



# Article Snapshot of Mycobacterium tuberculosis Phylogenetics from an Indian State of Arunachal Pradesh Bordering China

Shiv kumar Rashmi Mudliar <sup>1,†</sup>, Umay Kulsum <sup>1,†</sup>, Syed Beenish Rufai <sup>2,3</sup>, Mika Umpo <sup>4</sup>, Moi Nyori <sup>5</sup> and Sarman Singh <sup>1,\*</sup>

- <sup>1</sup> Department of Microbiology, All India Institute of Medical Sciences, Bhopal 462020, Madhya Pradesh, India; srashmim1@gmail.com (S.k.R.M.); kulsum.umay@gmail.com (U.K.)
- <sup>2</sup> Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Center, Montreal, QC H4A 3J1, Canada; beenish.aiims@gmail.com
- <sup>3</sup> McGill International TB Center, Montreal, QC H4A 3J1, Canada
- <sup>4</sup> Tomo Riba Institute of Health & Medical Sciences, Naharlagun 791110, Arunachal Pradesh, India; athupopu@gmail.com
- <sup>5</sup> State TB Cell, Naharlagun 791110, Arunachal Pradesh, India; stoar@rntcp.org
- \* Correspondence: sarman\_singh@yahoo.com; Tel.: +91-9810813435
- + These authors contributed equally to this work.

Abstract: Uncontrolled transmission of Mycobacterium tuberculosis (M. tuberculosis, MTB) drug resistant strains is a challenge to control efforts of the global tuberculosis program. Due to increasing multi-drug resistant (MDR) cases in Arunachal Pradesh, a northeastern state of India, the tracking and tracing of these resistant MTB strains is crucial for infection control and spread of drug resistance. This study aims to correlate the phenotypic DST, genomic DST (gDST) and phylogenetic analysis of MDR-MTB strains in the region. Of the total 200 samples 22 (11%) patients suspected of MDR-TB and 160 (80%) previously treated MDR-TB cases, 125 (62.5%) were identified as MTB. MGIT-960 SIRE DST detected 71/125 (56.8%) isolates as MDR/RR-MTB of which 22 (30.9%) were detected resistant to second-line drugs. Whole-genome sequencing of 65 isolates and their gDST found Ser315Thr mutation in katG (35/45; 77.8%) and Ser531Leu mutation in rpoB (21/41; 51.2%) associated with drug resistance. SNP barcoding categorized the dataset with Lineage2 (41; 63.1%) being predominant followed by Lineage3 (10; 15.4%), Lineage1 (8; 12.3%) and Lineage4 (6; 9.2%) respectively. Phylogenetic assignment by cgMLST gave insights of two Beijing sub-lineages viz; 2.2.1 (SNP difference < 19) and 2.2.1.2 (SNP difference < 9) associated with recent ongoing transmission in Arunachal Pradesh. This study provides insights in identifying two virulent Beijing sub-lineages (sub-lineage 2.2.1 and 2.2.1.2) with ongoing transmission of TB drug resistance in Arunachal Pradesh.

Keywords: DST; MGIT 960; WGS; cgMLST; SNP barcoding; phylogeny

# 1. Introduction

India is leading in the highest rates of tuberculosis (TB) incidence and mortality globally, with an estimate of 2.69 million cases [1,2]. Although drug-resistant tuberculosis (DR-TB) is a major public health concern globally, it represents an alarming situation in India, with 135,000 MDR-TB cases contributing to 27% of global DR-TB cases [1]. Patients with DR-TB often require profound changes in their drug regimens, which are invariably linked to poor treatment adherence and sub-optimal treatment outcomes compared to drug-sensitive TB. Higher drug-resistant TB cases remain a challenge for clinicians and National Tuberculosis Elimination Programme (NTEP) for accurate and effective TB treatment in India [3,4]. In India, the paucity of rapid diagnosis in locations having low resources and high endemicity areas where access to health care centers is difficult, remains a major constraint in treating DR-TB cases. It is estimated that around 56% of MDR-TB cases in India remain undiagnosed [4]. Arunachal Pradesh, one of the states in the northeastern



Citation: Mudliar, S.k.R.; Kulsum, U.; Rufai, S.B.; Umpo, M.; Nyori, M.; Singh, S. Snapshot of *Mycobacterium tuberculosis* Phylogenetics from an Indian State of Arunachal Pradesh Bordering China. *Genes* **2022**, *13*, 263. https://doi.org/10.3390/ genes13020263

Academic Editor: Michael McClelland

Received: 16 December 2021 Accepted: 25 January 2022 Published: 29 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). region of India bordering China with 80% area covered with forest, mostly with hilly terrains, has awakened consciousness of NTEPs due to the high prevalence of around 78.8% MDR-TB cases [5]. The reason for undiagnosed drug-resistant cases is the location of most villages in impoverished forest zones, poor connectivity of roads to health centers leading to inadequate access to health services.

There is increasing evidence that the inter-strain variation in *M. tuberculosis* exists due to variation in gene expression profiles and is biologically significant [6]. Several molecular epidemiological studies have proposed that certain types of *M. tuberculosis* strains that are prone to drug resistance may have rapid transmission rates and a higher rate of recurrence due to relapse [7]. Understanding the role of strain variation of *M. tuberculosis* to clinical phenotypes requires an approach to categorize *M. tuberculosis* isolates into groups that share most of the genotypic and phenotypic traits. For this, studies based on phylogenetic analysis, which organize clinical isolates into genetically related groups, are needed. Such studies provide an evolutionary framework for investigating polymorphisms and their potential biological relevance [8].

Molecular strain typing using the 24-loci MIRU-VNTR along with Spoligotyping is widely used for characterizing ongoing transmission of *M. tuberculosis* strains in a particular geographical location and has been shown to provide crucial information effective for the public health interventions [9]. However, these typing methods are reported to miss considerable amounts of genetic diversity, and where the overall diversity of circulating clones is limited, these approaches are insufficient to differentiate among strains [10].

With the advancement of Next-generation sequencing, Whole-genome sequencing (WGS) data are appraised for its use in epidemiological studies, strain typing for outbreak investigations, and surveillance of infectious disease due to higher sensitivity and rapid turnaround time [11]. Different genotyping methods such as spoligotyping and MIRU-VNTR were previously used for the categorization of lineages and sub-lineages. However, comparative analysis led to the use of single nucleotide polymorphisms (SNPs), which provided valuable insights into the epidemiology of circulating strains and were used as robust genetic signatures for phylogenetic categorization [12–14]. A well-established SNP barcode approach is already well known for analyzing 60 loci in *M. tuberculosis sensu* stricto genomes and has efficiently been used for categorization of major Lineages 1–7 and sub-lineages [13]. Moreover, core genome multilocus sequence typing (cgMLST) based on entire allele change is widely being used with confidence for assessing phylogenetic position to genomic data sets within a single species The sum of all core genes and their alleles for a species comprise the species cgMLST schema [14]. Phylogenetic heterogeneity, drug-resistant patterns, and association of lineages with drug resistance are not well known from the Arunachal Pradesh region of India. This study was aimed to utilize approaches of cgMLST and SNP barcoding to envision circulation of *M. tuberculosis sensu stricto* lineages. We also aim to perform phenotypic DST and genomic DST (gDST) to see the patterns of drug resistance in clinical strains of *M. tuberculosis*.

