Insight into the PZA resistance whole genome of Mycobacterium tuberculosis isolates from Khyber Pakhtunkhwa, Pakistan

Muhammad Tahir Khan
Capital University of science and Technology islamabad Pakistan
Sajid Ali
Quaid-i-Azam University
anwar Sheed Khan
Kohat University of Science and Technology
Arif Ali
Shanghai Jiao Tong University
Abbas Khan
Shanghai Jiao Tong University
Aman Chandra Kaushak
Jiangnan University
Muhammad Irfan
University of Florida
Satheshkumar Chinnasamy
Shanghai Jiao Tong University
Shulin Zhang
Shanghai Jiao Tong University
Yu-Juan Zhang
Chongqing Normal University
Zhelie Cui
Shanghai Jiao Tong University
Amie Wei
Shanghai Jiao Tong University
Yanjie Wang
Shanghai Jiao Tong University
Mingzhu Zhao
Shanghai Jiao Tong University
Kejia Liu
Shanghai Jiao Tong University
Muhammad Tariq Zeb

Veterinary research institute pakistan

Heng wang

Shanghai Jiao Tong University

Dong Qing Wei (dqwei@sjtu.edu.cn)

Shanghai Jiao Tong University School of Life Sciences and Biotechnology https://orcid.org/0000-0003-4200-7502

Research article

Keywords: Genome, PZA, resistance

Posted Date: June 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29899/v1

Abstract

Background

Tuberculosis (TB) is a global public health issue, getting worse due to emergence of resistance. Pyrazinamide (PZA) is first-line antimicrobial drugs used against non-replicated *Mycobacterium tuberculosis* (MTB). Data is scarce about whole genome sequencing of PZA resistance (PZA-R) in Khyber Pakhtunkhwa (KP) province of high burden country, Pakistan. In the current study we aimed to find the most common mutations in PZA-R MTB isolates in association with other candidate genes in a whole genome sequence (WGS). Samples were collected from TB suspects and drug susceptibility testing (DST) was performed according according to the WHO standards. The resistant samples were subjected for whole genome sequencing (WGS). The sequence data was through MTBseq and Total Genotyping Solution for *Mycobacterium tuberculosis* (TGS-TB). Metabolic model was analyzed, using RAST server.

Results

Among the three whole genome sequences, (NCBI BioProject Accession: PRJNA629298, PRJNA629388) 1997, 1162, and 2053 mutations including indel, was detected. Diverse variability has been detected in the membrane proteins PE and PPE, modulating the host immune response. Nine mutations in coding and promotor region have been detected in *pncA* with one novel (T-4C) variants. Mutations in the other drug candidate genes, KatG, rpoB have also been detected.

Conclusion

The metabolic model shows a distinct property. Diversity of variants has been detected in majority of MTB essential genes, functions from cell growth to cell signaling. The current study provides useful information, associated with geographic specific strains for biomarkers development and better management of drug resistance isolates.

Background

Tuberculosis (TB) is a global public health issue, getting worse due to emergence of resistance against first and second-line drugs. Under Directly observed treatment, short-course (DOTS), TB takes 6 to 8 months while the treatment of drug resistance takes a variable length of time upto 28 months depending upon drug resistance profile of bacilli [1, 2].

The most common mechanism behind drug resistance has been accounted as mutations in the target proteins [3–11]. Among the first line anti-TB drugs, pyrazinamide (PZA) is the only drug which effective against latent *Mycobacterium tuberculosis* (MTB). Resistance to pyrazinamide (PZA) has been associated with *pncA* encoded Pyrazinamidase (PZase) in more than 70% cases [12–14]. Three regions

in PZase 3 to 17, 61 to 85, and 132 to 142 has been associated with catalytic activity [14, 15]. Another minor cause of resistance has been associated with RpsA mutations [2, 19–25]. Residues in the fourth domain, have been associated with catalytic conversion of PZA into the active form pyrazinoic acid (POA).

Although in our recent studies, the prevalence and mechanism of PZA resistance behind mutations in pncA and rpsA have been investigated, [19, 28–31] however, whole genome sequencing (WGS) is one-stop method for epidemiological and molecular study of drug susceptibility testing. The potential clinical value and public health impact in the areas of DST for patient management and tracing of transmission chains for timely public health intervention have also been discussed [32–34]. The next-generation sequencing (NGS) is more affordable, fast, screening approach which allows for genetic variation in geographic specific MTB isolates. The WGS assists in phylogeny of the global distribution and tracing of transmission chains [35–38].

Numerous studies have highlighted the importance of WGS in the fields to record mutation rate, resistance, novel drugs, and evolution of the *Mycobacterium tuberculosis* complex (MTBC) [32, 34, 39–41]. The WGS lets for the examination of diverse mutations in the genome of circulating isolates in the geographic specific location conferring variations in pathogenicity and resistance. Mutations level as well as their occurrence inside and outside coding regions depicts the overall pathogenic variation among the circulating isolates that could be recognized from entire sequencing of genomes. The finding of novel targets, and their involvement in drug resistance could be accessed through mutations from entire genomes. The genome sequences may help to unveil the genetic variations in coding regions which are not directly confer resistance (Compensatory mutations) but rather compensate for the fitness cost of resistance [42]. In previous studies, we have investigated the PZA drug sensitivity testing (DST) followed by sequencing to find mutations in *pncA* and *rpsA* genes associated with PZA-resistance [19, 43] (Accession No. MH461111). In the current paper, we aimed to analyze the whole genome of PZA-R MTB to investigate the mutations in *pncA, rpsA, panD,* and other candidate genes along with the diversity of variations in the circulating isolates for better management of TB in this high burden area.

Results

Analyzing whole-genome sequencing data of MTB isolates enables both comprehensive antibiotic resistance, summarizing, outbreak surveillance, and also the identification of recent transmission chains. A total of five WGS results were analyzed, however, the two samples (ba1, ba2) were not suitable to be analyzed broadly through MTBSeq pipeline. Therefore, three samples (ba4, ba5, ba9) have been investigated comprehensively (NCBI BioProject Accession No: PRJNA629298, PRJNA629388). Large number of variants have been detected in majority of important genes including antituberculosis drug targets (Fig 1). Majority of these variant were single nucleotide polymorphism including synonymous and non-synonymous. However, insertion and deletion were also detected in significant numbers. About one over third part of whole genome (about 4000 genes) harbored mutations.

Mutations in drug targets

Mutations in *pncA, rpsA* and other first-line drug targets in coding as well as in promotor region have been depicted in table 1, 2, and 3.

The PZA-R samples harbored numerous mutations in the drug target genes. Sample Ba-1 has four promotor region mutation with one novel (T-4C) which has not been reported earlier. NGS strain typing could offer Mtb outbreak and surveillance. The genotype information obtained may help for better management of local TB infections.

Table 1. Mutations detected in the pncA, rpsA genes of PZA resistance isolates.

Sample	PncA Mutation	rpsA Mutation	Location	Lineage	literature
	T-11C	Nil	promoter	4	Reported
	A-7C				Reported
	A-12T				Reported
	T-4C				**Novel**
Ba-2		Indel (s) found		4	
Ba-4	L120R	Nil	Coding	3	Reported
Ba-5	G97D	Nil	Coding	4	Reported
	392-Indel: ACC/ACCCC				
Ba-9	V130G	Nil	coding	4	Reported

Mutations in the other antituberculosis drug targets have been detected (Table 2). The *rpoB* gene encoding the β-subunit of RNA polymerase, has been associated with rifampicin resistance caused by mutations in the 81-base pair region. In the current study four non-synonymous mutations has been detected in the β-subunit of RNA polymerase. SNPs S450L and D435Y, associated with RIF resistance, were outside RRDR of rpoB gene and rare in the previous studies. Among the isoniazid resistance, katG, S315T was a common mutation in all the three samples while inhA promoter region mutation were also detected. A total of 1996 variants have been detected in the resistance samples (Table 3), among which 188, 336, and 1472 were insertion, deletion, and SNPs respectively.

Table 2 Mutations detected in first-line drug targets

Sample	Phylo	Drug Resistance	Gene	Mutation
ba4	Lineage 3	Rifampicin	SNP764817 TG=rpoC	V483G
		Isoniazid	SNP1673425 CT	inhA promoter
		Isoniazid	SNP2155168 CG=katG	S315T
		Ethambutol	SNP4247429AC=embB	M306L
		Rifampicin	(RRDR) 761155CT=rpoB	S450L
		Fluoroquinolones	(QRDR) 7582AG=gyrA	D94G
ba5	Lineage 4	Rifampicin	(RRDR) 761155CT=rpoB	S450L
		Isoniazid	SNP2155168CG=katG	S315T
		Ethambutol	SNP4247431GA=embB	M306I
		Fluoroquinolones	(QRDR) 7570CT=gyrA	A90V
ba9	Lineage 3	Rifampicin	(RRDR) [761109GT=rpoB	D435Y
		Streptomycin	SNP1472359AT=rrs	S172C
		Isoniazid mutation -15	SNP1673425CT	inhA promoter
		Isoniazid	SNP2155168CG=katG	S315T
		Ethambutol	SNP4247431GC=embB	M306I

Table 3. Bases comparison altered in PZA-R isolate and types of polymorphism

Ref Sequence (H37Rv)		PZA-R sequences		SNP Type	
Base	**Frequency**	**Base**	**Frequency**	**Type**	**Freq**
A	411	GAP	336	Insertion	188
T	413	T	353	Deletion	336
G	555	A	359	SNPs	1472
C	617	C	471	**Total**	**1996**
		G	477		
		Total mutations	**1996**		

Sample ba9 and ba5 is unique among the other. The length and branch are also unique in pattern in the phylogenetic tree (colored red). Phylogenetically the sequenced MTB strains showed a unique characteristic, where the spoligotypes pattern is also novel in this study as shown in figures 2, 3, 4. The resultant pattern may have more or less severe drug resistance. However, these strains should be investigated through phenotypical DST to find it minimum inhibitory concentration against first and second-line drugs.

The subsystem difference has been shown through arrow pointing. Interestingly in majority of subsystems, the MTB isolates circulating in Khyber Pakhtunkhwa geography shows a little high number of protein/genes as compared to reference (H37Rv). The metabolic model in the figure 4 has clearly shown the difference between PZA-R and H37Rv suggesting the distinct metabolic characteristics of isolates circulating in this geography.

Circular diagrams have been generated through CGViewer server is shown (Figure 5) in comparison with reference genome. The genomic size of the MTB isolates circulating in this geography is more than the reference H37Rv. The CGView generates graphical maps of circular genomes. Sequences can be supplied in raw, FASTA, GenBank or EMBL format. The server uses BLAST to compare the primary sequence to compare genomes or sequence sets. The BLAST results are converted to a graphical map showing the entire sequence. The CGView Server can aid in the identification of conserved genome segments and differences in gene copy number [49].

