide sequence (5' to 3')	Microsphere set
sp1	ATAGAGGGTCGCCGGTTCTGGATCA	MC10013
sp2	CCTCATAATTGGGCGACAGCTTTTG	MC10014
sp3	CCGTGCTTCCAGTGATCGCCTTCTA	MC10015
sp4	ACGTCATACGCCGACCAATCATCAG	MC10018
sp5	TTTTCTGACCACTTGTGCGGGATTA	MC10019
sp6	CGTCGTCATTTCCGGCTTCAATTTC	MC10020
sp7	GAGGAGAGCGAGTACTCGGGGCTGC	MC10021
sp8	CGTGAAACCGCCCCCAGCCTCGCCG	MC10022
sp9	ACTCGGAATCCCATGTGCTGACAGC	MC10025
sp10	TCGACACCCGCTCTAGTTGACTTCC	MC10026
sp11	GTGAGCAACGGCGGCGGCAACCTGG	MC10027
sp12	ATATCTGCTGCCCGCCCGGGGAGAT	MC10073
sp13	GACCATCATTGCCATTCCCTCTCCC	MC10029
sp14	GGTGTGATGCGGATGGTCGGCTCGG	MC10030
sp15	CTTGAATAACGCGCAGTGAATTTCG	MC10074
sp16	CGAGTTCCCGTCAGCGTCGTAAATC	MC10034
sp17	GCGCCGGCCCGCGCGGATGACTCCG	MC10035
sp18	CATGGACCCGGGCGAGCTGCAGATG	MC10036
sp19	TAACTGGCTTGGCGCTGATCCTGGT	MC10037
sp20	TTGACCTCGCCAGGAGAGAAGATCA	MC10038
sp21	TCGATGTCGATGTCCCAATCGTCGA	MC10039
sp22	ACCGCAGACGGCACGATTGAGACAA	MC10075
sp23	AGCATCGCTGATGCGGTCCAGCTCG	MC10043
sp24	CCGCCTGCTGGGTGAGACGTGCTCG	MC10044
sp25	GATCAGCGACCACCGCACCCTGTCA	MC10045
sp26	CTTCAGCACCACCATCATCCGGCGC	MC10046
sp27	GGATTCGTGATCTCTTCCCGCGGAT	MC10047
sp28	TGCCCCGGCGTTTAGCGATCACAAC	MC10048
sp29	AAATACAGGCTCCACGACACGACCA	MC10051
sp30	GGTTGCCCCGCGCCCTTTTCCAGCC	MC10052
sp31	TCAGACAGGTTCGCGTCGATCAAGT	MC10053
sp32	GACCAAATAGGTATCGGCGTGTTCA	MC10054
sp33	GACATGACGGCGGTGCCGCACTTGA	MC10055
sp34	AAGTCACCTCGCCCACACCGTCGAA	MC10056
sp35	TCCGTACGCTCGAAACGCTTCCAAC	MC10057
sp36	CGAAATCCAGCACCACATCCGCAGC	MC10061
sp37	CGCGAACTCGTCCACAGTCCCCCTT	MC10062
sp38	CGTGGATGGCGGATGCGTTGTGCGC	MC10063
sp39	GACGATGGCCAGTAAATCGGCGTGG	MC10064
sp40	CGCCATCTGTGCCTCATACAGGTCC	MC10065
sp41	GGAGCTTTCCGGCTTCTATCAGGTA	MC10066
sp42	ATGGTGGGACATGGACGAGCGCGAC	MC10067
sp43	CGCAGAATCGCACCGGGTGCGGGAG	MC10072

Appendix D

Reagent preparation for TB-SPRINT hybridization

1. TMAC 1.5X (25mL)

Chemicals	Vol (mL)
TMAC 5M	22.5
Sarkosyl 20%	0.188
Tris-HCL 1M	1.875
EDTA 0.5M pH8	0.3
H2O	0.137
Total Volume	25

2. TMAC 1X (25mL)

Chemicals	Vol (mL)
TMAC 5M	15
Sarkosyl 20%	0.125
Tris-HCL 1M	1.25
EDTA 0.5M pH8	0.2
H2O	8.425
Total Volume	25