This study will help in the identification of hypervirulent strains/clones that are circulating among drug-resistant TB cases in Arunachal Pradesh and will provide crucial information to NTEP and public health programmers. Fast and accurate tracking of hypervirulent *M. tuberculosis* strains is, therefore, essential to keep track of ongoing circulating clones, which is decisive for infection control and can help in the prediction of potential future outbreaks.

#### 2. Materials and Methods

#### 2.1. Study Setting and Sample Collection

A total of 200 sputum samples (1 sample per patient) 22 (11%) patients suspected of MDR-TB, and 160 (80%) previously treated MDR-TB cases were collected from 6 districts of Arunachal Pradesh (Papum Pare, East Kameng, Kurung Kumen, Tirap, Lower Dibang Valley and Kra Daadi) by Department of Microbiology, Tomo Riba Institute of Health and Medical Sciences from September 2019 to September 2021. Samples were transported in

triple package cold chain within 72 h of collection to TB laboratory, Department of Microbiology, AIIMS Bhopal for liquid culture and DST. Informed consent was collected from all study participants. Ethical clearance was obtained for carrying out the study at TRIHMS Arunachal Pradesh and AIIMS Bhopal under reference number DME (T&R)/IEC/2015/1 and IHEC-LOP/2018/EF0104, respectively.

## 2.2. Decontamination of Samples and Bactec MGIT 960 Culture Inoculation

Sputum samples were processed using NALC-NaOH method [15]. Briefly, a minimum 3 mL of sputum sample was mixed with an equal amount of 0.5% NALC-4% NaOH, vortexed and incubated at 37 °C for 10 min. Samples were then neutralized and washed with phosphate buffer (PH 6.8) by centrifuging at 10,000 rpm for 10 min. Pellet was resuspended in 2 mL of phosphate buffer and mixed well. Smears were prepared for Ziehl-Neelsen (ZN) staining.

Five hundred microliter of the decontaminated sample was inoculated in Bactec MGIT 960 culture tubes containing 800  $\mu$ L mixture of oleic acid, albumin, dextrose, and catalase (OADC) and polymyxin B, amphotericin B, nalidixic acid, trimethoprim, azlocillin (PANTA) supplement as per the manufacturer's instructions (Becton Dickinson Diagnostic Instrument Systems, Sparks, MD, USA). Leftover decontaminated sample aliquots were stored at -80 °C for future use.

### 2.3. Identification of Cultures Using In-House Multiplex PCR

Bactec MGIT 960 cultures were identified as *M. tuberculosis* complex using in-house multiplex PCR, which targets *hsp-65* (genus-specific), *esat-6* (MTB specific), and internal transcribed spacer (ITS) MAC region (MAC specific) [16]. Amplified products were resolved through 2% agarose gel in Tris–acetate buffer [16].

Identified cultures of *M. tuberculosis* complex were subcultured on slants of Lowenstein Jensen (LJ) medium and incubated at 37 °C for 21–28 days. Growth from LJ medium was used for Bactec MGIT 960 DST and DNA extraction for WGS [17].

## 2.4. Bactec MGIT 960 SIRE DST

Single colony with the help of sterile inoculating loop from each LJ medium was inoculated in each MGIT 960 system tube and incubated in Bactec instrument until flagged positive. First-line SIRE DST was performed as per the manufacturer's protocol [18]. DST was performed on Day 1 and Day 2 by single dilution (0.5 mL of 1:100 dilution inoculum for growth control (GC) and 0.5 mL of inoculums directly in four drug panel tubes) and Day 3 to Day 5 (0.5 mL of 1:4 mL dilution inoculums directly for four drug panels and further dilution in 1:100 for GC) from the day of flagged positive of MGIT 960 instrument tube [18,19]. *M. tuberculosis* H37Rv [ATCC (American Type Culture Collection) number 2799] and known MDR-TB strain were used as quality control. The inoculated MGIT tubes with DST racks were loaded in the automated Bactec MGIT 960 system and the growth was continuously monitored by BD Epi-center.

#### 2.5. Second Line DST Using Bactec MGIT 960

Isolates identified as MDR-TB were tested against second line drugs viz., moxifloxacin (MOX), levofloxacin (LEV), amikacin (AMK), and linezolid (LNZ). All drugs were purchased from Sigma-Aldrich Corporation (St. Louis MO, USA) and were chemically in the form of powder. Stock solution of drugs AMK (1 mg/mL), MFX (1 mg/mL), LFX (1 mg/mL) and LNZ (1 mg/mL) were prepared as per the instruction and sterilized through 0.22 µm pore-size Millex-GS filter units (Millipore Bedford, MA, USA). *M. tuberculosis* H37Rv [ATCC (American Type Culture Collection) number 2799] and known fluoroquinolone (FQ) resistant strains were used as quality control strains. Second-line DST was performed as per the protocol [19–21]. As AST carrier rack for second-line drug (SLDs) panels were not available commercially for the MGIT-960 system, it was registered as one of the SIRE (Streptomycin,

Isoniazid, Rifampicin, Ethambutol) panel in order to obtain a printable report and drug susceptibility testing results.

#### 2.6. Whole Genome Sequencing

Genomic DNA extraction from 65 clinical isolates (representatives of sensitive and resistant groups as described in Figure 1) of *M. tuberculosis* complex was carried out using the Chloroform-Isoamyl alcohol (CI) method [22], and quantification of genomic DNA was performed using Qubit Fluorometer (Thermo Fisher). The genomic DNA was sequenced using plexWell WGS-24 Library Preparation kit (Illumina, San Diego, CA, USA). Library pools were subjected to paired-end sequencing on a HiSeq platform (Illumina, San Diego, CA, USA).



**Figure 1.** Workflow showing the phenotypic DST result and bioinformatics analysis of the isolates selected for WGS. FQ<sup>R</sup>: Fluoroquinolone Resistant; AMK<sup>R</sup>: Amikacin Resistant.

The quality of sequenced reads was screened using FastQC v0.11.9, and the reads with an average quality score of >=20 were retained [23]. Reads that were shorter than 36bp and possible adapter contaminating sequences were removed using Trim Galore Version 0.6.4 [24]. The output of contigs/genomes were assembled using the SPAdes genome assembler (version 3.9.0) using the default k-mer size [25] and annotated using Prokaryotic genome annotation pipeline (PGAP) [26].

#### 2.7. Identification of SNPs

Raw reads of each genome were mapped to *M. tuberculosis* H37Rv reference genome (Accession NC\_000962.3) using Burrows–Wheeler Aligner (BWA-MEM algorithm) bwa-0.7.12 [27]. SAM to BAM format conversion and sorting of mapped sequences, filling in mate coordinates to keep best reads using mate score tags was performed using Samtools 1.10 [28]. Duplicate alignments were marked and removed using samtools markdup command. BAM files were indexed for piling up variants by samtools mpileup. Mutations

with read depth above 10 reads were considered true mutations. Annotation and filtering of variants were conducted using SNPEFF 5.0e [29] and SnpSift [30].

#### 2.8. Assignment of Principal Genetic Groups

To assign a principal genetic group (PGG), each sequenced isolate was manually screened for polymorphisms in *gyrA* codon 95 and *KatG* codon 463 and were categorized accordingly to PGG as 1, 2, or 3, respectively, as explained previously [31].

