Article

Snapshot of *Mycobacterium tuberculosis* phylogenetics from an Indian State of Arunachal Pradesh bordering China.

S. Rashmi Mudliar[#1] Umay Kulsum[#1], Syed Beenish Rufai[2,3], Mika Umpo[4], Moi Nyori[5], Sarman Singh[1,*]

 # Contributed Equally

[1] Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India.
[2] Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Center, Montreal, QC H4A 3J1, Canada.
[3] McGill International TB Center, Montreal, QC H4A 3J1, Canada.
[4] Tomo Riba Institute of Health & Medical Sciences, Arunachal Pradesh, India.
[5] State TB Cell, Arunachal Pradesh India.
* Correspondence: Prof. Sarman Singh, MD, Department of Microbiology, All India Institute of Medical Sciences, Bhopal-462020 (India)
Phone: +91-0755- 2672317
Email: sarman_singh@yahoo.com, sarman.singh@gmail.com

Abstract: Uncontrolled transmission of *Mycobacterium tuberculosis* (*M. tuberculosis*, MTB) drug resistant strains is a challenge to control efforts of global tuberculosis programme. 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 total 200 suspected MDR-MTB isolates, 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 *Ser*531Leu 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 first insight in identifying the ongoing transmission of two virulent Beijing sub-lineages associated with TB drug resistance.

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

1. Introduction

India is leading in highest rates of tuberculosis (TB) incidence and mortality globally with estimation of 2.69 million cases [1,2]. Although, drug resistant tuberculosis (DR-TB) is a major public health disquiet 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 remains 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 difficult to reach areas with high endemicity present a major constraint on DR-TB treatment. It is estimated that around 56% of MDR-TB cases in India remain un-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

diagnosed [4]. Arunachal Pradesh, one of the states in the northeastern region of India bordering China with 80% area covered with forest, mostly with hilly terrains, has awakened consciousness of NTEPs due to 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 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]. In human TB, several molecular epidemiological studies have proposed that certain types of *M. tuberculosis* strains can be especially prone to drug resistance acquisition and may have rapid transmission rates [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 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 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 discriminate between strains [10].

With advancement of Next generation sequencing, Whole genome sequencing (WGS) data is 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 approaches were previously used for categorization of lineages and sub lineages however comparative analysis has led to the use of single nucleotide polymorphisms (SNPs) provides valuable insights into the epidemiology of circulating strains and are 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 multi locus sequence typing (cgMLST) based on entire allele change are 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 is not well known from 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 identification of hyper virulent strains/clones that are circulating 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

Study setting and Sample collection:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

A total of 200 sputum samples (one sample per patient) suspected of MDR-TB were collected from six 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. 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 consents were 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.

Decontamination of samples and Bactec MGIT 960 culture inoculation:

Sputum samples were processed using NALC-NaOH method [15]. Briefly, minimum 3ml 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 2ml of phosphate buffer and mixed well. Smears were prepared for Ziehl-Neelson (ZN) staining.

Five hundred microliter of decontaminated sample was inoculated in Bactec MGIT 960 culture tubes containing 800µl 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). Left over decontaminated sample aliquots were stored at -80^{0}C for future use.

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].

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 manufacturer's protocol [18]. DST was performed on Day 1 and Day 2 by single dilution (0.5ml of 1:100 dilution inoculum for growth control (GC) and 0.5ml of inoculums directly in four drug panel tubes) and Day 3 to Day 5 (0.5ml 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.

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 pur-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

chased from Sigma-Aldrich Corporation (St. Louis MO, USA) and were chemically in the form of powder. Stock solution of drugs AMK (1mg/ml), MFX(1mg/ml), LFX (1mg/ml) and LNZ (1mg/ml) were prepared as per the instruction and sterilized through 0.22µm 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 strain were used as quality control strain. Second line DST was performed as per the protocol [19-21]. As second line drug (SLD) panels do not exist in MGIT 960 system thus was registered as one of SIRE panels in order to get printable report and user manually entered drug-testing results on the reports obtained.