Diversity of variation in ESX

A large number of synonymous and non-synonymous mutations have been detected in ESX secretory system proteins (ESX-SSP) (Table 4). However, the effect of these mutations on virulency is needed to be investigated. Mutation colored yellow (Table 4) have been reported earlier [50] present in Beijing strains, where they have been characterized. The presence of such diverse type of mutations in drug resistance strains might be important of geographic specific diversity in ESX secretory system proteins. Six regions have been identified where 4 have been shown, represent T7SS and 2 of prophage regions (Fig 5). The component genes have also been shown in the figure, shows a high variability in mutations.

Table 4: Diversity of variation in ESX protein of MTB isolates, from Khyber Pakhtunkhwa

Sample	Position	Ref	Alle	Substitution	Freq	Rv	Gene*	Gene Product
ba4, ba9	4343784	G	A	Lys157Lys (aag/aaA)	2	Rv3868	eccA1	ESX-1 SSP EccA1
ba4	4366195	T	C	Glu215Gly (gag/gGg)	1	Rv3884c	eccA2	ESX-2 SSP EccA
ba4, ba5, ba9	4366272	G	C	Ala189Ala (gcc/gcG)	3	Rv3884c	eccA2	ESX-2 SSP EccA
ba4, ba9	342146	A	C	Glu6Ala (gaa/gCa)	2	Rv0282	eccA3	ESX-3 SSP EccA
ba4, ba9	342873	C	T	Val248Val (gtc/gtT)	2	Rv0282	eccA3	ESX-3 SSP EccA
ba4, ba9	4345548	G	A	Met170Ile (atg/atA)	2	Rv3869	eccB1	ESX-1 SSP EccB
ba4, ba9	3871246	T	C	Gly417Gly (gga/ggG)	2	Rv3450c	eccB4	ESX-4 SSP EccB4
ba4, ba5, ba9	4379680	C	G	Arg258Pro (cgc/cCc)	3	Rv3894c	eccC2	ESX-2 type VII SSP EccC
ba5	4377447	G	A	Asp1002Asp (gac/gaT)	1	Rv3894c	eccC2	ESX-2 type VII SSP EccC
ba4, ba5, ba9	346275	C	G	Pro214Arg (ccg/cGg)	3	Rv0284	eccC3	ESX-3 SSP EccC3
ba9	347388	A	T	His585Leu (cat/cTt)	1	Rv0284	eccC3	ESX-3 SSP EccC3
ba4, ba5, ba9	3864995	T	C	Ser1082Gly (agc/Ggc)	3	Rv3447c	eccC4	ESX-4 SSP EccC4
ba4, ba5, ba9	2022868	T	C	Ser1204Ser (agt/agC)	3	Rv1783	eccC5	ESX-5 type VII SSP EccC5
ba5	2019942	A	G	Gln229Arg (cag/cGg)	1	Rv1783	eccC5	ESX-5 type VII SSP EccC5
ba4	4355319	C	G	Leu105Val (ctg/Gtg)	1	Rv3877	eccD1	ESX-1 SSP EccD1
ba4, ba5, ba9	4356110	G	C	Leu368Leu (ctg/ctC)	3	Rv3877	eccD1	ESX-1 SSP EccD1
ba9	4371132	C	T	Val185Met (gtg/Atg)	1	Rv3887c	eccD2	ESX-2 SSP EccD

ba4, ba9	353197	C	T	Arg39Cys (cgc/Tgc)	2	Rv0290	eccD3	ESX-3 SSP EccD
ba4, ba9	3869355	T	C	Ile335Thr (atc/aCc)	2	Rv3448	eccD4	ESX-4 SSP EccD4
ba4, ba5, ba9	356528	A	G	Asn217Asp (aat/Gat)	3	Rv0292	eccE3	ESX-3 SSP EccE
ba4, ba5 ba9	4342187	A	G	Asp103Gly (gac/gGc)	3	Rv3866	espG1	ESX-1 EspG
ba4, ba5 ba10	4357597	C	G	Cys729Ser (tgc/tCc)	1	Rv3879c	espK	ESX-1 EspK
ba4, ba5 ba11	4359165	G	C	Thr206Thr (acc/acG)	2	Rv3879c	espK	ESX-1 EspK

Alle; Allele, SSP; secretory system proteins, *; part of T7SS

At least three of the five ESX systems (ESX-1, ESX-2, ESX-3, ESX-4 and ESX-5) are required for full virulence. The first ESX system (ESX-1) is used as attenuated vaccine strains *Mycobacterium bovis* bacille Calmette–Guérin (BCG) and *Mycobacterium microti* [54–59].

Type-VII secretory system (T7SS)

The essential genes of T7SS has been shown in figure 6 while mutations detected in these genes has given in the table 4. Mutations may affect host immune response. Studies should be carried to address the effect of these mutations on the immune response for better management of TB.

Diversity of PE/PPE

PE/PPE genes are highly polymorphic, present only in pathogenic mycobacterial species (Fig. 6). Hypervariable genes have been detected in recent sublineages (IV and V) of the PE and even within the sublineages. Sequence comparison of MTB isolates, circulating in this high burden country, shows a high variability in comparison with H37Rv. This sequence diversity has been specifically observed in PPE55 (Table 5) that may affect the immune response. The PPE55 gene is specific to the *M. tuberculosis* complex and is a strongly immunogenic protein whose expression correlates with active infection in humans with active clinical TB.

Mycobacterium tuberculosis has evolved from environmental mycobacteria, and some PE/PPE family proteins inherited from environmental mycobacteria that MTB utilizes for survival intracellularly now. Examining immune responses to PE/PPE suggests that PE/PPE epitopes evoking host immune responses, beneficial to the pathogen for survival. In the current study highly variable region in PE and PPE have been detected (Table 5).

Sequence variability in PE_PGRS family gene.

Among the PE-PGRS subfamily, PE_PGRS49 and PE_PGRS 21 exhibited the highest variability (48, 37) followed by PE_PGRS28 and PE_PGRS42 (Fig. 7) A sequence fragment (genome position: 1212336 -1212362) in PE_PGRS 21 of local strains is seemed, to be deleted (Table 6).

In our local strains, the PPE55 when compared with reference strain H37Rv, harbored numerous mutations with 13 non-synonymous, throughout its entire coding region (Table 5). In a previous investigation, PPE55 protein was found highly immunogenic that may be useful for distinguishing amongst latent and subclinical TB [67]. This high variability may contribute a weak immune response, resulting a weak eradication of MTB isolates. However, the complete diversity of mutations among major geographical regions and their affect the immune response is still unclear.

Similarly, PE_PGRS21 and PE_PGRS28 also exhibited a high variability in the local strains (Table 6). A sequence region from 1212336 to 1212362 (CGGCGGTGCCGGCGGTGTCGGCG GTGC) was deleted in PE_PGRS21. While one synonymous and non-synonymous mutation was also detected at amino acid position 268 and 471 (Ala268Ala and Val471Ala). Three non-synonymous mutations Ala438Thr (gcc/Acc), Arg412Gly (cgc/**G**gc), and Ser16Leu (tcg/t**T**g) have also been detected in PE_PGRS28.

Table: 5. Diversity of mutations detected in PPE 55 of PZA resistance isolates

Sample	Position (Genome)	Ref	Type	Allele	Amino Acid alteration	Rv	Gene	Freq*
ba9, ba4	3746409	A	SNP	G	Leu2259Pro (ctg/cCg)	Rv3347c	PPE55	2
ba9, ba4	3749653	G	SNP	A	Gln1178_ (caa/Taa)	Rv3347c	PPE55	2
ba9, ba4	3750177	A	SNP	T	Phe1003Tyr (ttc/tAc)	Rv3347c	PPE55	2
ba9, ba4	3750178	A	SNP	C	Phe1003Val (ttc/Gtc)	Rv3347c	PPE55	2
ba9, ba4	3750185	C	SNP	G	Ser1000Ser (tcg/tcC)	Rv3347c	PPE55	2
ba9, ba4	3750187	A	SNP	T	Ser1000Thr (tcg/Acg)	Rv3347c	PPE55	2
ba9, ba4	3750188	C	SNP	G	Met999Ile (atg/atC)	Rv3347c	PPE55	2
ba9, ba4	3750193	G	SNP	A	Leu998Phe (ctc/Ttc)	Rv3347c	PPE55	2
ba9, ba4	3750205	C	SNP	T	Asp994Asn (gac/Aac)	Rv3347c	PPE55	2
ba9, ba4	3750209	A	SNP	G	Asn992Asn (aat/aaC)	Rv3347c	PPE55	2
ba9, ba4	3750210	T	SNP	G	Asn992Thr (aat/aCt)	Rv3347c	PPE55	2
ba9, ba4	3750407	G	SNP	C	Gly926Gly (ggc/ggG)	Rv3347c	PPE55	2
ba9, ba4	3750417	A	SNP	T	Phe923Tyr (ttc/tAc)	Rv3347c	PPE55	2
ba9, ba4	3750421	T	SNP	C	Ser922Gly (agc/Ggc)	Rv3347c	PPE55	2
ba9, ba4	3750584	G	SNP	A	Asn867Asn (aac/aaT)	Rv3347c	PPE55	2
ba4	3751281	G	SNP	A	*Pro635Leu (ccg/cTg)	Rv3347c	PPE55	1
ba9	3751706	G	SNP	A	Ile493Ile (atc/atT)	Rv3347c	PPE55	1
ba9, ba4	3752003	G	SNP	C	Val394Val (gtC/gtG)	Rv3347c	PPE55	2

ba9, ba4	3752006	G	SNP	A	Asn393Asn (aac/aaT)	Rv3347c	PPE55	2
ba9, ba4	3752007	T	SNP	C	Asn393Ser (aac/agC)	Rv3347c	PPE55	2
ba9, ba4	3752008	T	SNP	C	Asn393Asp (aac/Gac)	Rv3347c	PPE55	2
ba9, ba4	3752012	C	SNP	G	Pro391Pro (ccg/ccC)	Rv3347c	PPE55	2
ba9, ba4	3752207	A	SNP	G	Ile326Ile (att/atC)	Rv3347c	PPE55	2
ba9, ba4	3753116	C	SNP	T	Pro23Pro (ccg/ccA)	Rv3347c	PPE55	2
ba9, ba4	3753164	T	SNP	G	Pro7Pro (cca/ccC)	Rv3347c	PPE55	2