## 2.9. Identification of Lineages and Sub-Lineages Using WGS SNP Barcoding

Isolates based on the patterns of SNP at the designated loci were categorized into phylogenetic lineages groups as lineage 1 (Indo-Oceanic), lineage 2 (East Asian), lineage 3 (East African Indian), lineage 4 (Euro-American), lineage 5 (West Africa 1), lineage 6 (West Africa 2), and lineage 7 (Horn of Africa) and lineage 8 [13]. After splitting WGS isolates into lineages, further categorization was conducted on the basis of SNP's [13]. Construction of the UPGMA tree was conducted by concatenating SNPs and visualized using iToL V.6 software [32].

## 2.10. Phylogenetic Analysis and Construction of cgMLST

Schema for cgMLST was set up with an efficient Workflow for a Blast Score Ratio Based Allele Calling Algorithm (Chew BBACA) [33]. For construction of cgMLST, first a wgMLST schema was created using M. tuberculosis H37Rv (accession: NC\_000962.3) as a training file generated by the prodigal algorithm. The wgMLST schema contained 10058 loci based on 104 genomes (65 genomes sequenced under this study, 39 complete and draft genomes downloaded from NCBI). Complete genome sequences of *M. tuberculosis* complex viz; M. bovis (NC\_002945.4), M. orygis (CP063804.1), M. africanum (FR878060.1), M. cannettii (NC\_015848.1), M. tuberculosis (NC\_000962.3) and M. caprae (NZ\_CDHG01000001.1) were extracted from NCBI. Draft genomes from NCBI database that belong to Lineage 1 (PR-JNA235648, PRJNA223559, PRJNA229273, PRJNA229212, PRJNA229320, PRJNA229266), Lineage 2 (PRJNA219760, PRJNA226779, PRJNA267047, CCDC5180, CCDC5079, PR-JNA229630, NDYV00000000, PRJNA360122), Lineage 3 (PRJNA229310, PRJNA229235, PRJNA229259, NDYU00000000), Lineage 4 (PRJNA229257, PRJNA223558, PRJNA229237, PRJNA228063, PRJNA228052, PRJNA229638, PRJNA218312, PRJNA233359, PRJNA233363), Lineage 5 (PRJNA211660), Lineage 6 (PRJNA211707, PRJNA211702), Lineage 7 (PRJEB8432) and Lineage 8 (PRJNA598991) and lineage B (PRJNA229213), were extracted and used as representative for *M. tuberculosis* lineage 1-8. These 39 publicly available complete genomes were used for validation of the cgMLST schema. The resulting loci was then subjected to AlleleCall, which identified and excluded 104 possible paralogous loci from further downstream analysis using the default BLAST Score Ratio (BSR) threshold of 0.6. Finally, cgMLST was extracted containing a set of 1443 core loci (present in 100% of the isolates). The resulting cgMLST matrix was uploaded in phyloviz 2.0 [34] to generate and visualize UPGMA Tree. The genome of *M. cannettii* was used to root the tree.

#### 2.11. Data Analysis

All data obtained from gDST, phenotypic DST, PGG, and categorization of lineage on the basis of SNP barcoding were maintained on MS Excel 2013 for further analysis.

## 3. Results

#### 3.1. Demographic Details and Characteristics of MDR-TB Patients

Of the total 200 patients included in the study, 91 (45.5%) were males and 109 females with mean age ( $\pm$  standard deviation) of 29.52  $\pm$  13.21 and 27.85  $\pm$  14.17 years, respectively. The majority of cases were adults, 183 (91.5%), and 17 (8.5%) were from the pediatric age group.

# 3.2. Bactec MGIT 960 Culture Results and Identification M. tuberculosis Complex Isolates

One hundred and twenty six of 200 (63%) samples were smear-positive, and 74 (37%) smear-negative. Of the total 200 cultures inoculated in Bactec MGIT 960 145 (72.5%) were flagged positive with an average turnaround time (TAT) of 18 days. All flagged positive cultures were further confirmed by ZN-stained smear examination, of which 131/145 (90.3%) were smear-positive for AFB while 14/145 (9.7%) were contaminated. Of 131 cultures, in-house multiplex PCR identified 6 (4.6%) cultures as Non-tuberculosis Mycobacterium and 125 (95.4%) cultures as *M. tuberculosis* complex (Figure 1).

# 3.3. Bactec MGIT 960 SIRE DST

Of total 125 cultures, Bactec MGIT 960 SIRE DST detected 66 (52.8%) as MDR-TB (resistant to both RIF and INH), 5 (4.0%) as mono-resistant to RIF, and 14 (11.2%) were drug-resistant isolates (resistant to any of first-line drug other than MDR/RR). 40 (32.0%) isolates were detected as pan-sensitive (Table 1).

SIRF Drug Suscentibility Pattern by MGIT 960						Second Line Drug Susceptibility Pattern by MGIT 960					
STREDING	INH	DIE	EMB	n	0/_	AMK	MEY	I EV	11 500		0/_
D	D	D	D	27	20 (9/	AWIK	D D	D	C	12	2E 10/
ĸ	ĸ	ĸ	К	37	29.6%	5	R	R	5	15	35.1%
						ĸ	R	K	5	2	5.4%
						5	R	S	5	2	5.4%
						5	5	S	5	20	54.1%
S	R	R	S	15	12.0%	S	R	R	S	2	13.3%
						S	S	S	S	13	86.7%
S	R	R	R	5	4.0%	S	R	R	S	2	40.0%
						S	S	S	S	3	60.0%
R	R	R	S	9	7.2%	S	R	R	S	1	11.1%
						S	S	S	S	8	88.9%
S	S	R	S	3	2.4%	S	S	S	S	3	100.%
S	R	S	S	8	6.4%	-	-	-	-	_	_
S	S	R	R	1	0.8%	S	S	S	S	1	100%
R	S	R	S	1	0.8%	S	S	S	S	1	100%
R	R	S	R	2	1.6%	-	-	-	-	-	-
R	S	S	S	1	0.8%	-	-	-	-	-	-
S	S	S	R	2	1.6%	-	-	-	-	-	-
R	R	S	S	1	0.8%	-	-	-	-	-	-
S	S	S	S	40	32.0%	-	-	-	-	-	-
$\frac{\text{STR}^{\text{S}} = 74}{\text{STR}^{\text{R}} = 51}$	$\frac{\text{INH}^{\text{S}}}{\text{INH}^{\text{R}}} = 48$ $\text{INH}^{\text{R}} = 77$	$\frac{\text{RIF}^{\text{S}} = 54}{\text{RIF}^{\text{R}} = 71}$	$\frac{\text{EMB}^{\text{s}}}{\text{EMB}^{\text{R}}} = 78$ $= 47$	n = 125		AMK <sup>S</sup> = 69 AMK <sup>R</sup> =2	$\frac{MFX^{S}}{MFX^{R}} = 49$ $MFX^{R} = 22$	$\frac{LFX^{S}}{LFX^{R}} = 51$ $LFX^{R} = 20$	$LNZ^S = 71$	n = 71	

Table 1. Association of first-line drug-resistant patterns with second-line DST on Bactec MGIT-960.