Whole Genome Sequencing

Genomic DNA extraction from 65 clinical isolates (representatives of sensitive and resistant groups as described in Fig 1) of *M. tuberculosis* complex were 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 plattform (Illumina, San Diego, CA, USA).

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].

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. Annotation and filtering of variants was done using SNPEFF 5.0e [29] and SnpSift [30].

Assignment of principle genetic groups:

To assign 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 principle genetic groups as 1, 2 or 3 respectively as explained previously [31].

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 done on the basis of SNP's [13]. Construction of UPGMA tree was done by concatenating SNPs and visualized using iToLv6 software [32].

Phylogenetic analysis and construction of cgMLST:

Schema for cgMLST was setup with efficient Workflow for a Blast Score Ratio Based Allele Calling Algorithm (ChewBBACA) [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 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 (PRJNA235648, PRJNA223559, PRJNA229273, PRJNA229212, PRJNA229320, PRJNA229266), Lineage 2 (PRJNA219760, PRJNA226779, PRJNA267047, CCDC5180, CCDC5079, PRJNA229630, 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.

Data analysis:

All of the gDST, phenotypic DST, PGG data 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 total 200 patients included in the study, 91 (45.5%) were male and 109 (54.5%) females with mean age (±standard deviation) of 29.52±13.21 and 27.85±14.17 years respectively. The majority of cases were adults 183 (91.5%) and 17 (8.5%) were from paediatric age group.

3.2. Bactec MGIT 960 culture results and identification M. tuberculosis complex isolates:

Of total 200 samples 126 (63%) were smear positive and 74 (37%) smear negative. Of total 200 cultures inoculated in Bactec MGIT 960 145 (72.5%) were flagged positive with average turnaround time (TAT) of eighteen 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. (**Fig 1**)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

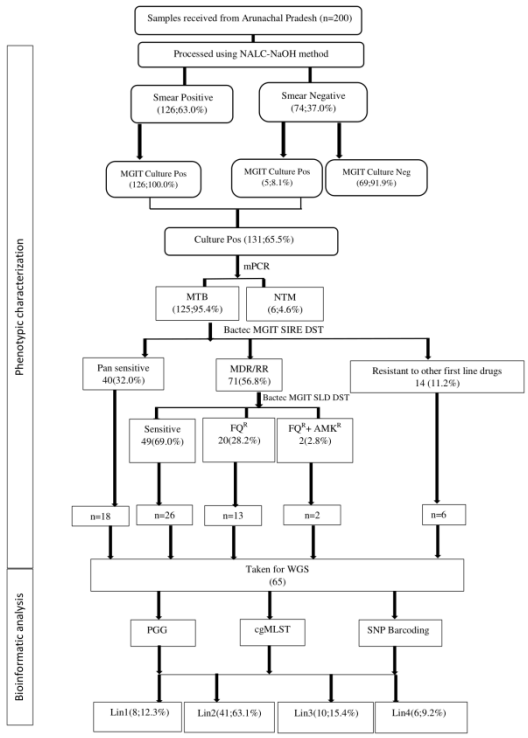

Fig 1: Workflow showing the phenotypic DST result and bioinformatics analysis of the isolates selected for WGS. FQ^R: Fluoroquinolone Resistant; AMK^R: Amikacin Resistant

SIRE drug susceptibility pattern by MGIT 960						Second line drug susceptibility pattern by MGIT 960					
STR	**INH**	**RIF**	**EMB**	**n**	**%**	**AMK**	**MFX**	**LFX**	**LNZ**	**n**	**%**
R	R	R	R	37	29.6%	S	R	R	S	13	35.1%
						R	R	R	S	2	5.4%
						S	R	S	S	2	5.4%
						S	S	S	S	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%	-	-	-	-	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

STR^S = 74	INH^S= 48	RIF^S =54	EMB^s = 78	n=125	AMK^S = 69	MFX^S = 49	LFX^S =51	LNZ^S=71	n=71
STR^R =51	INH^R = 77	RIF^R =71	EMB^R =47		AMK^R =2	MFX^R = 22	LFX^R = 20		