*Freq; frequency, *Pro635Leu; novel

Table: 6. Mutation in PE_PGRS21 and PE_PGRS28 gene

Position	H37Rv (Ref)	MutationType	Local Allele	Substitution/Deletion	Rv	Gene
1212336	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212337	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212338	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212339	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212340	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212341	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212342	T	Del	GAP	————————	Rv1087	PE_PGRS21
1212343	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212344	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212345	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212346	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212347	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212348	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212349	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212350	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212351	T	Del	GAP	————————	Rv1087	PE_PGRS21
1212352	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212353	T	Del	GAP	————————	Rv1087	PE_PGRS21
1212354	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212355	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212356	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212357	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212358	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212359	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212360	T	Del	GAP	————————	Rv1087	PE_PGRS21
1212361	G	Del	GAP	————————	Rv1087	PE_PGRS21
1212362	C	Del	GAP	————————	Rv1087	PE_PGRS21
1212363	C	SNP	T	Ala268Ala (gcc/gcT)	Rv1087	PE_PGRS21

1212971	T	SNP	C	Val471Ala (gtc/gCc)	Rv1087	PE_PGRS21
1636826	C	SNP	A	Gly468Gly (ggg/ggT)	Rv1452c	PE_PGRS28
1636918	C	SNP	T	Ala438Thr (gcc/Acc)	Rv1452c	PE_PGRS28
1636991	T	SNP	C	Gly413Gly (gga/ggG)	Rv1452c	PE_PGRS28
1636996	G	SNP	C	*Arg412Gly (cgc/Ggc)	Rv1452c	PE_PGRS28
1637006	G	SNP	A	Val408Val (gtc/gtT)	Rv1452c	PE_PGRS28
1637009	G	SNP	A	Gly407Gly (ggc/ggT)	Rv1452c	PE_PGRS28
1638182	C	SNP	T	Ser16Ser (tcg/tcA)	Rv1452c	PE_PGRS28
1638183	G	SNP	A	Ser16Leu (tcg/tTg)	Rv1452c	PE_PGRS28
1638188	C	SNP	T	Ala14Ala (gcg/gcA)	Rv1452c	PE_PGRS28
1638191	C	SNP	G	Ala13Ala (gcg/gcC)	Rv1452c	PE_PGRS28
1638194	G	SNP	C	Ala12Ala (gcc/gcG)	Rv1452c	PE_PGRS28

*Arg412Gly; novel

Variants in forkhead-associated domains of MTB

In the current study, a novel deletion of sequence, CCGCGTTGCTCGGGGTAA in **Forkhead-associated (**fhaA) (Rv0020c) domain at genome position 24699-24716 was detected. Further its interacting signaling partner PknG also harbored a novel mutation at location 613 (Arg613Gln (cgg/cAg)). Similarly, mutation (Ser385Arg) in pknA at genomic position 17608 has been observed, reflecting mutations in diverse genes functions from regulation the MTB cell growth to cell signaling under stress conditions.

Variants in pncB1

A novel deletion (TCAGGCCG) and two nonsynonymous mutations (Pro447Ser and Gly429Ala) have been detected in PncB1 at genomic position 1499213- 1499220. PncB1 (Rv1330c) and PncB2 (Rv0573c) are two putative nicotinic acid phosphoribosyl transferases. Inhibitors designed against nicotinamide adenine dinucleotide (NAD) synthetase, may induced bactericidal effect as abrupt starvation of NAD in actively growing and nonreplicating *MTB.*

Discussion

It is very practical to investigate the drug resistance level in novel genotypes to uncover the reason of high prevalence drug resistance. The frequencies of resistance in Beijing genotype have been reported, higher against rifampin, ofloxacin and also multidrug-resistance were significantly higher than non-Beijing

strains. In Fujian and Guangdong, a novel genotype named “China Southern genotype (CS)” was found [44, 45]. Studies in high burden countries including Pakistan are needed to find a correlation of drug resistance and novel spoligotypes circulating in distinct geographic regions. According to the previous reports [45], drug-resistance and Beijing genotype has significant associations which might be responsible for the spread and emergence of MDR-TB. These results have been confirmed in Vietnam, Germany, Cuba and America, indicating high incidence of the Beijing genotype that might be correlated with high transmission and virulence. In a more recent study in Khyber Pakhtunkhwa province of Pakistan [44], the majority of the MTB isolates were found of unknown pattern that were evolutionary linked to L3/CAS strain, where nine isolates (5.4%) of the unknown strains were tentatively named as L3/CAS-KP. Drug resistance in these geographically distinct strains should be investigated for better management of global TB control program 2030. Although the number of samples are very limited in a recent study [46], variation in key genes’ suggests some more investigation on the pathogenesis and single-nucleotide polymorphisms to find any correlation with high transmission and virulence.

In the current investigation, the diversity of variation in ESX protein may suggest a different type of virulency which need to be investigated through murine models. MTB uses early secretory antigenic target (ESAT6) protein family secretion (ESX) systems (T7SS), to export effector proteins that helps the pathogen to resist or evade the host immune response [51, 52]. The ESX-1 plays an important function in virulence, linked to secreted effector proteins, EsxA, inducing phagosomal rupture in phagocytes of host. Interestingly the ESX-1 effectors EspA, EspC and EspD, absent in non-pathogenic species, are present in pathogenic mycobacteria. Besides ESX-1, MTB, has four additional ESX systems: ESX-3 function as iron acquisition; ESX-5 is involved in the secretion of members of PE and PPE having Pro-Glu and Pro-Pro-Glu amino-terminal motifs [53]. ESX-2 and ESX-4 are systems for which the functions are currently unknown. ESX-1 and ESX-5 are both crucial virulence determinants of MTB.

The earlier studies demonstrated that the ESX-1 (T7SS) is essential for bacterial access to the host cytosol [60–62]. The frequency of mutations in each ESX, ESX-1 (7), ESX-2 (5), ESX-3 (6), ESX-4 (3), and ESX-5 (2) is different and the highest has been detected in ESX-1 followed by ESX-3. ESX-3 are important in inflammatory and pathological responses during infection [63]. The variation in PE/PPE could be linked to a functional role pathogenicity [64–66]. To distinguish between latent and active *M. tuberculosis* infection, the anti-PPE55 antibodies may be applied as biomarkers [67]. The PE/PPE gene family exposes the virulence-associated regulators in most recent sublineage genes like SigD and SigB. Before the pathogenicity of mycobacteria arose, regulators of earlier sublineage PE/PPE genes may have been present. Functional studies demonstrate that the PE/PPE proteins are commonly localize to the surface in mycobacteria and are involved in secretory systems. Lipolytic enzymes of this family a role in virulency, as key components of the cell wall interacting with the host for survival purposes. The presence of confirmed lipolytic enzymes in the PE/PPE families of proteins are limited to sublineage V; suggesting an essential role in lipid metabolism in MTB virulency [68–70]. Together with apparent abundance during in vivo infection, suggests that these members play essential roles in pathogenicity [71–73]. About 10% of MTB coding genome is comprised of PE and PPE families [74]. Members of the PE and PPE are known by

the presence of a Pro-Glu (PE) and Pro-Pro-Glu (PPE) conserved motif at N-terminal [75]. PE and PPE have 110 and 180 amino acids conserved region at N-terminal. however, the C-terminal is highly variable.

The characteristic PE and PPE domains is very crucial for targeting the family members of PE/PPE to the cell wall, eliciting potent B- and T-cell responses. The PGRS domain of PE_PGRS30, a virulence factor, is responsible for targeting the protein to mycobacterial cell poles in M. tuberculosis and M. bovis [76, 77]. In PPE17, the PPE domain is crucial for cell surface localization [78]. The PE/PPE family of proteins influence macrophage signaling and pathogenesis of mycobacteria [79–81].

The PE/PPE protein family contributes pathogenicity of MTB that may be targeted for novel drug or immune-modulatory approaches. However, a better understanding of PE/PPE protein function is required before it could be realized as potential. This high diversity and variation in genomic sequence in the PE/PPE proteins in locally circulation strains may affect the immunity against MTB. This variation should be investigated in geographic specific isolates among high burden countries for better management of TB.

The PE-PGRS subfamily is largest class of the PE family, with 67 members in MTB H37Rv consist of the PE domain followed by a C-terminal extension encoded by polymorphic GC-rich repetitive sequences (PGRS) motif. Currently PE-PGRS proteins contain about 1900 amino acids with glycine up to 50% [79]. The C terminals of subfamilies of PE_PGRS and PPE_MPTR have GC-rich stretches. These stretches are considered as hotspots for recombination events and mutations [82]. This high variability results in antigenic diversity that may help the pathogen to evade host-protective immune responses, [76]. The high variability may help the pathogen in survival to evade host-protective immune responses [76]. Among the PPE family, PE_PGRS47 suppresses autophagy and impairs antigen presentation (Saini et al. 2016). In a more recent study, depletion of PknA, resulted in increased drug susceptibility to key tuberculosis drug, rifampin and β-lactam antibiotics [83].

PknG is a Serine/Threonine protein kinase recognized as a key player in mycobacterial physiology and pathogenesis. Recent studies show that the enzyme glutamine synthetase and the protein FhaA are the two substrates are phosphorylated by PknG in vitro specific residues in both [84]. In another study the fhaA deletion mutant were created and different classes of antibiotics were screened to check their sensitivity. MTB isolates that were fhaA deletion mutant, have been detected as sensitive to multiple antibiotics showing an overall permeability flaw [85]. However, the effect of substitutions has not been evaluated so for, which need to be investigated for future drug development against this target.

A novel deletion (TCAGGCCG) and two nonsynonymous mutations (Pro447Ser and Gly429Ala) in PncB1 at genomic position 1499213-1499220. PncB1 (Rv1330c) and PncB2 (Rv0573c) are phosphoribosyl transferases, playing role in cofactor salvage [86]. Cell death is caused by disruption of cellular redox homeostasis and cofactor starvation and also as electron transport is weakened by limiting NAD. *PncB2 plays a more important role in the adaptation to nonreplicating persistence and also known as member of the DosR regulon, important during hypoxic conditions* [87, 88]. *However, the effect of these mutations is needed to be investigated before future drug designing for better management of treatment.*

In conclusion the metabolic model shows a diversity of variants in essential genes, functions from cell growth to cell signaling. The physiology and pathogenesis determinants genes have novel type of mutations in this distinct geographic region. The high variability in antigenic diversity that may help the pathogen to evade host-protective immune responses. The diversity of mutations in PE_PGRS47 family may help the pathogen to suppresses autophagy and impairs antigen presentation. The current study provides useful information, associated with geographic specific strains for biomarkers development and better management of drug resistance isolates.