3. TE (50mL) (to be prepared fresh)

Chemicals	Vol (mL)
Tris-HCL 1M pH8	0.5
EDTA 0.5M pH8	0.1
H2O	49.4
Total Volume	50

Appendix E

Reagent preparation for buffer Q

Chemicals / Reagents	Final [10X]	End [/rxn] mM
Tris-HCl 1M pH 8.75 (MW=121.1 g/mol)	2.1 mL	20 mM
KCl 3M	0.330 mL	10 mM
(NH4)2SO4 2M (MW = 132.14 g/mol)	0.500 mL	10 mM
MgSO4 1M (MW MgSO4,7H2O=246.47 g/mol)	0.2mL	2 mM
MgCl2 1M (MW=95.21 g/mol)	0.15mL	1.5 mM
Triton 100X	0.1mL	0.1%
H2O	6.62 mL	

Appendix F

PCR thermocycling profile for duplex 24-loci MIRU-VNTR

Programme 1 Temperature	Programme 2 Temperature	Programme 3 Temperature	Time	Cycle
94°C			5 minutes	1x
94°C			30 seconds	35x
61°C	**55°C**		30 seconds	
72°C		**70°C**	90 seconds	
72°C		**70°C**	5 minutes	1x
20°C			∞	∞

Note: 1 PCR Run duration is 1:58 hours

Appendix G

Occupation by nationality in Sabah, East Malaysia

Occupation	Indonesian (%)	Malaysian (%)	Filipino (%)	Total (%)
Armed Forces		3 (0.4)		3 (0.3)
Automotive			1 (0.4)	1 (0.1)
Blue Colar Labourer	20 (22.7)	35 (5.2)	58 (22.6)	113 (11.1)
Business Owner		10 (1.5)	1 (0.4)	11 (1.1)
Carpentry		1 (0.1)	2 (0.8)	3 (0.3)
Clerical Support workers		16 (2.4)	1 (0.4)	17 (1.7)
Craft & related Trade	3 (3.4)	13 (1.9)	3 (1.2)	19 (1.9)
Driver	3 (3.4)	30 (4.5)	1 (0.4)	34 (3.3)
Elementary Occupation	2 (2.3)	16 (2.4)	2 (0.8)	20 (2.0)
Farming & Agriculture	5 (5.7)	61 (9.1)	6 (2.3)	72 (7.1)
Fisheries		16 (2.4)	14 (5.4)	30 (2.9)
Food & Beverage	1 (1.1)	12 (1.8)	2 (0.8)	15 (1.5)
Healthcare		6 (0.9)		6 (0.6)
Home Maker	13 (14.8)	85 (12.6)	38 (14.8)	136 (13.3)
Minor	1 (1.1)	3 (0.4)	5 (1.9)	9 (0.9)
Prison Inmate	8 (9.1)	4 (0.6)	10 (3.9)	22 (2.2)
Professional		20 (3.0)		20 (2.0)
Religious		1 (0.1)		1 (0.1)
Security Guard	1 (1.1)	13 (1.9)	1 (0.4)	15 (1.5)
Self Employed	5 (5.7)	44 (6.5)	4 (1.6)	53 (5.2)
Senior Citizen		17 (2.5)		17 (1.7)
Service & Sales		31 (4.6)	4 (1.6)	35 (3.4)
Student		37 (5.5)		37 (3.6)
Tourism		2 (0.3)	1 (0.4)	3 (0.3)
Unemployed	19 (21.6)	104 (15.4)	56 (21.8)	179 (17.6)
Unknown	7 (8.0)	94 (13.9)	47 (18.3)	148 (14.5)
Total	**88 (100.0)**	**674 (100.0)**	**257 (100.0)**	**1019 (100.0)**