STR^S: Streptomycin susceptible; STR^R: Streptomycin resistant; INH^S: Isoniazid susceptible; INH^R: Isoniazid resistant; RIF^S: Rifampicin susceptible; RIF^R: Rifampicin resistant; EMB^s: Ethambutol susceptible; EMB^R: Ethambutol resistant; AMK^S : Amikacin susceptible; AMK^R: Amikacin resistant; MFX^S : Moxifloxacin susceptible; MFX^R: Moxifloxacin resistant; LFX^S : Levofloxacin susceptible; LFX^R: Levofloxacin resistant; LNZ^S: Linezolid susceptible

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**)

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 (1mg/ml), LFX(1mg/ml), LNZ(1mg/ml) and MOX(1mg/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. Pattern 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:

Total sixty-five isolates sequenced were analysed for mutations conferring drug resistance in genes associated with first-line and second-line drug resistance (**Table 2**). Genomes sequences of each isolate was 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.

Table2: Patterns of mutation resulting from WGS analysis and their association with phenotypic DST

Drug	Gene	Phenotypic DST	Whole Genome Sequencing Results					
		MGIT Result	Mutation	No of Isolates (n=41)	Lineages			
					Lin1	Lin2	Lin3	Lin4
RIF	*rpoB*	R	Ser531Leu	21(51.2%)	4(4/24)	15(15/24)	2(2/24)	-
		R	Leu511Pro Phe505Leu*	3(7.3%)	-	3(3/3)	-	-
		R	Leu511Pro His526Gln	1(2.4%)	-	1(1/1)	-	-
		R	Leu511Pro His526Gln Phe505Leu*	2(4.9%)	-	2(1/1)	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

		R	Asp516Val	2(4.9%)	-	1(1/2)	-	1(1/1)
		R	Asp516Tyr	1(2.4%)	-	1(1/1))	-	-
		R	His526Tyr	3(7.3%)	1(1/3)	2(2/3)	-	-
		R	His526Asp	3(7.3%)	-	3(3/3)	-	-
		R	Gln513Pro	1(2.4%)	-		-	1(1/1)
	rpoB	R	Ser531Leu	1(2.4%)	-	1(1/1)	-	-
			Ile561Val*					
	rpoC		Ile572Thr					
	rpoB	R	Ser531Leu	1(2.4%)	1(1/1)	-	-	-
	rpoA		Gly112Ser					
	rpoB	R	Ser531Leu	1(2.4%)	-	1(1/1)	-	-
	rpoA		Gly319Lys					
	rpoB	R	Ser531Leu	1(2.4%)	-	-	-	1(1/1)
	rpoA		Val264Gly					
	rpoB	S	Leu545Met*	1(2.4%)	-	1(1/1)	-	-
Total				**n=42 (R=41, S=1)**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (n=45)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
INH	*KatG*	R	Ser315Thr	35(77.8%)	3(3/35)	27(27/35)	3(3/35)	2(2/35)
		R	Ser450Leu	1(2.2%)	1(1/1)	-	-	-
	inhA	R	Ser94Ala	2(4.4%)	2(2/2)	-	-	-
	ahp	R	52C>T	2(4.4%)	2(2/2)	-	-	-
	Fab	R	15C>T	1(2.2%)		1(1/1)	-	-
	Kat	R	Ser140Gly	1(2.2%)		1(1/1)	-	-
			15C>T					
	Fab	R	Ser315Thr	1(2.2%)		1(1/1)	-	-
			17C>T					
		R	Ser315Thr	2(4.4%)		2(2/2)	-	-
			15C>T					
Total				**n=45 (R=45)**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (N=30)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
EMB	*embB*	R	Met306Val	18(60.0%)	-	16(16/18)	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