STR<sup>S</sup>: Streptomycin susceptible; STR<sup>R</sup>: Streptomycin resistant; INH<sup>S</sup>: Isoniazid susceptible; INH<sup>R</sup>: Isoniazid resistant; RIF<sup>S</sup>: Rifampicin susceptible; RIF<sup>R</sup>: Rifampicin resistant; EMB<sup>s</sup>: Ethambutol susceptible; EMB<sup>R</sup>: Ethambutol resistant; AMK<sup>S</sup>: Amikacin susceptible; AMK<sup>R</sup>: Amikacin resistant; MFX<sup>S</sup>: Moxifloxacin susceptible; MFX<sup>R</sup>: Moxifloxacin resistant; LFX<sup>S</sup>: Levofloxacin susceptible; LFX<sup>R</sup>: Levofloxacin resistant; LNZ<sup>S</sup>: Linezolid susceptible.

## 3.4. Bactec MGIT 960 Second Line DST

Of 71 (56.8%) MDR-TB and RIF mono-resistant isolates subjected to second-line DST for drugs AMK (1 mg/mL), LFX(1 mg/mL), LNZ(1 mg/mL), and MOX(1 mg/mL), 49 (69.0%) were found to be susceptible, 20 (28.2%) were mono-resistant to FQ, [2(2.8%) was resistant to only MOX and 18 (25.4%) to (MFX+LFX)]; 2 (2.8%) isolates were found to be resistant to (FQ+AMK) while no isolates were found mono-resistant to AMK or resistant to LNZ. Patterns of first- and second-line drugs are shown in Table 1.

#### 3.5. Mutations in Genes Associated with First- and Second-Line Drugs Using WGS

A total of 65 isolates sequenced were analyzed for mutations conferring drug resistance in genes associated with first-line and second-line drug resistance (Table 2). Genomes sequences of each isolate were screened for mutations in genes conferring resistance to first-line anti-tuberculosis drugs viz; rpsL, rrs, gidB for STR; KatG, inhA, ahpA, fab, ndh for INH; rpoA, rpoB, rpoC for RIF; embABC for EMB and pncA for PZA.