Appendix H

Distribution of MTB SIT among Malaysian, Filipino and Indonesian in Sabah

NATIONALITY	TOTAL	MTB SIT		
PHP-INA-MAS	10	BEIJING-Beijing-1 EAI-EAI2-Manila-4187 EAI-EAI2-Manila-897 EAI-EAI2-Manila-287 EAI-EAI2-Manila-19 Unknown-Unknown-4186 EAI-EAI2-Manila-4196 EAI-EAI2-Manila-4181 Unknown-Unknown-4204		
PHP-MAS	18	EAI-EAI2-Manila-756 EAI-EAI2-Manila-4200 EAI-EAI2-Manila-1781 T-T1-1494 EAI-EAI2-Manila-4197 LAM-LAM2-17 EAI-EAI2-Manila-4184 EAI-EAI5-4203 EAI-EAI2-Manila-483 EAI-EAI2-Manila-894	EAI-EAI5-1365 EAI-EAI2-Manila-1169 Unknown-Unknown-4195 EAI-EAI2-Manila-895 EAI-EAI5-4211 EAI-EAI5-4190 EAI-EAI2-Manila-4193 Unknown-Unknown-4209	
INA-MAS	5	EAI-EAI2-Manila-4202 T-T1-1079 T-T1-53 EAI-EAI5-1489 LAM-LAM8-290		
MAS	52	EAI-EAI6-BGD1-3484 T-T1-51 EAI-EAI3-IND-11 EAI-EAI2-Manila-1501 MANU-Manu3-1378 EAI-EAI2-Manila-4216 BEIJING-Beijing-2610 EAI-EAI2-Manila-4210 Unknown-Unknown-4185 Unknown-Unknown-821 Unknown-Unknown-4208 EAI-EAI5-28 Unknown-Unknown-4192 EAI-EAI6-BGD1-882 EAI-EAI5-892 Unknown-Unknown-32 EAI-EAI2-Manila-4191 EAI-EAI1-SOM-3355 EAI-EAI2-Manila-4213 EAI-EAI1-SOM-349	EAI-EAI5-236 EAI-EAI3-IND-1097 Unknown-Unknown-4198 H-H3-50 T-T4-CEU1-39 EAI-EAI1-SOM-48 EAI-EAI2-Manila-1511 Unknown-Unknown-1196 EAI-EAI2-4182 Unknown-Unknown-623 LAM-LAM-249 EAI-EAI2-Manila-758 EAI-EAI5-152 EAI-EAI2-4214 LAM-LAM9-42 EAI-EAI2-Manila-4189 LAM-LAM9-891 EAI-EAI5-1408 LAM-LAM9-866 EAI-EAI5-3993	EAI-EAI5-4199 T-T1-612 X-X1-336 T-T-H37Rv-4206 EAI-EAI6-BGD1-591 X-X3-92 EAI-EAI2-Manila-4183 EAI-EAI2-Manila-1490 MANU-Manu_ancestor-523 AFRI-AFRI-4215 EAI-EAI3-IND-4212 EAI-EAI5-413

NATIONALITY	TOTAL	MTB SIT	
PHP	5	EAI-EAI5-4188	
		EAI-EAI2-Manila-4194	
		EAI-EAI2-Manila-1775	
		Unknown-Unknown-1509	
		EAI-EAI2-Manila-1992	
INA	12	LAM-LAM-29	T-T3-37
		EAI-EAI2-Manila-4201	T-T1-2786
		EAI-EAI1-SOM-744	
		H-H3-512	
		H-H3-655	
		EAI-EA1-SOM-734	
		EAI-EAI2-Manila-2704	
		T-T-102	
		EAI-EAI5-3953	
		EAI-EAI2-Manila-1171	