Abbreviations

BSL-III: Bbiosafety level-III laboratory

DOTS: Directly observed treatment, short-course

DST: Drug susceptibility testing

ESX-SSP: ESX secretory system proteins

fhaA: Forkhead-associated

CS: China Southern genotype

ESAT6: Early secretory antigenic target

INH: Isoniazid

LJ: Lowenstein Jensen medium

MTB: *Mycobacterium tuberculosis*

MGIT: Mycobacterium growth indicator tube

NAD: Nicotinamide adenine dinucleotide

NALC-NaOH: N-acetyl-L-cysteine sodium hydroxide

PZA: Pyrazinamide

PPE: Pro-Pro-Glu

PGRS: Polymorphic GC-rich repetitive sequences

PZA-R: Pyrazinamide resistance

TB: Tuberculosis

T7SS: Type-VII secretory system

WGS: Whole genome sequencing

Material And Methods

Ethical approval

A local ethics committee ruled that no formal ethics approval was required in this particular case.

Sample collection

Random samples have been collected from TB suspects with basic information from their guardians.

All the samples were subjected to decontamination and digestion processing.

Sample processing

All received samples were digested and decontaminated using standard N-acetyl-L-cysteine sodium hydroxide (NALC-NaOH) method (GLI, 2014) in a biosafety level-III laboratory (BSL-III) at Provincial TB Reference Laboratory, Peshawar. Briefly one aliquot was inoculated on Lowenstein Jensen medium (LJ) and in a mycobacterium growth indicator tube (MGIT). Positive growth in the tubes was confirmed by Tbc ID device (Ref: 245159, Becton, Dickinson).

Drug susceptibility testing

All confirmed mycobacterial isolates were processed for both phenotypic DST and molecular resistance assay. DST was performed using a BD BACTEC MGIT 960 SIRE kit (Ref: 245123, Becton, Dickinson) according to the policy WHO second line drug susceptibility testing [89]. The standard minimum inhibitory concentration (MIC) for ofloxacin (OFX), levofloxacin (LEV), and moxifloxacin (MOX) was taken as 2ug/mL, 1ug/mL, and 1ug/mL respectively. One sample aliquot was processed for acid fast bacilli (AFB) microscopy using primostar-LED fluorescent microscopy.

Whole genome sequencing

Samples resistance to PZA were subjected for extraction of DNA through CTAB method [40, 90] followed by whole-genome sequencing (WGS), a more effective, accessible and reasonable method for deep insight into the genome of MTB isolates, involving single nucleotide variations (SNVs). The extracted DNA from PZA-resistance samples were sequenced and partial analysis has been performed through Illumina-X10 sequencer, based on the Solexa sequencing principle at center of Shanghai Jiao Tong University, China.

Ssequences analysis

To analyze the WGS data, an all-in-one web-based tool, Total Genotyping Solution for TB (TGS-TB) [91], using NGS for spoligotyping and the detection of phylogenies, IS6110 insertion sites, with core genomic SNVs, and 43 customized loci for variable number tandem repeat (VNTR). The tool is a user-friendly, performed all these analysis on a simple click interface. The methodology was implemented with a KvarQ script, predicting MTBC lineages and antimicrobial resistance.

Quality and mutation detection

The quality of the sequence was checked with FASTQC and after trimming the raw reads, the genome was mapped against the reference strain H37Rv using MTBSeq software package [92]. The TGS-TB and MTBseq detect variant positions with known associated with antibiotic resistance and performs also lineage classification. When comparing multiple datasets, MTBseq provides a joint list of variants and a FASTA alignment of SNP positions for use in phylogenomic analysis, and identifies groups of related isolates, provides a more accurate strain typing for epidemiological and clinical investigations [91, 92]. In brief, the alignment was carried out using BWA-mem algorithm. The resulting output in Sam/Bam format was analyzed using Sam tools package. Statistical analysis, variant and positioning was done in gatk software package. For further annotation, samples in FASTA format have been subjected to RAST server [47], assigning functions to the genes, identifying the protein-encoding, rRNA and tRNA genes, predicts subsystems in the genome.

Whole genome sequencing comparison

Whole genome sequence results in comparison with H37Rv were uploaded to RAST server subsystem technology. The server achieves accuracy and completeness on the use of a subsystems that are manually curated as protein families largely derived from the subsystems. RAST server implements two classes of asserted gene functions: subsystem-based and non-subsystem-based assertions. The earlier is based on recognition of functional variants of subsystems, while the later are using integration of evidence from a number of other tools. The SEED servers inside RAST offer access to data [47, 48] which have the ability to annotate prokaryotic genomes and creating metabolic reconstructions and detailed model of metabolism.

Mapping whole genome into circular diagrams

The whole genome sequence was mapped in the form of circular structure using CGView Server [49], generating graphical maps that show sequence features, base composition plots, and sequence similarity plots. The server can visualize the genome features of bacteria, plasmid, chloroplast or mitochondria.

Analysis of type-VII secretory system

In mycobacteria, the ESX-1 type VII secretion system (T7SS) is vital for access to the host cytosol [65, 93]. In order to study the variability in the T7SS, the member genes were analyzed in the whole genome of MTB through VRprofile server [94], a web server facilitating in fast screening of virulence and antibiotic

resistance genes using a backend MobilomeDB database, built on sets gene cluster loci of bacterial type III/IV/VI/VII secretion systems.

Declaration

Ethical approval and consent to participate

A local ethics committee ruled that no formal ethics approval is required in this particular case (PTP/PTRL-402/16). Prior to research study, an informed written consent was gained.

Image attribution

The graphical abstract in the current study is our own created image.

Consent for publication

Not applicable

Availability of data and materials

The datasets in the present study is available in public database under accession No. (NCBI BioProject Accession: PRJNA629298, PRJNA629388).

Competing Interest

All the authors have no conflicts of interest

Funding

The research was supported by Dong-Qing Wei through grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China for experimental work, the National Natural Science Foundation of China (Contract no. 61832019, 61503244) for arrangement of chemicals and disposables, the Science and Technology Commission of Shanghai Municipality (Grant: 19430750600) for DNA extraction and sequencing, the Natural Science Foundation of Henan Province (162300410060), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14) for culturing and data analysis. The funding agencies has no role in writing the manuscript, or collection, analysis, design of the study, or interpretation of data.

Authors' contributions

Conception: **DQW,** MTK, SA

Experiment was design by **DQW,** MTK, SA, AA, SC

Experimental work was conducted by MTK, SA, ASK, MI.

Data analysis: MTK, SA, AK, ACK, HW, SZ, YZ, ZC, AW, YW, MZ, KL.

Manuscript writing: MTK, MTZ, SA, AA, SC, HW

Funding and supervision: DQW

All authors have read and approved the manuscript.

Acknowledgments

The present study was ethically supported by Prof. Dr. Shaukat Iqbal Malik, Dr. Sahar Fazal and Aamer Iqbal Bhatti, Capital University of Science and Technology Islamabad Capital University of Science and Technology Islamabad.

References

1. African SE. Controlled clinical trial of four 6-month regimens of chemotherapy for pulmonary tuberculosis. Am Rev Respir Dis. 2015.
2. Gu Y, Yu X, Jiang G, Wang X, Ma Y, Li Y, et al. Pyrazinamide resistance among multidrug-resistant tuberculosis clinical isolates in a national referral center of China and its correlations with pncA, rpsA, and panD gene mutations. Diagn Microbiol Infect Dis. 2016;84:207–11.
3. Palzkill T, Palzkill T. Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC β-lactamases. Front Mol Biosci. 2018;5. doi:10.3389/fmolb.2018.00016.
4. Reiche MA, Warner DF, Mizrahi V. Targeting DNA Replication and Repair for the Development of Novel Therapeutics against Tuberculosis. Front Mol Biosci. 2017;4. doi:10.3389/fmolb.2017.00075.
5. Ashworth A. Drug Resistance Caused by Reversion Mutation. Cancer Res. 2008;68:10021–3.
6. Tyagi RK, Das MK, Singh SS, Sharma YD. Discordance in drug resistance-associated mutation patterns in marker genes of Plasmodium falciparum and Plasmodium knowlesi during coinfections. J Antimicrob Chemother. 2013;68:1081–8.
7. Bell DW, Gore I, Okimoto RA, Godin-Heymann N, Sordella R, Mulloy R, et al. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. Nat Genet. 2005;37:1315–6.
8. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. Genome Res. 2007;17:1195–201.
9. Thomas A, El SR, Reed JC, Krajewski S, Silber R, Potmesil M, et al. Drug-induced apoptosis in B-cell chronic lymphocytic leukemia: relationship between p53 gene mutation and bcl-2/bax proteins in drug resistance. Oncogene. 1996;12:1055–62.

10. Yun C-H, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong K-K, et al. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci. 2008;105:2070–5.
11. Yang F, Zheng G, Fu T, Li X, Tu G, Li YH, et al. Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. Phys Chem Chem Phys. 2018;20:23873–84.
12. Yoon J-H, Nam J-S, Kim K-J, Ro Y-T. Characterization of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis isolates from Korea and analysis of the correlation between the mutations and pyrazinamidase activity. World J Microbiol Biotechnol. 2014;30:2821–8.
13. Yadon AN, Maharaj K, Adamson JH, Lai Y-P, Sacchettini JC, Ioerger TR, et al. A comprehensive characterization of PncA polymorphisms that confer resistance to pyrazinamide. Nat Commun. 2017;8. doi:10.1038/s41467-017-00721-2.
14. Lemaitre N, Callebaut I, Frenois F, Jarleir V, Sougakof W. Study of the structure–activity relationships for the pyrazinamidase (PncA) from Mycobacterium tuberculosis. Biochem J. 2001;353:453–8.
15. Sheen P, Ferrer P, Gilman RH, López-Llano J, Fuentes P, Valencia E, et al. Effect of pyrazinamidase activity on pyrazinamide resistance in Mycobacterium tuberculosis. Tuberculosis. 2009;89:109–13.
16. Worth CL, Gong S, Blundell TL. Structural and functional constraints in the evolution of protein families. Nat Rev Mol Cell Biol. 2009;10:709–20.
17. Ganesan P, Ramalingam R. Investigation of structural stability and functionality of homodimeric gramicidin towards peptide-based drug: a molecular simulation approach. J Cell Biochem. 2018.
18. Kosloff M, Kolodny R. Sequence-similar, structure-dissimilar protein pairs in the PDB. Proteins. 2008;71:891–902.
19. Khan MT, Malik SI, Bhatti AI, Ali S, Khan AS, Zeb MT, et al. Pyrazinamide-resistant mycobacterium tuberculosis isolates from Khyber Pakhtunkhwa and rpsA mutations. J Biol Regul Homeost Agents. 2018;32:705–9.
20. Ramirez-Busby SM, Rodwell TC, Fink L, Catanzaro D, Jackson RL, Pettigrove M, et al. A Multinational Analysis of Mutations and Heterogeneity in PZase, RpsA , and PanD Associated with Pyrazinamide Resistance in M/XDR Mycobacterium tuberculosis. Sci Rep. 2017;7:1–9.
21. Singh A, Somvanshi P, Grover A. Pyrazinamide drug resistance in RpsA mutant (Δ438A) of Mycobacterium tuberculosis: Dynamics of essential motions and free-energy landscape analysis. J Cell Biochem. 2019;120:7386–402.
22. Zhang S, Chen J, Shi W, Liu W, Zhang W, Zhang Y. Mutations in panD encoding aspartate decarboxylase are associated with pyrazinamide resistance in Mycobacterium tuberculosis. Emerg Microbes Infect. 2013;2:e34.
23. Iwamoto T, Murase Y, Yoshida S, Aono A, Kuroda M, Sekizuka T, et al. Overcoming the pitfalls of automatic interpretation of whole genome sequencing data by online tools for the prediction of pyrazinamide resistance in Mycobacterium tuberculosis. PLOS ONE. 2019;14:e0212798.