		R	Gly406Asp	1(3.3%)	-	-	1(1/1)	-
		R	Met306Ile	4(13.3%)	1(1/1)	-	-	-
		R	Asp354Ala	1(3.3%)	-	1(1/1)	-	-
		R	Met306Leu	1(3.3%)	-	-	1(1/1)	-
	embB	R	Glu405Asp	1(3.3%)	-	1(1/1)	-	-
		R	Gln853Pro Met306Val	2(6.7%)	-	2(2/2)	-	-
	embA embA	R	Met306Val 12C>T	2(6.7%)	-	2(2/2)	-	-
Total				**n=30 (R=30)**				
Drug	**Gene**	**Phenotypic DST**		**Whole Genome Sequencing Results**				
		MGIT Result	**Mutation**	**No of Isolates (N=26)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
STR	*rpsL*	R	Lys43Arg	22(84.6%)	-	21(21/22)	1(1/1)	-
		R	Lys88Arg	2(7.7%)	-	2(2/2)	-	-
	rrs	R	514 A>C	2(7.7%)	-	2(2/2)	-	-
Total				**n=26 (R=26)**				
Drug	**Gene**	**Phenotypic DST**		**Whole Genome Sequencing Results**				
		MGIT Result	**Mutation**	**No of Isolates (n=16)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
FQ	*gyrA*	R	Asp94Gly	6(37.5%)	-	5(5/6)	1(1/6)	-
		R	Asp94Tyr	2(12.5%)	-	2(2/2)	-	-
		R	Asp94Asn	2(12.5%)	-	2(2/2)1	-	-
		R	Asp94Ala	1(6.3%)	-	1(1/1)	-	-
		R	Ala90Val	1(6.3%)	-	-	-	1(1/1)
		R	Asp94His	1(6.3%)	-	1(1/1)	-	-
	gyrB	R	Ile486Leu	1(6.3%)	-	1(1/1)	-	-
		R	Asp461His	1(6.3%)	-	-	-	1(1/1)
		R	Ala504Val	1(6.3%)	-	-	-	-
Total				**n=16 (R=16)**				
Drug	**Gene**	**Phenotypic DST**		**Whole Genome Sequencing Results**				
		MGIT	**Mutation**	**No of**			**Lineages**	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

		Result		Isolates (n=10)	Lin1	Lin2	Lin3	Lin4
PZA	*pncA*	R	Asp49Ala	5(50.0%)	-	5(5/5)	-	-
		R	Gly108Arg	2(20.0%)	-	2(2/2)	-	-
		R	11A>G	2(20.0%)	-	2(2/2)	-	-
		R	Asp136Tyr	1(10.0%)	1(1/1)	-	-	-
Total				**n=10 (R=10)**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (n=2)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
AMK	*rrs*	R	1484 G>T	1(50.0%)	-	1(1/1)	-	-
		R	1401 A>G	1(50.0%)	-	1(1/1)	-	-
Total				**n=2 (R=2)**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (n=8)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
ETH	*inha*	NA	Ser94Ala	2(25.0%)	2(2/2)	-	-	-
	fab	NA	15C>T	3(37.5%)	-	3(3/3)	-	-
		NA	17G>T	1(12.5%)	-	-	-	1(1/1)
	ethA	NA	886_886del	2(25.0%)	-	1(1/2)	-	1(1/2)
Total				**n=8**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (n=2)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
Cysr.	*alr*	NA	Met343Thr	2(100.0%)	-	2(2/2)	-	-
Total				**n=2**				

Drug	**Gene**	**Phenotypic DST**	**Whole Genome Sequencing Results**					
		MGIT Result	**Mutation**	**No of Isolates (n=2)**	**Lineages**			
					Lin1	**Lin2**	**Lin3**	**Lin4**
PAS	*thy*	NA	16C>T	1(50.0%)	-	1(1/1)	-	-
	folC	NA	Ile43Thr	1(50.0%)	-	1(1/1)	-	-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

Total	**n=2**

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)

Of total sixty-five 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 sixty-five sequenced genomes were also analysed 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 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 **Table2.** Variant densities of each genome against M. tuberculosis H37Rv was generated using Blast Ring Image Generator BRIGv0.95 and is shown in **Fig 2.**