Appendix I

40 GenTypes (Spoligotype+24-loci MIRU-VNTR) clusters of MTB from Sabah, n=129

No	Sub-lineage	SIT	Total	24MIRU Profile	MIT24
1	EAI2-Manila	19	2	2145243A2843266222342712	391
2	EAI2-Manila	19	2	2145243A2923276222342713	392
3	EAI2-Manila	1169	2	2145243A2923276222342713	392
4	EAI2-Manila	19	7	2145243A2973276223342713	394
5	EAI2-Manila	19	11	2145243A2943266223342712	396
6	EAI2-Manila	897	6	2145243A2943266223342712	396
7	EAI2-Manila	19	12	2145243A2923276223342713	397
8	EAI2-Manila	19	2	2145233A2843266223342713	398
9	EAI2-Manila	19	2	2145243A2923276223342613	399
10	EAI2-Manila	287	7	2145243A1943266223342713	400
11	EAI2-Manila	483	3	2145243A2943266223332713	401
12	EAI2-Manila	19	3	2145243A2843266223332813	402
13	EAI2-Manila	19	3	21452439294326622334 2712	403
14	EAI2-Manila	19	3	2145243A2973276213342713	404
15	EAI2-Manila	19	3	2145243A2923276213342713	405
16	EAI2-Manila	19	4	2A45243A2973246223342713	406
17	EAI2-Manila	19	2	2145243A2923276223322713	407
18	EAI2-Manila	19	2	2145243929232662233427 13	408
19	EAI2-Manila	19	5	2145243A2823276223342713	409
20	EAI2-Manila	19	3	2145243A2773266223342713	410
21	EAI2-Manila	19	2	2145243829232762233427 13	411
22	EAI2-Manila	19	2	2145243A2973266223342713	412
23	EAI5	1365	4	2145233A2943266223342712	413
24	EAI2-Manila	19	3	2145243A2963266223342712	414
25	EAI2-Manila	19	2	2145243A2943266223342713	416
26	EAI2-Manila	19	3	2144223627432562233427 13	417
27	EAI2-Manila	756	2	2145243A2443266223342712	419
28	EAI2-Manila	287	2	2145243516432662233427 13	421
29	EAI5	1365	2	2144243A2943266223342712	422
30	EAI2-Manila	19	2	2145243A2873266223332813	424
31	Unknown	Orphan	2	2145233A2844266243342712	425
32	EAI2-Manila	19	2	2145243A2673276223342513	428
33	EAI2-Manila	287	2	2145243928232762233427 13	430
34	EAI2-Manila	19	2	2145243A2A73266223342713	431
35	EAI2-Manila	19	2	2145253A2963266223342712	432
36	EAI2-Manila	19	2	2145243A2743266223342713	433
37	Beijing	1	3	24423335254442517335382 3	434
38	EAI2-Manila	19	2	2145243A2923286213342713	435
39	AFRI	4215	2	2145243A2942266223342713	440
40	EAI5	1365	2	21452535284326622334 2712	441

LIST OF PUBLICATIONS AND PAPERS PRESENTED

The findings from this study generated one manuscript. The manuscript has been published in the Journal of Tuberculosis.

1. Paul, D.C., Ngeow, Y.F., Yap, S.F., Dony, J.F., Avoi, R., Mohammad, R. and Ng, H.F., 2022. Concentrated specimen smear microscopy utilising a polymer membrane sandwich filtration vessel for the detection of acid-fast bacilli in health facilities in Sabah, East Malaysia. Tuberculosis, 133, p.102183.

In addition, a part of this study's data was presented at the Inaugural FMHS Scientific Meeting (UTAR Sungai Long, Malaysia), 2019. The references of these papers or publications are as follows:

1. Paul, D.C., Dony, J.F., Avoi, R., Yap, S.F., Ngeow, Y.F. 2019. Easy Spoligotyping for Mycobacterium tuberculosis Complex (MTB) Utilizing the Luminex Platform. Malaysian J Pathol 2019; 41(3):446

2. Paul, D.C. 2019. Luminex Microbead-Based Spoligotyping Assay (Beamedex, France) for Mycobacterium Tuberculosis Complex (MTB). Malaysian J Pathol 2019; 41(3):440

Other presentations are listed as follows:

1. ICPaLM 2018 (Consolation prize Poster presentation)
   Microbead-based spoligotyping of *Mycobacterium tuberculosis* complex isolates from Sabah, East Malaysia (abstract in Malaysian J Pathol 2018; 40(2):248)

2. 13th Sabah Public Health Colloquium (15-16 October 2018, 3rd prize Oral presentation)
   PRELIMINARY STUDY ON *MYCOBACTERIUM TUBERCULOSIS* COMPLEX GENOTYPES IN SABAH

3. Academia Sinica-Taiwan (3-4 June 2019)
   SPOLIGOTYPING OF MYCOBACTERIUM TUBERCULOSIS COMPLEX (MTB) UTILISING LUMINEX TECHNOLOGY: AN EFFICIENT ALTERNATIVE