24. Liu J, Shi W, Zhang S, Hao X, Maslov DA, Shur KV, et al. Mutations in Efflux Pump Rv1258c (Tap) Cause Resistance to Pyrazinamide, Isoniazid, and Streptomycin in M. tuberculosis. Front Microbiol. 2019;10:5836.
25. Simons SO, Mulder A, van Ingen J, Boeree MJ, van Soolingen D. Role of rpsA gene sequencing in diagnosis of pyrazinamide resistance. J Clin Microbiol. 2013;51:382–382.
26. Salah P, Bisaglia M, Aliprandi P, Uzan M, Sizun C, Bontems F. Probing the relationship between Gram-negative and Gram-positive S1 proteins by sequence analysis. Nucleic Acids Res. 2009;37:5578–88.
27. Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. Cell. 1997;88:235–42.
28. Khan M. Pyrazinamide resistance and mutations L19R, R140H, and E144K in Pyrazinamidase of Mycobacterium tuberculosis. J Cell Biochem P. 2019;120 5 SRC-BaiduScholar:7154–66.
29. Khan MT, Malik SI. Structural dynamics behind variants in pyrazinamidase and pyrazinamide resistance. J Biomol Struct Dyn. 2019;:1–15.
30. Khan MT, Malik SI, Ali S, Sheed Khan A, Nadeem T, Zeb MT, et al. Prevalence of Pyrazinamide Resistance in Khyber Pakhtunkhwa, Pakistan. Microb Drug Resist Larchmt N. 2018.
31. Khan MT, Rehaman AU, Junaid M, Malik SI, Wei D-Q. Insight into novel clinical mutants of RpsA-S324F, E325K, and G341R of Mycobacterium tuberculosis associated with pyrazinamide resistance. Comput Struct Biotechnol J. 2018. doi:10.1016/j.csbj.2018.09.004.
32. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, 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. doi:10.3389/fmicb.2019.00309.
33. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, et al. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J Bacteriol. 2002;184:5479–90.
34. Iketleng T, Lessells R, Dlamini MT, Mogashoa T, Mupfumi L, Moyo S, et al. Mycobacterium tuberculosis Next-Generation Whole Genome Sequencing: Opportunities and Challenges. Tuberculosis Research and Treatment. 2018. doi:10.1155/2018/1298542.
35. Ford C, Yusim K, Ioerger T, Feng S, Chase M, Greene M, et al. Mycobacterium tuberculosis – Heterogeneity revealed through whole genome sequencing. Tuberculosis. 2012;92:194–201.
36. Zakham F, Laurent S, Esteves Carreira AL, Corbaz A, Bertelli C, Masserey E, et al. Whole-genome sequencing for rapid, reliable and routine investigation of Mycobacterium tuberculosis transmission in local communities. New Microbes New Infect. 2019;31. doi:10.1016/j.nmni.2019.100582.
37. Nikolayevskyy V, Niemann S, Anthony R, van Soolingen D, Tagliani E, Ködmön C, et al. Role and value of whole genome sequencing in studying tuberculosis transmission. Clin Microbiol Infect. 2019;25:1377–82.
38. Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al. High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic

modelling of a cohort from Valencia Region, Spain. PLOS Med. 2019;16:e1002961.

39. Takiff HE, Feo O. Clinical value of whole-genome sequencing of Mycobacterium tuberculosis. Lancet Infect Dis. 2015;15:1077–90.
40. Nimmo C, Shaw LP, Doyle R, Williams R, Brien K, Burgess C, et al. Whole genome sequencing Mycobacterium tuberculosis directly from sputum identifies more genetic diversity than sequencing from culture. BMC Genomics. 2019;20. doi:10.1186/s12864-019-5782-2.
41. Cohen KA, Manson AL, Desjardins CA, Abeel T, Earl AM. Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges. Genome Med. 2019;11:45.
42. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet. 2012;44:106–10.
43. khan MT, Malik SI, Ali S, Masood N, Nadeem T, Khan AS, et al. Pyrazinamide resistance and mutations in pncA among isolates of Mycobacterium tuberculosis from Khyber Pakhtunkhwa, Pakistan. BMC Infect Dis. 2019;19:116.
44. Ali S, Khan MT, Anwar Sheed K, Khan MM, Hasan F. Spoligotyping analysis of Mycobacterium tuberculosis in Khyber Pakhtunkhwa area, Pakistan. Infect Drug Resist. 2019;12:1363–9.
45. Pang Y, Zhou Y, Zhao B, Liu G, Jiang G, Xia H, et al. Spoligotyping and Drug Resistance Analysis of Mycobacterium tuberculosis Strains from National Survey in China. PLOS ONE. 2012;7:e32976.
46. Yar AM, Zaman G, Hussain A, Changhui Y, Rasul A, Hussain A, et al. Comparative Genome Analysis of 2 Mycobacterium Tuberculosis Strains from Pakistan: Insights Globally Into Drug Resistance, Virulence, and Niche Adaptation. Evol Bioinforma Online. 2018;14. doi:10.1177/1176934318790252.
47. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 2008;9:75.
48. Aziz RK, Devoid S, Disz T, Edwards RA, Henry CS, Olsen GJ, et al. SEED Servers: High-Performance Access to the SEED Genomes, Annotations, and Metabolic Models. PLOS ONE. 2012;7:e48053.
49. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36 Web Server issue:W181.
50. Sayes F, Blanc C, Ates LS, Deboosere N, Orgeur M, Le Chevalier F, et al. Multiplexed Quantitation of Intraphagocyte Mycobacterium tuberculosis Secreted Protein Effectors. Cell Rep. 2018;23:1072–84.
51. Famelis N, Rivera-Calzada A, Degliesposti G, Wingender M, Mietrach N, Skehel JM, et al. Architecture of the mycobacterial type VII secretion system. Nature. 2019;576:321–5.
52. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol. 2016;14:677–91.
53. Simeone R, Bottai D, Frigui W, Majlessi L, Brosch R. ESX/type VII secretion systems of mycobacteria: Insights into evolution, pathogenicity and protection. Tuberc Edinb Scotl. 2015;95 Suppl 1:S150-154.

54. Pym AS, Brodin P, Brosch R, Huerre M, Cole ST. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol Microbiol. 2002;46:709–17.
55. Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, et al. Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med. 2003;9:533–9.
56. Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, et al. The primary mechanism of attenuation of bacillus Calmette–Guérin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci. 2003;100:12420–5.
57. Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, et al. Deletion of RD1 from Mycobacterium tuberculosis mimics bacille Calmette-Guérin attenuation. J Infect Dis. 2003;187:117–23.
58. Stanley SA, Raghavan S, Hwang WW, Cox JS. Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. Proc Natl Acad Sci U S A. 2003;100:13001–6.
59. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci. 2003;100:12989–94.
60. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, et al. Phagosomal Rupture by Mycobacterium tuberculosis Results in Toxicity and Host Cell Death. PLoS Pathog. 2012;8. doi:10.1371/journal.ppat.1002507.
61. McDonough KA, Kress Y, Bloom BR. Pathogenesis of tuberculosis: interaction of Mycobacterium tuberculosis with macrophages. Infect Immun. 1993;61:2763–73.
62. Smith I. Mycobacterium tuberculosis Pathogenesis and Molecular Determinants of Virulence. Clin Microbiol Rev. 2003;16:463–96.
63. Kim YS, Yang C-S, Nguyen LT, Kim JK, Jin HS, Choe JH, et al. Mycobacterium abscessus ESX-3 plays an important role in host inflammatory and pathological responses during infection. Microbes Infect. 2017;19:5–17.
64. Fishbein S, van Wyk N, Warren RM, Sampson SL. Phylogeny to function: PE/PPE protein evolution and impact on Mycobacterium tuberculosis pathogenicity. Mol Microbiol. 2015;96:901–16.
65. Abdallah AM, Gey van Pittius NC, DiGiuseppe Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CMJE, et al. Type VII secretion — mycobacteria show the way. Nat Rev Microbiol. 2007;5:883–91.
66. Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, et al. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol Microbiol. 2006;62:667–79.
67. Singh KK, Dong Y, Patibandla SA, McMurray DN, Arora VK, Laal S. Immunogenicity of the Mycobacterium tuberculosis PPE55 (Rv3347c) Protein during Incipient and Clinical Tuberculosis. Infect Immun. 2005;73:5004–14.
68. Bhat KH, Das A, Srikantam A, Mukhopadhyay S. PPE2 protein of Mycobacterium tuberculosis may inhibit nitric oxide in activated macrophages. Ann N Y Acad Sci. 2013;1283:97–101.