D		Phenotypic DST	Whole Genome Sequencing Results						
Drug	Gene	MGIT Result	Mutation	No of Isolates		Lineages			
					Lin1	Lin2	Lin3	Lin4	
		R	Ser531Leu	21/41 (51.2%)	4 (19%)	15 (71.4)	2)	-	
		R	Leu511Pro Phe505Leu *	3/41 (7.3%)	-	3 (100%)	_	_	
		R	Leu511Pro His526Gln	1/41 (2.4%)	-	1 (100%)	_	_	
	rpoB	R	Leu511Pro His526Gln Phe505Leu *	2/41 (4.9%)	_	2 (100%)	_	_	
		R	Asp516Val	2/41 (4.9%)	_	1 (50%)	_	- (50%)	
RIF		R	Asp516Tyr	1/41 (2.4%)	-	1 (100%)	_	_	
(n = 41)		R	His526Tyr	3 /41(7.3%)	1 (33.3%)	2 (66.7%)	_	_	
		R	His526Asp	3/41 (7.3%)	-	3 (100%)	_	_	
		R	Gln513Pro	1/41 (2.4%)	_		_	1 (100%)	
	rpoB rpoC	R	Ser531Leu Ile561Val * Ile572Thr	1/41 (2.4%)	_	1 (100%)	_	_	
	rpoB rpoA	R	Ser531Leu Gly112Ser	1/41 (2.4%)	1 (100%)	_	_	_	
	rpoB rpoA	R	Ser531Leu Gly319Lys	1/41 (2.4%)	-	1 (100%)	_	_	
	rpoB rpoA	R	Ser531Leu Val264Gly	1/41 (2.4%)	-	-	_	1 (100%)	
	rpoB	S	Leu545Met *	1/41 (2.4%)	-	1 (100%)	_	_	
	KatG	R	Ser315Thr	35/45 (77.8%)	3 (8.6%)	27 (77.1%)	3 (8.6%)	2 (5.8%)	
		R	Ser450Leu	1/45 (2.2%)	1 (100%)	_	_	_	
	inhA	R	Ser94Ala	2/45 (4.4%)	2 (100%)	_	_	_	
	ahp	R	52C>T	2/45 (4.4%)	2 (100%)	_	-	-	
INH	Fab	R	15C>T	1/45 (2.2%)		1 (100%)	_	_	
(n = 45)	Kat Fab	R	Ser140Gly 15C>T	1/45 (2.2%)		1 (100%)	_	_	
		R	Ser315Thr 17C>T	1/45 (2.2%)		1 (100%)	_	_	
		R	Ser315Thr 15C>T	2/45 (4.4%)		2 (100%)	_	_	
		R	Met306Val	18/30 (60.0%)	_	16 (88.8%)	-	2 (12.2%)	
		R	Gly406Asp	1/30 (3.3%)	-	-	1 (100%)	-	
	embB	R	Met306Ile	4/30 (13.3%)	1 (25%)	3 (75%)		-	
EMD		R	Asp354Ala	1/30 (3.3%)	-	1 (100%)	-	-	
(n = 30)		R	Met306Leu	1/30 (3.3%)	-	-	1 (100%)	-	
	1.D	R	Glu405Asp	1/30 (3.3%)	_	1 (100%)	_	-	
	етвв	R	Gln853ProMet306Val	2/30 (6.7%)	_	2 (100%)	-	-	
	embA embA	R	Met306Val 12C>T	2/30 (6.7%)	_	2 (100%)	_	-	
		R	Lys43Arg	22/26 (84.6%)	-	21 (95.4%)	1 (19.1%)	-	
STR (n = 26)	rpsL	R	Lys88Arg	2/26 (7.7%)	_	2 (100%)	_	-	
	rrs	R	514 A>C	2/26 (7.7%)	_	2 (100%)	_	_	
		R	Asp94Gly	6/16 (37.5%)	_	5 (83.3%)	1 (16.6%)	-	
	~~~~4	R	Asp94Tyr	2/16 (12.5%)	_	2 (100%)	-	-	
		R	Asp94Asn	2/16 (12.5%)	_	2 (100%)	-	-	
FO	gyrA	R	Asp94Ala	1/16 (6.3%)		1 (100%)	-	-	
(n = 16)		R	Ala90Val	1/16 (6.3%)			_	- (100%)	
		R	Asp94His	1/16 (6.3%)	_	1 (100%)	_	_	
		R	Ile486Leu	1/16 (6.3%)		1 (100%)	_	_	
	gyrB	R	Asp461His	1/16 (6.3%)				1 (100%)	
		R	Ala504Val	1/16 (6.3%)	_	_	_	_	

Table 2. Patterns of mutation resulting from WGS analysis and their association with phenotypic DST.

	Gene	Phenotypic DST	Whole Genome Sequencing Results							
Drug		MGIT Result	Mutation		Lineages					
				No of isolates	Lin1	Lin2	Lin3	Lin4		
	pncA	R	Asp49Ala	5/10 (50.0%)	_	5 (100%)	_	_		
PZA		R	Gly108Arg	2/10 (20.0%)	_	2 (100%)	-	_		
(n = 10)		R	11A>G	2/10 (20.0%)	_	2 (100%)	_	_		
		R	Asp136Tyr	1/10 (10.0%)	1 (100%)	-	_	_		
	rrs	R	1484 G>T	1/2 (50.0%)	_	1 (100%)	_	_		
AMK $(n = 2)$		R	1401 A>G	1/2 (50.0%)	_	1 (100%)	_	_		
	inha	NA	Ser94Ala	2/8 (25.0%)	2 (100%)	_	_	_		
ETH	fab	NA	15C>T	3/8 (37.5%)	_	3 (100%)	_	_		
(n = 8)		NA	17G>T	1/8 (12.5%)	_	-	_	1 (100%)		
	ethA	NA	886_886del	2/8(25.0%)	_	1 (50%)	-	1 (50%)		
Cysr (n = 2)	alr	NA	Met343Thr	2/2 (100.0%)	-	2 (100%)	-	_		
PAS	thy	NA	16C>T	1/2 (50.0%)	_	1 (100%)	-	_		
(n = 2)	folC	NA	Ile43Thr	1/2 (50.0%)	_	1 (100%)	_	_		

## Table 2. Cont.

STR: Streptomycin; INH: Isoniazid; RIF: Rifampicin; EMB: Ethambutol; PZA: Pyrazinamide; ETH: Ethionamide; FQ: Fluoroquinolone; AMK: Amikacin; PAS: Para-aminosalicylic acid; Cysr.: Cycloserine; S: susceptible; R: resistant; \* outside Rifampicin Resistance Determining Region (RRDR); n=number of isolates.

Of the total 65 isolates, Lys43Arg (22/26; 84.6%) in rpsL, Ser315Thr (35/45; 77.8%) in katG, Ser531Leu (21/41; 51.2%) in rpoB, Met306Val (18/30; 60%) in embB, and Asp49Ala (5/10; 50%) in pncA was found to be predominantly present in genes known to confer drug resistance for first-line drugs STR, INH, RIF, EMB, and PZA, respectively.

All 65 sequenced genomes were also analyzed for mutations in genes conferring resistance to second-line anti-tuberculosis drugs viz gyrA, gyrB for FQ; rrl and rplC for linezolid; Rv0678, Rv2535c, Rv1979c and mmpl5 for clofazimine; alr, ddl, ald and cycA for cycloserine; rrs for AMK; Rv0678 and atpE for bedaquiline; thyA, ribD and folC for PAS; fgd, ddn, fbiA, fbiB and fbiC for delamanid.

In case of FQ, Asp94Gly (6/16; 37.5%) was found to be the predominant mutation in gyrA gene region, of which two genomes were also found to have mutation 1484G>T and 1401 A>G in rrs gene region conferring drug resistance to injectable class of drugs known to confer drug resistance. No mutations were found for genes associated with linezolid, clofazimine, delamanid, and bedaquiline.

Patterns of mutation resulting from WGS analysis and their association with phenotypic DST are shown in Table 2. Variant densities of each genome against *M. tuberculosis* H37Rv were generated using Blast Ring Image Generator BRIGv0.95 and is shown in Figure 2.

## 3.6. Phylogenetic Analysis and Identification of Lineages Based on cgMLST and SNP Barcoding

All 65 sequenced genomes were used along with 39 publicly available genomes (including 33 genomes representing lineage 1-8 of *M. tuberculosis* and 6 representatives from *M. tuberculosis* complex) to generate a phylogeny. The resulting tree showed the 65 isolates clustering with the publicly available lineage-defined genomes of *M. tuberculosis* (Figure 3). Lineage 2 (East-Asian) dominated the dataset with 41 (63.1%) genomes. Ten genomes (15.4%) were grouped in lineage 3 (East-African Indian) while 8 (12.3%) and 6 (9.2%) genomes were clustered with lineage 1 (Indo-Oceanic) and 4 (Euro-American), respectively. No genomes were clustered with lineage 5, 6, 7, and 8.



**Figure 2.** Variant densities of the 65 sequenced genomes against M. tuberculosis H37Rv using BRIG v0.95.

For all the 65 sequenced genomes, SNP barcoding was carried out on the basis of concatenated SNP's. This barcoding analysis revealed that lineage 2 dominated the dataset, followed by lineage 3, 1, and 4, respectively. 34 out of 41 (82.9%) genomes of lineage 2 belonged to sub-lineage 2.2.1 (Beijing), while 7 (17.1%) genomes belonged to sub-lineage 2.2.1.2 (Beijing). Seven (87.5%) out of 8 lineage 1 genomes belonged to sub-lineage 1.1.3 (EAI6) while one genome (12.5%) was assigned to sub-lineage 1.1.2 (EAI5). of the total ten genomes of lineage 3, only one (10%) belonged to sub-lineage 3.1.2.1 (CAS2) while the remaining 9 (90%) were not assigned any sub-lineage. Lineage 4 consisted of total six genomes with 2 (33.3%), 1 (16.7%), 2 (33.3%) genomes belonging to sub-lineage 4.5 (H3; H4; T), 4.3 (LAM) and 4.1.1.1 (X2), respectively, whereas one genome (16.7%) was not assigned to any of the sub-lineage. The UPGMA tree generated using WGS SNP barcoding was congruent to the cgMLST tree based on lineage distribution (Figure 4).