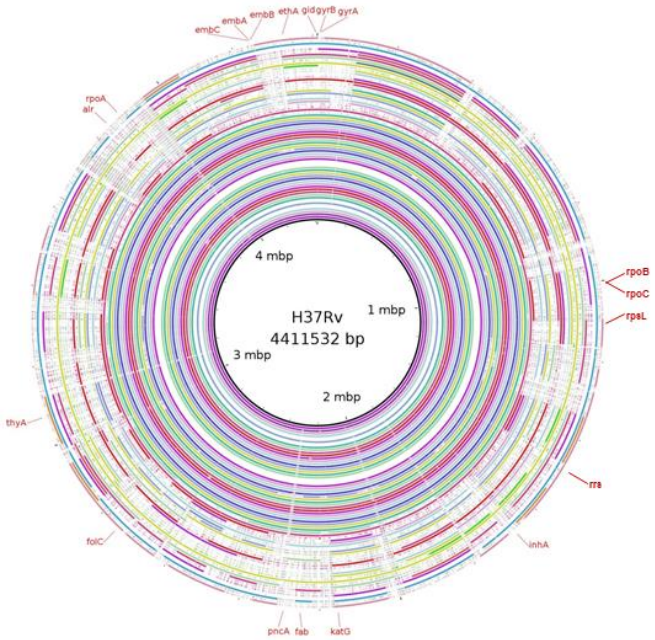

Fig 2. Variant densities of the 65 sequenced genomes against *M. tuberculosis* H37Rv using BRIG v0.95.

3.6. Phylogenetic analysis and identification of lineages based on cgMLST and SNP barcoding:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

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 (**Fig 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.

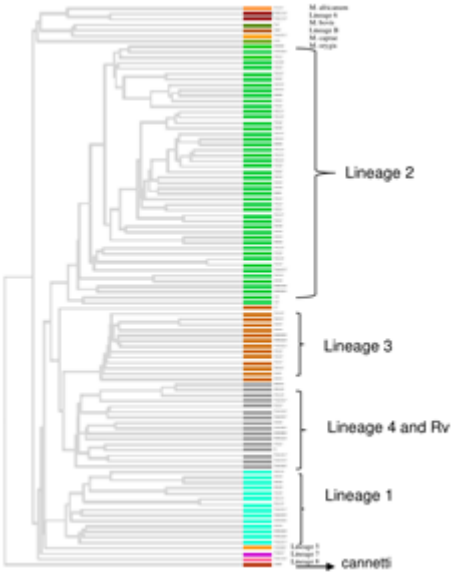

Fig 3: Phylogenetic analysis (UPGMA tree) based on cgMLST association of 104 genomes (65 isolates under this study and 33 lineage-defined genomes and 6 reference genomes of MTBC complex viz. *M. africanum, M. bovis, M. caprae, M. orygis, M. tuberculosis and M. cannettii*). *M. cannettii* was used to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

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 (**Fig 4**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 December 2021 doi:10.20944/preprints202112.0283.v1

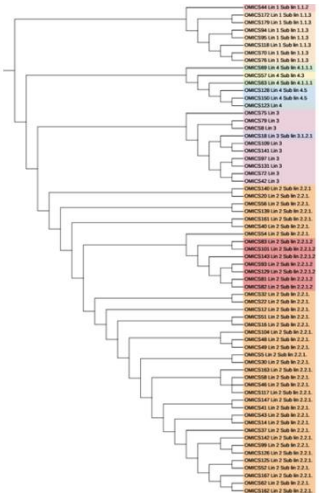

Fig 4: Phylogenetic analysis (UPGMA tree) of 65 sequenced genomes using SNP barcoding and assigned lineages as well as sublineages shown with different colours. Lineage 2 with sublineage 2.2.1 was found to be predominant in the state of Arunachal Pradesh.

3.7. Phylogenetic analysis based on principle genetic 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 sublineage classification, all the genomes belonging to sublineage of lineage 1, 2 and 3 were clustered in PGG1 group while sublineages 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 sublineage classification for each isolate is shown in **Supplementary Table1**.