69. Chatrath S, Gupta VK, Dixit A, Garg LC. The Rv1651c-encoded PE-PGRS30 protein expressed in Mycobacterium smegmatis exhibits polar localization and modulates its growth profile. FEMS Microbiol Lett. 2011;322:194–9.
70. Cascioferro A, Daleke MH, Ventura M, Donà V, Delogu G, Palù G, et al. Functional dissection of the PE domain responsible for translocation of PE_PGRS33 across the mycobacterial cell wall. PloS One. 2011;6:e27713.
71. Daim S, Kawamura I, Tsuchiya K, Hara H, Kurenuma T, Shen Y, et al. Expression of the Mycobacterium tuberculosis PPE37 protein in Mycobacterium smegmatis induces low tumour necrosis factor alpha and interleukin 6 production in murine macrophages. J Med Microbiol. 2011;60 Pt 5:582–91.
72. Copin R, Coscollá M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G, et al. Sequence diversity in the pe_pgrs genes of Mycobacterium tuberculosis is independent of human T cell recognition. mBio. 2014;5:e00960-00913.
73. Dong D, Wang D, Li M, Wang H, Yu J, Wang C, et al. PPE38 modulates the innate immune response and is required for Mycobacterium marinum virulence. Infect Immun. 2012;80:43–54.
74. Cole St, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 1998;393:537–44.
75. Sampson SL. Mycobacterial PE/PPE proteins at the host-pathogen interface. Clin Dev Immunol. 2011;2011:497203.
76. Karboul A, Mazza A, Gey van Pittius NC, Ho JL, Brousseau R, Mardassi H. Frequent homologous recombination events in Mycobacterium tuberculosis PE/PPE multigene families: potential role in antigenic variability. J Bacteriol. 2008;190:7838–46.
77. De Maio F, Maulucci G, Minerva M, Anoosheh S, Palucci I, Iantomasi R, et al. Impact of protein domains on PE_PGRS30 polar localization in Mycobacteria. PloS One. 2014;9:e112482.
78. Donà V, Ventura M, Sali M, Cascioferro A, Provvedi R, Palù G, et al. The PPE domain of PPE17 is responsible for its surface localization and can be used to express heterologous proteins on the mycobacterial surface. PloS One. 2013;8:e57517.
79. Banu S, Honoré N, Saint-Joanis B, Philpott D, Prévost M-C, Cole ST. Are the PE-PGRS proteins of Mycobacterium tuberculosis variable surface antigens? Mol Microbiol. 2002;44:9–19.
80. Brennan MJ, Delogu G. The PE multigene family: a “molecular mantra” for mycobacteria. Trends Microbiol. 2002;10:246–9.
81. Delogu G, Brennan MJ. Comparative immune response to PE and PE_PGRS antigens of Mycobacterium tuberculosis. Infect Immun. 2001;69:5606–11.
82. McEvoy CRE, van Helden PD, Warren RM, Gey van Pittius NC. Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic Mycobacterium tuberculosis PPE38 gene region. BMC Evol Biol. 2009;9:237.
83. Zeng J, Platig J, Cheng T-Y, Ahmed S, Skaf Y, Potluri L-P, et al. Protein kinases PknA and PknB independently and coordinately regulate essential Mycobacterium tuberculosis physiologies and

antimicrobial susceptibility. PLOS Pathog. 2020;16:e1008452.
84. Gil M, Lima A, Rivera B, Rossello J, Urdániz E, Cascioferro A, et al. New substrates and interactors of the mycobacterial Serine/Threonine protein kinase PknG identified by a tailored interactomic approach. J Proteomics. 2019;192:321–33.
85. Viswanathan G, Yadav S, Joshi SV, Raghunand TR. Insights into the function of FhaA, a cell division-associated protein in mycobacteria. FEMS Microbiol Lett. 2017;364.
86. Boshoff HIM, Xu X, Tahlan K, Dowd CS, Pethe K, Camacho LR, et al. Biosynthesis and Recycling of Nicotinamide Cofactors in Mycobacterium tuberculosis. J Biol Chem. 2008;283:19329–41.
87. Voskuil MI, Schnappinger D, Visconti KC, Harrell MI, Dolganov GM, Sherman DR, et al. Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. J Exp Med. 2003;198:705–13.
88. Via LE, Lin PL, Ray SM, Carrillo J, Allen SS, Eum SY, et al. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. Infect Immun. 2008;76:2333–40.
89. Barrera L, Cooreman E, de Dieu Iragena J, Drobniewski F, Duda P, Havelkova M, et al. Policy Guidance on Drug-Susceptibility Testing (DST) of Second-Line Antituberculosis Drugs. Geneva: World Health Organization; 2008. http://www.ncbi.nlm.nih.gov/books/NBK310860/.
90. Mohammadi S, Esfahani BN, Moghim S, Mirhendi H, Zaniani FR, Safaei HG, et al. Optimal DNA Isolation Method for Detection of Nontuberculous Mycobacteria by Polymerase Chain Reaction. Adv Biomed Res. 2017;6:133.
91. Sekizuka T, Yamashita A, Murase Y, Iwamoto T, Mitarai S, Kato S, et al. TGS-TB: Total Genotyping Solution for Mycobacterium tuberculosis Using Short-Read Whole-Genome Sequencing. PLOS ONE. 2015;10:e0142951.
92. Kohl TA, Utpatel C, Schleusener V, Filippo MRD, Beckert P, Cirillo DM, et al. MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. PeerJ. 2018;6:e5895.
93. Mittal E, Skowyra ML, Uwase G, Tinaztepe E, Mehra A, Köster S, et al. Mycobacterium tuberculosis Type VII Secretion System Effectors Differentially Impact the ESCRT Endomembrane Damage Response. mBio. 2018;9. doi:10.1128/mBio.01765-18.
94. Li J, Tai C, Deng Z, Zhong W, He Y, Ou H-Y. VRprofile: gene-cluster-detection-based profiling of virulence and antibiotic resistance traits encoded within genome sequences of pathogenic bacteria. Brief Bioinform. 2018;19:566–74.

Figures

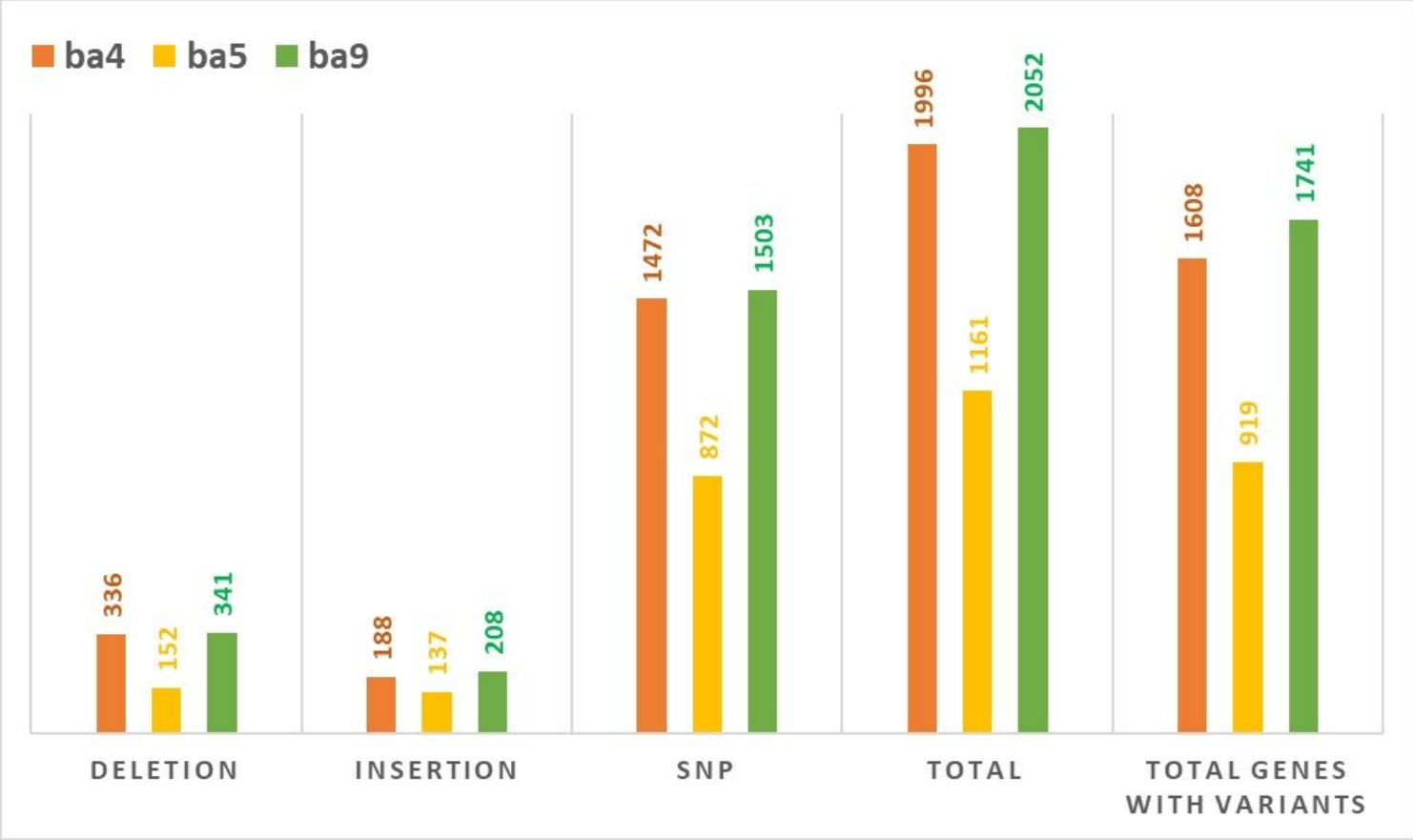


Figure 1

Total number of variants in three WGS of MTB.

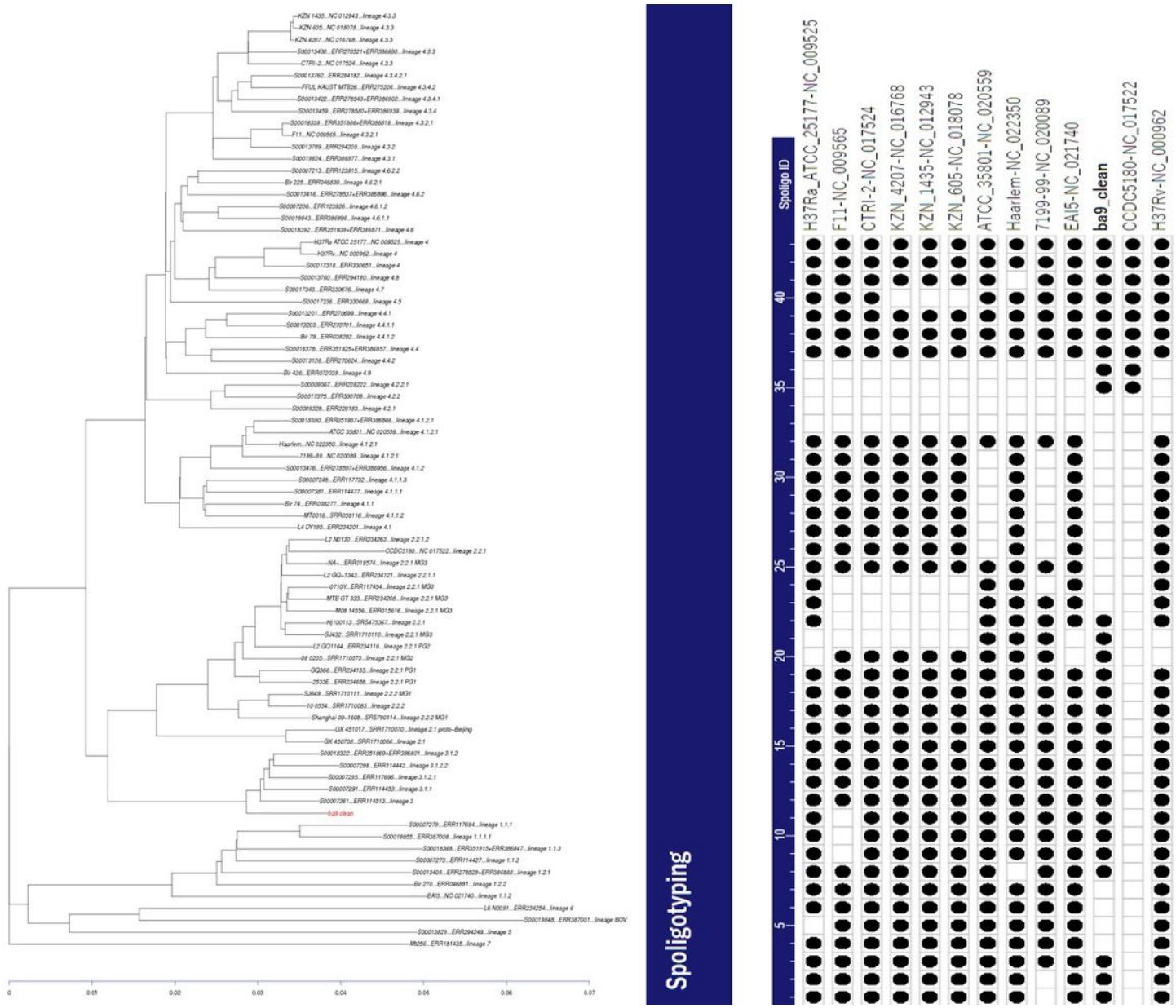


Figure 2

A unique pattern of sample ba9. Total Genotyping Solution for TB (TGS-TB) has been used for spoligotyping and the detection of phylogenies, IS6110 insertion sites, with core genomic SNVs, and 43 customized loci for variable number tandem repeat (VNTR).