**Figure 3.** Phylogenetic analysis (UPGMA tree) based on cgMLST association of 104 genomes (65 isolates under this study and 33 lineage-defined genomes and 6 reference genomes of MTBC complex viz. *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium orygis, M. tuberculosis* and *Mycobacterium cannettii*). *Mycobacterium cannettii* was used to root the tree. Each lineage is shown with different colours.





## 3.7. Phylogenetic Analysis Based on PrincipalGenetic Group (PGG)

Based on the constitution of amino acids at loci 95 and 493, the PGG informative sites within the genes gyrA and katG, we found each isolate was designated to a PGG. Out of 65 sequenced genomes, 60 were clustered in PGG1, 4 in PGG2, and only 1 in PGG3. On comparing the data based on sub-lineage classification, all the genomes belonging to sub-lineage 1, 2, and 3 were clustered in PGG1 group, while sub-lineages of lineage 4 dominated PGG2 group. Only 1 genome from lineage 4 was assigned PGG3, which was

a pre-XDR isolate. Assignment of PGG and sub-lineage classification for each isolate is shown in Supplementary Table S1.

#### 4. Discussion

With increasing drug-resistant TB cases in India, it becomes pivotal to recognize the clonal expansion of lineages or clones contributing to drug resistance, specifically in geographical regions where drug resistance is suspected [35]. Northeastern states of India have higher rates of MDR-TB, around 32.7% of which Arunachal Pradesh region is known to have 78.8% of TB drug resistance [5,36]. Most of the land in Arunachal Pradesh is under forest area, and villages located in such impoverished forest zones have very inadequate or non-existent access to health services, including the hampered efforts of the national TB elimination program. Further, proximity to China, difficult terrains, and heavy rainfalls and landslides impact health services, leading to a high MDR rate. Nonetheless, host genetics also might be contributing to high resistance in this area, but that aspect was not studied in this work, and we are not making any conclusions on the genetic reasons for high MDR. Whole-genome sequencing (WGS) has become a standard for typing of *M. tuberculosis* isolates and is known to have higher resolution over MIRU-VNTR-based clustering [37]. In this study we planned to utilize the approach of WGS and SNP barcoding for typing and clustering of *M. tuberculosis* lineages circulating in the Arunachal Pradesh region of India. This study will help in defining the transmission of strains associated with drug resistance circulating in the Arunachal Pradesh region of India.

Our report for MDR-TB from Arunachal Pradesh is 56.8%, slightly lower than previously reported 79.3%. The variation may be due to a lower number of sample sizes in the previous study, results based on DNA-based line probe Assay rather than Bactec MGIT 960, which detects viable bacilli and incorrect denominators used while calculating drug resistance rates [36]. In our study, 80% of the samples were obtained from previously treated TB cases resulting in an increased rate of drug resistance. However, no reports of FQ resistance were reported from Arunachal Pradesh earlier.

There is mounting evidence that strain diversity plays a role in the transmission of disease [38]. In order to spot strain transmission in Arunachal Pradesh, we randomly selected cultures for WGS from each category of varying data sets of drug-resistant patterns and used gene by gene PGG, SNP barcoding, and core genome MLST (cgMLST) method to allocate strains in well-defined phylogenetic groupings (Figure 1). All these methods were used in various phylogenetic studies for standardized phylogenetic assignment of diseases and outbreak resolutions [13,39,40]. The PGG results also correlated with results of lineage grouping by SNP barcoding as 60 (92.3%) isolates belong to PGG group 1 (KatG Leu463Leu, gyrA Thr95Thr) of which 8 (13.3%) were Lineage 1 (Indo Oceanic), 41 (68.3%) Lineage 2 (East Asian) and 10 (16.7%) Lineage 3 (East-Arican Indian) and 1 (1.7%) Lineage 4 (Euro American). PGG group 2 (KatG Leu463Arg, gyrA Thr95Thr) and PGG group 3 (KatG Leu463Arg, gyrA Thr95Ser) included 4 (6.2%) and 1 (1.5%) strain belonging to lineage 4. Categorization of lineage as per the PGG group reported from other studies was consistent with our finding [31,40] (Supplementary Table S1).

This study provided for the first time a complete picture of TB phylogenetics across Arunachal Pradesh region based on cgMLST compared to phylogenetics that is based on SNP calling methods. The resulting cgMLST phylogenetic tree contained all reference lineages and four major *M. tuberculosis* lineages from our dataset and matched with results of SNP-based methods. None of the genomes belong to Lineage 5-8, and all were *M. tuberculosis sensu stricto* (Figure 1).

Using SNP barcoding method, we found 41 (63.1%) isolates grouped with lineage 2 (East Asian); 10 (15.4%) isolates with lineage 3 (Central Asian), 8 (12.3%) isolates with lineage 1, and 6 (9.2%) isolates to be lineage 4 (Supplementary Table S1). We found Lineage 2 (East Asian) as the predominant lineage (63.1%) circulating in Arunachal Pradesh among suspected drug-resistant TB cases. To gain further insight, we looked for SNP-based markers to differentiate sub-lineages. Among Lineage 2 only two clones circulating in Arunachal

Pradesh were found viz: sub-lineage 2.2.1 (82.9%) and 2.2.1.2 (17.1%), respectively. Both these sub-lineages 2.2.1 and 2.2.1.2 belonged to modern Beijing clade, which was depicted by the presence of SNP markers at mutT2 codon Gly58Ala, ogt Gly12Gly specific to modern Beijing as reported in earlier studies [40,41]. Of a total of 6535 SNP's, Beijing clone 2.2.1 was responsible for >80% of transmission and clusters using thresholds of up to 19 SNPs showing recent transmission of strains. Another clone of Beijing 2.2.1.2 showed clusters with SNP differences of 9 SNPs, showing ongoing transmission of the strains specifically associated with drug resistance. All these 7 (100.0%) isolates of clone 2.2.1.2 were multidrugresistant, and 2 (28.6%) were resistant to FQ. Beijing sub-lineages 2.2.1 and 2.2.1.2 was reported from various parts of the world associated with outbreaks and drug resistance from Vietnam and Southern China [42–47]. We also observed Lineage 1 clone 1.1.3 with SNP differences of 101, showing this clone as endemic in Arunachal Pradesh for longer time periods. Lineage 1 is known to be associated with activation of long-term latent infection compared to that of Lineage 2 (modern Beijing) strains, which are known for more likely to progress to active disease in various host populations, more virulent, and thus highly transmissible [48,49]. A total of 6 isolates of lineage 4 were identified and were unclustered. Of Lineage 3, one clone including 9 (90%) isolates was found showing SNP differences of 18, also showing ongoing transmission. Out of 9 isolates, 3 (33.3%) were MDR-TB. One isolate (10%) of sub-lineage 3.1.2.1 was also found and was MDR-TB as well as resistant to FQ. New clades of lineage 3 were also reported to be circulating in the Assam region of India by Devi et al., which is consistent with our study [50].