4. Discussion

With increasing drug resistant TB cases in India, it becomes pivotal to recognize 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]. Inhospitable topography and challenging climatic conditions make rendering of health services and access to TB centers difficult in Arunachal Pradesh. 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 Arunachal Pradesh region of India. This will be the first kind of study from this region which will help in defining the transmission of strains associated with drug resistance circulating in Arunachal Pradesh region of India.

Our report for MDR-TB from Arunachal Pradesh is 52.8%, slightly lower than previously reported 79.3%, the variation may be due to a lower number of sample sizes in previous study and results based on DNA based line probe Assay rather than Bactec MGIT 960 which detects viable bacilli [36]. However, no reports of FQ resistance have been reported from Arunachal Pradesh earlier.

There is mounting evidence that strain diversity plays a role in transmission of disease. 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 (Fig 1). All these methods have been used in various phylogenetic studies for standardized phylogenetic assignment of diseases and outbreak resolutions [13,38,39]. 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 is also reported from previous studies and are consistent with our finding [31,39].(Supplementary Table1).

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* (Fig1).

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 Table1). We found Lineage 2 (East Asian) as 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 presence of SNP markers at mutT2 codon Gly58Ala, ogt Gly12Gly specific to modern Beijing as reported in earlier studies [40,41]. Of total 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 the clone 2.2.1.2 were multi drug resistant and 2 (28.6%) were resistant to FQ. Beijing sub-lineage 2.2.1 and 2.2.1.2 is reported from various parts of 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 since 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]. Total 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 sublineage 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 Assam region of India by Devi *et al*, which is consistent with our study [50].

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 may prevent transmission of these strain across other parts of India. Clonal

expansion of these strains in Arunachal Pradesh in 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 presence of Lineage 1 as endemic in Arunachal Pradesh. These findings may have important implications for control and prevention of TB in northeastern part of India, Arunachal Pradesh.

Author Contributions: Mudliar RS standardized and performed all experiments and analyzed the data. Kulsum U performed bioinformatic analysis and contributed to manuscript writing. Rufai SB contributed to reviewing and writing of manuscript. Umpo M and Nyori M helped in collection of TB samples and managed transportation of samples to AIIMS Bhopal. Singh S supervised and coordinated the work, finalized the manuscript, arranged reagents and chemicals.

Funding: This research and APC was funded by 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 Ms Payal Soni and Mr Mukesh Patel for technical help.

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

References:

1. World Health Organization. 2020. GlobalTuberculosis Report 2020.World Health Organization, Geneva, Switzerland.
2. National Tuberculosis Elimination Programme. India TB Report 2020. National Tuberculosis Elimination Programme, New Delhi, India.
3. Lange C, Dheda K, Chesov D, Mandalakas AM, Udwadia Z, Horsburgh CR. 2019. Management of drug-resistant tuberculosis. Lancet 394:953–966.
4. Husain AA, Kupz A, Kashyap RS. 2021. Controlling the drug-resistant tuberculosis epidemic in India: challenges and implications. Epidemiol Health 43:e2021022.
5. Singh S. 2014. Early detection of multi-drug resistant tuberculosis in India using GenoType MTBDRplus assay & profile of resistance mutations in Mycobacterium tuberculosis. Indian J Med Res 140:477–479.
6. Gao Q, Kripke KE, Saldanha AJ, Yan W, Holmes S, Small PM. 2005. Gene expression diversity among Mycobacterium tuberculosis clinical isolates. Microbiology (Reading) 151:5–14.
7. Anh DD, Borgdorff MW, Van LN, Lan NT, van Gorkom T, Kremer K, van Soolingen D. 2000. Mycobacterium tuberculosis Beijing genotype emerging in Vietnam. Emerg Infect Dis 6:302–305.
8. Kato-Maeda M, Bifani PJ, Kreiswirth BN, Small PM. 2001. The nature and consequence of genetic variability within Mycobacterium tuberculosis. J Clin Invest 107:533–537.
9. Mekonnen A, Merker M, Collins JM, Addise D, Aseffa A, Petros B, Ameni G, Niemann S. 2018. Molecular epidemiology and drug resistance patterns of Mycobacterium tuberculosis complex isolates from university students and the local community in Eastern Ethiopia. PLoS One 13:e0198054.
10. Ford C, Yusim K, Ioerger T, Feng S, Chase M, Greene M, Korber B, Fortune S. 2012. Mycobacterium tuberculosis--heterogeneity revealed through whole genome sequencing. Tuberculosis (Edinb) 92:194–201.
11. Cohen KA, Manson AL, Desjardins CA, Abeel T, Earl AM. 2019. Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges. Genome Med 11:45.
12. Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM. 2006. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci U S A 103:2869–2873.
13. Kohl TA, Diel R, Harmsen D, Rothgänger J, Walter KM, Merker M, Weniger T, Niemann S. 2014. Whole-genome-based Mycobacterium tuberculosis surveillance: a standardized, portable, and expandable approach. J Clin Microbiol 52:2479–2486.