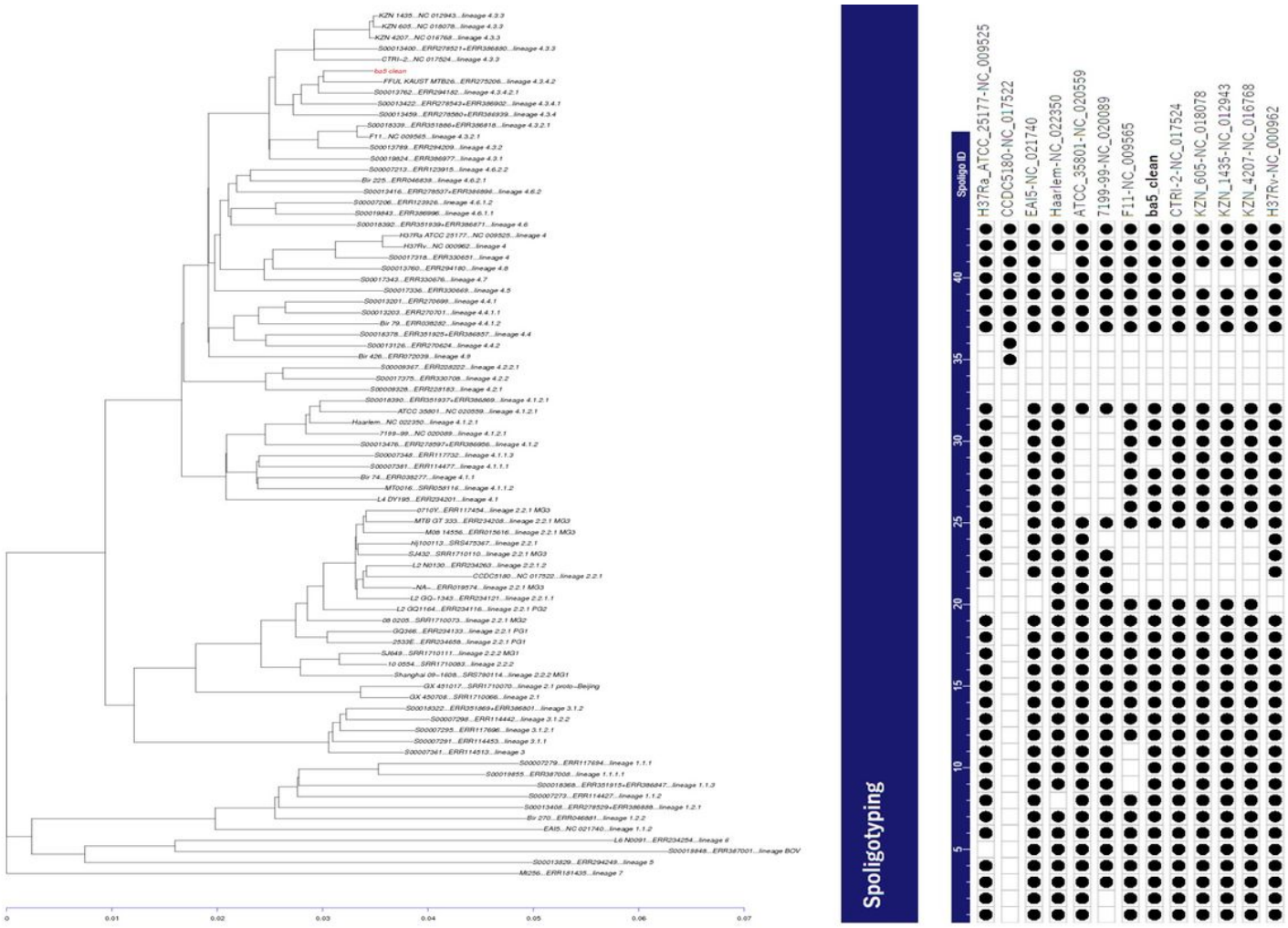


Figure 3

Spoligotyping pattern of sample ba5. Sample ba5 is unique in pattern in the phylogenetic tree (colored red).

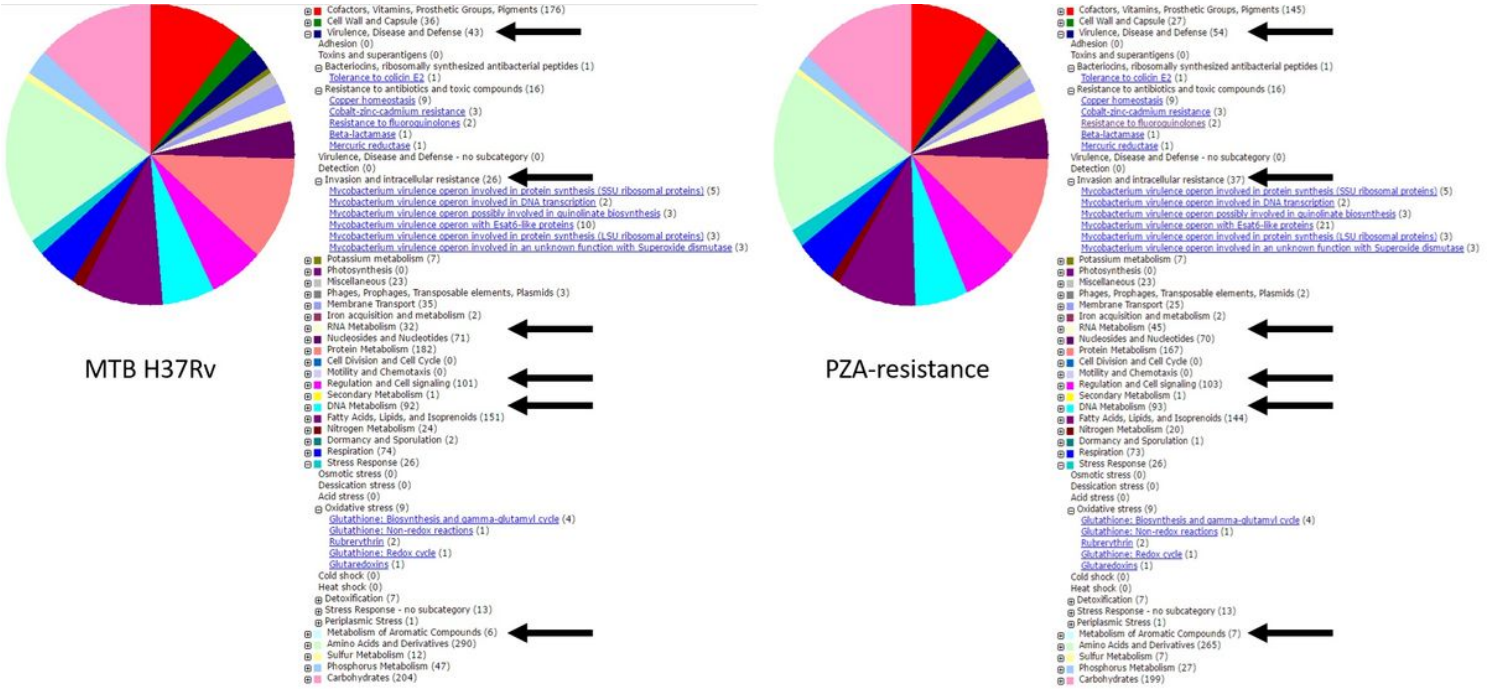


Figure 4

Comparison of WG MTB reference genome and PZA-resistance isolate through RAST. The RAST server provides an environment for browsing the annotated genome and comparing it to the hundreds of genomes maintained within the SEED integration [47, 48]. The genome viewer included in RAST supports comparison against existing genomes in detail. It also determines and displaying genomic context around specific genes. The differences in subsystem of H37Rv and PZA resistance whole genome has been shown through an arrow pointing towards a specific category and their number of proteins/genes included.