The main strength of our study is that it highlights the higher rates of drug resistance in the Indian state of Arunachal Pradesh, which has a common border with China, and provides insights into the phylogenetic diversity of MDR-TB isolates from this state using cgMLST. These findings may have important implications in understanding the molecular epidemiology of DR-TB and for its control and prevention, particularly in the state of Arunachal Pradesh. However, our study has some limitations. The main limitation was that we could not include a control group from other states/regions of India in order to compare the strain diversity, specifically the Beijing sub-lineage 2.2.1 and 2.2.1.2, its association with drug resistance.

## 5. Conclusions

Our findings show dissemination of clusters of Beijing clones associated with drug resistance and regional spread may be emerging and aggressive. Approaches to contain Beijing strains (sub-lineage 2.2.1 and 2.2.1.2) may prevent transmission of these strains across other parts of India. Clonal expansion of these strains in Arunachal Pradesh in the future may lead to an outbreak of Beijing strains and underline the need for surveillance studies incorporating epidemiological information and a track of ongoing transmission to prevent drug-resistant TB outbreaks. We also found transmission of lineage 3 clade and the presence of Lineage 1 as endemic in Arunachal Pradesh. These findings may have important implications for control and prevention of TB in the northeastern part of India, Arunachal Pradesh.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13020263/s1, Table S1.

**Author Contributions:** S.k.R.M. standardized and performed all experiments and analyzed the data. U.K. performed bioinformatic analysis and contributed to manuscript writing. S.B.R. contributed to reviewing and writing of manuscripts. M.U. and M.N. helped in the collection of TB samples and managed the transportation of samples to AIIMS Bhopal. S.S. supervised and coordinated the work, finalized the manuscript, arranged reagents and chemicals. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research and APC was funded by the Department of Biotechnology (DBT), Government of India grant number MDR-TB/2017/11 provided to Prof. Sarman Singh.

**Institutional Review Board Statement:** The study was conducted in accordance with the ethical clearance committee at TRIHMS Arunachal Pradesh and AIIMS Bhopal under reference number DME (T&R)/IEC/2015/1 and IHEC-LOP/2018/EF0104, respectively.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Genomes of sixty-five *M. tuberculosis* isolates have been deposited in GenBank under BioProject accession no. PRJNA717132. Raw reads of all sixty-five *M. tuberculosis* have been made available in the Sequence Read Archive (SRA) under study number SRR331414 linked with BioProject number PRJNA717132.

Acknowledgments: We wish to thank Payal Soni and Mukesh Patel for their technical help.

**Conflicts of Interest:** The authors declare no conflict of interest.

# References

- 1. World Health Organization. Global Tuberculosis Report 2020; World Health Organization: Geneva, Switzerland, 2020.
- National Tuberculosis Elimination Programme. India TB Report 2020; National Tuberculosis Elimination Programme: New Delhi, India, 2020.
- Lange, C.; Dheda, K.; Chesov, D.; Mandalakas, A.M.; Udwadia, Z.; Horsburgh, C.R. Management of drug-resistant tuberculosis. Lancet 2019, 394, 953–966. [CrossRef]
- 4. Husain, A.A.; Kupz, A.; Kashyap, R.S. Controlling the drug-resistant tuberculosis epidemic in India: Challenges and impli-cations. *Epidemiol. Health* **2021**, *43*, e2021022. [CrossRef] [PubMed]
- 5. Singh, S. Early detection of multi-drug resistant tuberculosis in India using GenoType MTBDRplus assay & profile of resistance mutations in *Mycobacterium tuberculosis*. *Indian J. Med Res.* **2014**, *140*, 477–479.
- Gao, Q.; Kripke, K.E.; Saldanha, A.J.; Yan, W.; Holmes, S.; Small, P.M. Gene expression diversity among *Mycobacterium tuber-culosis* clinical isolates. *Microbiology* 2005, 151, 5–14. [CrossRef]
- Anh, D.D.; Borgdorff, M.W.; Van, L.N.; Lan, N.T.; van Gorkom, T.; Kremer, K.; van Soolingen, D. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg. Infect. Dis.* 2000, *6*, 302–305. [PubMed]
- 8. Kato-Maeda, M.; Bifani, P.J.; Kreiswirth, B.N.; Small, P.M. The nature and consequence of genetic variability within *Mycobacterium tuberculosis. J. Clin. Invest.* **2001**, *107*, 533–537. [CrossRef] [PubMed]
- Mekonnen, A.; Merker, M.; Collins, J.; Addise, D.; Aseffa, A.; Petros, B.; Ameni, G.; Niemann, S. Molecular epidemiology and drug resistance patterns of *Mycobacterium tuberculosis* complex isolates from university students and the local community in Eastern Ethiopia. *PLoS ONE* 2018, 13, e0198054. [CrossRef] [PubMed]
- Ford, C.; Yusim, K.; Ioerger, T.; Feng, S.; Chase, M.; Greene, M.; Korber, B.; Fortune, S. *Mycobacterium tuberculosis*—Heterogeneity revealed through whole genome sequencing. *Tuberculosis* 2012, 92, 194–201. [CrossRef]
- 11. Cohen, K.A.; Manson, A.L.; Desjardins, C.A.; Abeel, T.; Earl, A.M. Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: Progress, promise, and challenges. *Genome Med.* **2019**, *11*, 45. [CrossRef]
- Gagneux, S.; DeRiemer, K.; Van, T.; Kato-Maeda, M.; de Jong, B.; Narayanan, S.; Nicol, M.; Niemann, S.; Kremer, K.; Gutierrez, M.C.; et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 2006, 103, 2869–2873. [CrossRef]
- Kohl, T.A.; Diel, R.; Harmsen, D.; Rothgänger, J.; Walter, K.M.; Merker, M.; Weniger, T.; Niemann, S. Whole-genome-based My-cobacterium tuberculosis surveillance: A standardized, portable, and expandable approach. J. Clin. Microbiol. 2014, 5, 2479–2486. [CrossRef]
- Jones, R.C.; Harris, L.G.; Morgan, S.; Ruddy, M.C.; Perry, M.; Williams, R.; Humphrey, T.; Temple, M.; Davies, A.P. Phylogenetic Analysis of *Mycobacterium tuberculosis* Strains in Wales by Use of Core Genome Multilocus Sequence Typing to Analyze Whole-Genome Sequencing Data. J. Clin. Microbiol. 2019, 57, 1–11. [CrossRef] [PubMed]
- 15. Rufai, S.B.; Singh, A.; Kumar, P.; Singh, J.; Singh, S. Performance of Xpert MTB/RIF Assay in Diagnosis of *Pleural Tuberculosis* by Use of Pleural Fluid Samples. *J. Clin. Microbiol.* **2015**, *53*, 3636–3638. [CrossRef]
- Gopinath, K.; Singh, S. Multiplex PCR assay for simultaneous detection and differentiation of *Mycobacterium tuberculosis*, *Mycobacterium avium* complexes and other Mycobacterial species directly from clinical specimens. J. Appl. Microbiol. 2009, 107, 425–435. [CrossRef]
- Advani, J.; Verma, R.; Chatterjee, O.; Pachouri, P.K.; Upadhyay, P.; Singh, R.; Yadav, J.; Naaz, F.; Ravikumar, R.; Buggi, S.; et al. Whole Genome Sequencing of *Mycobacterium tuberculosis* Clinical Isolates from India Reveals Genetic Heterogeneity and Region-Specific Variations That Might Affect Drug Susceptibility. *Front. Microbiol.* 2019, *10*, 309. [CrossRef] [PubMed]
- Siddiqi, S.; Ahmed, A.; Asif, S.; Behera, D.; Javaid, M.; Jani, J.; Jyoti, A.; Mahatre, R.; Mahto, D.; Richter, E.; et al. Direct drug susceptibility testing of *Mycobacterium tuberculosis* for rapid detection of multidrug re-sistance using the Bactec MGIT 960 system: A multicenter study. *J. Clin. Microbiol.* 2012, *50*, 435–440. [CrossRef]