4. UTAR FMHS Biennial Conference (July 2019)
   Easy spoligotyping for *Mycobacterium tuberculosis* Complex (MTB) utilizing the Luminex platform (oral & poster presentation)

5. ICPaLM 2020 (best poster in Medical Microbiology Category) - Comparing the Concentrated Sputum Smear Method (CSSM) with Direct Sputum Smear Microscopy (DSSM) for the Detection of Acid-Fast Bacilli (TB) in Health Facilities in Sabah, Malaysia

6. National TB and Lung Disease Conference 2024 (21-23 June 2024) Detection of *Mycobacterium tuberculosis* (MTB) And Drug-Resistant TB (DRTB) in Sabah, East Malaysia Using a Two-Tiered Classical Genotyping Approach (Spoligotyping and Duplex 24-loci MIRU-VNTR) and TB-RINT, A Microbead-Based Resistotyping Assay (oral presentation)

Abstract in International Symposia

Paul, D.C., Dony, J.F., Avoi, R., Yap, S.F., Ngeow, Y.F. **2019**. Spoligotyping of Mycobacterium tuberculosis Complex (MTB) utilising Luminex technology: An Efficient Alternative. Academia Sinica-Taiwan (3-4 June 2019)

Abstract & Proceeding for National & State / Local Symposia

1. **Dawn Carmel Paul,** Richard Avoi, Jiloris F. Dony, Sook-Fan Yap and Yun-Fong Ngeow. 2018**.** Microbead-based spoligotyping of *Mycobacterium tuberculosis* complex isolates from Sabah, East Malaysia (abstract in Malaysian J Pathol 2018; 40(2):248). ICPaLM 2018 (Abstract: Poster. poster presentation - consolation prize)

2. **Dawn Carmel Paul**, Rashidah Mohammad, *Richard* Avoi & Jiloris F. Dony. 2018. Preliminary Study on MTB Complex Genotypes in Sabah. Kolokium Kesihatan Awam ke-13, Jabatan Kesihatan Negeri Sabah (**Abstract**: Oral, 3rd prize)

3. **Paul, D.C**., Dony, J.F., Avoi, R., Yap, S.F., Ngeow, Y.F. **2019**. Easy Spoligotyping for Mycobacterium tuberculosis Complex (MTB) Utilizing the Luminex Platform. *Malaysian J Pathol 2019; 41(3):446*. UTAR FMHS Biennial Conference (July 2019)

4. **Paul, D.C**. **2019**. Luminex Microbead-Based Spoligotyping Assay (Beamedex, France) for *Mycobacterium Tuberculosis* Complex (MTB). *Malaysian J Pathol 2019; 41(3):440*. UTAR FMHS Biennial Conference (July 2019)

5. **Dawn Carmel Paul**, Rashidah Mohammad, Hien Fuh Ng, Jiloris F. Dony, Richard Avoi, Sook Fan Yap, Yun Fong Ngeow. **2020**. Comparing the Concentrated Sputum Smear Method (CSSM) with Direct Sputum Smear Microscopy (DSSM) for the Detection of Acid-Fast Bacilli (TB) in Health Facilities in Sabah, Malaysia. ICPaLM 2020 (Abstract: Poster, best poster in Medical Microbiology Category)

6. **Dawn Carmel Paul**, Ngeow Yun Fong, Emilyn Costa Conceição, Yap Sook Fan, Jiloris F. Dony, Richard Avoi. **2023**. A Snapshot of Mycobacterium tuberculosis Complex Genotypes in Sabah, East Malaysia Utilising a Microbead-Based Spoligotyping and Duplex 24-loci MIRU VNTR. In *ASOMH SCIENTIFIC CONFERENCE (ASiC V)*. Kuala Lumpur

7. **Paul, D.C**., Yap, S.F., Ngeow, Y.F., Dony, J.F., Avoi, R., **2024**. Detection of *Mycobacterium tuberculosis* (MTB) And Drug-Resistant TB (DRTB) in Sabah, East Malaysia Using a Two-Tiered Classical Genotyping Approach (Spoligotyping and Duplex 24-loci MIRU-VNTR) and TB-RINT, A Microbead-Based Resistotyping Assay. In *NTBLD Conference 2024*. Kuala Lumpur