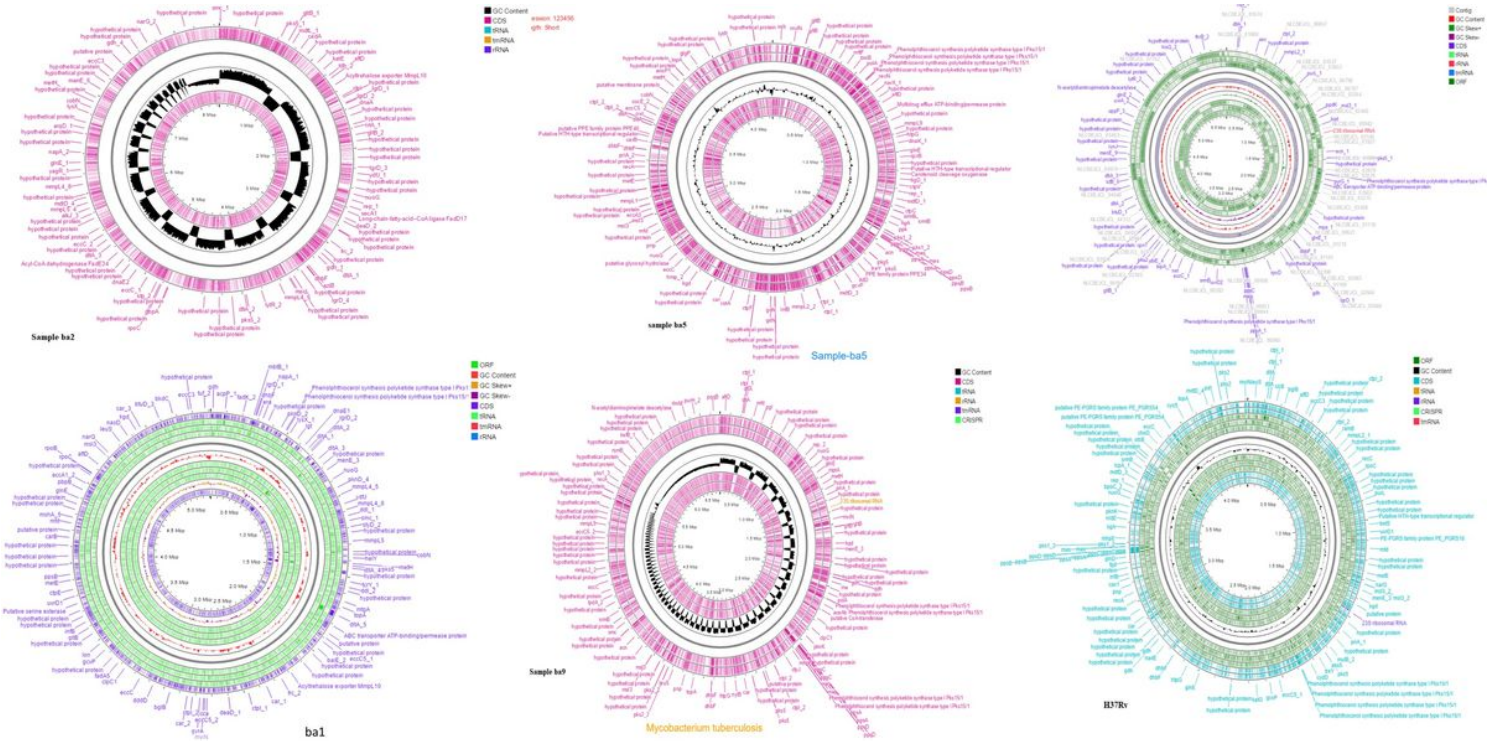
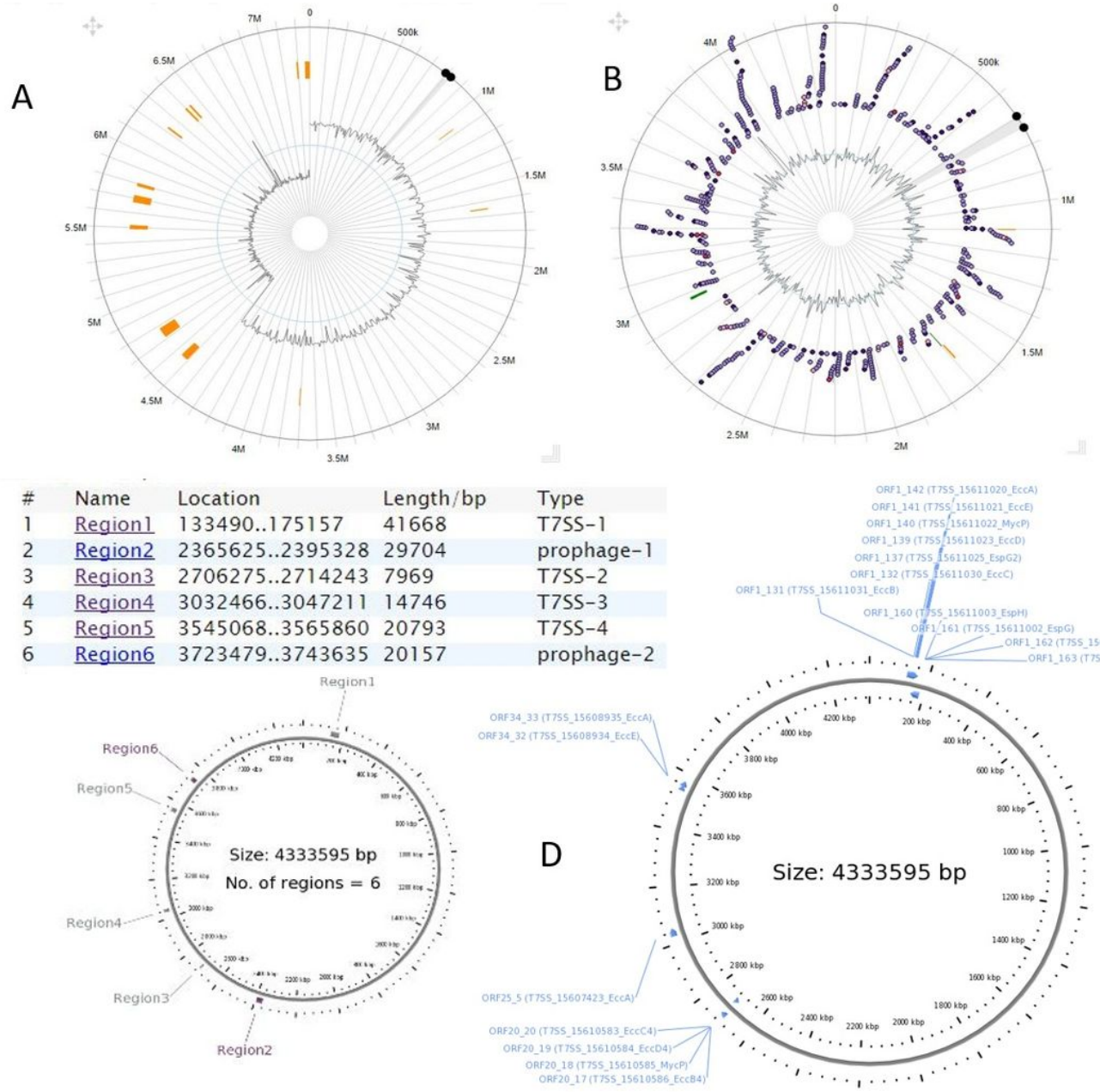


Figure 5

MTB whole genome circular representation in comparison with reference (H37Rv).



#	Name	Location	Length/bp	Type
1	Region1	133490..175157	41668	T7SS-1
2	Region2	2365625..2395328	29704	prophage-1
3	Region3	2706275..2714243	7969	T7SS-2
4	Region4	3032466..3047211	14746	T7SS-3
5	Region5	3545068..3565860	20793	T7SS-4
6	Region6	3723479..3743635	20157	prophage-2

Figure 6

Type-VII secretory system (T7SS) of local MTB isolates. (A) virulency factors in the current isolates. Yellow shows the location. (B) virulency factor of reference strain (H37Rv). (7) T7SS systems and their location in the genome. (D) Member genes in T7SS and their location in the local genome.

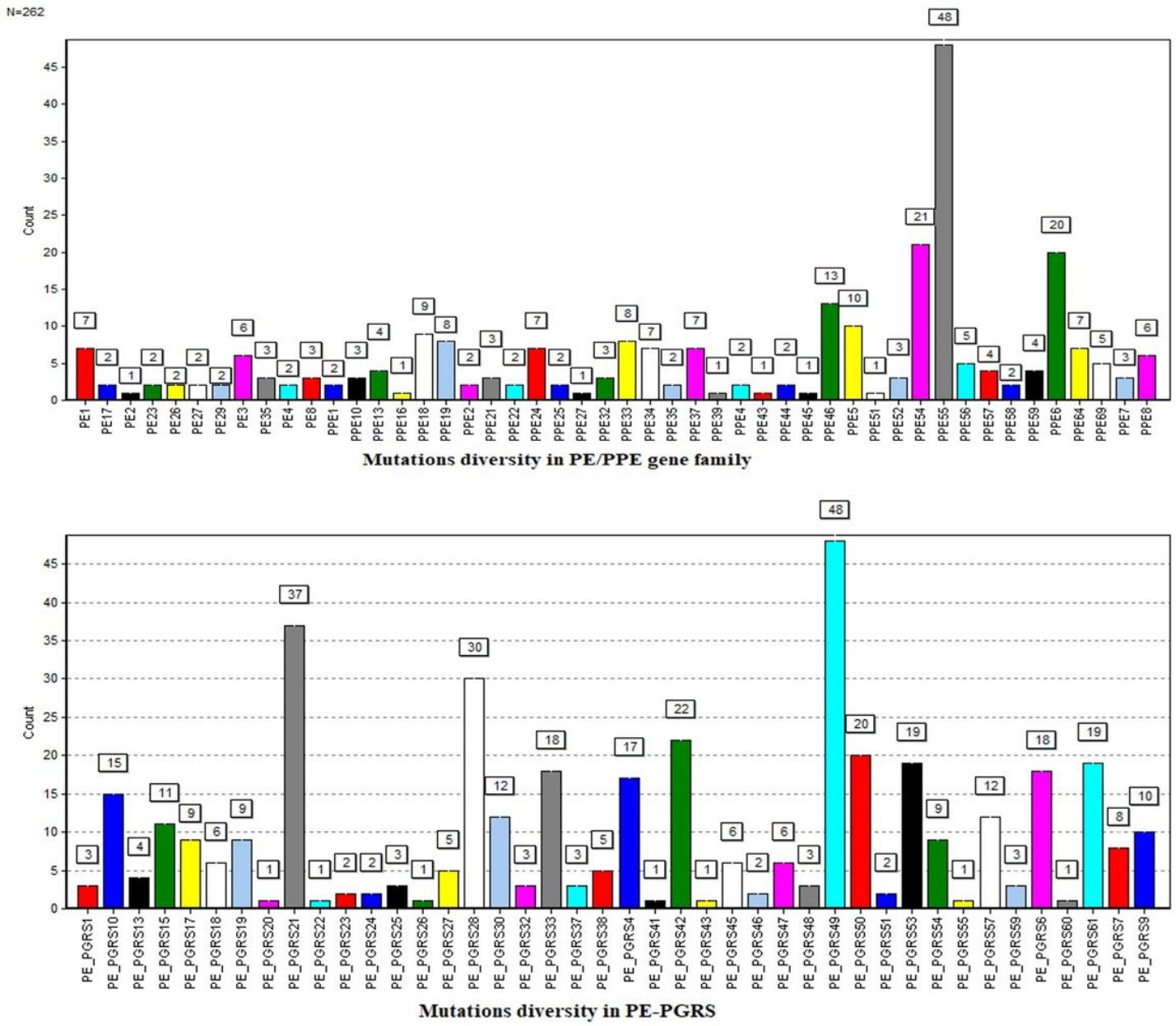


Figure 7

Diversity in mutation PE_PGRS proteins in all samples

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary3.xlsx
- Supplementary2.xlsx
- Supplementary1.xlsx
- Graphicalabstract.jpg