- Rufai, S.B.; Singh, J.; Kumar, P.; Mathur, P.; Singh, S. Association of gyrA and rrs gene mutations detected by MTBDRsl V1 on *Mycobacterium tuberculosis* strains of diverse genetic background from India. *Sci. Rep.* 2018, *8*, 1–15. [CrossRef] [PubMed]

- 20. World Health Organization. *Technical Manual for Drug Susceptibility Testing of Medicines Used in the Treatment of Tuberculosis;* World Health Organization: Geneva, Switzerland, 2020.
- Kim, H.; Seo, M.; Kil Park, Y.; Yoo, J.-I.; Lee, Y.S.; Chung, G.T.; Ryoo, S. Evaluation of MGIT 960 System for the Second-Line Drugs Susceptibility Testing of *Mycobacterium tuberculosis*. *Tuberc. Res. Treat.* 2013, 2013, 1–6. [CrossRef]
- Somerville, W.; Thibert, L.; Schwartzman, K.; Behr, M.A. Extraction of *Mycobacterium tuberculosis* DNA: A Question of Containment. J. Clin. Microbiol. 2005, 43, 2996–2997. [CrossRef] [PubMed]
- Black, P.A.; de Vos, M.; Louw, G.E.; van der Merwe, R.G.; Dippenaar, A.; Streicher, E.M.; Abdallah, A.M.; Sampson, S.L.; Victor, T.C.; Dolby, T.; et al. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC Genom.* 2013, *16*, 857. [CrossRef]
- 24. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 2011, 17, 10–12. [CrossRef]
- Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 2012, 19, 455–477. [CrossRef] [PubMed]
- Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016, 44, 6614–6624. [CrossRef] [PubMed]
- Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009, 25, 1754–1760. [CrossRef] [PubMed]
- Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079. [CrossRef] [PubMed]
- Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 2012, *6*, 80–92. [CrossRef]
- Cingolani, P.; Patel, V.M.; Coon, M.; Nguyen, T.; Land, S.J.; Ruden, D.M.; Lu, X. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* 2012, 3, 35. [CrossRef]
- Sreevatsan, S.; Pan, X.; Stockbauer, K.E.; Connell, N.D.; Kreiswirth, B.N.; Whittam, T.S.; Musser, J.M. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl. Acad. Sci. USA* 1997, 94, 9869–9874. [CrossRef]
- 32. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 2019, 47, W256–W259. [CrossRef]
- Silva, M.; Machado, M.P.; Silva, D.N.; Rossi, M.; Moran-Gilad, J.; Santos, S.; Ramirez, M.; Carriço, J.A. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 2018, 4, e000166. [CrossRef]
- 34. Francisco, A.P.; Vaz, C.; Monteiro, P.T.; Melo-Cristino, J.; Ramirez, M.; Carriço, J.A. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinform.* **2012**, *13*, 87. [CrossRef]
- 35. Chatterjee, S.; Poonawala, H.; Jain, Y. Drug-resistant tuberculosis: Is India ready for the challenge? *BMJ Glob. Health* **2018**, *3*, e000971. [CrossRef] [PubMed]
- Singhal, R.; Myneedu, V.P.; Arora, J.; Singh, N.; Sah, G.C.; Sarin, R. Detection of multi-drug resistance & characterization of mutations in *Mycobacterium tuberculosis* isolates from North- Eastern States of India using GenoType MTBDRplus assay. *Indian J. Med. Res.* 2014, 140, 501–506. [PubMed]
- Roetzer, A.; Diel, R.; Kohl, T.A.; Rückert, C.; Nübel, U.; Blom, J.; Wirth, T.; Jaenicke, S.; Schuback, S.; Rüsch-Gerdes, S.; et al. Whole Genome Sequencing versus Traditional Genotyping for Investigation of a *Mycobacterium tuberculosis* Outbreak: A Longitudinal Molecular Epidemiological Study. *PLoS Med.* 2013, *10*, e1001387. [CrossRef]
- Chizimu, J.Y.; Solo, E.S.; Bwalya, P.; Kapalamula, T.F.; Akapelwa, M.L.; Lungu, P.; Shrestha, D.; Fukushima, Y.; Mukonka, V.; Thapa, J.; et al. Genetic Diversity and Transmission of Multidrug-Resistant *Mycobacterium tuberculosis* strains in Lusaka, Zambia. *Int. J. Infect. Dis.* 2021, 114, 142–150. [CrossRef] [PubMed]
- Coll, F.; McNerney, R.; Guerra-Assunção, J.A.; Glynn, J.R.; Perdigão, J.; Viveiros, M.; Portugal, I.; Pain, A.; Martin, N.; Clark, T.G. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* 2014, 5, 4812. [CrossRef]
- 40. Kohl, T.A.; Harmsen, D.; Rothgänger, J.; Walker, T.; Diel, R.; Niemann, S. Harmonized Genome Wide Typing of *Tubercle Bacilli* Using a Web-Based Gene-By-Gene Nomenclature System. *EBioMedicine* **2018**, *34*, 131–138. [CrossRef]
- Bergval, I.; Sengstake, S.; Brankova, N.; Levterova, V.; Abadía, E.; Tadumaze, N.; Bablishvili, N.; Akhalaia, M.; Tuin, K.; Schuitema, A.; et al. Combined species identification, genotyping, and drug resistance detection of *Mycobacterium tuberculosis* cultures by MLPA on a bead-based array. *PLoS ONE* 2012, 7, e43240. [CrossRef]
- Nieto Ramirez, L.M.; Ferro, B.E.; Diaz, G.; Anthony, R.M.; de Beer, J.; van Soolingen, D. Genetic profiling of *Mycobacterium tuberculosis* revealed "modern" Beijing strains linked to MDR-TB from Southwestern Colombia. *PLoS ONE* 2020, 15, e0224908. [CrossRef]
- Rufai, S.B.; Sankar, M.M.; Singh, J.; Singh, S. Predominance of Beijing lineage among pre-extensively drug-resistant and extensively drug-resistant strains of *Mycobacterium tuberculosis*: A tertiary care center experience. *Int. J. Mycobacteriol.* 2016, 5 (Suppl 1), S197–S198. [CrossRef]

- Gupta, A.; Sinha, P.; Nema, V.; Gupta, P.K.; Chakraborty, P.; Kulkarni, S.; Rastogi, N.; Anupurba, S. Detection of Beijing strains of MDR M. tuberculosis and their association with drug resistance mutations in katG, rpoB, and embB genes. *BMC Infect. Dis.* 2020, 20, 1–7. [CrossRef] [PubMed]
- 45. Liu, Y.; Jiang, X.; Li, W.; Zhang, X.; Wang, W.; Li, C. The study on the association between Beijing genotype family and drug susceptibility phenotypes of *Mycobacterium tuberculosis* in Beijing. *Sci. Rep.* **2017**, *7*, 1–7. [CrossRef] [PubMed]
- 46. San, L.L.; Aye, K.S.; Oo, N.A.T.; Shwe, M.M.; Fukushima, Y.; Gordon, S.; Suzuki, Y.; Nakajima, C. Insight into multidrug-resistant Beijing genotype *Mycobacterium tuberculosis* isolates in Myanmar. *Int. J. Infect. Dis.* **2018**, *76*, 109–119. [CrossRef] [PubMed]
- Holt, K.E.; McAdam, P.; Thai, P.V.K.; Thuong, N.T.T.; Ha, D.T.M.; Lan, N.N.; Lan, N.H.; Nhu, N.T.Q.; Hai, H.T.; Ha, V.T.N.; et al. Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat. Genet.* 2018, 50, 849–856. [CrossRef] [PubMed]
- Ajawatanawong, P.; Yanai, H.; Smittipat, N.; Disratthakit, A.; Yamada, N.; Miyahara, R.; Nedsuwan, S.; Imasanguan, W.; Kantipong, P.; Chaiyasirinroje, B.; et al. A novel Ancestral Beijing sublineage of *Mycobacterium tuberculosis* suggests the transition site to Modern Beijing sublineages. *Sci. Rep.* 2019, *9*, 13718. [CrossRef]
- Hanekom, M.; van Pittius, N.G.; McEvoy, C.; Victor, T.; Van Helden, P.; Warren, R. Mycobacterium tuberculosis Beijing genotype: A template for success. Tuberculosis 2011, 91, 510–523. [CrossRef]
- Devi, K.R.; Bhutia, R.; Bhowmick, S.; Mukherjee, K.; Mahanta, J.; Narain, K. Genetic Diversity of *Mycobacterium tuberculosis* Isolates from Assam, India: Dominance of Beijing Family and Discovery of Two New Clades Related to CAS1\_Delhi and EAI Family Based on Spoligotyping and MIRU-VNTR Typing. *PLoS ONE* 2015, 10, e0145860. [CrossRef]