14. Jones RC, Harris LG, Morgan S, Ruddy MC, Perry M, Williams R, Humphrey T, Temple M, Davies AP. 2019. 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 57:e02025-18.
15. Rufai SB, Singh A, Kumar P, Singh J, Singh S. 2015. Performance of Xpert MTB/RIF Assay in Diagnosis of Pleural Tuberculosis by Use of Pleural Fluid Samples. J Clin Microbiol 53:3636–3638.
16. Gopinath K, Singh S. 2009. 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 107:425–435.
17. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, Yadav J, Naaz F, Ravikumar R, Buggi S, Suar M, Gupta UD, Pandey A, Chauhan DS, Tripathy SP, Gowda H, Prasad TSK. 2019. Whole Genome Sequencing of Mycobacterium tuberculosis Clinical Isolates From India Reveals Genetic Heterogeneity and Region-Specific Variations That Might Affect Drug Susceptibility. Frontiers in Microbiology 10:309.
18. Siddiqi S, Ahmed A, Asif S, Behera D, Javaid M, Jani J, Jyoti A, Mahatre R, Mahto D, Richter E, Rodrigues C, Visalakshi P, Rüsch-Gerdes S. 2012. Direct drug susceptibility testing of Mycobacterium tuberculosis for rapid detection of multidrug resistance using the Bactec MGIT 960 system: a multicenter study. J Clin Microbiol 50:435–440.
19. Rufai SB, Singh J, Kumar P, Mathur P, Singh S. 2018. Association of gyrA and rrs gene mutations detected by MTBDRsl V1 on Mycobacterium tuberculosis strains of diverse genetic background from India. Sci Rep 8.
20. World Health Organization 2020.Technical Manual for drug susceptibility testing of medicines used in the treatment of tuberculosis. World Health Organization, Geneva, Switzerland.
21. Kim H, Seo M, Park YK, Yoo J-I, Lee YS, Chung GT, Ryoo S. 2013. Evaluation of MGIT 960 System for the Second-Line Drugs Susceptibility Testing of Mycobacterium tuberculosis. Tuberc Res Treat 2013:108401.
22. Somerville W, Thibert L, Schwartzman K, Behr MA. 2005. Extraction of Mycobacterium tuberculosis DNA: a question of containment. J Clin Microbiol 43:2996–2997.
23. Black PA, de Vos M, Louw GE, van der Merwe RG, Dippenaar A, Streicher EM, Abdallah AM, Sampson SL, Victor TC, Dolby T, Simpson JA, van Helden PD, Warren RM, Pain A. 2015. Whole genome sequencing reveals genomic heterogeneity and antibiotic purification in Mycobacterium tuberculosis isolates. BMC Genomics 16:857.
24. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 1. EMBnet.journal 17:10–12.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.
26. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624.
27. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
29. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92.
30. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 3:35.
31. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A 94:9869–9874.
32. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259.
33. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carriço JA. 2018. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microb Genom 4.
34. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13:87.
35. Chatterjee S, Poonawala H, Jain Y. 2018. Drug-resistant tuberculosis: is India ready for the challenge? BMJ Glob Health 3:e000971.
36. Singhal R, Myneedu VP, Arora J, Singh N, Sah GC, Sarin R. 2014. 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 140:501–506.
37. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rüsch-Gerdes S, Supply P, Kalinowski J, Niemann S. 2013. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS Med 10:e1001387.
38. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun 5:4812.

39. Kohl TA, Harmsen D, Rothgänger J, Walker T, Diel R, Niemann S. 2018. Harmonized Genome Wide Typing of Tubercle Bacilli Using a Web-Based Gene-By-Gene Nomenclature System. EBioMedicine 34:131–138.
40. Bergval I, Sengstake S, Brankova N, Levterova V, Abadía E, Tadumaze N, Bablishvili N, Akhalaia M, Tuin K, Schuitema A, Panaiotov S, Bachiyska E, Kantardjiev T, de Zwaan R, Schürch A, van Soolingen D, van 't Hoog A, Cobelens F, Aspindzelashvili R, Sola C, Klatser P, Anthony R. 2012. Combined species identification, genotyping, and drug resistance detection of Mycobacterium tuberculosis cultures by MLPA on a bead-based array. PLoS One 7:e43240.
41. Nieto Ramirez LM, Ferro BE, Diaz G, Anthony RM, de Beer J, van Soolingen D. 2020. Genetic profiling of Mycobacterium tuberculosis revealed "modern" Beijing strains linked to MDR-TB from Southwestern Colombia. PLoS One 15:e0224908.
42. Rufai SB, Sankar MM, Singh J, Singh S. 2016. 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 5 Suppl 1:S197–S198.
43. Gupta A, Sinha P, Nema V, Gupta PK, Chakraborty P, Kulkarni S, Rastogi N, Anupurba S. 2020. Detection of Beijing strains of MDR M. tuberculosis and their association with drug resistance mutations in katG, rpoB, and embB genes. BMC Infect Dis 20:752.
44. Liu Y, Jiang X, Li W, Zhang X, Wang W, Li C. 2017. The study on the association between Beijing genotype family and drug susceptibility phenotypes of Mycobacterium tuberculosis in Beijing. Sci Rep 7:15076.
45. San LL, Aye KS, Oo NAT, Shwe MM, Fukushima Y, Gordon SV, Suzuki Y, Nakajima C. 2018. Insight into multidrug-resistant Beijing genotype Mycobacterium tuberculosis isolates in Myanmar. Int J Infect Dis 76:109–119.
46. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, Lan NH, Nhu NTQ, Hai HT, Ha VTN, Thwaites G, Edwards DJ, Nath AP, Pham K, Ascher DB, Farrar J, Khor CC, Teo YY, Inouye M, Caws M, Dunstan SJ. 2018. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat Genet 50:849–856.
47. Ajawatanawong P, Yanai H, Smittipat N, Disratthakit A, Yamada N, Miyahara R, Nedsuwan S, Imasanguan W, Kantipong P, Chaiyasirinroje B, Wongyai J, Plitphonganphim S, Tantivitayakul P, Phelan J, Parkhill J, Clark TG, Hibberd ML, Ruangchai W, Palittapongarnpim P, Juthayothin T, Thawornwattana Y, Viratyosin W, Tongsima S, Mahasirimongkol S, Tokunaga K, Palittapongarnpim P. 2019. A novel Ancestral Beijing sublineage of Mycobacterium tuberculosis suggests the transition site to Modern Beijing sublineages. Sci Rep 9:13718.
48. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, Lan NH, Nhu NTQ, Hai HT, Ha VTN, Thwaites G, Edwards DJ, Nath AP, Pham K, Ascher DB, Farrar J, Khor CC, Teo YY, Inouye M, Caws M, Dunstan SJ. 2018. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat Genet 50:849–856.
49. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. 2011. Mycobacterium tuberculosis Beijing genotype: a template for success. Tuberculosis (Edinb) 91:510–523.
50. Devi KR, Bhutia R, Bhowmick S, Mukherjee K, Mahanta J, Narain K. 2015. 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 10:e0145860.